GRAPH LEARNING FOR BRAIN IMAGING

EDITED BY: Feng Liu, Yu Zhang, Jordi Solé-Casals, Islem Rekik and Yehia Massoud

PUBLISHED IN: Frontiers in Neuroscience

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-83250-134-4
DOI 10.3389/978-2-83250-134-4

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

GRAPH LEARNING FOR BRAIN IMAGING

Topic Editors:
Feng Liu, Stevens Institute of Technology, United States
Yu Zhang, Lehigh University, United States
Jordi Solé-Casals, Universitat de Vic - Universitat Central de Catalunya, Spain
Islem Rekik, Istanbul Technical University, Turkey
Yehia Massoud, King Abdullah University of Science and Technology, Saudi Arabia

Citation: Liu, F., Zhang, Y., Solé-Casals, J., Rekik, I., Massoud, Y., eds. (2022). Graph Learning for Brain Imaging. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-134-4

Table of Contents

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 22 August 2022
DOI 10.3389/fnins.2022.1001818

OPEN ACCESS

EDITED AND REVIEWED BY
Minjeong Kim,
University of North Carolina at Greensboro, United States

*CORRESPONDENCE
Feng Liu
fliu22@stevens.edu
Yu Zhang
yuzi20@lehigh.edu
Islem Rekik
islem.rekik@gmail.com
Yehia Massoud
yehia.massoud4@gmail.com
Jordi Solé-Casals
jordi.sole@uvic.cat

SPECIALTY SECTION
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

RECEIVED 24 July 2022
ACCEPTED 02 August 2022
PUBLISHED 22 August 2022

CITATION
Liu F, Zhang Y, Rekik I, Massoud Y and Solé-Casals J (2022) Editorial: Graph learning for brain imaging. *Front. Neurosci.* 16:1001818. doi: 10.3389/fnins.2022.1001818

Editorial: Graph learning for brain imaging

Feng Liu[1]*, Yu Zhang[2,3]*, Islem Rekik[4]*, Yehia Massoud[5]* and Jordi Solé-Casals[6,7]*

[1]School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, United States, [2]Department of Bioengineering, Lehigh University, Bethlehem, PA, United States, [3]Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA, United States, [4]BASIRA Lab, Faculty of Computer and Informatics Engineering, Istanbul Technical University, Istanbul, Turkey, [5]Department of Electrical and Computer Engineering, King Abdullah University of Science and Technology, Makkah, Saudi Arabia, [6]Department of Engineering, University of Vic-Central University of Catalonia, Vic, Spain, [7]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

KEYWORDS

graph learning, brain networks, deep learning, brain imaging, graph neural networks (GNN), multimodality

Editorial on the Research Topic
Graph learning for brain imaging

Unprecedented collections of large-scale brain imaging data, such as MRI, PET, fMRI, M/EEG, DTI, etc. provide a unique opportunity to deepen our understanding of the brain working mechanisms, improve prognostic predictions for mental disorders, and tailor personalized treatment plans for brain diseases. Recent advances in machine learning and large-scale brain imaging data collection, storage, and sharing lead to a series of novel interdisciplinary approaches in the fields of computational neuroscience, signal processing, deep learning, brain imaging, cognitive science, and computational psychiatry, among which graph learning provides a valuable means to address important questions in brain imaging.

Graph learning refers to designing effective machine learning and deep learning methods to extract important information from graphs or exploiting the graph structure in the data to guide knowledge discovery. Given the complex data structure in different imaging modalities as well as the networked organizational structure of the human brain, novel learning methods based on graphs inferred from imaging data, graph regularizations for the data, and graph embedding of the recorded data, have shown great promise in modeling the interactions of multiple brain regions, information fusion among networks derived from different brain imaging modalities, latent space modeling of the high dimensional brain networks, and quantifying topological neurobiomarkers. This Research Topic synergizes the state-of-the-art discoveries in terms of new computational brain imaging models and insights into brain mechanisms through the lens of brain networks and graph learning.

We accepted 10 manuscripts recommended by the reviewers after evaluating the novelty and quality of the contributions. In order to introduce these works in more detail,

we highlight three domains in this Editorial that emerge from the 10 contributions to this Research Topic.

(1) *Leveraging graph theory and network analysis to identify the biomarkers of brain disorders.* Specifically, Cui et al. used graph theory analysis based on fMRI to investigate alterations of brain functional networks in profound bilateral congenital sensorineural hearing loss (SNHL) in infants, and this study also provided novel insights into functional network alterations in the early stage of profound bilateral congenital SNHL. Zhu et al. explored the aberrant functional connectivity of sensory motor networks in BD-I (bipolar disorder type I) patients and its associations with executive dysfunction. The authors found a significant relationship between the abnormal intranetwork and internetwork functional connectivity values, clinical symptoms and executive function, which provides new information for exploring the neural physiopathology of executive dysfunction in BD-I patients. Chen Y. et al. proposed an invertible dynamic Graph Convolutional Network (GCN) model to identify Autism Spectrum Disorder (ASD) and investigate the alterations of connectivity patterns associated with the disorder. Their proposed method achieves superior classification performance, which provides an interpretable deep learning model for brain connectivity analysis and is of great potential in studying other brain-related disorders.

(2) *Using new machine learning frameworks to understand the functional and structural brain maps, and an integration of both functional and structural brain networks.* In this category, Jon Albers et al. presented a novel approach for quantifying the relationship between brain function and structure and the integration of these in terms of processing units. Their proposed framework naturally can be extended to a general multimodal modeling framework. Eschenburg et al. proposed a cortical segmentation method that, given resting-state connectivity features readily computed during conventional MRI pre-processing and a set of corresponding training labels, can generate cortical parcellations for new MRI data. They found that, in all cases, graph neural networks consistently and significantly outperformed a baseline neural network. Qiu et al. proposed an individualized cortical parcellation based on graph neural networks to learn the reliable functional characteristics of each brain parcel on a large fMRI dataset and to infer the areal probability of each vertex on unseen subjects. This study provides new avenues for precise mapping of cortical areas onto individual brains, and shows potential applications in locating personalized functional areas in the diagnosis and treatment of neurological disorders.

(3) *Methodology oriented papers for data augmentation, multimodal fusion, and graph signal processing.* For example, Zhang et al. proposed a novel approach to generate a fused cognitive network with the optimal performance in discriminating cognitive states by using graph learning. Their findings suggest that the fused cognitive network provides the potential to develop new mind decoding approaches. Chen X. et al. proposed to use a data augmentation method by adding artificial samples generated using graph empirical mode decomposition, which can improve the average classification performance. Furthermore, their augmentation method can be extended to other similar small datasets. Jiao et al. proposed to use the low-frequency components to approximate the extended source activation after graph Fourier transform (GFT) and built a bidirectional long-short term memory (BiLSTM) neural network to solve the Electrophysiological source imaging problem. Chan et al. proposed a new framework called Joining Omics and Imaging Networks *via* Graph Convolutional Layers and Attention (JOIN-GCLA), which consists of multiple graph convolution layers and an attention mechanism to combine multi-modal imaging data and multi-omics datasets for the prediction of PD. The JOIN-GCLA architecture makes it possible to analyze multi-modal imaging data along with multi-omics datasets.

These collected articles have made outstanding contributions to the field of brain science and brain imaging. The research can make a broader impact on the brain disorder diagnostic and prognostic analysis by using network theory, deep learning, and graph signal processing.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers in Neuroscience

METHODS
published: 24 December 2021
doi: 10.3389/fnins.2021.797500

Learning Cortical Parcellations Using Graph Neural Networks

Kristian M. Eschenburg[1,2], *Thomas J. Grabowski*[2,3,4] *and David R. Haynor*[1,2,3]*

[1] Department of Bioengineering, University of Washington, Seattle, WA, United States, [2] Integrated Brain Imaging Center, University of Washington Medical Center, Seattle, WA, United States, [3] Department of Radiology, University of Washington Medical Center, Seattle, WA, United States, [4] Department of Neurology, University of Washington Medical Center, Seattle, WA, United States

Deep learning has been applied to magnetic resonance imaging (MRI) for a variety of purposes, ranging from the acceleration of image acquisition and image denoising to tissue segmentation and disease diagnosis. Convolutional neural networks have been particularly useful for analyzing MRI data due to the regularly sampled spatial and temporal nature of the data. However, advances in the field of brain imaging have led to network- and surface-based analyses that are often better represented in the graph domain. In this analysis, we propose a general purpose cortical segmentation method that, given resting-state connectivity features readily computed during conventional MRI pre-processing and a set of corresponding training labels, can generate cortical parcellations for new MRI data. We applied recent advances in the field of graph neural networks to the problem of cortical surface segmentation, using resting-state connectivity to learn discrete maps of the human neocortex. We found that graph neural networks accurately learn low-dimensional representations of functional brain connectivity that can be naturally extended to map the cortices of new datasets. After optimizing over algorithm type, network architecture, and training features, our approach yielded mean classification accuracies of 79.91% relative to a previously published parcellation. We describe how some hyperparameter choices including training and testing data duration, network architecture, and algorithm choice affect model performance.

Keywords: graph neural network, parcellation, functional connectivity, representation learning, segmentation, brain, human

OPEN ACCESS

Edited by:
Jordi Solé-Casals, Universitat de Vic - Universitat Central de Catalunya, Spain

Reviewed by:
Sun Zhe, RIKEN, Japan
Shijie Zhao, Northwestern Polytechnical University, China

***Correspondence:**
David R. Haynor
haynor@uw.edu

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *18 October 2021*
Accepted: *03 December 2021*
Published: *24 December 2021*

Citation:
Eschenburg KM, Grabowski TJ and Haynor DR (2021) Learning Cortical Parcellations Using Graph Neural Networks. Front. Neurosci. 15:797500. doi: 10.3389/fnins.2021.797500

1. INTRODUCTION

Neural network approaches such as multi-layer feed-forward networks have been applied to a wide variety of tasks in medical imaging, ranging from disease classification to tissue segmentation. However, these networks do not always take into account the true spatial relationships between data points. Convolutional neural network approaches, such as those applied to static images or dynamic video streams, learn translationally-invariant, multidimensional kernel filters over the data domain. Both these methods assume that the data is sampled regularly in space, allowing convolution and pooling of information from fixed neighborhood topologies. However, real-world data, such as graph-structured data, is often sampled on irregular domains. Data sampled from graph domains often contains non-uniform topology—individual data points can vary in their neighborhood structure, and notions of direction (e.g., up, down, left, right) do not generalize

well to graphs. This makes learning filters to process graph-structured data very difficult with conventional neural network approaches.

Graph neural networks are a class of neural network models that operate on data distributed over a graph domain. Data are sampled from a graph with an explicit structure defined by a set of nodes and edges. These models have been shown to be useful for graph and node classification tasks, along with learning generative models of data distributed over graphs (Kipf and Welling, 2016b; Hamilton et al., 2017; Zhao et al., 2019; Zeng et al., 2020). Graph convolution networks (GCN), proposed in Defferrard et al. (2016), generalized the idea of convolutional networks on grid-like data to data distributed over irregular domains by applying Chebyshev polynomial approximations of spectral filters to graph data. Graph attention networks (GAT) are based on the idea of an attention function, a learned global function that selectively aggregates information across node neighborhoods. The attention function maps a query and set of key-value pairs to an output (Vaswani et al., 2017). The output is defined as a weighted sum of the values, where weights are computed using some similarity or kernel function of the key-value pairs.

It is believed that biological signals distributed over the cortical manifold are locally stationary. Given a small cortical patch, voxels sampled from the patch will display similar functional and structural connectivity patterns, cortical thickness and myelin density measures, and gene expression profiles, among various other signals (Glasser and van Essen, 2011; Amunts et al., 2020; Wagstyl et al., 2020). Prior studies have attempted to delineate and map the cortex by identifying contiguous cortical subregions that are characterized by relative uniformity of these signals (Blumensath et al., 2013; Arslan et al., 2015; Baldassano et al., 2015; Gordon et al., 2016). This work is based on the fundamental idea that contiguous regions of the cortex with similar connectivity and histological properties will tend to function as coherent units. Biological signals distributed over the cortex exhibit local but not global stationarity, so any attempt to parcellate the cortex must take both properties into account.

Most brain imaging studies utilize cortical atlases—template maps of the cortex that can be deformed and mapped to individual subjects' brains—to discretize the cortical manifold and simplify downstream analyses (Fischl et al., 2004; Bullmore and Sporns, 2012). However, it remains an open question how to "apply" existing cortical maps to unmapped data. A recent study identified considerable variability in the size, topological organization, and existence of cortical areas defined by functional connectivity across individuals, raising the question of how best to utilize the biological properties of any given unmapped dataset to drive the application of a cortical atlas to this new data (Glasser et al., 2016).

Here, we developed an approach to perform cortical segmentation—a node classification problem—using graph neural networks. The cerebral cortex is often represented as a folded sheet, and a usable parcellation approach must be applicable to this sort of data. Neural networks can be extended to account for non-stationarity in MRI volumes by incorporating 3D-volumetric convolution kernels. However, these approaches are not easily applied to data distributed over 2-D manifolds like the cortical surface. Additionally, more recent large-scale studies interpolate neurological signals, like cortical activation patterns or various histological scalar measures, onto the cortical manifold to mitigate the potential for mixing signals from anatomically close yet geodesically distant cortical regions, e.g., across sulci (Yeo et al., 2011; Glasser et al., 2013). These studies could also benefit from methods that operate directly on graphs.

With the growth of large-scale open-source brain imaging databases [ADNI (Petersen et al., 2010), ABCD (Hagler et al., 2019), HCP (Glasser et al., 2013)], neuroscientists now have access to high-quality data that can be used for training models that can then be applied to new datasets. We leveraged the statistical properties of these high-quality datasets to inform the segmentation of new data using multiple variants of graph neural networks. We considered graph convolution networks and two variants of graph attention networks: standard attention networks (Velickovic et al., 2018), and attention networks with adaptive network depth weighting (a.k.a. jumping-knowledge networks, Xu et al., 2018). We examined how algorithm choice and network parameterization affect cortical segmentation performance. We trained our classification models on high-quality open-source imaging data, and tested them on two datasets with unique spatial and temporal resolutions and different pre-processing pipelines. Other methods have been proposed for delineating the cortex using various registration (Fischl et al., 2004; Robinson et al., 2018), neural network (Hacker et al., 2013; Glasser et al., 2016), label fusion (Asman and Landman, 2012, 2014; Liu et al., 2016), and even graph neural network approaches (Cucurull et al., 2018; Gopinath et al., 2019). To the best of our knowledge, this is the first attempt to examine the performance of common variants of graph neural networks in a whole-brain cortical classification setting and explore their ability to generalize to new datasets using functional magnetic resonance imaging (fMRI). While other studies have proposed the use of graph neural networks to delineate cortical areas, these studies did not perform in-depth analyses on how network architecture, algorithm parameter choices, feature type, and training and testing data parameters impact the predicted cortical maps (Cucurull et al., 2018; Gopinath et al., 2019). To this end, we studied how each of these different variables impacts model performance and prediction reliability.

2. BACKGROUND

2.1. Graph Convolution Networks

Convolution filters over graphs using spectral graph theory were introduced by Defferrard et al. (2016). For a graph $G = (V, E)$ with N nodes and symmetric normalized graph Laplacian, L, define the eigendecomposition of $L = U\Lambda U^T$, where the columns of U are the spectral eigenfunctions of G. Given a graph signal $x \in \mathbb{R}^N$ distributed over G, the graph Fourier transform of x is defined as $\tilde{x} = U^T x$, and its inverse graph Fourier transform as $x = U\tilde{x}$. Graph filtering of x is then defined as $g_\theta(L)x = Ug_\theta(\Lambda)U^T x$, where g_θ is an arbitrary function of the eigenvalues.

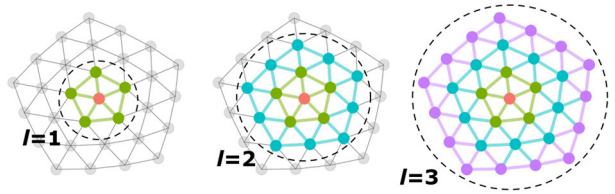

FIGURE 1 | Each layer, *l*, implicitly aggregates more distant neighborhood signals into a node update. The first layer aggregates information over immediately adjacent neighbors, while the second, third, etc. layers incorporate signals from increasingly larger neighborhoods.

Because these filters are not localized in space, Defferrard et al. (2016) proposed to use a Chebyshev polynomial approximation to learn spatially localized filters directly from the Laplacian, reducing the filtering operation of a x to

$$g_\theta(L)x = \sum_{k=0}^{K-1} \theta_k T_k(L)x \tag{1}$$

where $T_k(L)$ and is the k-th polynomial and θ_k the k-th learnable Chebyshev coefficient. The polynomial order, K, determines the local spatial extent of the filter. If two nodes i and j are more than K hops apart, the filter value $g_\theta(L)_{i,j} = 0$.

In Kipf and Welling (2016a), the polynomial order is set to $K = 1$ so that the spatial extent of the filter is limited to directly adjacent nodes and only one coefficient weight is learned per feature component in each layer of the network. Given $H^l \in \mathbb{R}^{N\times k_l}$, the input feature matrix for layer l, the model learns k_l Chebyshev coefficients, in addition to any additional mixing weights. The model incorporates signals from the l-ring neighborhood into the update of a node—each layer implicitly aggregates over a larger neighborhood than the previous layer (**Figure 1**).

2.2. Graph Attention Networks

Whereas graph convolution networks uniformly aggregate local neighborhood signals, attention networks learn optimized weights for each node neighbor using an attention mechanism. Assume we have data distributed over a graph with N nodes. Inputs to the network are characterized by matrix $X \in \mathbb{R}^{N\times F}$, where F is the number of features. Assume that at any given layer, the inputs to layer l are represented as $H^l \in \mathbb{R}^{N\times k_l}$, where $H^0 = X$. We define the immediate neighborhood of node i as $\mathcal{N}_i$. For two vectors $\vec{n}, \vec{p} \in R^k$, we define their feature-wise concatenation as $n||p \in R^{2k}$. In Velickovic et al. (2018), the attention paid by node i to node $j \in \mathcal{N}_i$ at layer l is computed using a single-layer perceptron as

$$\alpha_{i,j} = \sigma(\vec{a}^T(\mathbf{W}^l\vec{h}_i^l||\mathbf{W}^l\vec{h}_j^l)) \tag{2}$$

where σ is a fixed non-linearity, $\mathbf{W}^l \in \mathbb{R}^{k_{l+1}\times k_l}$ is a learned layer-specific global linear projection matrix and $\vec{a}$, the attention function, is also learned. The attention weights for $j \in \mathcal{N}_i$ are then normalized by a softmax operation. To update the features of node i at the $(l+1)$-st layer, we compute the weighted sum over the neighborhood $\mathcal{N}_i$ with weights defined by the normalized attentions.

Velickovic et al. (2018) propose an ensemble ("multi-head") attention mechanism, such that, for each layer, M different attention functions are learned, each with their own weight vector $\vec{a}_m^l$. The outputs of each attention head are concatenated feature-wise. In the last layer, the number of hidden channels is the number of output classes, C—rather than concatenating across attention heads, the outputs of all attention heads are averaged to generate the final network output.

2.3. Jumping-Knowledge Networks

While graph neural networks have been instrumental in applying principles of deep learning to graph-structured domains, they are not without pitfalls (Kipf and Welling, 2016a; Velickovic et al., 2018; Xu et al., 2018; Wang et al., 2019). Graph neural networks are prone to over-fitting of model parameters and over-smoothing of learned embeddings as network depth increases (Wang et al., 2019). One approach to alleviate this over-smoothing is to adaptively learn optimized network depths for each node in the graph, a method (Xu et al., 2018) describe as "jumping-knowledge networks."

Suppose we have a network with L layers, such that the l-th layer embedding h_i^l for node i is learned by incorporating signals from up to l hops away from node i. The layer aggregation function described by Xu et al. (2018) learns a unique output embedding by optimally combining the embeddings of each hidden layer as

$$y_i = \sigma(g(h_i^1, h_i^2, \ldots, h_i^L)) \tag{3}$$

Xu et al. (2018) propose three permutation-invariant aggregation functions for $g(x)$: concatenation, max-pooling, and long-short term memory (LSTM) (Hochreiter and Schmidhuber, 1997). The output, y, is then passed through a linear feed-forward layer to generate the network probabilities. Concatenation is a *global* aggregator (i.e., the same function is applied to all graph nodes) whereas max-pooling and LSTM both learn node-specific aggregations. Further, by utilizing a bi-directional LSTM layer, jumping-knowledge networks learn layer-specific attention weights *for each node* which can then be interrogated *post-hoc* (**Figure 2**). In this analysis, we incorporated the jumping-knowledge mechanism into an attention network framework and examine cortical segmentation performance using both the LSTM and the concatenation functions.

Given a sequence of samples $x_1, x_2, \ldots x_t$, an LSTM layer maintains a memory of previously observed samples in the sequence in order to learn dependencies between elements. Here, the "sequence" consists of the embeddings learned at each consecutive hidden layer, $h^1, h^2 \ldots h^L$, representing increasingly-abstract representations of functional connectivity. We hypothesized that, because the jumping knowledge networks learn optimized node-specific network depths, these networks would be able to more-accurately segment the cortex of new data.

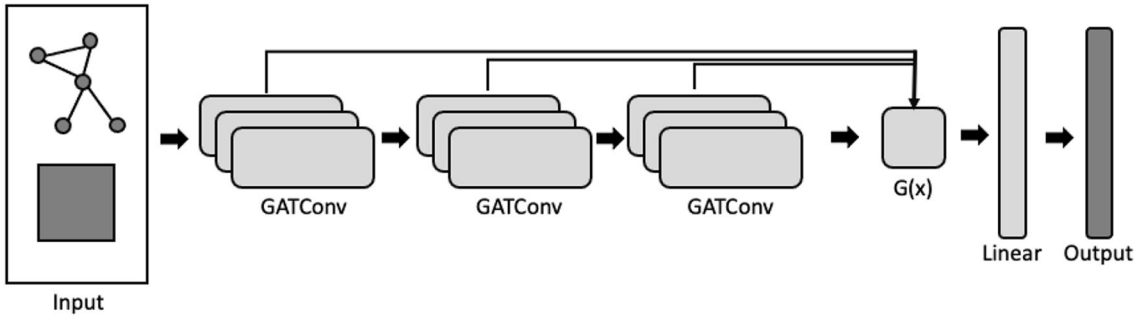

FIGURE 2 | Graph attention network employing a jumping-knowledge mechanism. The network takes as input the graph adjacency structure and the nodewise feature matrix, and outputs a node-by-label logit matrix. Each GATConv block is composed of multiple attention heads. Arrows indicate the direction of processing. Aggregation function, $g(x)$, which takes as input the embeddings from each GATConv block, learns a convex combination of the layer-wise embeddings.

3. DATA

The data used in this study come from the Human Connectome Project (HCP) (Glasser et al., 2013, 2016) and from the Midnight Scan Club (MSC) (Gordon et al., 2017). We were specifically interested in examining how models trained on one dataset would perform on another dataset. Specifically, we trained models on data from the HCP (Glasser et al., 2013), one of the highest quality MRI datasets to date in terms of spatial and temporal sampling of brain signals. We then tested our models on images from both the HCP and MSC datasets.

3.1. HCP Dataset

The HCP consortium collected data on a set of 1,200 young adult subjects 21–35 years of age. We utilized a subset of 268 of these datasets (22–35 years; 153 female) from the S500 data release. The HCP acquired high-resolution 0.7 mm isotropic T1w (TI = 1,000 ms, TR = 2,400 ms, TE = 2.14 ms, FA = 8°, FOV = 224 mm, matrix = 320, 256 saggital slices) and T2w images (TR = 3,200 ms, TE = 565 ms, FOV = 224 mm, matrix = 320). T1w and T2w data were pre-processed using a custom pipeline developed by the HCP (Glasser et al., 2013) using FreeSurfer (Fischl et al., 2004) to generate highly refined cortical surface meshes at the white/gray and pial/CSF interfaces. The surface meshes were spatially normalized to Montreal Neurological Institute (MNI) space and resampled to have 32k vertices. The pipeline also generated four surface-based scalar maps: cortical thickness, Gaussian curvature along the cortical manifold, sulcal depth of the cortical gyri and sulci, and a myelin density map characterizing the spatially-varying myelin content of the gray matter (Glasser and van Essen, 2011).

For each subject, the HCP acquired four resting-state functional MRI (rs-fMRI) images: TR = 0.720 s, TE = 33 ms, multi-band factor = 8, FA = 52°, FOV = 208 × 180 mm, Matrix = 104 × 90 × 72, voxel size: 2 × 2 × 2 mm. The authors refer to these four acquisitions as: `REST1_LR`, `REST1_RL`, `REST2_LR`, `REST2_RL`. The images were acquired over two separate days, such that `REST1_LR` / `REST1_RL` were acquired on 1 day, and `REST2_LR` / `REST2_RL` were acquired on another. Each session acquired 1,200 time-points, such that each BOLD session was roughly 15-min in length. These images were pre-processed using a custom pipeline developed by the HCP (Glasser et al., 2013). BOLD images were denoised using subject-ICA (Beckmann et al., 2005) and FIX (Salimi-Khorshidi et al., 2014) to automatically identify and remove spurious noise components, and motion parameters were regressed out. No additional global signal regression, tissue regression, temporal filtering, or motion scrubbing were performed. Denoised voxel time series were interpolated onto the fsaverage_LR32k surface mesh using a barycentric averaging algorithm, and then smoothed at FWHM = 2 mm to avoid the mixing of signals across gyri. Surface-mapped BOLD signals were brought into register across subjects using a multi-modal surface matching algorithm (Robinson et al., 2014) to the fsaverage_LR32 space and vectorized to CIFTI format, mapping each surface vertex to an index in a vector (toward the end of this work, we learned that different HCP data releases were processed using different versions of this surface registration algorithm; we discuss this in more depth in section 5.5). CIFTI vector indices, referred to as "grayordinates" by the HCP, are in spatial correspondence across subjects (i.e., index i in subjects s and t correspond to roughly the same anatomical location), such that each subject shares the same mesh topology and adjacency structure. Time-series for each session were demeaned and temporally concatenated.

The HCP consortium developed a pipeline to generate high-resolution multi-modal cortical parcellations (MMP) with 180 cortical areas using a spatial derivative based algorithm (Glasser et al., 2016) computed from resting and task-based fMRI signals, cortical thickness, myelin content, and cortical curvature. Manual editing was performed on the group-average gradient-based parcellation to ensure that boundaries conformed across feature types. Using a set of 210 independent subjects as training data, the authors trained a 3-layer neural network model to learn these boundary-based regions. The authors trained 180 classifiers, one for each cortical area, to distinguish a single cortical area from its immediately adjacent neighborhood (using a 30 mm radius neighborhood size) in a binary classification setting. At test time, the authors compared the probabilities of the predicted areal class across all classifiers in a single find-the-biggest operation. Label predictions were regularized to minimize

spurious predictions and "holes" in the final parcellation. Apart from the 30 mm radius around each group-level area, the classifiers did not incorporate any spatial information at training or test time. Predictions generated from subjects in the training set were used to compute a group-average multi-modal parcellation which can be freely downloaded here: https://balsa.wustl.edu/DLabel/show/nn6K. The individual parcellations and the classifier itself have not yet been publically released.

We utilized the subject-level cortical parcellations generated by the HCP as the training set for our models. Subject-level parcellations for a subset of 449 subjects were made available by an HCP investigator (see Acknowledgements).

3.2. Midnight Scan Club Dataset

The Midnight Scan Club dataset consists of MRI data acquired on ten individual subjects (5 female) ranging in age from 24 to 34 years of age: https://openneuro.org/datasets/ds000224/versions/1.0.3 (Gordon et al., 2017). The MCP study acquired 5 h of resting-state data on each participant in ten 30-min acquisitions, with the goal being to develop high-precision, individual-specific functional connectomes to yield deeper insight into the reproducibility and inter-subject differences in functional connectivity.

The MSC dataset preprocessing followed a roughly similar pipeline to that of the HCP dataset. Four 0.8 mm isotropic T1w images (TI = 1,000 ms, TR = 2,400 ms, TE = 3.74 ms, FA = 8°, matrix = 224, saggital) and four 0.8 mm isotropic T2w images (TR = 3,200 ms, TE = 479 ms, matrix = 224 slices, saggital) were acquired. T1w images were processed using FreeSurfer to generate refined cortical mesh representations of the white/gray and pial/CSF tissue interfaces, which were subsequently warped to the fsaverage_LR brain surface using the FreeSurfer shape-based spherical registration method, and resampled to 164K and 32k vertex resolutions. The authors performed myelin mapping by computing the volumetric T1/T2 ratio and interpolating the voxel-wise myelin densities onto the 32k surface mesh.

MSC resting-state data were acquired using gradient-echo EPI sequences with the following parameters: TR = 2.2 s, TE = 27 ms, FA = 90°, voxel size = 4 × 4 × 4 mm. The MSC applied slice timing correction, and distortion correction using subject-specific mean field maps. Images were demeaned and detrended, and global, ventricular, and white matter signals were regressed out. Images were interpolated using least squares spectral estimation and band-pass filtered (0.009 Hz $<$ f $<$ 0.08 Hz), and then scrubbed of high-motion volumes. Denoised volumetric resting-state data were then interpolated onto the midthickness 32k vertex mesh. The MSC study did not perform subject-ICA and FIX to remove spurious noise components from the temporal signals.

4. METHODS

Here, we describe processing steps applied to the HCP and MSC fMRI datasets for this analysis. We begin with the minimally pre-processed BOLD and scalar data interpolated onto the 32k surface mesh.

4.1. Regional Functional Connectivity

As mentioned above in sections 3.1 and 3.2, the MSC and HCP studies aligned cortical surfaces to the fsaverage_LR surface space. The result is such that, given two meshes S and T, the anatomical location of grayordinate i in mesh S corresponds to generally the same anatomical location as grayordinate i in mesh T, allowing for direct comparisons between the same grayordinates across individual surfaces.

In cases where spatial normalization of surfaces has not been performed, it would be incorrect to assume that two grayordinate indices correspond to the same anatomical locations across subjects. In order to alleviate the requirement of explicit vertex-wise correspondence across training, validation, and testing datasets, we assume that most imaging studies will first run FreeSurfer to generate subject-specific folding-based cortical parcellations (Desikan et al., 2006; Destrieux et al., 2010). We can then aggregate the high-dimensional vertex-wise connectivity features over one of these cortical atlases, as in Eschenburg et al. (2018), and simultaneously reduce the feature vector dimension. This guarantees that column indices of feature vectors represent anatomically comparable variables across individuals corresponding to connectivity to whole cortical areas rather than explicit vertex-vertex connections. These low-dimensional vectors are agnostic to the original mesh resolution and degree of spatial normalization. As long as resting-state data are collected for a given study, and that good spatial correspondence between the T1w and BOLD image can be achieved, we can apply our processing steps to this data.

Given a BOLD time series matrix $T \in \mathbb{R}^{32k \times t}$ and cortical atlas with k regions, we consider the set of vertices assigned to region k and compute the mean time-series of region k as:

$$\hat{T}_{k,t} = \frac{1}{|k|} \sum_{i \in k} T_{i,t} \tag{4}$$

where $\hat{T} \in \mathbb{R}^{K \times t}$ is the matrix of mean regional time-series. We compute $R \in \mathbb{R}^{32k \times K}$, the Pearson cross-correlation between T and $\hat{T}$, where $R_{i,k}$ represents the temporal correlation between a vertex i and cortical region k. These cross-correlation vectors are used as features to train our models.

In this analysis, we generated connectivity features using the Destrieux atlas (Destrieux et al., 2010) with 75 regions per hemisphere, as it is computed by FreeSurfer and represents a reasonably high-resolution partition of the cortical surface that we hypothesize captures vertex-to-vertex functional variability well. In section 5.5, we show how classification performance depends on which cortical atlas we regionalize over, and on which representation of functional connectivity models are trained on.

We also examined segmentation performance when models were trained on continuous representations of functional connectivity, computed by group-ICA and dual regression. As part of their preprocessing, the HCP applied group-ICA to a set of 1,003 subjects using MELODIC's Incremental Group PCA (MIGP) algorithm to compute group-ICA components of dimensions 15, 25, 50, and 100 (Smith et al., 2014). We dual-regressed these group-level components onto each subject's resting-state data to generate subject-level ICA components.

These subject-level regression coefficients were fed into our models as alternative representations of functional connectivity.

4.2. Markers of Global Spatial Position

We also included measures of position in grayordinate space (global spatial position) as model features (Cucurull et al., 2018; Gopinath et al., 2019). Surface mesh Laplacian eigenvectors represent a spatial variance decomposition of the cortical mesh into orthogonal bases along the cortical manifold. We retained the first three eigenvectors corresponding to eigenvalues $\lambda_1, \lambda_2, \lambda_3$. The eigenfunctions represent an intrinsic coordinate system of the surface that is invariant to rotations and translations of the surface mesh.

The eigendecomposition computes eigenvectors up to a sign flip (that is, the positive/negative direction of an eigenvector is arbitrary), and eigenvector ordering is not guaranteed to be equivalent across individuals. We chose a template subject and flipped (multiplied by -1) and reordered the eigenvectors of all remaining subjects with respect to this template subject via the Hungarian algorithm, to identify the lowest cost vector matching for every template-test pair (here, we minimized the Pearson correlation distance).

4.3. Incorporating a Spatial Prior

The models trained in this analysis represent multi-class classifiers. By default, each vertex considers every label (out of a total of 180 possible labels) as a viable candidate. This approach, however, does not take advantage of the fact that training and testing data are in spatial correspondence with one another. For example, if we know a vertex is likely to be assigned a label in the occipital lobe, we can restrict the set of candidate labels for this vertex to a subset of the possible 180 areas e.g., only those areas in the primary and higher-order visual areas. We restricted the label search space of a test vertex to only those labels with non-zero probabilities in the training set. If a given vertex i is never assigned label k in the training data, we set the estimated network probability of label k for vertex i to 0, such that it is never assigned label k in the test set. We implemented the application of the spatial prior by multiplying the network logits with a binary masking matrix *at test time* (e.g., the prior is not included in the model training phase).

Applying the spatial prior is only feasible if the test image surface mesh has been spatially normalized to the fsaverage_LR32 space. Given that many studies will be interested in performing multi-subject inference over surface-based maps, we believe this is a reasonable assumption to make. We examine classification performance when excluding and including a spatial prior.

4.4. Regional Homogeneity

We examined whether our models learned parcellations in which the features of each parcel were homogenous. We defined homogeneity for a given parcellation as in Gordon et al. (2016). Assume we are given a resting-state fMRI BOLD time series matrix $T \in \mathbb{R}^{32k \times t}$ and precomputed cortical parcellation with L cortical areas. For each parcel $l \in L$ with n_l vertices, we computed the Pearson correlation matrix, $R_l \in \mathbb{R}^{n_l \times 32k}$, between the parcel BOLD signals with the BOLD signals of the entire cortex. We then applied the singular value decomposition as $R = USV^T$, where S is the diagonal matrix of singular values $\sigma_1, \sigma_2 \ldots \sigma_N$. Gordon et al. (2016) defined homogeneity as $\rho_l = 100 * (\sigma_1^2 \Big/ \sum_{i=1}^{k} \sigma_i^2)$, the percent of variance explained by the first principal component. The variance captured by the first component describes how well a single vector explains the functional connectivity profiles of a given cortical parcel—the larger the variance explained, the more homogeneous the parcel connectivity. We computed an estimate of functional homogeneity for each parcel and averaged the estimates across all parcels.

For scalar features (e.g., myelin density), we estimated homogeneity as the ratio of within-parcel variance to between-parcel variance. For each parcel $l \in L$ and feature $F \in \mathbb{R}^{32k}$, we computed the mean, μ_l, and variance, σ_l^2 of the parcel-wise features. Homogeneity is estimated as $\sum_{i=1}^{L}(\sigma_l^2 - \bar{\sigma^2}) \Big/ \sum_{i=1}^{L}(\mu_l - \bar{\mu})^2$, where $\bar{\sigma^2}$ and $\bar{\mu}$ are the average variance and average mean estimates across all parcels. A smaller value represents more homogeneous parcels. This measure of homogeneity is a dimensionless quantity that allows for the comparison of estimates across datasets and features.

4.5. Model Training and Parameter Selection

We implemented each graph neural network model using the Python package Deep Graph Library (DGL) and PyTorch (Wang et al., 2020). Code developed for this analysis for training these models can be found here: https://github.com/kristianeschenburg/parcellearning/.

We split the 268 HCP subjects into 100 training samples, 20 validation samples, and 148 test samples. For parameter optimization, we trained models on three types of datasets: (1) 100 15-min images (`REST1_LR` session for each subject), (2) 100 60-min images (temporal concatenation of all four rfMRI sessions), and (3), 400 15-min images (four independent rfMRI sessions per subject). We used a validation dataset of 20 subjects of the same scanning duration as the training data to determine when to stop training. We examined the performance of each model on test hold-out test set of different scanning durations: 15-min (four independent rfMRI sessions), 30-min (concatenation of two 15-min rfMRI sessions acquired on the same day), and 60-min (temporal concatenation of all four 15-min rfMRI sessions). The outcome variable to be predicted was the subject-level parcellation provided to us by MG. We performed similar temporal concatenation of the MSC data, concatenating the original ten 30-min sessions into five 60-min sessions, two 150-min sessions, and one 300-min session.

The features used for parameter optimization were the regionalized functional correlations between each cortical vertex and all regions in the Destrieux atlas, the first three Laplacian eigenvector embeddings capturing global location information, and four scalar maps corresponding to sulcal depth, Gaussian curvature, myelin density, and cortical thickness for a total of 81 features at each vertex. We concatenated these features column-wise into a matrix for each subject.

We refer to the graph convolution network, graph attention network, and jumping knowledge network as "GCN," "GAT," and "JKGAT," respectively. We compared the performance of these algorithm variants to a simple linear feed-forward neural network ("baseline") where only the features at each vertex were used to classify cortical nodes (no adjacency information is incorporated into the learning process). We optimized model performance over network depth, number of hidden channels per layer, feature dropout rate, number of attention heads (GAT and JKGAT only), and aggregation function (JKGAT only). The "default" parameters are 3 layers, a dropout rate of 0.1, 32 hidden channels, 4 attention heads per layer, and an LSTM aggregation function. We varied one parameter at a time: for example, when comparing networks with 3 layers vs. 6 layers, all other parameters are fixed to the default values.

For training, we used the cross entropy loss implemented in Pytorch, a LeakyReLU activation function with a negative slope of 0.2, and Adam optimization with a weight decay rate of 0.0005 and L2 weight regularization of 0.005. We trained in mini-batches of size $s = 10$ graphs and accumulated the gradients for each batch before computing the gradient update. We trained for 1,000 epochs using an early stopping criteria evaluated on the validation loss. At each iteration, we retained the model if the current validation loss was lower than the previous validation loss. If validation loss did not decrease for 150 epochs, training was terminated and the best performing model was saved. In practice, we found that few of the models trained for more than 1,000 epochs.

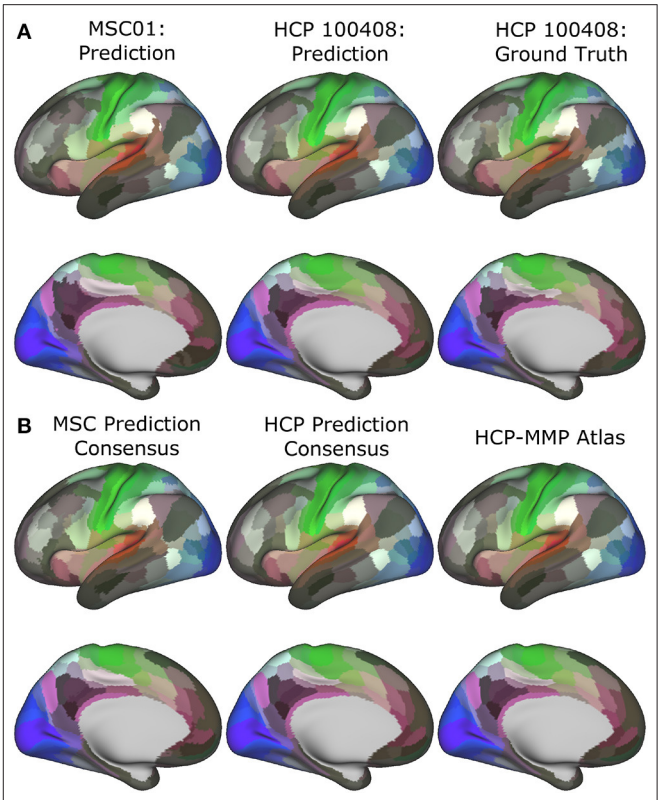

FIGURE 3 | Subject-level **(A)** and group-level **(B)** predictions generated by the optimal model in the MSC (left) and HCP (middle) datasets.

5. RESULTS

We first examine the best performing model of those we considered in our analysis, and discuss the classification accuracy and reproducibility of parcellations predicted by this model in relation to parcellations computed by Glasser et al. (2016), which we call "ground truth" in what follows. We define classification accuracy as the percentage of correctly predicted vertex labels relative to the ground truth maps. We then show broadly how algorithm choice, network architecture, and training and testing image scan duration affect overall model performance. Finally, we illustrate how classification performance is related to the features used during model training and testing.

5.1. Prediction Accuracy in the Best Performing Model

Network optimization was performed using labels provided by Matthew Glasser (see section Acknowledgments) using subject data from the S500 HCP release. As mentioned in section 3.1, the S1200 data release uses a different surface registration algorithm, producing subject-level resting-state data that is better aligned with the labels provided by Glasser. Final model evaluation was performed using this S1200 data. The best performing model was the 6-layer graph attention network (GAT), with 4 attention heads per layer, 32 hidden channels per layer, and a dropout rate of 0.1, and incorporated a spatial prior at test time. When trained on features computed using ICA, this model achieved a mean classification accuracy of 79.91% on the S1200 subjects. We henceforth refer to this model as the "optimal" model, and discuss results associated with this model below.

In **Figure 3A**, we show predicted parcellations computed using this model for exemplar HCP and MSC test subjects. Predicted subject-level parcellations closely resemble the "ground truth" maps generated by Glasser et al. (2016) (see **Supplementary Material** for additional examples of predictions generated by each model). No specific contiguity constraint was imposed on the parcellations; it is inherent in the graph neural network models. Subjects from the MSC dataset do not have corresponding ground truth maps against which to compare their predictions. In **Figure 3B**, we show consensus predictions for each dataset, compared against the publicly released HCP-MMP atlas. Consensus predictions were computed by assigning a vertex to the label most frequently assigned to that vertex across the individual test subject predictions. We see that both consensus predictions closely resemble the HCP-MMP atlas—however, the consensus map derived from the MSC subjects shows noisy parcel boundaries and disconnected areal components (lateral and medial prefrontal areas).

Figure 4 shows the spatial distribution of classification accuracy rates averaged across all subjects in the HCP test set. Average accuracy is shown as a map distributed over the cortex, with values ranging between 0 (blue; vertex incorrectly

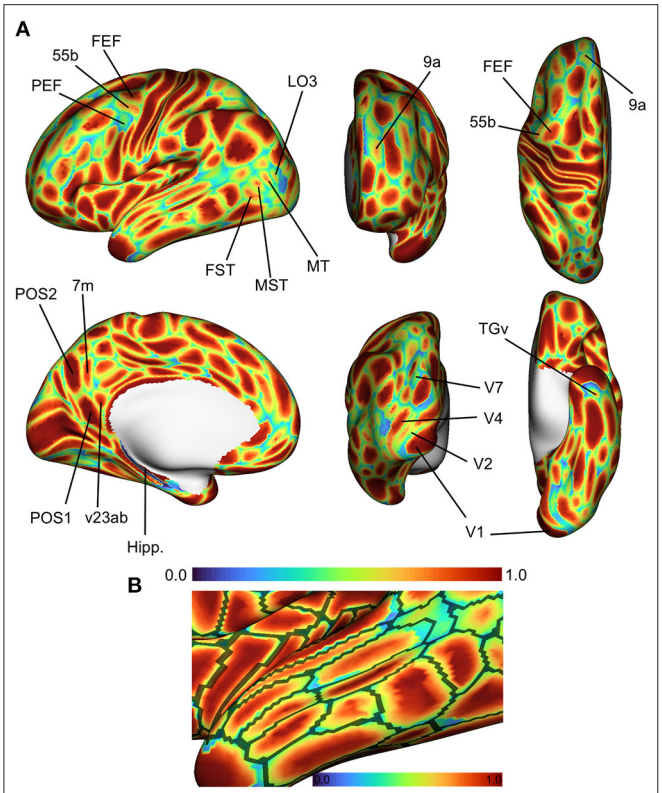

FIGURE 4 | Average accuracy maps for the HCP test set using the optimal model, computed by averaging the classification error maps across all HCP dataset test subjects. **(A)** Blue (0.0) = vertex incorrectly classified in all test subjects; Red (1.0) = vertex correctly classified in all test subjects. Areas in the lateral prefrontal and ventral/dorsal occipital areas showed the highest error rates. **(B)** Errors occur most frequently at the boundaries of cortical regions. Black lines represent areal boundaries of the consensus prediction parcellation.

classified in all subjects) and 1 (red; vertex classified correctly in all subjects). Vertices near the centers of cortical regions were classified correctly more frequently, while prediction errors tended to be distributed near the boundaries of cortical regions. To some degree, this effect can be attributed to the idea that boundaries between putative cortical areas represent segments of the cortex with changing biological properties. In developing a statistical model to assign a vertex to one cortical area or another, vertices at region boundaries will have more ambiguous label assignments simply due to the fact that their feature vectors are sampled from a space with greater distributional overlap across various cortical areas. However, another explanation is that MRI resolution is low with respect to cortical functional features like cell columns. Consequently, this means that voxel-wise measurements reflect mixtures of connectivity patterns due to partial volume effects, thereby reducing the ability of a statistical model to distinguish between two cortical areas at parcel boundaries.

While errors globally tended to be concentrated at region boundaries, some cortical areas showed higher error rates than others. Of note are higher error rates for cortical areas in the superior temporal areas in the fundus and medial superior temporal regions (FST, MST, MT, and V4t), and lateral higher-order visual areas (LO1, LO2, LO3). In the lateral prefrontal area, we found that the premotor eye field (PEF) shows higher error rates relative to adjacent regions (55b and frontal eye field, FEF). Glasser et al. (2016) identified three unique topologies (typical, shifted, and split) for area 55b that varied across subjects, which might to some degree explain the higher error rates in area PEF.

We quantified the relationship between the spatial distribution of errors and their distance to cortical areal boundaries. We computed the fraction of misclassified vertices that occurred at a geodesic distance of k edges (geodesic hops) from any cortical areal boundary. Using the default model parameters and regionalized features, we examined this distribution of errors as function of distance (**Supplementary Material**). Over 50% of misclassified vertices occurred at the region boundaries i.e., those vertices in the ground-truth parcellations that are directly adjacent to different regions, and roughly 30 and 12% of misclassified vertices were 1 and 2 edges away from areal boundaries, respectively. The simple feed-forward network misclassified vertices further away from region boundaries, while the three graph neural networks tended to misclassify only vertices close to the boundary.

Although the MSC subjects do not have corresponding ground truth maps, the data is in spatial correspondence with the fsaverage_LR32 map. We computed the correspondence of maps predicted on the MSC subjects with the HCP-MMP atlas in order to gain insight into the accuracy of these predictions. Mean correspondence of predictions computed on the MSC and HCP datasets with the HCP-MMP atlas was 70.04 and 84.35%, respectively (**Supplementary Material**).

Mean model probabilities computed by the optimal model for a set of cortical areas are illustrated in **Figure 5**, showing that areal probabilities are local in nature and restricted to precise anatomical locations. Individual areal probabilities computed by Glasser et al. (2016) and Coalson et al. (2018) using their binary classifier are shown in the bottom row. Probability estimates in the HCP dataset mirror those estimated by the original HCP classifier (Glasser et al., 2016), indicating that our model faithfully learns the proper spatial extent of each cortical areal. Estimates in the MSC dataset were slightly more diffuse and less confident (see areas V1 and 46), such that probability mass was assigned to more disparate areas of the cortex, relative to probabilities estimated in the HCP dataset.

5.2. Model Predictions Are Reproducible Across Scanning Sessions

The HCP acquired four 15-min resting-state acquisitions per subject, while the MSC acquired ten 30-min resting-state acquisitions per subject. We examined how reliable predictions generated from each resting-state session were within subjects, and how this reliability related to the scanning duration. For a given subject, we estimated session-specific reproducibility using datasets of the same scan duration. We defined reproducibility using the Dice coefficient, which measures the similarity of two

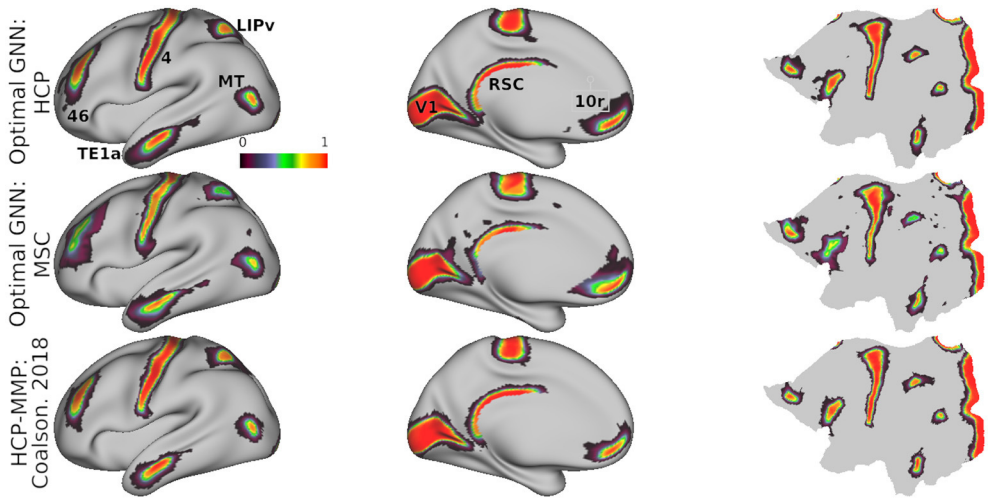

FIGURE 5 | Mean model probabilities for a subset of cortical areas for the HCP **(top)** and MSC **(middle)** datasets computed using the optimal model, and the MMP binary class probabilities from Glasser et al. (2016) and Coalson et al. (2018) **(bottom)**. Probabilistic maps are illustrated for areas V1, 46, TE1a, LIPv, MT, RSC, and 10r. These maps are thresholded at a minimum probability value of 0.005, the probability of randomly assigning a vertex to one of the 180 cortical areas.

images. The Dice coefficient between sets J and K is defined as

$$Dice(J, K) = \frac{2 * |J \cap K|}{|J| + |K|} \tag{5}$$

Figure 6 shows the mean areal Dice coefficients for each dataset from predictions computed using the optimal model. Predictions made on the HCP dataset were more reproducible across the entire cortex than predictions on the MSC dataset. In both datasets, sensory/motor and areas near the angular and supramarginal gyri were most reproducible. The visual cortex showed high reproducibility in area V1, while areas V2-V4 were less reproducible.

Figure 7A, shows mean reproducibility estimates computed on the HCP and MSC datasets. Predictions for both datasets were highly reproducible across repeated scanning sessions, and reproducibility increased with increasing scan duration. Mean Dice coefficient estimates in the HCP dataset were 0.81 and 0.86 for the 15- and 30-min durations. In the MSC dataset, the mean Dice coefficients were 0.69, 0.76, and 0.82 for the 30-, 60-, and 150-min durations. When fixing scan duration (e.g., 30-min durations), HCP data were more reproducible than the MSC data. One feature that we could not evaluate directly was the reproducibility of the ground truth maps. Glasser et al. (2016) reported maximum and median Dice coefficient estimates of 0.75 and 0.72 for repeated scans on HCP participants, indicating that our classifier learned parcellations that were more reproducible than those generated by the binary classifier.

Figure 7B illustrates subject-level reproducibility estimates in the MSC dataset. Predictions for subject MSC08 were significantly less reliable, relative to the other subjects. Gordon et al. (2017) also identified MSC08 as having low reproducibility with respect to various graph theoretical metrics computed from the functional connectivity matrices. They noted that subject MSC08 reported restlessness, displayed considerable head motion, and repeatedly fell asleep during the scanning sessions.

Area-level topologies were also reproducible across scanning sessions (**Supplementary Material**). Glasser et al. (2016) identified three unique topologies of area 55b, corresponding to a "typical," "shifted," or "split" organization pattern, relative to the group-average cortical map. We were able to identify these same unique topologies in individual subjects, indicating that graph neural networks are identifying the unique connectivity fingerprints of each cortical area, and not simply learning where the parcel is. When we examined the predictions generated by the optimal model on the four independent 15-min scanning sessions, we found that, within a given subject, the topological organization of area 55b was reproducible. Allowing for some variability in prediction boundaries and location due to resampling of the connectivity data and partial volume effects, this indicates that the graph neural networks are learning subject-specific topological layouts that incorporate their unique connectivity and histology patterns.

5.3. Parcellations Learned by GNNs Are Homogeneous in Their Scalar and Connectivity Measures

If a model is in fact learning unique, discrete areas, the distribution of biological features in these areas should be relatively homogeneous. Unsupervised learning clustering algorithms designed to parcellate the cortex often incorporate objective functions that attempt to maximize within-parcel similarity and minimize between-parcel similarity. On the other hand, gradient-based approaches, like those proposed in Gordon et al. (2016), Wig et al. (2014), and Schaefer et al. (2018), do not directly maximize an objective function in this manner, but rather identify putative areal boundaries by identifying where biological

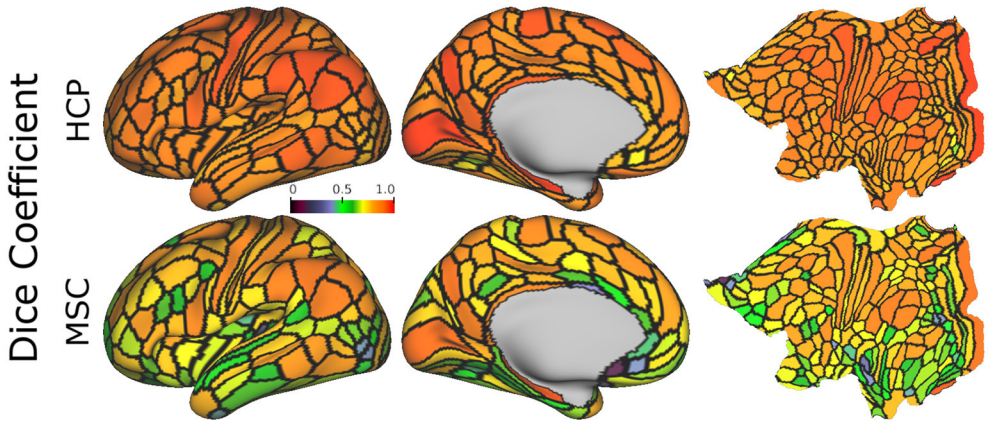

FIGURE 6 | Mean areal Dice coefficient estimates, computed using the optimal model on 15-min HCP data (4 repeated sessions) and 30-min MSC data (10 repeated sessions), normalized with the same color map. Estimates are computed for each area, and averaged across all subjects.

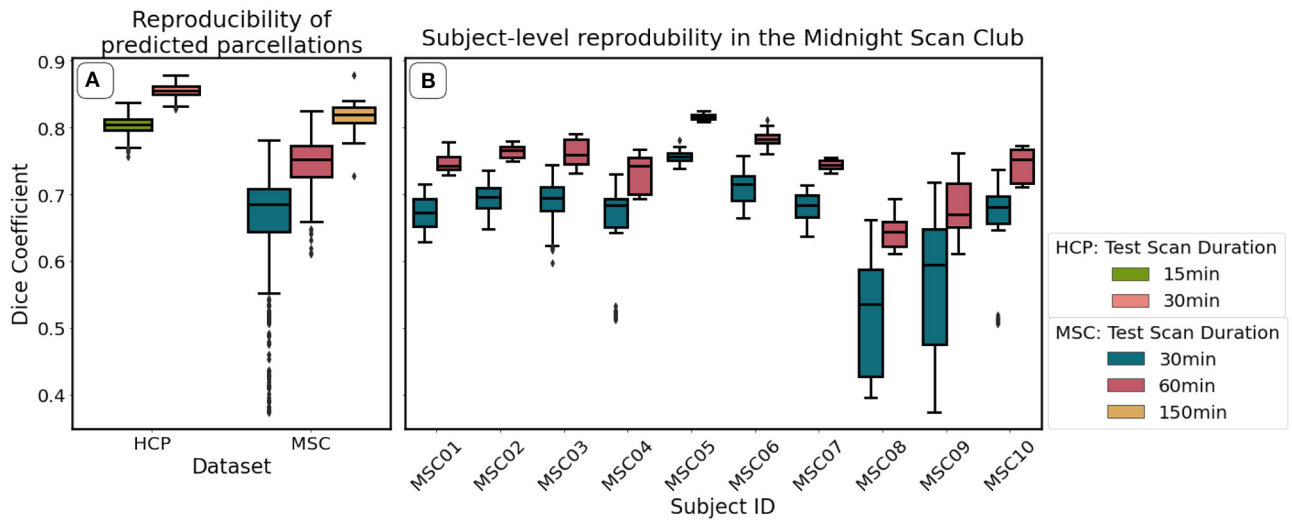

FIGURE 7 | Reproducibility of predicted maps generated by the optimal model, as measured using the Dice coefficient. We show mean reproducibility estimates for each dataset **(A)**, and subject-level estimates in the Midnight Scan Club **(B)**. Estimates for 60 min (HCP) and 300 min (MSC) durations are not shown in **(A)** because there is only one image per subject for these durations. Similarly, estimates for 150 min durations are not shown in **(B)** because there is only a single scalar estimate per subject.

properties change dramatically in a small local neighborhood. It is assumed that this biological gradient captures differences in homogeneity between adjacent cortical areas. In order to group cortical voxels together, these voxels must inherently share some physical or biological traits.

We computed homogeneity estimates as described in section 4.4. In order to compare the homogeneity and variance estimates between predicted parcellations, we fixed the features used to compute these estimates. For a given subject, we computed functional homogeneity using that subject's 60-min BOLD signal (HCP), or the 300-min BOLD signal (MSC). In this way, the only variable that changed with respect to the homogeneity estimate is the cortical map itself. We could then make meaningful quantitative comparisons between estimates for different maps, with respect to a given dataset.

Cortical maps predicted in the HCP dataset explained, on average, 67.03% of the functional variation while MSC predictions explained 72.90% (t: −3.137, p: 0.007) (**Figure 8**). We hypothesized that parcellations predicted in the HCP dataset would be more homogeneous, relative to those learned in the MSC dataset, due to the fact that the MSC imaging data were acquired with lower spatial resolution than that acquired by the HCP and therefore subject to greater partial volume effects. Homogeneity of myelin (t: −0.910, p: 0.377) and sulcal depth (t: 1.043, p: 0.320) was not statistically different between the two datasets, while curvature was less variable in the HCP dataset (t: −2.423, p: 0.029). Contrary to our hypothesis, cortical thickness was less variable in the MSC dataset (t: 11.562, p: 0.000). This is likely a consequence of using a dimensionless representation of homogeneity, which is internally normalized for each dataset

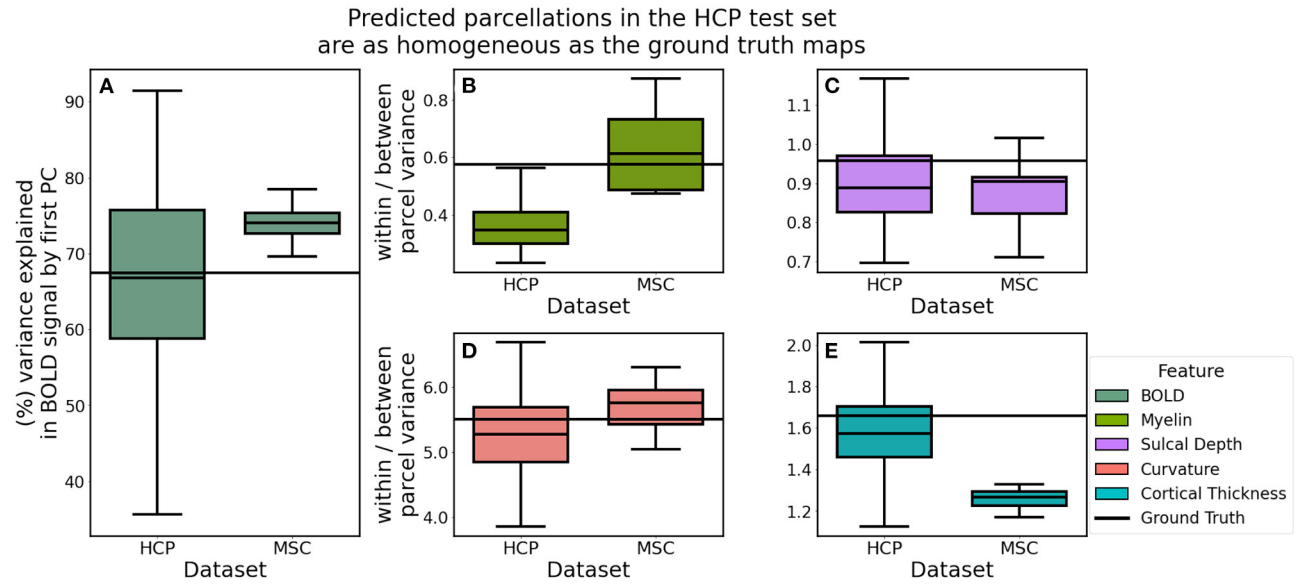

FIGURE 8 | Homogeneity of predicted parcellations in the HCP and MSC datasets using the optimal model. **(A)** Predicted parcels in the HCP test set explained as much variability in the functional connectivity as the ground truth parcels. **(B–E)** Predictions in the MSC had more variable myelin content and less variable cortical thickness estimates, relative to the HCP predictions.

as a ratio of the within-to-between parcel variances. This metric allows for the direct comparison of homogeneity estimates across datasets, instead of representing the raw variance estimates.

We compared homogeneity estimates in the predicated HCP parcellations to estimates computed for the ground truth maps using paired t-tests. Predicted and ground truth maps both explained roughly 67% of the functional variation (t: −0.305, p: 0.761). Myelin (t: 0.176, p: 0.860) and curvature: (t: −1.746, p: 0.083) variation were not statistically different between the two groups. However, predictions were more homogeneous than the ground truth maps with respect to sulcal depth (t: −4.442, p: 0.000) and cortical thickness: (t: −2.553, p: 0.012).

5.4. Network Architecture Impacts Model Performance

As noted in section 5, we first optimized over network algorithms and architectures using the S500 dataset, and then utilized the S1200 dataset for model evaluation. We fixed the features used for network optimization to the regionalized connectivity features. We examined how varying each network parameter impacted model classification accuracy (**Table 1**). As mentioned in section 5.1, the best performing model was the GAT network with 6 layers with a classification accuracy of 67.60% on the S500 dataset (significantly inferior to the performance of the same network on S1200 data, with an accuracy of 79.91%). We found that optimal performance for the GAT and GCN networks was achieved with 6 layers, 9 layers for the JKGAT, and 3 layers for the baseline model. In general, classification accuracy increased with the number of attention heads, and number of hidden channels, while classification accuracy decreased with increasing feature dropout rates. Using an LSTM aggregation function rather than a simple concatenation marginally decreased classification accuracy for the jumping-knowledge networks. In contrast to our predictions, we found that the GAT networks slightly outperformed the more flexible JKGAT networks for most parameterizations.

We used a fixed validation dataset of 20 subjects to determine when to stop model training and evaluated the performance of our models using a fixed test dataset of 148 subjects. In order to determine the reliability of our accuracy estimates, we computed the standard error of classification accuracy for each model using a bootstrapped approach (**Supplementary Material**). We randomly sampled 100 test subjects, with replacement, out of the 148, and computed the mean accuracy for each sample, for each model. We repeated this process 1,000 times, and computed the variability of these bootstrapped estimates. Standard error estimates were less than 0.5%, indicating that test set accuracy estimates are robust with respect to resampling of the test dataset.

We examined how classification accuracy in the HCP dataset was related to the scanning duration of training and testing datasets using the default model parameters (as defined in section 5). When fixing test scan duration, classification accuracy improved as the training dataset size increased for all model types, with maximum accuracy achieved by graph attention network models trained on 400 15-min duration datasets (**Supplementary Material**). When training dataset size and training scan duration were fixed, longer test image duration yielded more accurate predictions across the board. Predictions on 60-min test data were more accurate than those computed on 30-min images, which in turn were more accurate than those generated from 15-min images (**Supplementary Material**). However, models trained on 15-min

TABLE 1 | Model classification accuracy as a function of network architecture and parameterization.

		Model			
Parameter	**Value**	**Baseline (%)**	**GCN (%)**	**GAT (%)**	**JKGAT (%)**
Network depth	[3]	62.64	64.93	67.02	66.71
	6	61.13	65.14	**67.60**	67.33
	9	57.72	64.76	67.36	67.42
Hidden channels	16	60.54	62.60	66.37	66.12
	[32]	62.64	64.93	67.02	66.71
	64	63.84	66.24	67.15	67.15
Dropout rate	[0.1]	62.64	64.93	67.02	66.71
	0.3	60.74	63.94	66.72	66.58
	0.5	58.34	63.10	65.45	65.39
	0.7	55.63	61.18	62.70	62.60
Attention heads	[4]			67.02	66.71
	8			67.39	67.30
	12			67.56	67.29
Aggregation function	concat				66.85
	[lstm]				66.71

Models were trained on 400 15-min datasets, and tested on 60-min test data using the S500 dataset. Boxed values indicate the default parameter values. The best performing model was the GAT network with 6 layers, achieving a mean classification accuracy of 67.60%. Values in bold are the mean classification accuracy of the best model, trained on resting-state connectivity features computed by regionalizing time-series over the Destrieux cortical atlas (see Section 4.1).

data performed best when tested on 15-min data, and models trained on 60-min data performed best when tested on 60-min data (**Supplementary Material**) indicating an interaction between training and testing scan duration. Similarly, when fixing training and testing scan duration, we found that including the spatial prior significantly improved classification accuracy in all architectures.

5.5. Incorporating Functional Connectivity Improves Model Performance Beyond Spatial Location and Scalar Metrics

After identifying the optimal network architecture, we examined how model performance varied as a function of which features the model was trained on. Briefly, we delineated three broad feature types: (1) scalar features corresponding to myelin, cortical thickness, sulcal depth, and cortical curvature (2) global location features corresponding to the spectral coordinates computed from the graph Laplacian and (3) connectivity features computed from the resting-state signal. In our primary analysis, we utilized connectivity features computed by regionalizing over the Destrieux atlas (75 folding-based cortical areas). We compared these features against those computed using the Desikan-Killiany atlas (35 folding-based cortical areas) and the Yeo-17 resting-state network atlas (Yeo et al., 2011). The Yeo-17 atlas is a functional atlas of discretized resting-state networks, computed via independent component analysis. We identified the connected components of each of the 17 resting-state networks and excluded component regions with sizes smaller than 10 vertices, resulting in a map of 55 discrete functionally-derived subregions of the cortex. We also examined the performance of models trained on continuous, overlapping connectivity features representing resting-state networks computed using group-ICA and dual regression.

Computing connectivity features over the Destrieux atlas yielded increased classification accuracy over the Desikan-Killiany atlas (72.01 vs. 70.08%; paired t: 25.197, p: 0.000; see models "Full-DX" and "Full-DK"). We hypothesized that computing connectivity features over a functionally-aware parcellation (Yeo-17) would yield a significant improvement in classification accuracy, relative to the Destrieux atlas, but this was not the case (see "Full-DX" vs. "Full-YEO" in **Figure 9**). Models trained on the Yeo-17 features had a mean classification accuracy of 71.58% (paired t: 1.916, p: 0.057). Training on spatial location or histological features alone yielded mean classification accuracies of 44.10 and 54.45%, respectively (**Figure 9A**). However, training on features defined by resting-state ICA components had clear performance benefits. Models trained on ICA dimensions of 15, 25, 50, and 100 generated mean classification accuracies of 75.34, 77.79, 79.68, and 79.91%, respectively (**Figure 9C**). Similarly, incorporating the prior mask also improved model performance. However, the mask added

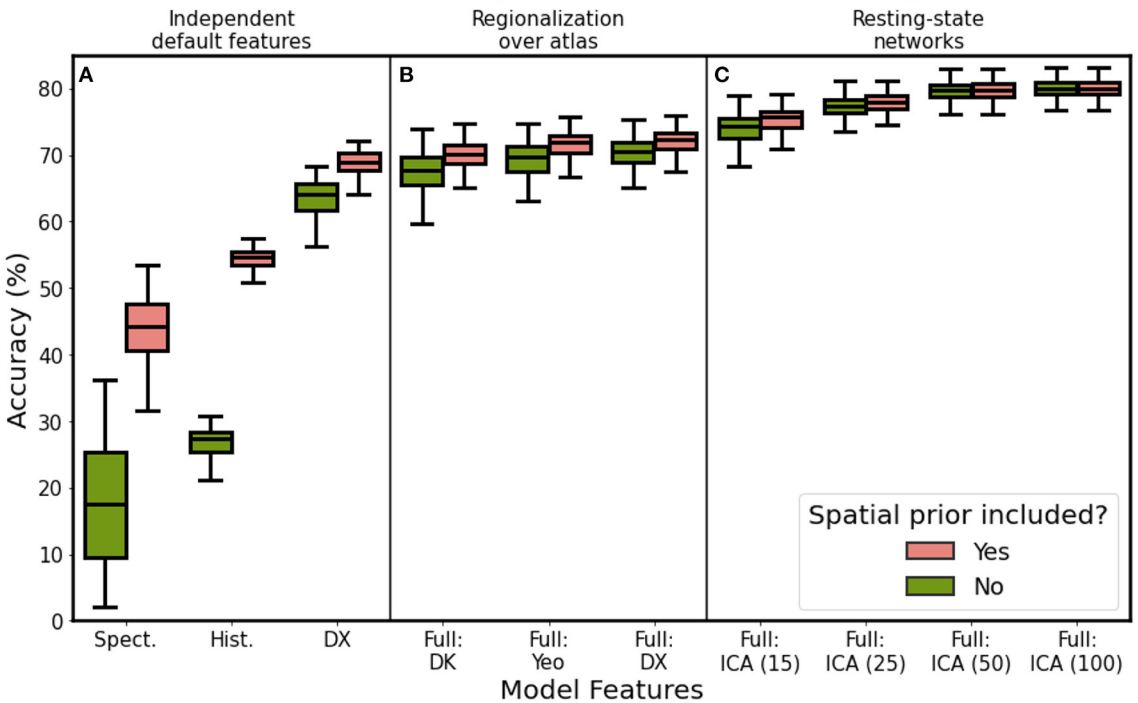

FIGURE 9 | Classification accuracy as a function of model features, using the optimal model architecture for **(A)** single feature types, **(B)** regionalization over different cortical atlases, and **(C)** independent component analysis features. Refer to **Table 2** for a description of each feature set.

TABLE 2 | Feature combinations tested by our optimal model.

	Feature sets						
	Full feature sets				**Connectivity**	**Scalar**	**Location**
	DK (F)	**DX (F)**	**YEO (F)**	**ICA (F)**	**DX**	**Hist.**	**Spect.**
Thickness	+	+	+	+		+	
Curvature	+	+	+	+		+	
Myelin	+	+	+	+		+	
Sulcal depth	+	+	+	+		+	
Laplacian	+	+	+	+			+
Desikan (DK)	+						
Destrieux (DX)		+			+		
Yeo-17 (YEO)			+				
ICA-RSN				+			

Features included in a model are marked by a "+." "Full" models include histological features, global position information, and functional connectivity signals.

diminishing returns, with the better-performing models benefiting less from its inclusion. Models trained on higher-dimensional ICA resting-state networks (50 and 100 networks), performed almost as well without the spatial prior as they did with it.

Late into our analysis, we learned of differences in the preprocessing steps used to generate the minimally-preprocessed HCP resting-state data, and to generate the subject- and group-level HCP-MMP parcellations. Specifically, the S500 and S1200 data releases were preprocessed using different surface registration algorithms: MSMSulc and MSMAll (Robinson et al., 2014, 2018). A consequence of these preprocessing differences is that data from the S1200 release is better aligned with the subject-level labels provided by Glasser. After performing network optimization using the S500 data, we evaluated final model performance on the S1200 dataset. **Figure 10** illustrates model performance after training on each independent dataset. We

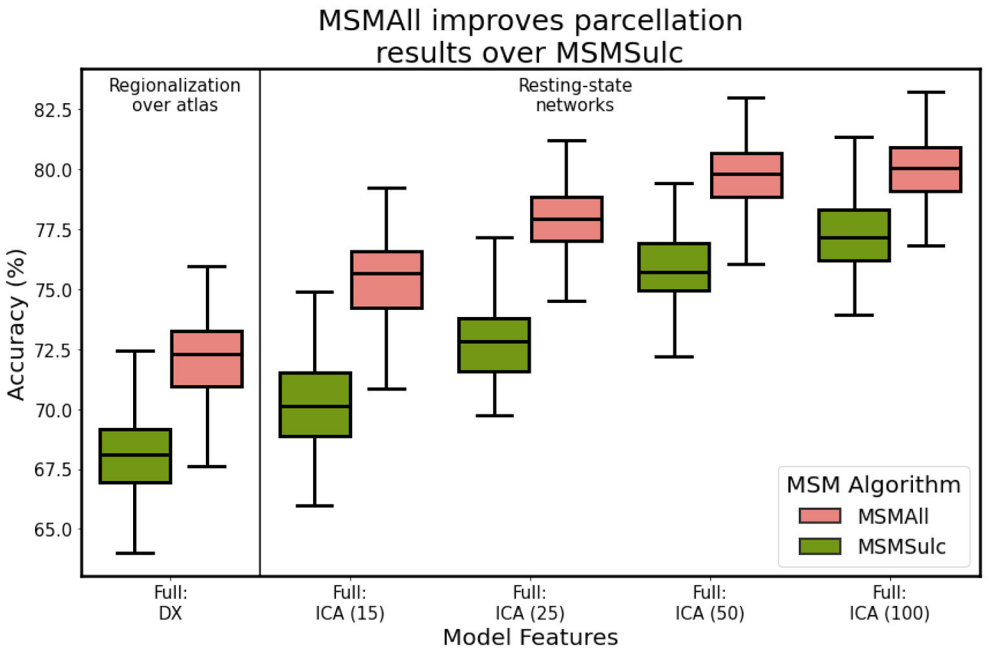

FIGURE 10 | Classification accuracy as a function of HCP data release and corresponding multi-modal surface matching algorithm. S500: MSMSulc (Robinson et al., 2014), S1200: MSMAll (Robinson et al., 2018).

found that utilizing the S1200 dataset showed significant improvements in mean classification accuracy by upwards 5%, relative to the S500 dataset. This indicates that the surface registration algorithm choice plays a critical role in cortical segmentation quality.

6. DISCUSSION

In this analysis, we presented a general cortical segmentation approach that, given functional connectivity information and a set of corresponding training labels, can generate cortical parcellations for individual participants. This approach to segmenting the cortex requires accessible MRI acquisition sequences and standard morphological parcellations as inputs. We compared three different graph neural network variants to a baseline fully-connected network. We found that, in all cases, graph neural networks consistently and significantly outperformed a baseline neural network that excluded adjacency information. We identified the best performing model and explored its performance with respect to various metrics like segmentation accuracy, prediction reliability, and areal homogeneity in two independent datasets.

Predictions generated for both the HCP and MSC datasets were highly reproducible. However, we found that nearly twice as much resting-state data was required in MSC subjects to achieve the same reproducibility estimates as in the HCP data. Predictions generated on the HCP dataset were more reproducible than the ground truth maps themselves (Glasser et al., 2016), while predictions in the MSC data were roughly as reproducible as the ground-truth parcellations. This may in part be due to the way we trained our models. Models were trained on repeated samples of BOLD images, such that for a given training subject, models were shown four BOLD datasets. This likely enabled the models to better learn the mapping between a given subject's unique BOLD signature, and its cortical map. Another possible explanation is that the ground truth maps were generated using a linear perceptron model, which does not take into account any spatial relationships between data points, while graph neural networks do take this spatial structure into account. It is likely the case that the perceptron model could not adapt to utilize spatial dependencies in the BOLD signal in local neighborhoods and thereby failed to fully learn unique subject-specific connectivity fingerprints, and consequently learned more variable parcellations.

The optimal model predicted parcellations that were as homogeneous as the ground truth maps when considering multidimensional connectivity features and univariate scalar features. Though the models considered in this analysis are capable of learning parcels that capture inter-areal variation of functional brain connectivity and other cortical features, it is worth noting that homogeneity as a measure of parcellation quality is an imperfect metric and should be used judiciously. For example, the primary sensory areas can be further divided into five somatotopic subregions corresponding to the upper and lower limbs, trunk, ocular, and face areas (Glasser et al., 2013). These subdivisions correspond well with task-based fMRI activity and gradients in myelin content, indicating that the parcels learned by GNNs in our analysis still incorporate significant variability due to the aggregation of signals from different somatosensory areas. While learning homogeneous regions is important in order to effectively capture spatial biological variation,

maximizing homogeneity was not the training criterion for this analysis.

As noted in section 3, the MSC study applied different preprocessing steps than the HCP. Specifically, the MSC did not perform FIX-ICA to remove noise components from the BOLD images and utilized the FreeSurfer spherical surface registration to bring surfaces into spatial correspondence with one another instead of the multi-modal surface matching algorithm (Robinson et al., 2014, 2018). Given that the MSC dataset did not have "ground truth" labels against which we could compare predictions made on the MSC data, we compared predictions against the HCP-MMP atlas (Glasser et al., 2016). As expected, predictions generated on the HCP dataset more closely resembled the HCP-MMP atlas than predictions made from the MSC dataset (the HCP-MMP atlas was derived as a group-average of individual ground truth parcellations). Nevertheless, we found that correspondence of MSC predictions with the atlas followed similar trends with respect to testing image duration. We believe some discrepancy in results between the HCP and MSC datasets can be attributed to the differences in dataset-specific preprocessing choices noted above, although the relationship between methodological choices and parcellation outcome requires future analyses. Performance differences across the two datasets are also possibly a result of the models learning characteristics inherent to the training (HCP) dataset, and thereby performing better on hold out subjects from that same dataset.

Our optimal model was the 6-layer graph attention network, trained and tested on resting-state network components computed using a 50-dimensional ICA. This model performed as well with the spatial prior as it did without. However, models trained on regionalized connectivity features benefited from including the spatial prior. We believe it would be prudent for future studies to include a spatial prior of some form into their classification frameworks. Interestingly, predictions on HCP test subjects resembled the HCP-MMP atlas more closely than they resembled their ground truth counterparts, which might in part be driven by the specific form of the prior. We made the assumption that cortical map topology is relatively conserved across individuals. This assumption may be too conservative and may reduce model sensitivity to atypical cortical connectivity patterns. Nevertheless, there is evidence our GNN models learn subject-specific topologies of cortical areas, rather than simply learning *where* a cortical parcel usually is. Importantly, we found that the optimal GAT model could identify three unique topologies for area 55b (typical, shifted, and split) and that predictions generated by our model replicated, with high fidelity, the same spatial organization patterns as identified in Glasser et al. (2016). This indicates that the model is capable of learning unique connectivity fingerprints of each cortical area on a subject-by-subject basis, rather than simply learning the group average fingerprint. As such, we do not believe that including the spatial prior in its current form inhibits the ability of the graph neural network models used in this analysis to identify atypical cortical topologies.

We compared three different graph neural networks: graph convolution networks, standard attention networks, and jumping-knowledge networks. We hypothesized that JKGAT networks would significantly outperform GAT networks due to the increased flexibility to learn optimized node-specific network depths. In their original formulation of the jumping-knowledge network architecture, Xu et al. (2018) found that including the jumping-knowledge mechanism improved model performance relative to the GAT in almost all of their comparisons. However, we found this not to be the case. This may be a consequence of the increased number of estimated parameters in the JKGAT networks, relative to the GAT—the jumping-knowledge aggregation layer learns the parameters for the aggregation function cells in addition to the attention head and projection matrix weights learned in the GAT networks. The lower classification accuracy at test time is possibly the result of model over-fitting, necessitating a larger training dataset. It is possible that the jumping-knowledge mechanism is generally more useful in the case where graph topologies vary considerably across a network, as opposed to more regular graphs such as cortical surface data.

As expected, network performance was dependent on both the size and duration of the training set, and duration of the testing data. Classification accuracy increased when models were trained on larger datasets consisting of shorter-duration images. Conversely, accuracy increased when models were deployed on longer-duration test data. It is important to note that we examined performance of our models on images of long scanning durations by concatenating multiple sessions together (30/60-min in the HCP, and 60/150/300-min in the MSC). It is unrealistic to expect study participants to be able to lay in an MRI scanner for single sessions of these lengths. However, it is useful to examine how model performance is impacted by tunable parameters like scan duration in order to best guide image acquisition in future studies. We found that utilizing repeated scans on individual subjects as independent training examples, rather than concatenating repeated scans together into single datasets, significantly improved our classification frameworks. This likely speaks to the ability of neural network models to generalize better to noise in the datasets. Training models on multiple samples of shorter-duration images more accurately captures the individual variability in the resting-state signal than fewer longer-duration images, thereby allowing the networks to more accurately learn a mapping between functional connectivity and cortical areal assignments.

Our methodology could be improved in a variety of ways. We chose not to perform intensive hyperparameter optimization, and instead focused our efforts on overall performance of the various network architectures as a function of network parameters and data parameters, and the applicability of trained models to new datasets. However, in the case where a classification model is meant to be distributed to the research community for open-source use, it would be prudent to perform a more extensive search over the best possible parameter choices.

The utility of functional connectivity has been shown in a variety of studies for delineating cortices (Blumensath et al., 2013; Arslan et al., 2015; Baldassano et al., 2015; Gordon et al., 2016). However, in recent years, using diffusion tractography for learning whole-brain cortical maps has been underutilized, relative to functional connectivity (Gorbach et al., 2011; Parisot et al., 2015; Bajada et al., 2017). Given cortical maps defined

independently by tractography and functional connectivity, it is difficult to "match" cortical areas across maps to compare biological properties, so heuristics are often applied. Few studies have simultaneously combined functional connectivity and tractography to better inform the prediction of cortical maps. Recent work has extended the idea of variational auto-encoders to the case of multi-modal data by training coupled auto-encoders to jointly learn embeddings of multiple data types. In Gala et al. (2021), the authors apply this approach to jointly learn embeddings defined by transcriptomics and electrophysiology that allow them to identify cell clusters with both similar transcriptomic and electrophysiology properties. Future work could apply similar ideas to aggregate functional and diffusion-based connectivity signals.

The majority of recent studies have approached the cortical mapping problem from the perspective of generating new parcellations from underlying neurobiological data using unsupervised clustering or spatial gradient methods. These approaches attempt to delineate areal boundaries by grouping cortical voxels together on the basis of similarity between their features. Spatial gradient-based methods explicitly define areal boundaries, while clustering methods define these boundaries implicitly. However, both approaches are distinct from methods that utilize pre-existing or pre-computed parcellations as templates for mapping *new* data. In the current analysis, we were concerned with the latter problem.

Clustering and spatial gradient approaches are often interested in relating newly-generated cortical maps to underlying *in vitro* measures, such as transcriptomics or cytoarchitectural results. Clearly, it is impossible to acquire this data in human subjects simultaneously with *in vivo* data. Various projects have attempted to build cytoarchitectural datasets from post-mortem subjects to use as a basis of comparisons for maps generated *in vivo* (Amunts et al., 2020). While some cortical areas have been recapitulated using both *in vitro* and *in vivo* features, this is not a general rule across the cortex. As such, cross-modal verification is often difficult, and leaves room for methods and datasets than can improve upon the validation of cortical mapping studies.

One limitation of our analysis concerns the use of different versions of the multi-modal surface matching for cortical surface alignment for the S500 HCP data release (Glasser et al., 2013; Robinson et al., 2014), the S1200 release (Robinson et al., 2018), and for the subject-level HCP-MMP parcellations (Glasser et al., 2016), which used a different regularization term. These differences between the three registration methods result in a slight spatial misalignment between the training labels and the cortical features. While the S500 data release utilized MSMSulc, a spherical surface registration driven by cortical folding patterns, the S1200 release utilized MSMAll, and incorporated functional connectivity into the spatial resampling step. Glasser et al. (2016) used a prototypical version of MSMAll in addition to MSMSulc, and thereby incorporated additional features derived from resting-state networks to drive the surface matching process. Importantly, this discrepancy between the training labels and training features is not a flaw in our methodology itself, and correcting for this difference in the registration approach would only improve the results of our analysis. As we showed in **Figure 10**, incorporating MSMAll-processed data from the S1200 dataset, instead of MSMSulc-processed data from the S500 dataset, improved model classification accuracy by nearly 5%. We hypothesize that this improvement would only increase if we had access to the data processed with the prototypical version of MSMAll. Based on the comparisons of subject-level predictions with the subject-level ground truth MMP maps, our models performed well in spite of these registration discrepancies. Our results lend evidence to the robustness of graph neural networks for learning cortical maps from functional connectivity.

Finally, participants in both the HCP and MSC studies were healthy young adults, and the datasets had been extensively quality controlled. Little to no work has been done on extending connectivity-based classifiers to atypical populations, such as to individuals with neurodegeneration. It is unknown how a model trained on connectivity properties from healthy individuals would perform in populations where connectivity is known to degrade. While our model (and that developed by Glasser et al., 2016) predict maps based on healthy individuals, it is possible that some studies would need to train population-specific models.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

KE conceptualized this study, developed the code, performed the analyses, and wrote the bulk of the document. TG provided comments and neuroscientific insight into the analysis, contributed to the editing, and organization structure of the manuscript. DH provided extensive neuroscientific and technical guidance for this work, contributed to the editing, and organizational structure of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This project was supported by grant NSF BCS 1734430, titled Collaborative Research: Relationship of Cortical Field Anatomy to Network Vulnerability and Behavior (TG, PI).

ACKNOWLEDGMENTS

We thank Matthew F. Glasser for making subject-level Human Connectome Project multi-modal parcellations available for this analysis, and for his helpful and extensive comments on a draft version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.797500/full#supplementary-material

REFERENCES

Amunts, K., Mohlberg, H., Bludau, S., and Zilles, K. (2020). Julich-Brain: a 3D probabilistic atlas of the human brain's cytoarchitecture. *Science* 369, 988–992. doi: 10.1126/science.abb4588

Arslan, S., Parisot, S., and Rueckert, D. (2015). Joint spectral decomposition for the parcellation of the human cerebral cortex using resting-state fMRI. *Inf. Process. Med. Imaging* 24, 85–97. doi: 10.1007/978-3-319-19992-4_7

Asman, A. J., and Landman, B. A. (2012). Non-local statistical label fusion for multi-atlas segmentation. *Med. Image Anal.* 17, 194–208. doi: 10.1016/j.media.2012.10.002

Asman, A. J., and Landman, B. A. (2014). Hierarchical performance estimation in the statistical label fusion framework. *Med. Image Anal.* 18, 1070–1081. doi: 10.1016/j.media.2014.06.005

Bajada, C. J., Jackson, R. L., Haroon, H. A., Azadbakht, H., Parker, G. J., Lambon Ralph, M. A., et al. (2017). A graded tractographic parcellation of the temporal lobe. *Neuroimage* 155, 503–512. doi: 10.1016/j.neuroimage.2017.04.016

Baldassano, C., Beck, D. M., and Fei-Fei, L. (2015). Parcellating connectivity in spatial maps. *PeerJ* 3:e784. doi: 10.7717/peerj.784

Beckmann, C. F., DeLuca, M., Devlin, J. T., and Smith, S. M. (2005). Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 1001–1013. doi: 10.1098/rstb.2005.1634

Blumensath, T., Jbabdi, S., Glasser, M. F., Van Essen, D. C., Ugurbil, K., Behrens, T. E., et al. (2013). Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. *Neuroimage* 76, 313–324. doi: 10.1016/j.neuroimage.2013.03.024

Bullmore, E., and Sporns, O. (2012). The economy of brain network organization. *Nat. Rev. Neurosci.* 13, 336–349. doi: 10.1038/nrn3214

Coalson, T. S., Van Essen, D. C., and Glasser, M. F. (2018). The impact of traditional neuroimaging methods on the spatial localization of cortical areas. *Proc. Natl. Acad. Sci. U.S.A.* 115, E6356–E6365. doi: 10.1073/pnas.1801582115

Cucurull, G., Wagstyl, K., Casanova, A., Velickovic, P., Jakobsen, E., Drozdzal, M., et al. (2018). "Convolutional neural networks for mesh-based parcellation of the cerebral cortex," in *Med. Imaging with Deep Learn*. Available online at: https://openreview.net/pdf?id=rkKvBAiiz.

Defferrard, M., Bresson, X., and Vandergheynst, P. (2016). "Convolutional neural networks on graphs with fast localized spectral filtering," in *NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems*. Barcelona: ACM, 187–98.

Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53, 1–15. doi: 10.1016/j.neuroimage.2010.06.010

Eschenburg, K., Haynor, D., and Grabowski, T. (2018). "Automated connectivity-based cortical mapping using registration-constrained classification," in *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging* (Houston, TX). doi: 10.1117/12.2293968

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D. H., et al. (2004). Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14, 11–22. doi: 10.1093/cercor/bhg087

Gala, R., Budzillo, A., Baftizadeh, F., Miller, J., Gouwens, N., Arkhipov, A., et al. (2021). Consistent cross-modal identification of cortical neurons with coupled autoencoders. *Nat. Comput. Sci.* 1, 120–127. doi: 10.1038/s43588-021-00030-1

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171–178. doi: 10.1038/nature18933

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J., et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* 80, 105–124. doi: 10.1016/j.neuroimage.2013.04.127

Glasser, M. F., and van Essen, D. C. (2011). Mapping human cortical areas *in vivo* based on myelin content as revealed by T1- and T2-weighted MRI. *J. Neurosci.* 31, 11597–11616. doi: 10.1523/JNEUROSCI.2180-11.2011

Gopinath, K., Desrosiers, C., and Lombaert, H. (2019). Graph convolutions on spectral embeddings for cortical surface parcellation. *Med. Image Anal.* 54, 297–305. doi: 10.1016/j.media.2019.03.012

Gorbach, N. S., Schutte, C., Melzer, C., Goldau, M., Sujazow, O., Jitsev, J., et al. (2011). Hierarchical information-based clustering for connectivity-based cortex parcellation. *Front. Neuroinform.* 5:18. doi: 10.3389/fninf.2011.00018

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., and Petersen, S. E. (2016). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* 26, 288–303. doi: 10.1093/cercor/bhu239

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., et al. (2017). Precision functional mapping of individual human brains. *Neuron* 95, 791.e7–807.e7. doi: 10.1016/j.neuron.2017.07.011

Hacker, C. D., Laumann, T. O., Szrama, N. P., Baldassarre, A., Snyder, A. Z., Leuthardt, E. C., et al. (2013). Resting state network estimation in individual subjects. *Neuroimage* 15, 616–633. doi: 10.1016/j.neuroimage.2013.05.108

Hagler, D. J., Hatton, S., Cornejo, M. D., Makowski, C., Fair, D. A., Dick, A. S., et al. (2019). Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *Neuroimage* 202:116091. doi: 10.1016/j.neuroimage.2019.116091

Hamilton, W. L., Ying, R., and Leskovec, J. (2017). "Representation learning on graphs: methods and applications," in *IEEE Data Engineering Bulletin* (California, CF: IEEE) arXiv:1709.05584.

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Kipf, T. N., and Welling, M. (2016a). *Semi-Supervised Classification With Graph Convolutional Networks*. Technical report, University of Amsterdam.

Kipf, T. N., and Welling, M. (2016b). *Variational Graph Auto-Encoders*. Technical report, University of Amsterdam.

Liu, M., Kitsch, A., Miller, S., Chau, V., Poskitt, K., Rousseau, F., et al. (2016). Patch-based augmentation of expectation-maximization for brain MRI tissue segmentation at arbitrary age after premature birth. *Neuroimage* 127, 387–408. doi: 10.1016/j.neuroimage.2015.12.009

Parisot, S., Arslan, S., Passerat-Palmbach, J., Wells, W. M. III, Rueckert, D., Wells, W. M. III, et al. (2015). Tractography-driven groupwise multi-scale parcellation of the cortex. *Inf. Process. Med. Imaging* 24, 600–612. doi: 10.1007/978-3-319-19992-4_47

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., et al. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI) clinical characterization. *Neurology* 74, 201–209. doi: 10.1212/WNL.0b013e3181cb3e25

Robinson, E. C., Garcia, K., Glasser, M. F., Chen, Z., Coalson, T. S., Makropoulos, A., et al. (2018). Multimodal surface matching with higher-order smoothness constraints. *Neuroimage* 167, 453–465. doi: 10.1016/j.neuroimage.2017.10.037

Robinson, E. C., Jbabdi, S., Glasser, M. F., Andersson, J., Burgess, G. C., Harms, M. P., et al. (2014). MSM: a new flexible framework for multimodal surface matching. *Neuroimage* 100, 414–426. doi: 10.1016/j.neuroimage.2014.05.069

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., and Smith, S. M. (2014). Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* 90, 449–468. doi: 10.1016/j.neuroimage.2013.11.046

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., et al. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* 28, 3095–3114. doi: 10.1093/cercor/bhx179

Smith, S. M., Hyvärinen, A., Varoquaux, G., Miller, K. L., and Beckmann, C. F. (2014). Group-PCA for very large fMRI datasets. *Neuroimage* 101:738. doi: 10.1016/j.neuroimage.2014.07.051

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., et al. (2017). *Attention Is All You Need*. Technical report, Google.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lí, P., and Bengio, Y. (2018). "Graph attention networks," in *International Conference on Learning Representations* (Vancouver, VC :ICLR) arXiv:1710.10903.

Wagstyl, K., Larocque, S., Cucurull, G., Lepage, C., Cohen, J. P., Bludau, S., et al. (2020). BigBrain 3D atlas of cortical layers: cortical and laminar thickness gradients diverge in sensory and motor cortices. *PLoS Biol.* 18:e3000678. doi: 10.1371/journal.pbio.3000678

Wang, G., Ying, R., Huang, J., and Leskovec, J. (2019). *Improving Graph Attention Networks with Large Margin-based Constraints*. Technical report, Stanford University, Mountain View.

Wang, M., Zheng, D., Ye, Z., Gan, Q., Li, M., Song, X., et al. (2020). *Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks*. Technical report, New York University.

Wig, G. S., Laumann, T. O., and Petersen, S. E. (2014). An approach for parcellating human cortical areas using resting-state correlations. *Neuroimage* 93(Pt 2), 276–291. doi: 10.1016/j.neuroimage.2013.07.035

Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K.-I., and Jegelka, S. (2018). *Representation Learning on Graphs with Jumping Knowledge Networks*. Technical report, MIT.

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 1125–1165. doi: 10.1152/jn.00338.2011

Zeng, H., Zhou, H., Srivastava, A., Kannan, R., and Prasanna, V. (2020). *GraphSAINT: Graph Sampling Based Inductive Learning Method*. Technical report, University of Southern California.

Zhao, Q., Adeli, E., Honnorat, N., Leng, T., and Pohl, K. M. (2019). Variational autoencoder for regression: application to brain aging analysis. Technical report, Stanford. doi: 10.1007/978-3-030-32245-8_91

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 14 January 2022
doi: 10.3389/fnins.2021.810833

Altered Functional Network in Infants With Profound Bilateral Congenital Sensorineural Hearing Loss: A Graph Theory Analysis

*Wenzhuo Cui, Shanshan Wang, Boyu Chen and Guoguang Fan**

Department of Radiology, The First Affiliated Hospital of China Medical University, Shenyang, China

Functional magnetic resonance imaging (fMRI) studies have suggested that there is a functional reorganization of brain areas in patients with sensorineural hearing loss (SNHL). Recently, graph theory analysis has brought a new understanding of the functional connectome and topological features in central neural system diseases. However, little is known about the functional network topology changes in SNHL patients, especially in infants. In this study, 34 infants with profound bilateral congenital SNHL and 28 infants with normal hearing aged 11–36 months were recruited. No difference was found in small-world parameters and network efficiency parameters. Differences in global and nodal topologic organization, hub distribution, and whole-brain functional connectivity were explored using graph theory analysis. Both normal-hearing infants and SNHL infants exhibited small-world topology. Furthermore, the SNHL group showed a decreased nodal degree in the bilateral thalamus. Six hubs in the SNHL group and seven hubs in the normal-hearing group were identified. The left middle temporal gyrus was a hub only in the SNHL group, while the right parahippocampal gyrus and bilateral temporal pole were hubs only in the normal-hearing group. Functional connectivity between auditory regions and motor regions, between auditory regions and default-mode-network (DMN) regions, and within DMN regions was found to be decreased in the SNHL group. These results indicate a functional reorganization of brain functional networks as a result of hearing loss. This study provides evidence that functional reorganization occurs in the early stage of life in infants with profound bilateral congenital SNHL from the perspective of complex networks.

Keywords: functional connectivity, graph theory, hub, infants, resting-state functional magnetic resonance imaging, sensorineural hearing loss

OPEN ACCESS

Edited by:
Islem Rekik,
Istanbul Technical University, Turkey

Reviewed by:
Han Lv,
Capital Medical University, China
Roberta Di Mauro,
Praxis Dr. Roser und Kollegen, Germany

****Correspondence:***
Guoguang Fan
fanguog@sina.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *08 November 2021*
Accepted: *22 December 2021*
Published: *14 January 2022*

Citation:
Cui W, Wang S, Chen B and Fan G (2022) Altered Functional Network in Infants With Profound Bilateral Congenital Sensorineural Hearing Loss: A Graph Theory Analysis. Front. Neurosci. 15:810833. doi: 10.3389/fnins.2021.810833

INTRODUCTION

Sensorineural hearing loss (SNHL) is caused by lesions in the hair cells, vestibulocochlear nerve, or auditory cortex of the brain. In children, the causes of SNHL are not completely understood, but genetic factors may play a part (Kvestad et al., 2014). Cochlear implantation (CI) is an important procedure for the restoration of hearing for children with severe-to-profound SNHL. The United States Food and Drug Administration (FDA)'s recommended age of CI is at least 12 months (Purcell et al., 2021). One study found that children with congenital SNHL benefit more if they receive CI at the age of 1–3 years (Zwolan et al., 2004). Another study suggested that children

with congenital SNHL who receive CI before the age of 3.5 benefit more than those who receive it after the age of seven (Sharma and Campbell, 2011). In humans, there is a rapid microstructural white matter development during the first 3 years of life (Lebel and Deoni, 2018). Therefore, the age before 3.5 years is considered as the sensitive or critical period, and the age of 7 as the end of this period (Sharma et al., 2015; Patton et al., 2019). During this period, the auditory cortex has maximum plasticity, and children with congenital hearing loss can benefit the most from CI.

Patients with SNHL need to undergo magnetic resonance imaging (MRI) examinations before CI to confirm whether there are lesions in the central neural system or malformations in the inner ear. In recent years, functional magnetic resonance imaging (fMRI) has been used to estimate the functional changes in the brain in SNHL patients. Task-based fMRI studies found that the auditory areas of SNHL patients can be activated by visual and somatosensory stimuli due to long-term hearing deprivation, which is usually called cross-modal reorganization (Finney et al., 2001; Cardon and Sharma, 2018). Seed-based functional connectivity studies also found alterations in auditory, visual, and language areas (Heggdal et al., 2016; Shi et al., 2016). Furthermore, independent component analysis (ICA) based on resting-state functional magnetic resonance imaging (rs-fMRI) research showed dysconnectivity within and between multiple resting-state networks of executive control, higher-order cognition, and attention (Luan et al., 2019), indicating that the effects of SNHL are not limited to only several systems and resting-state networks but may result in changes in the whole brain network. The study of Stolzberg et al. on neonatal deaf cats using ICA method also found widely altered functional connectivity in both auditory and non-auditory regions (Stolzberg et al., 2018). However, evidence also showed different cross-modal reorganization pattern in primary and high-order auditory cortex (Kral et al., 2003; Berger et al., 2017).Therefore, the changes in the whole-brain functional network in SNHL patients are worth studying.

Complex network analysis, which originates from graph theory, has been used to characterize the functional connectomes within the whole-brain network (Benito-León et al., 2019). This approach provides a topological method to describe network properties (Rubinov and Sporns, 2010) and could be a good complement to seed-based functional connectivity. In recent years, the brain structural and functional network has been found to be changed in many neurological and psychological disorders (Suo et al., 2017a,b, 2019; Benito-León et al., 2019; Openneer et al., 2020; Radetz et al., 2020). Therefore, complex network analysis may be a new approach to evaluating brain functional plasticity in SNHL. Using graph theory analysis, Bonna et al. (2021) found reduced modularity and disrupted functional connectivity in adults who became deaf in childhood when compared with normal hearing siblings. Li et al. (2016) found altered functional connectivity and hub distribution in SNHL adolescents. However, the changes in topological properties and whole-brain functional connectivity in congenital SNHL infants within the developing critical period remain unknown.

In the present study, we used graph theory analysis to compare the alterations in the functional connectome in bilateral congenital SNHL infants and normal-hearing infants within the critical period. Based on previous evidence from both neuroimaging and clinical studies, we hypothesized that the functional network topology of SNHL infants would have similar alterations to that in adults, compared to normal-hearing siblings. We also hoped that this study could provide a new understanding of neuroplasticity in bilateral congenital SNHL infants within the critical period. The University of North Carolina (UNC) 2-year-old infant atlas (Shi et al., 2011) based on Automated Anatomical Labeling (AAL) was used to improve the accuracy of the analysis of infant brain images.

MATERIALS AND METHODS

Participants

This study included 38 infants with congenital bilateral SNHL and 31 with normal hearing. All SNHL infants failed to pass the newborn hearing screening examinations after birth. The auditory brainstem response (ABR) results were symmetrically greater than 90 dB, indicating profound hearing loss. These infants were candidates for cochlear implant surgery, and an MRI scan was used as a presurgical evaluation. None of the patients indicated a history of head surgery, ototoxic drug use, trauma, or any central nervous system disease. The participants of SNHL group had no history of wearing hearing aids. In addition, high-resolution computed tomography (HRCT) scans showed no inner ear malformation. In the normal-hearing (NH) group, participants underwent MRI examinations for non-hearing indications, and MRI scans showed no central nervous system lesions. Infants in the NH group had no history of neurological disease, head injury, surgery, or hearing problems. The data of two SNHL infants and one NH infant were excluded from the analysis due to excessive head movement during scanning. Another two SNHL infants were excluded because they did not finish the scan. Consequently, 34 infants with congenital bilateral SNHL (16 males and 18 females; mean age 23.41 $\pm$ 8.26 months; age range 11–36 months) and 30 infants with normal hearing (16 males and 14 females, mean age 24.03 $\pm$ 7.18 months; age range 11–36 months) were included in this study. The SNHL group and NH group were matched in age ($p = 0.089$) and gender (chi-squared $p = 0.616$). All the examinations were approved by the ethics committee of the hospital. Informed consent was obtained from every subject's parent.

Data Acquisition

MR images were collected on a Siemens Verio Tim 3.0 T MR scanner (Siemens Medical Solutions, Erlangen, Germany) with a 16-channel head coil. To acquire resting-state fMRI data, all the subjects were sedated with 50–60 mg/kg of 10% chloral hydrate orally 15 min before scanning. Earplugs and headphones were used to protect hearing and reduce the impact of scanner noise. Infants were observed closely by a pediatrician during the scans. Anatomical MRI and fMRI acquisitions were obtained from all infants using the protocol detailed below.

Routine MRI scans consisted of axial and sagittal T1-weighted images (T1WI) and axial and coronal T2-weighted images (T2WI) to observe whether there were lesions and/or abnormalities in the brain. After that, the echo-planar imaging sequence was used to acquire resting-state fMRI data. The parameters were as follows: repetition time (TR), 2000 ms; echo time (TE), 30 ms; field of view (FOV), 220 mm × 220 mm; flip angle, 90°; slice thickness, 4 mm; matrix, 70 × 70; voxel size, 3.1 mm × 3.1 mm × 4 mm; 30 slices were acquired covering the whole brain, and the total volume was 190. High-resolution T1-weighted structural brain images were collected using a 3D SPGR sequence with the following parameters: TR, 2400 ms; TE, 3.16 ms; inversion time (TI), 900 ms; FOV, 220 mm × 220 mm; flip angle, 9°; slice thickness, 1 mm; matrix, 224 × 256; voxel size, 1.0 mm × 1.0 mm × 1.0 mm; 144 sagittal slices were acquired covering the whole brain.

Data Preprocessing and Network Construction

Data were preprocessed using the DPARSF toolbox (Yan and Zang, 2010) in SPM12[1]. The first 10 functional volumes of each subject were removed from analysis because of the instability of the equipment and the subjects' adaptation to the environment. Then, images were corrected for section timing and head movement. Infants were excluded with a threshold of 2.0 mm in translation or 2.0° in rotation. Subsequently, the functional images of each subject were coregistered to their corresponding T1WI high-resolution image, and T1WI images were segmented with UNC 2-year-old tissue probability maps (Shi et al., 2011). After that, images were normalized to the 2-year-old brain template (Shi et al., 2011) and resampled to 3 mm × 3 mm × 3 mm. Further steps were performed, including temporal band-pass filtering (0.01–0.08 Hz), linear detrending, and regression of the cerebrospinal fluid signal, white matter signal, whole-brain averaged signal, and head motion parameters in 24 directions.

The functional network was constructed and analyzed using GRETNA[2]. The functional images were parcellated into 90 brain regions based on the UNC 2-year-old infant atlas (Shi et al., 2011). Each region was considered as a node of the network. Then, the mean time series of each region were acquired, and Pearson's correlation coefficients of the mean time series between each pair of nodes were calculated, resulting in a 90 × 90 correlation matrix. Fisher's R-to-Z transformation was performed for the graph theory analysis. The group-averaged matrix of the two groups were showed in **Figure 1**.

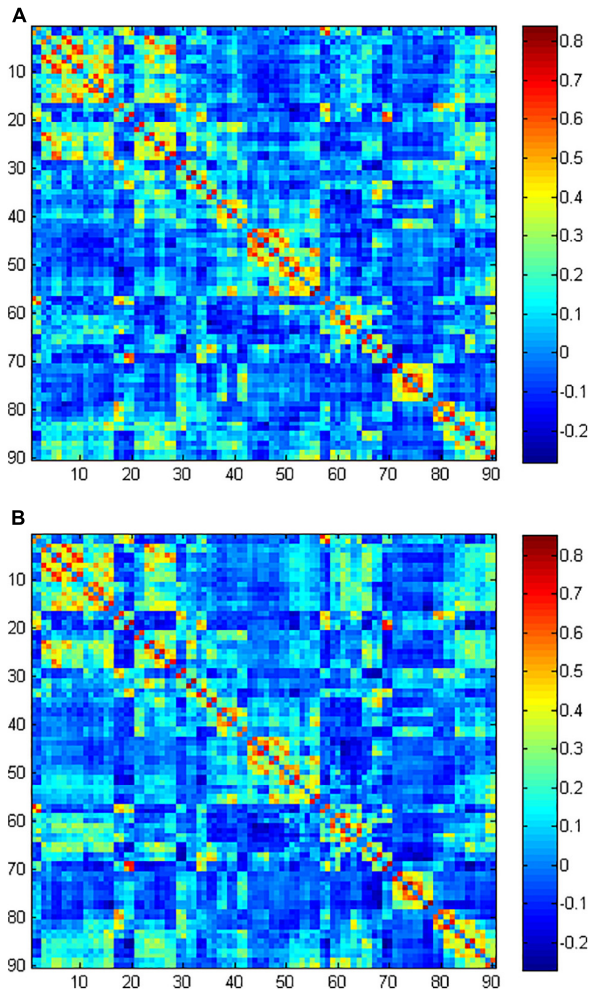

FIGURE 1 | The group-averaged functional connectivity matrix of normal-hearing (NH) group **(A)** and sensorineural hearing loss (SNHL) group **(B)**.

Global and Nodal Topological Properties

We used a wide range of sparsity thresholds (0.1 to 0.4, step 0.01) to avoid deviations from using a single-sparsity threshold. The area under the curve (AUC) of both global and nodal network metrics was calculated and compared. Global parameters included small-world parameters (clustering coefficient Cp, characteristic path length Lp, normalized clustering coefficient γ, normalized characteristic path length λ, and small-worldness σ) and network efficiency parameters (local efficiency Eloc and global efficiency Eglob). Cp, γ, and Eloc were used to measure segregation, while Lp, λ, and Eglob were used to measure integration (Rubinov and Sporns, 2010). A network with high Cp and γ, as well as low Lp and λ, was considered to have a small-word organization (Suo et al., 2019). A network was defined as a small-world network when the small-worldness ($\sigma = \gamma/\lambda$) > 1, which reflects an economic model to transform information efficiently (Achard and Bullmore, 2007). Nodal parameters included nodal degree, nodal efficiency, and nodal betweenness. The details and interpretations of global and nodal parameters mentioned above were explained in a previous study (Rubinov and Sporns, 2010). Hubs are important for information transmission, which is variably defined in the literature. In this study, a node was considered to be a hub in a network if its nodal degree, nodal efficiency, and nodal betweenness were at least 1 standard deviation greater than the mean value (Xu et al., 2016).

[1]http://www.fil.ion.ucl.ac.uk/spm/software/spm12/

[2]http://www.nitrc.org/projects/gretna/

Edge-Wise Connectivity Strength Analysis

The edge-wise connectivity strength difference was evaluated using the network-based statistic (NBS) (Zalesky et al., 2010) by looking for different clusters of connections between the two groups. The initial threshold was set to be a p-value less than 0.001. Then, the non-parametric permutation method (5,000 permutations) was used to estimate the statistical significance. Subnetworks with corrected p-values less than 0.05 were considered to be significant.

Statistical Analysis

Demographic data were compared using the Statistical Package for the Social Sciences (SPSS), version 22.0 (IBM Corp., Armonk, NY, United States). Graph theory properties were compared using GRETNA software. Between-group differences in the AUC of global and nodal network metrics were tested using two-sample t-test. A p-value less than 0.05 was identified as significant. When comparing nodal parameters, the Bonferroni correction was applied to correct for multiple comparisons with a significant $p < 0.05$.

RESULTS

Demographic and Clinical Characteristics

Demographic and clinical data for the subjects in the NH group and SNHL group are shown in **Table 1**. No significant differences were found in age and gender.

Global and Nodal Topologic Organization of the Functional Network

We found $\lambda > 1$, $\gamma > 1$, and $\sigma > 1$ across all the sparsities in both the SNHL group and the NH group (**Figure 2**). There was no significant difference in the global parameters (Cp, Lp, γ, λ, σ, Eg, and Eloc) of the functional network (**Table 2**). After Bonferroni correction, compared with the NH group, the SNHL group showed an increased nodal degree in the left inferior marginal angular gyrus and a decreased nodal degree in the bilateral thalamus (**Table 3**). There was no significant difference in nodal betweenness and nodal efficiency. The receiver operating characteristic (ROC) curve based on binary logistic regression using the nodal degree of the left inferior marginal angular gyrus and the bilateral thalamus was showed in **Figure 3**.

TABLE 1 | Summary of the demographic and clinical data.

	NH group	SNHL group	p-value
Number (n)	30	34	–
Age (months)[a] (mean ± SD)	24.03 ± 7.18	23.41 ± 8.26	0.089
Age range (months)	11-36	11-36	–
Gender (male/female)[b]	16/14	16/18	0.616
ABR of left ear (dB HL)	>90	–	–
ABR of right ear (dB HL)	>90	–	–

ABR, auditory brainstem response; NH, normal hearing; SNHL, sensorineural hearing loss; SD, standard deviation.
[a]Statistical analyses for comparisons between groups were carried out with t-tests.
[b]Statistical analyses for comparisons between groups were carried out with $\chi 2$ tests.

Hubs in the Brain Network

We found six hubs in the SNHL group and seven hubs in the NH group. As shown in **Figure 4**, the left medial superior frontal gyrus and right inferior temporal gyrus were hubs in both the SNHL group and NH group. The right dorsolateral superior frontal gyrus, bilateral orbital part of the inferior frontal gyrus, and left middle temporal gyrus were hubs only in the SNHL group, while the left medial orbital superior frontal gyrus, right parahippocampal gyrus, bilateral temporal pole of the superior temporal gyrus, and left temporal pole of the middle temporal gyrus were hubs only in the NH group.

Functional Connectivity

Compared with the NH group, the SNHL group presented a significantly decreased subnetwork component in the NBS analysis results ($p < 0.001$, NBS corrected). As shown in **Figure 5**, the subnetwork consisted of seven nodes and six edges. The involved regions included the bilateral median cingulate and paracingulate gyri, left precuneus, right paracentral lobule, bilateral middle temporal gyrus, and right temporal pole of the middle temporal gyrus.

DISCUSSION

This study investigated the complex brain network organization in infants with profound bilateral congenital SNHL within a critical period of development based on rs-fMRI using graph theory analysis. We found differences between the SNHL group and NH group in nodal degree, hub distribution, and whole-brain functional connectivity. These results indicated that functional reorganization occurred in the first few years of hearing deprivation.

Topological Parameter Alterations

The human brain network exhibits small-world topology from the fetal stage (Turk et al., 2019). In this study, both the SNHL group and NH group showed small-word organization, with no significant differences. This result was consistent with previous studies on sudden unilateral sensorineural hearing loss in adults and prelingually deaf adolescents (Li et al., 2016; Xu et al., 2016), indicating that the efficiency of information transmission in the brain was not affected by hearing deprivation.

However, compared with normal-hearing infants, SNHL infants showed no significant difference in global topological parameters, which was consistent with the study of Zhang et al. (2018) on long-term unilateral hearing loss but different from the study of Xu et al. (2016) on sudden unilateral sensorineural hearing loss. Alteration of functional network global topological parameters has been observed in many neurological disorders, including epilepsy (Zhang et al., 2011;

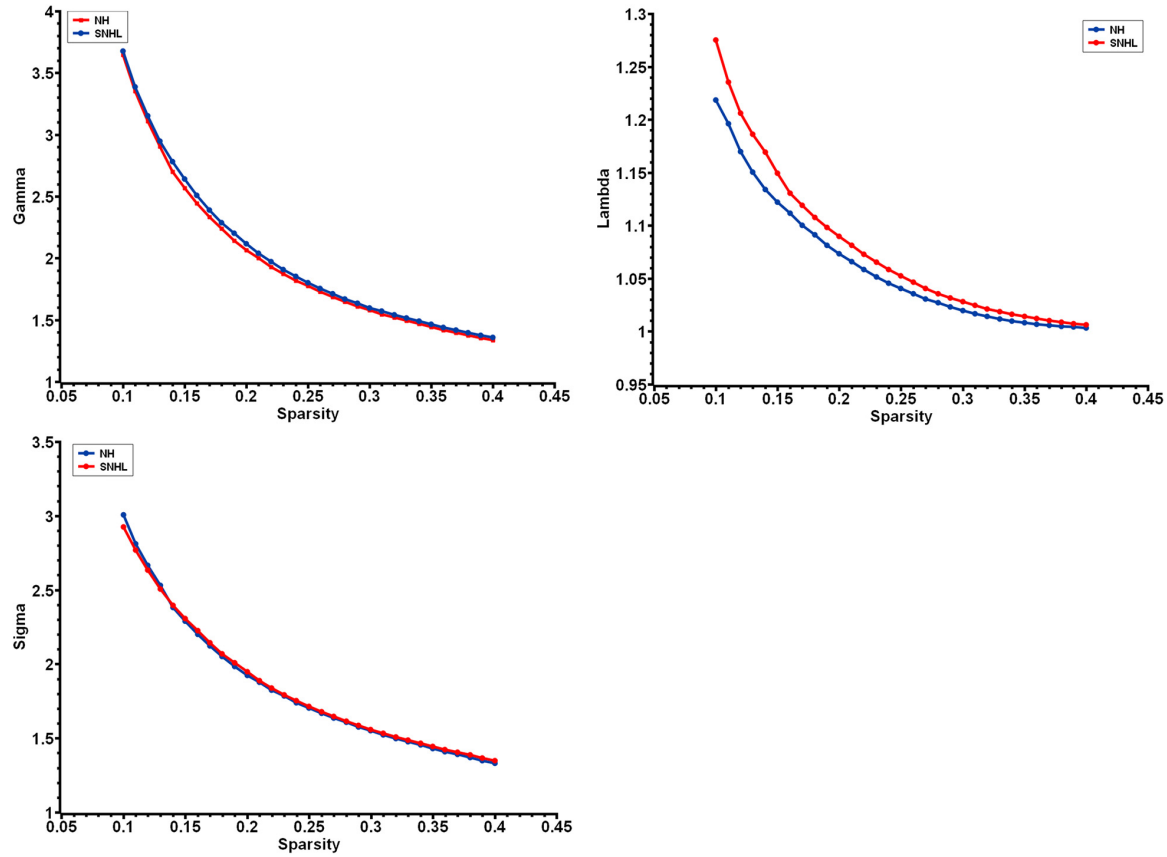

FIGURE 2 | The small-world parameters of functional networks as a function of sparsity threshold from 0.1 to 0.4. Both the NH group and SNHL group exhibited γ, λ, and σ larger than 1, which indicates that both groups exhibited the small-word organization. (NH, normal hearing; SNHL, sensorineural hearing loss; γ, normalized clustering coefficient; λ, normalized characteristic path length; σ, small-worldness).

Li et al., 2020), Alzheimer's disease (Pereira et al., 2016), Parkinson's disease (Suo et al., 2017b), and essential tremor (Benito-León et al., 2019). These patients usually have obvious neurologic symptoms, including cognitive impairment or motor disorder, with a long duration. The duration of sudden sensorineural hearing loss is less than 3 days, which was significantly shorter than that of patients in Xu et al.'s study (at least 2 months) and infants with congenital hearing loss in our study. Furthermore, SNHL patients usually do not have neurological symptoms as severe as those with neurological disorders. Based on the above analysis, we speculate that the sudden loss of sensory import could change the pattern of information transmission of the functional network, but the brain will adjust to adapt and compensate for the loss of sensory in order to maintain the topological organization if the situation persists over a long period of time. This may be a result of neuroplasticity. Moreover, considering the special age of our participants, the maintenance of global topological organization may be the

TABLE 2 | Global network parameters in NH group and SNHL group.

Global network measures	NH group (n = 34)	SNHL group (n = 30)	t-value	p-value
Cp	0.17 ± 0.009	0.17 ± 0.011	−1.332	0.190
Lp	0.53 ± 0.009	0.53 ± 0.018	−1.563	0.125
γ	0.59 ± 0.068	0.60 ± 0.075	−0.466	0.643
λ	0.32 ± 0.004	0.32 ± 0.008	−2.096	0.052
σ	0.55 ± 0.061	0.55 ± 0.072	−0.111	0.912
Eglob	0.17 ± 0.002	0.17 ± 0.004	1.510	0.138
Eloc	0.23 ± 0.005	0.23 ± 0.006	−1.384	0.173

NH, normal hearing; SNHL, sensorineural hearing loss; Cp, clustering coefficient; Lp, characteristic path length; γ, normalized clustering coefficient; λ, normalized characteristic path length; σ, small-worldness; Eglob, global efficiency; Eloc, local efficiency.

TABLE 3 | Regions of altered nodal degree in NH group and SNHL group ($p < 0.05$, Bonferroni corrected).

Brain regions	NH group	SNHL group	t-value	p-value
THA.L	5.24 ± 1.799	3.37 ± 1.138	4.207	0.00013
THA.R	5.25 ± 1.667	3.55 ± 1.152	3.993	0.00024
IPL.L	6.07 ± 1.390	7.81 ± 1.470	−3.860	0.00037

NH, normal hearing; SNHL, sensorineural hearing loss; THA.L, left thalamus; THA.R, right thalamus; IPL.L, left inferior marginal angular gyrus.

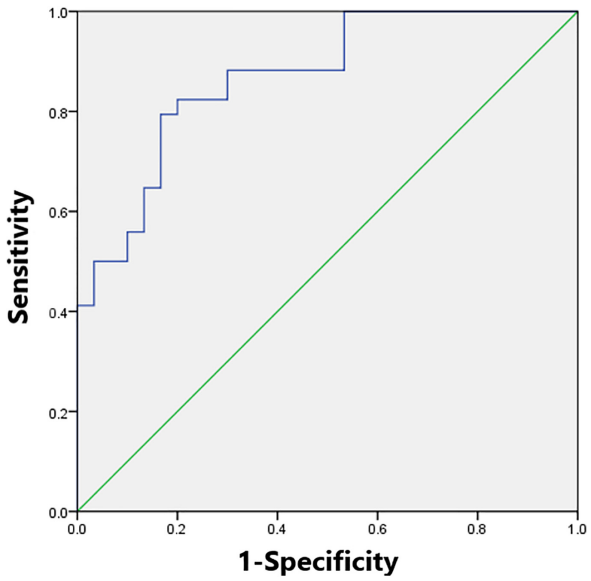

FIGURE 3 | The receiver operating characteristic (ROC) curve based on binary logistic regression using the nodal degree of the left inferior marginal angular gyrus and the bilateral thalamus. The module reached the highest area under the curve of 0.87.

basement of the maximal plasticity in the critical period, which leads to the rapid reconstruction of hearing and speaking ability after CI.

In the bilateral thalamus, a decreased nodal degree was observed in the functional network of SNHL infants, which means less or decreased functional connectivity with other regions. This finding likely underlies the key role of the thalamus in auditory information transfer. The thalamus, which consists of many nuclei, is the relaying nucleus of the sensory pathway and receives input from the cortex (Hwang et al., 2017; Martini et al., 2021). It also mediates important functions including memory, emotion, and attention (Arend et al., 2015; Geier et al., 2020). Imaging evidence has also shown decreased thalamic functional connectivity with both auditory and non-auditory regions in long-term SNHL patients (Xu et al., 2019). Our finding was consistent with Xu et al.'s (2019) result. Moreover, we found an increased nodal degree in the left inferior marginal angular gyrus, which was included in the inferior parietal lobule. The inferior parietal lobule was associated with sensory integration, body image, self-concept, and executive function (Torrey, 2007). The ICA method has also found the inferior parietal lobule to be included in the DMN, the frontoparietal control network, and the cingulo-opercular network (Igelström and Graziano, 2017). In the study of Luan et al. (2019), the right inferior parietal lobule showed increased intra-network connectivity within the right frontoparietal network, and this alteration was correlated with the Symbol Digit Modalities Test scores of SNHL patients. The increased nodal degree in the inferior marginal angular gyrus might be related to the alteration of high-order congenital functions in SNHL infants.

Hub Distribution Differences

In a brain network, regions that are more interconnected with other regions and possess a high nodal degree are considered as hubs (Xu et al., 2016; Medaglia, 2017; Zhao et al., 2019). Hub changes are considered to be related to neurological and psychological diseases (Tian et al., 2016; Lin et al., 2018). We found fewer hubs in infants than a previous study in adults using the same method to identify hubs (Xu et al., 2016), which may be due to the fact that the number of hubs increases with age (Cao et al., 2017). In this study, the hubs of both the SNHL group and NH group were located in the prefrontal lobe and temporal lobe. This distribution was consistent with previous studies on the early developing brain (Grayson and Fair, 2017; Oldham and Fornito, 2019; Turk et al., 2019; Wen et al., 2019).

The middle temporal gyrus is involved in both auditory and visual processes, namely word processing and action observation (Van Essen et al., 1986; Papeo et al., 2019). The middle temporal gyrus acting as a hub in the SNHL group in our study might be a result of deaf infants relying more on visual information. However, the functional connectivity between the thalamus and middle temporal gyrus was previously found to be decreased in SNHL adults (Xu et al., 2019). The cortical thickness of the middle temporal gyrus was found to be increased in postlingually deaf adults (Pereira-Jorge et al., 2018) but decreased in prelingually profound SNHL children (Qu et al., 2020) and postlingually deaf adults (Sun et al., 2021). Thus, we speculated that the functional and structural changes in the middle temporal gyrus in SNHL patients will differ between children and adults, and between prelingually deaf patients and postlingually deaf patients. The mechanism of the alterations in the middle temporal gyrus in different conditions needs to be studied further. The parahippocampal gyrus is a key region of memory formation (Düzel et al., 2003). The parahippocampal gyrus and temporal pole are involved in auditory memory (Muñoz-López et al., 2015; Córcoles-Parada et al., 2019). In the study of Rönnberg et al. (2011), hearing loss patients showed impaired episodic memory. Accordingly, in another study, the gray matter volume of the right parahippocampal gyrus was found to be decreased in unilateral hearing loss patients (Yang et al., 2014). In our study, the parahippocampal gyrus and temporal pole were hubs only in the NH group. The absence of these two regions as hubs in the SNHL group suggested weakened auditory memory ability, which could be a consequence of hearing deprivation.

Functional Connectivity Differences

The main regions involved in the decreased subnetwork of the SNHL group were the DMN, auditory network, and sensorimotor network. Decreased functional connectivity between the auditory and motor regions in hearing loss patients has been reported by previous studies, which was considered a result of failure in imitating speech sounds due to hearing deprivation (Shi et al., 2016; Bonna et al., 2021). In our study, we detected a weaker connection between the middle temporal gyrus and paracentral lobule. The middle temporal gyrus is a key region of

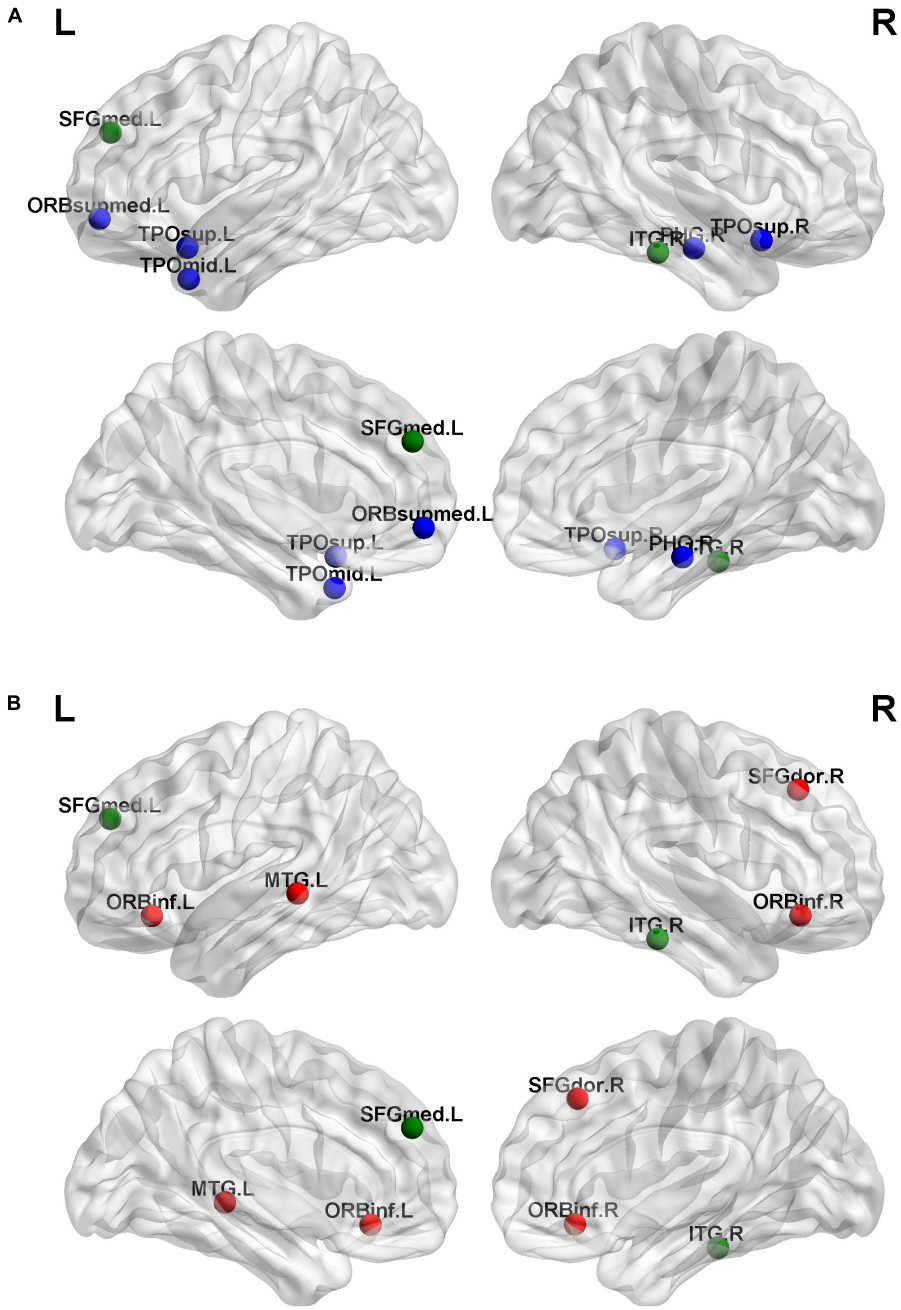

FIGURE 4 | Hubs for brain functional networks in normal-hearing (NH) group **(A)** and sensorineural hearing loss (SNHL) group **(B)**. The hubs identified in both groups are shown in green color. While the hubs for the SNHL group only are present with red color, the hubs for the NH group only are shown in blue color.

auditory perception and is involved in word processing (Papeo et al., 2019; Feng et al., 2021). Clinical studies also found poor balance and gait performance of SNHL children (Melo, 2017; Melo et al., 2018; Sokolov et al., 2019). Decreased functional connectivity between the middle temporal gyrus and paracentral lobe may reflect the separation of auditory and motor function in SNHL infants.

We also found decreased functional connectivity between the DMN and auditory regions. Similar changes were also found in a previous study of Li et al. on congenital severe sensorineural hearing loss infants under 2 years old (Li et al., 2019). The DMN is identified in infancy and becomes similar to that in adults at 2 years of age (Gao et al., 2009, 2015). The functional connectivity between DMN regions was found to increase with age during childhood and adolescence (Fan et al., 2021). The DMN consists of brain regions that are more active when there are no external tasks (Anticevic et al., 2012). It is involved in multiple high-order cognitive functions,

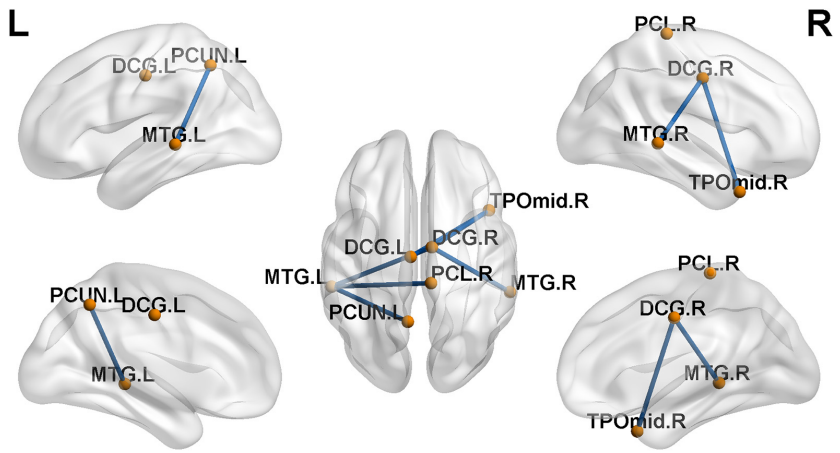

FIGURE 5 | Decreased functional connections in sensorineural hearing loss (SNHL) group when compared with the normal-hearing (NH) group ($p < 0.001$, NBS corrected). The subnetwork is consisted with seven nodes and six edges.

including self-related cognition, working memory, and emotion processing (Buckner et al., 2008; Anticevic et al., 2012; Chai et al., 2014). Moreover, the development of theory of mind, which is considered to be correlated with the functional connectivity strength of DMN regions, was found to be affected by the loss of linguistic experience caused by hearing loss (Hughes et al., 2019; Richardson et al., 2020). Decreased functional connectivity within the DMN was found in a previous ICA study by our team (Wang et al., 2021). Combined with the findings of Wang et al. (2021), this suggests that decreased functional connectivity between DMN regions and auditory regions and within DMN regions may suggest impaired high-order cognitive functions caused by hearing deprivation in the developing brain of SNHL infants.

Limitations

There are several limitations to the present study. First, the small sample size may decrease the statistical power of the results. Second, the human brain develops rapidly in the first few years of life, and there may be individual differences in brain development due to the impact of living environment. Third, there is no universal infant atlas, and using different templates can affect the accuracy of the results. These factors may reduce the detection power. Furthermore, infants who participated in the present study were sedated with chloral hydrate, which is usually thought to affect the resting-state functional connectivity compared with natural sleep. However, previous studies in SNHL infants using rs-fMRI under sedation have delivered meaningful results, which showed the feasibility of the use of sedation (Xia et al., 2017; Li et al., 2019; Wang et al., 2019). Moreover, Fransson et al. (2007) showed that there was no significant effect on functional connectivity in infants and children under sedation during scanning. Another factor which must be considered is that although we have used the hearing protection measure, the NH group was still exposed to a greater level of scanner noise than the SNHL group due to massive threshold shifts in the SNHL sample. This may bring some inevitable confounding factors. Finally, the results drawn from graph theory analysis are often not intuitive and may be difficult to interpret (Lv et al., 2018). In future studies, the sample size should be increased and further grouped according to age. Combined studies using other methods are needed as a complementary. Long-term follow-up should be conducted to explore the value of complex network analysis as a biomarker for predicting CI outcomes.

CONCLUSION

This study used graph theory analysis based on fMRI to investigate alterations of brain functional networks in profound bilateral congenital SNHL in infants in a critical period of development. We found that the functional brain network of SNHL infants within the critical period still maintains the balance of integration and segregation. Compared with NH infants, we found an increased nodal degree in the left inferior marginal angular gyrus and decreased nodal degree in the bilateral thalamus in SNHL infants. We also found a different hub distribution and functional connectivity in both auditory regions and high-order cognitive regions in SNHL infants. These changes reflect a functional network reorganization and potential changes in high-order cognitive function in SNHL infants. This study also provided novel insights into functional network alterations in the early stage of profound bilateral congenital SNHL.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the data also forms part of an ongoing study. Requests to access the datasets should be directed to GF, fanguog@sina.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Medical Science Research Ethics Committee, The First Affiliated Hospital of China Medical University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

WC, SW, and GF designed the study. WC and SW performed the research. WC and BC analyzed the data. WC wrote the draft. All authors contributed to the article and approved the submitted version.

REFERENCES

Achard, S., and Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS Comput. Biol.* 3:e17. doi: 10.1371/journal.pcbi.0030017

Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X. J., and Krystal, J. H. (2012). The role of default network deactivation in cognition and disease. *Trends Cogn. Sci.* 16, 584–592. doi: 10.1016/j.tics.2012.10.008

Arend, I., Henik, A., and Okon-Singer, H. (2015). Dissociating emotion and attention functions in the pulvinar nucleus of the thalamus. *Neuropsychology* 29, 191–196. doi: 10.1037/neu0000139

Benito-León, J., Sanz-Morales, E., Melero, H., Louis, E. D., Romero, J. P., Rocon, E., et al. (2019). Graph theory analysis of resting-state functional magnetic resonance imaging in essential tremor. *Hum. Brain Mapp.* 40, 4686–4702. doi: 10.1002/hbm.24730

Berger, C., Kühne, D., Scheper, V., and Kral, A. (2017). Congenital deafness affects deep layers in primary and secondary auditory cortex. *J. Comp. Neurol.* 525, 3110–3125. doi: 10.1002/cne.24267

Bonna, K., Finc, K., Zimmermann, M., Bola, L., Mostowski, P., Szul, M., et al. (2021). Early deafness leads to re-shaping of functional connectivity beyond the auditory cortex. *Brain Imaging Behav.* 15, 1469–1482. doi: 10.1007/s11682-020-00346-y

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Cao, M., He, Y., Dai, Z., Liao, X., Jeon, T., Ouyang, M., et al. (2017). Early development of functional network segregation revealed by connectomic analysis of the preterm human brain. *Cereb. Cortex* 27, 1949–1963. doi: 10.1093/cercor/bhw038

Cardon, G., and Sharma, A. (2018). Somatosensory cross-modal reorganization in adults with age-related, early-stage hearing loss. *Front. Hum. Neurosci.* 12:172. doi: 10.3389/fnhum.2018.00172

Chai, X. J., Ofen, N., Gabrieli, J. D. E., and Whitfield-Gabrieli, S. (2014). Development of deactivation of the default-mode network during episodic memory formation. *Neuroimage* 84, 932–938. doi: 10.1016/j.neuroimage.2013.09.032

Córcoles-Parada, M., Ubero-Martínez, M., Morris, R. G. M., Insausti, R., Mishkin, M., and Muñoz-López, M. (2019). Frontal and insular input to the dorsolateral temporal pole in primates: implications for auditory memory. *Front. Neurosci.* 13:1099. doi: 10.3389/fnins.2019.01099

Düzel, E., Habib, R., Rotte, M., Guderian, S., Tulving, E., and Heinze, H. J. (2003). Human hippocampal and parahippocampal activity during visual associative recognition memory for spatial and nonspatial stimulus configurations. *J. Neurosci.* 23, 9439–9444. doi: 10.1523/jneurosci.23-28-09439.2003

Fan, F., Liao, X., Lei, T., Zhao, T., Xia, M., Men, W., et al. (2021). Development of the default-mode network during childhood and adolescence: a longitudinal resting-state fMRI study. *Neuroimage* 226:117581. doi: 10.1016/j.neuroimage.2020.117581

Feng, G., Gan, Z., Llanos, F., Meng, D., Wang, S., Wong, P. C. M., et al. (2021). A distributed dynamic brain network mediates linguistic tone representation and categorization. *Neuroimage* 224:117410. doi: 10.1016/j.neuroimage.2020.117410

Finney, E. M., Fine, I., and Dobkins, K. R. (2001). Visual stimuli activate auditory cortex in the deaf. *Nat. Neurosci.* 4, 1171–1173. doi: 10.1038/nn763

Fransson, P., Skiöld, B., Horsch, S., Nordell, A., Blennow, M., Lagercrantz, H., et al. (2007). Resting-state networks in the infant brain. *Proc. Natl. Acad. Sci. U.S.A.* 104, 15531–15536. doi: 10.1073/pnas.0704380104

Gao, W., Alcauter, S., Smith, J. K., Gilmore, J. H., and Lin, W. (2015). Development of human brain cortical network architecture during infancy. *Brain Struct. Funct.* 220, 1173–1186. doi: 10.1007/s00429-014-0710-3

Gao, W., Zhu, H., Giovanello, K. S., Smith, J. K., Shen, D., Gilmore, J. H., et al. (2009). Evidence on the emergence of the brain's default network from 2-week-old to 2-year-old healthy pediatric subjects. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6790–6795. doi: 10.1073/pnas.0811221106

Geier, K. T., Buchsbaum, B. R., Parimoo, S., and Olsen, R. K. (2020). The role of anterior and medial dorsal thalamus in associative memory encoding and retrieval. *Neuropsychologia* 148:107623. doi: 10.1016/j.neuropsychologia.2020.107623

Grayson, D. S., and Fair, D. A. (2017). Development of large-scale functional networks from birth to adulthood: a guide to the neuroimaging literature. *Neuroimage* 160, 15–31. doi: 10.1016/j.neuroimage.2017.01.079

Heggdal, P. O. L., Brännström, J., Aarstad, H. J., Vassbotn, F. S., and Specht, K. (2016). Functional-structural reorganisation of the neuronal network for auditory perception in subjects with unilateral hearing loss: review of neuroimaging studies. *Hear. Res.* 332, 73–79. doi: 10.1016/j.heares.2015.11.015

Hughes, C., Cassidy, B. S., Faskowitz, J., Avena-Koenigsberger, A., Sporns, O., and Krendl, A. C. (2019). Age differences in specific neural connections within the Default Mode Network underlie theory of mind. *Neuroimage* 191, 269–277. doi: 10.1016/j.neuroimage.2019.02.024

Hwang, K., Bertolero, M. A., Liu, W. B., and D'Esposito, M. (2017). The human thalamus is an integrative hub for functional brain networks. *J. Neurosci.* 37, 5594–5607. doi: 10.1523/JNEUROSCI.0067-17.2017

Igelström, K. M., and Graziano, M. S. A. (2017). The inferior parietal lobule and temporoparietal junction: a network perspective. *Neuropsychologia* 105, 70–83. doi: 10.1016/j.neuropsychologia.2017.01.001

Kral, A., Schröder, J. H., Klinke, R., and Engel, A. K. (2003). Absence of cross-modal reorganization in the primary auditory cortex of congenitally deaf cats. *Exp. Brain Res.* 153, 605–613. doi: 10.1007/s00221-003-1609-z

Kvestad, E., Lie, K. K., Eskild, A., and Engdahl, B. (2014). Sensorineural hearing loss in children: the association with Apgar score. A registry-based study of 392371 children in Norway. *Int. J. Pediatr. Otorhinolaryngol.* 78, 1940–1944. doi: 10.1016/j.ijporl.2014.08.032

Lebel, C., and Deoni, S. (2018). The development of brain white matter microstructure. *Neuroimage* 182, 207–218. doi: 10.1016/j.neuroimage.2017.12.097

Li, Q., Guo, H., Liu, L., and Xia, S. (2019). Changes in the functional connectivity of auditory and language-related brain regions in children with congenital severe sensorineural hearing loss: an fMRI study. *J. Neurolinguistics* 51, 84–95. doi: 10.1016/j.jneuroling.2019.01.005

Li, W., Li, J., Wang, J., Zhou, P., Wang, Z., Xian, J., et al. (2016). Functional reorganizations of brain network in prelingually deaf adolescents. *Neural Plast.* 2016:9849087. doi: 10.1155/2016/9849087

Li, Y., Chen, Q., and Huang, W. (2020). Disrupted topological properties of functional networks in epileptic children with generalized tonic-clonic seizures. *Brain Behav.* 10, 1–14. doi: 10.1002/brb3.1890

Lin, S. J., Baumeister, T. R., Garg, S., and McKeown, M. J. (2018). Cognitive profiles and hub vulnerability in Parkinson's disease. *Front. Neurol.* 9:482. doi: 10.3389/fneur.2018.00482

Luan, Y., Wang, C., Jiao, Y., Tang, T., Zhang, J., and Teng, G. J. (2019). Dysconnectivity of multiple resting-state networks associated with higher-order functions in sensorineural hearing loss. *Front. Neurosci.* 13:55. doi: 10.3389/fnins.2019.00055

Lv, H., Wang, Z., Tong, E., Williams, L. M., Zaharchuk, G., Zeineh, M., et al. (2018). Resting-state functional MRI: everything that nonexperts have always wanted to know. *Am. J. Neuroradiol.* 39, 1390–1399. doi: 10.3174/ajnr.A5527

Martini, F. J., Guillamón-Vivancos, T., Moreno-Juan, V., Valdeolmillos, M., and López-Bendito, G. (2021). Spontaneous activity in developing thalamic and cortical sensory networks. *Neuron* 109, 2519–2534. doi: 10.1016/j.neuron.2021.06.026

Medaglia, J. D. (2017). Graph theoretic analysis of resting state functional MR imaging. *Neuroimaging Clin. N. Am.* 27, 593–607. doi: 10.1016/j.nic.2017.06.008

Melo, R. S. (2017). Gait performance of children and adolescents with sensorineural hearing loss. *Gait Posture* 57, 109–114. doi: 10.1016/j.gaitpost.2017.05.031

Melo, R. S., Lemos, A., Raposo, M. C. F., Belian, R. B., and Ferraz, K. M. (2018). Balance performance of children and adolescents with sensorineural hearing loss: repercussions of hearing loss degrees and etiological factors. *Int. J. Pediatr. Otorhinolaryngol.* 110, 16–21. doi: 10.1016/j.ijporl.2018.04.016

Muñoz-López, M., Insausti, R., Mohedano-Moriano, A., Mishkin, M., and Saunders, R. C. (2015). Anatomical pathways for auditory memory II: information from rostral superior temporal gyrus to dorsolateral temporal pole and medial temporal cortex. *Front. Neurosci.* 9:158. doi: 10.3389/fnins.2015.00158

Oldham, S., and Fornito, A. (2019). The development of brain network hubs. *Dev. Cogn. Neurosci.* 36:100607. doi: 10.1016/j.dcn.2018.12.005

Openneer, T. J. C., Marsman, J. B. C., van der Meer, D., Forde, N. J., Akkermans, S. E. A., Naaijen, J., et al. (2020). A graph theory study of resting-state functional connectivity in children with Tourette syndrome. *Cortex* 126, 63–72. doi: 10.1016/j.cortex.2020.01.006

Papeo, L., Agostini, B., and Lingnau, A. (2019). The large-scale organization of gestures and words in the middle temporal gyrus. *J. Neurosci.* 39, 5966–5974. doi: 10.1523/JNEUROSCI.2668-18.2019

Patton, M. H., Blundon, J. A., and Zakharenko, S. S. (2019). Rejuvenation of plasticity in the brain: opening the critical period. *Curr. Opin. Neurobiol.* 54, 83–89. doi: 10.1016/j.conb.2018.09.003

Pereira, J. B., Mijalkov, M., Kakaei, E., Mecocci, P., Vellas, B., Tsolaki, M., et al. (2016). Disrupted network topology in patients with stable and progressive mild cognitive impairment and Alzheimer's disease. *Cereb. Cortex* 26, 3476–3493. doi: 10.1093/cercor/bhw128

Pereira-Jorge, M. R., Andrade, K. C., Palhano-Fontes, F. X., Diniz, P. R. B., Sturzbecher, M., Santos, A. C., et al. (2018). Anatomical and functional MRI changes after one year of auditory rehabilitation with hearing aids. *Neural Plast.* 2018:9303674. doi: 10.1155/2018/9303674

Purcell, P. L., Deep, N. L., Waltzman, S. B., Roland, J. T., Cushing, S. L., Papsin, B. C., et al. (2021). Cochlear implantation in infants: why and how. *Trends Hear.* 25, 1–10. doi: 10.1177/23312165211031751

Qu, H., Tang, H., Pan, J., Zhao, Y., and Wang, W. (2020). Alteration of cortical and subcortical structures in children with profound sensorineural hearing loss. *Front. Hum. Neurosci.* 14:565445. doi: 10.3389/fnhum.2020.565445

Radetz, A., Koirala, N., Krämer, J., Johnen, A., Fleischer, V., Gonzalez-Escamilla, G., et al. (2020). Gray matter integrity predicts white matter network reorganization in multiple sclerosis. *Hum. Brain Mapp.* 41, 917–927. doi: 10.1002/hbm.24849

Richardson, H., Koster-Hale, J., Caselli, N., Magid, R., Benedict, R., Olson, H., et al. (2020). Reduced neural selectivity for mental states in deaf children with delayed exposure to sign language. *Nat. Commun.* 11:3246. doi: 10.1038/s41467-020-17004-y

Rönnberg, J., Danielsson, H., Rudner, M., Arlinger, S., Sternäng, O., Wahlin, A., et al. (2011). Hearing loss is negatively related to episodic and semantic long-term memory but not to short-term memory. *J. Speech Lang. Hear. Res.* 54, 705–726. doi: 10.1044/1092-4388(2010/09-0088)

Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069. doi: 10.1016/j.neuroimage.2009.10.003

Sharma, A., and Campbell, J. (2011). A sensitive period for cochlear implantation in deaf children. *J. Matern. Neonatal Med.* 24, 151–153. doi: 10.3109/14767058.2011.607614.A

Sharma, A., Campbell, J., and Cardon, G. (2015). Developmental and cross-modal plasticity in deafness: evidence from the P1 and N1 event related potentials in cochlear implanted children. *Int. J. Psychophysiol.* 95, 135–144. doi: 10.1016/j.ijpsycho.2014.04.007

Shi, B., Yang, L. Z., Liu, Y., Zhao, S. L., Wang, Y., Gu, F., et al. (2016). Early-onset hearing loss reorganizes the visual and auditory network in children without cochlear implantation. *Neuroreport* 27, 197–202. doi: 10.1097/WNR.0000000000000524

Shi, F., Yap, P. T., Wu, G., Jia, H., Gilmore, J. H., Lin, W., et al. (2011). Infant brain atlases from neonates to 1- and 2-year-olds. *PLoS One* 6:e18746. doi: 10.1371/journal.pone.0018746

Sokolov, M., Gordon, K. A., Polonenko, M., Blaser, S. I., Papsin, B. C., and Cushing, S. L. (2019). Vestibular and balance function is often impaired in children with profound unilateral sensorineural hearing loss. *Hear. Res.* 372, 52–61. doi: 10.1016/j.heares.2018.03.032

Stolzberg, D., Butler, B. E., and Lomber, S. G. (2018). Effects of neonatal deafness on resting-state functional network connectivity. *Neuroimage* 165, 69–82. doi: 10.1016/j.neuroimage.2017.10.002

Sun, Z., Seo, J. W., Park, H. J., Lee, J. Y., Kwak, M. Y., Kim, Y., et al. (2021). Cortical reorganization following auditory deprivation predicts cochlear implant performance in postlingually deaf adults. *Hum. Brain Mapp.* 42, 233–244. doi: 10.1002/hbm.25219

Suo, X., Lei, D., Li, N., Cheng, L., Chen, F., Wang, M., et al. (2017b). Functional brain connectome and its relation to hoehn and yahr stage in Parkinson disease. *Radiology* 285, 904–913. doi: 10.1148/radiol.2017162929

Suo, X., Lei, D., Chen, F., Wu, M., Li, L., Sun, L., et al. (2017a). Anatomic insights into disrupted small-world networks in pediatric posttraumatic stress disorder. *Radiology* 282, 826–834. doi: 10.1148/radiol.2016160907

Suo, X., Lei, D., Li, W., Chen, F., Niu, R., Kuang, W., et al. (2019). Large-scale white matter network reorganization in posttraumatic stress disorder. *Hum. Brain Mapp.* 40, 4801–4812. doi: 10.1002/hbm.24738

Tian, L., Meng, C., Jiang, Y., Tang, Q., Wang, S., Xie, X., et al. (2016). Abnormal functional connectivity of brain network hubs associated with symptom severity in treatment-naive patients with obsessive-compulsive disorder: a resting-state functional MRI study. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 66, 104–111. doi: 10.1016/j.pnpbp.2015.12.003

Torrey, E. F. (2007). Schizophrenia and the inferior parietal lobule. *Schizophr. Res.* 97, 215–225. doi: 10.1016/j.schres.2007.08.023

Turk, E., van den Heuvel, M. I., Benders, M. J., de Heus, R., Franx, A., Manning, J. H., et al. (2019). Functional connectome of the fetal brain. *J. Neurosci.* 39, 9716–9724. doi: 10.1523/JNEUROSCI.2891-18.2019

Van Essen, D. C., Newsome, W. T., Maunsell, J. H. R., and Bixby, J. L. (1986). The projections from striate cortex (V1) to areas V2 and V3 in the macaque monkey: asymmetries, areal boundaries, and patchy connections. *J. Comp. Neurol.* 244, 451–480. doi: 10.1002/cne.902440405

Wang, S., Chen, B., Yu, Y., Yang, H., Cui, W., Fan, G., et al. (2021). Altered resting-state functional network connectivity in profound sensorineural hearing loss infants within an early sensitive period: a group ICA study. *Hum. Brain Mapp.* 42, 4314–4326. doi: 10.1002/hbm.25548

Wang, S., Chen, B., Yu, Y., Yang, H., Cui, W., Li, J., et al. (2019). Alterations of structural and functional connectivity in profound sensorineural hearing loss infants within an early sensitive period: a combined DTI and fMRI study. *Dev. Cogn. Neurosci.* 38:100654. doi: 10.1016/j.dcn.2019.100654

Wen, X., Zhang, H., Li, G., Liu, M., Yin, W., Lin, W., et al. (2019). First-year development of modules and hubs in infant brain functional networks. *Neuroimage* 185, 222–235. doi: 10.1016/j.neuroimage.2018.10.019

Xia, S., Song, T., Che, J., Li, Q., Chai, C., Zheng, M., et al. (2017). Altered brain functional activity in infants with congenital bilateral severe sensorineural hearing loss: a resting-state functional MRI study under sedation. *Neural Plast.* 2017:8986362. doi: 10.1155/2017/8986362

Xu, H., Fan, W., Zhao, X., Li, J., Zhang, W., Lei, P., et al. (2016). Disrupted functional brain connectome in unilateral sudden sensorineural hearing loss. *Hear. Res.* 335, 138–148. doi: 10.1016/j.heares.2016.02.016

Xu, X. M., Jiao, Y., Tang, T. Y., Zhang, J., Lu, C. Q., Salvi, R., et al. (2019). Sensorineural hearing loss and cognitive impairments: contributions of thalamus using multiparametric MRI. *J. Magn. Reson. Imaging* 50, 787–797. doi: 10.1002/jmri.26665

Yan, C.-G., and Zang, Y.-F. (2010). DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. *Front. Syst. Neurosci.* 4:13. doi: 10.3389/fnsys.2010.00013

Yang, M., Chen, H. J., Liu, B., Huang, Z. C., Feng, Y., Li, J., et al. (2014). Brain structural and functional alterations in patients with unilateral hearing loss. *Hear. Res.* 316, 37–43. doi: 10.1016/j.heares.2014.07.006

Zalesky, A., Fornito, A., and Bullmore, E. T. (2010). Network-based statistic: Identifying differences in brain networks. *Neuroimage* 53, 1197–1207. doi: 10.1016/j.neuroimage.2010.06.041

Zhang, Y., Mao, Z., Feng, S., Liu, X., Lan, L., Zhang, J., et al. (2018). Altered functional networks in long-term unilateral hearing loss: a connectome analysis. *Brain Behav.* 8:e00912. doi: 10.1002/brb3.912

Zhang, Z., Liao, W., Chen, H., Mantini, D., Ding, J. R., Xu, Q., et al. (2011). Altered functional-structural coupling of large-scale brain networks in idiopathic generalized epilepsy. *Brain* 134, 2912–2928. doi: 10.1093/brain/awr223

Zhao, T., Mishra, V., Jeon, T., Ouyang, M., Peng, Q., Chalak, L., et al. (2019). Structural network maturation of the preterm human brain. *Neuroimage* 185, 699–710. doi: 10.1016/j.neuroimage.2018.06.047

Zwolan, T. A., Ashbaugh, C. M., Alarfaj, A., Kileny, P. R., Arts, H. A., El-Kashlan, H. K., et al. (2004). Pediatric cochlear implant patient performance as a function of age at implantation. *Otol. Neurotol.* 25, 112–120. doi: 10.1097/00129492-200403000-00006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 04 February 2022
doi: 10.3389/fnins.2021.828512

An Invertible Dynamic Graph Convolutional Network for Multi-Center ASD Classification

Yueying Chen[1,2], *Aiping Liu*[1,2]*, *Xueyang Fu*[1,2], *Jie Wen*[3] and *Xun Chen*[1,2]

[1] School of Information Science and Technology, University of Science and Technology of China, Hefei, China, [2] USTC IAT-Huami Joint Laboratory for Brain-Machine Intelligence, Institute of Advanced Technology, University of Science and Technology of China, Hefei, China, [3] Division of Life Sciences and Medicine, Department of Radiology, The First Affiliated Hospital of USTC (Anhui Provincial Hospital), University of Science and Technology of China, Hefei, China

Autism Spectrum Disorder (ASD) is one common developmental disorder with great variations in symptoms and severity, making the diagnosis of ASD a challenging task. Existing deep learning models using brain connectivity features to classify ASD still suffer from degraded performance for multi-center data due to limited feature representation ability and insufficient interpretability. Given that Graph Convolutional Network (GCN) has demonstrated superiority in learning discriminative representations of brain connectivity networks, in this paper, we propose an invertible dynamic GCN model to identify ASD and investigate the alterations of connectivity patterns associated with the disease. In order to select explainable features from the model, invertible blocks are introduced in the whole network, and we are able to reconstruct the input dynamic features from the network's output. A pre-screening of connectivity features is adopted to reduce the redundancy of the input information, and a fully-connected layer is added to perform classification. The experimental results on 867 subjects show that our proposed method achieves superior disease classification performance. It provides an interpretable deep learning model for brain connectivity analysis and is of great potential in studying brain-related disorders.

Keywords: fMRI, graph convolutional networks, invertible networks, brain connectivity networks, autism spectrum disorder, disease classification

OPEN ACCESS

Edited by:
Yu Zhang,
Lehigh University, United States

Reviewed by:
Qibin Zhao,
RIKEN, Japan
Yulun Zhang,
ETH Zürich, Switzerland

***Correspondence:**
Aiping Liu
aipingl@ustc.edu.cn

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *03 December 2021*
Accepted: *23 December 2021*
Published: *04 February 2022*

Citation:
Chen Y, Liu A, Fu X, Wen J and Chen X (2022) An Invertible Dynamic Graph Convolutional Network for Multi-Center ASD Classification. Front. Neurosci. 15:828512. doi: 10.3389/fnins.2021.828512

1. INTRODUCTION

As one of the most common neurodevelopmental disorders, the exact etiology of Autism Spectrum Disorder (ASD) remains unknown. In the past 50 years, ASD has gone from a narrowly defined, rare disorder of childhood to a well-publicized disease, and recognized as a very common and heritable brain disorder. The major characteristic of ASD is being deficit in social interaction and social communication with repetitive and unusual behaviors and activities (Lord et al., 2018). Despite medical progress, the diagnosis of ASD still depends on the symptom-based clinical criteria with complex diagnostic steps. However, with increasing recognition of the importance of early diagnosis for effective intervention, more effort has been made on exploring other possible modalities and biomarkers for ASD identification.

With the development of neuroimaging technologies, resting-state functional Magnetic Resonance Imaging (rs-fMRI) has attracted increasing interest in ASD studies, which enjoys advantages of superior spatial resolution to accurately locate the active areas in the whole brain, overcoming the limitations of earlier tools such as positron emission tomography (PET),

electroencephalography (EEG), and magnetoencephalography (MEG). By computing the correlation between fMRI time series of different regions of interests (ROIs), we can construct a functional connectivity network and many disorders may lead to the alterations in it (Li et al., 2016; Miller et al., 2016; Bachmann et al., 2018; Chandra et al., 2019; Zhang et al., 2020). For example, a widespread decrease of functional connectivity strengths was reported in patients with Alzheimer's Disease (AD) (Demirtaş et al., 2017). Studies showed that regional connectivity changes (both increase and decrease) of dopaminergic cortico-striatal and mesolimbic-striatal loops have been found in PD subjects (Filippi et al., 2018). ASD has also been suggested to be related to altered brain connectivity in the development of disease and has been extensively investigated (Kleinhans et al., 2008; Monk et al., 2009; Yerys et al., 2015; Dajani and Uddin, 2016; Xu et al., 2020). While a wide range of connectivity changes are reported, inconsistent conclusions have been observed in studies of functional connectivity in ASD, indicating the importance to thoroughly investigate the connectivity patterns with a large population of ASD.

Based on brain connectivity networks, machine learning, especially deep learning methods have further provided powerful tools to extract representative features associated with ASD and have greatly deepened our understanding of the disease (Chan et al., 2020). The classical machine learning techniques such as Support Vector Machines (SVM) are most widely used to identify patients from healthy controls in various studies (Subbaraju et al., 2017). For instance, Abraham et al. (2017) achieved 66.8% classification accuracy on 871 subjects obtained from ABIDE dataset.

Neural networks and deep learning methods such as autoencoder, Deep Neural Network (DNN) (Guo et al., 2017), Long Short Term Memory (LSTM) (Dvornek et al., 2018), and Convolutional Neural Network (CNN) (Haweel et al., 2021) have generated better performance in ASD classification. For instance, Yin et al. (2021) applied a DNN model and achieved the classification accuracy of 76.2% on 871 subjects of ABIDE dataset, and further improved the performance to an accuracy of 79.2% by combining DNN with an autoencoder.

Compared with traditional deep learning models, Graph Convolutional Network (GCN) can deal with data of non-Euclidean structure, which may be more suitable, and more interpretable for brain connectivity graph generated by fMRI. GCN has been used to classify ASD and select biomarkers from typical developing subjects (Ktena et al., 2018; Parisot et al., 2018). Recently, with a connectivity-based GCN model, 70.7% accuracy for classifying 1057 subjects (525 ASD and 532 healthy controls) has been reported (Wang et al., 2021). It's worth noting that when integrating information from more modalities, we may obtain higher classification accuracy. For instance, 85.06% of accuracy in ASD classification has been reported in Rakić et al. (2020) based on both structural MRI (sMRI) and fMRI features of 368 ASD and 449 healthy control subjects using an autoencoder model. While more modalities are beneficial to disease identification, it requires extra resources on data collection. In this paper, we are more interested in resting-state fMRI and focus on the ASD classification using brain connectivity features based on fMRI signals.

However, most deep learning models are limited in interpretation because of their black box representation. Although the classification performances of most deep learning networks are superior to those of traditional or interpretable methods, the features they finally generate can hardly be corresponded to the inputs, challenging the selection of helpful biomarkers. To overcome this shortcoming, Jacobsen et al. (2018) proposed an invertible network using a fully-connected layer as an inner trainable network, which can accurately reconstruct the inputs to a layer from its outputs without any degradation of classification accuracy. Given its superiority, Zhuang et al. (2019) proposed an invertible network for ASD classification, and gained 71% accuracy on the whole ABIDE dataset.

To improve the model interpretability and to better utilize structural, spatial, and temporal characteristics of brain connectivity networks, in this paper, we propose an invertible dynamic GCN (ID-GCN) model for ASD classification. More specifically, invertible blocks are utilized in the whole network, capable of reconstructing the input features from the output of the network, followed by a fully-connected layer to perform classification. Additionally, we select the connectivity features with a pre-screening operation to reduce the redundancy of the input information. The proposed method is verified on multi-center ABIDE datasets and the results demonstrate its effectiveness for disease classification and potential for studying the disease-related connectivity features. The contributions of this paper are summarized as:

- An invertible graph convolutional network is designed for disease classification based on brain connectivity networks. It is capable of generating disease-related interpretable connectivity features and improving classification accuracy.
- The proposed model integrates the structural, spatial, and dynamic information of the brain connectivity networks, and a prior selection of the features is adopted to reduce the redundancy of the input features.
- The proposed method has been validated on ABIDE dataset with superior performance.

2. METHODS

In this section, we first provide the notations and their definitions used in this paper, then we introduce our proposed invertible dynamic GCN model in detail.

2.1. Notations and Definitions

In this paper, we use $G(V, E)$ to represent a graph, where $V = \{v_1, v_2, ..., v_n\}$ is the set of nodes, and $E = \{e_{ij}\}$ is the set of edges. In the spatial connectivity graph, e_{ij} represents the Euclidean distance of two connected nodes, and in the functional connectivity graph, e_{ij} represents their connectivity strength. Additionally, let A denote the adjacency matrix of the graph and X denote the correlation matrix, in which every row represents a node's features.

2.2. Graph Convolutional Network

Graph Convolutional Network is a deep learning architecture, which can not only use the data itself but also the relationship between data represented as a graph. Through the adjacency matrix A of the graph, we can first calculate the normalized Laplacian matrix of X, which calculation formula is:

$$L = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \tag{1}$$

Where I is an identity matrix and D is the diagonal degree matrix of X. Then, we get an eigendecomposition of the Laplacian matrix, $L = U \Lambda U^T$, where U is a set of orthonormal eigenvectors, and $\Lambda = \text{diag}(\lambda_0, ..., \lambda_{n-1})$ is the matrix's non-negative eigenvalues. Based on these formulas, we get the propagation rule of graph convolution layers is:

$$X^l = \sigma(U \Theta(\Lambda) U^T X^{l-1}) \tag{2}$$

Where σ is the activation function of the layer, and $\Theta(\cdot)$ is the GCN convolution kernel. To simplify the calculation, we then fit the kernel by Chebyshev polynomials of order k (Hammond et al., 2011), which can be derived from:

$$T_k(c) = 2cT_{k-1}(c) - T_{k-2}(c) \tag{3}$$

$$T_0(c) = I, T_1(c) = c \tag{4}$$

And the fitting formula is:

$$\Theta(\Lambda) = \sum_{k=0}^{K-1} \beta_k T_k(\tilde{\Lambda}) \tag{5}$$

$$\tilde{\Lambda} = \frac{2}{\lambda_{max}} \Lambda - I \tag{6}$$

Where β_k is the weight coefficient of the kth Chebyshev polynomial, and λ_{max} is the max eigenvalue of the Laplacian matrix. Since the calculation of Chebyshev polynomials is performed only on eigenvectors Λ, it does not affect other matrix operations like doing eigendecomposition. So the Equation (2) can be expressed as:

$$X^l = \sigma((\sum_{k=0}^{K-1} \beta_k T_k(\tilde{L})) X^{l-1}) \tag{7}$$

Where $\tilde{L}$ is defined as $\tilde{L} = \frac{2}{\lambda_{max}} L - I$. Then we substitute the trainable weight matrix W for β_k, and get the final propagation rule of graph convolution layers as:

$$X^l = \sigma(\sum_{k=0}^{K-1} T_k(\tilde{L})(X^{l-1}) W) \tag{8}$$

2.3. Invertible Block

The architecture of the invertible block is shown in **Figure 1**, where the inputs are x_1 and x_2, and the outputs are denoted as z_1 and z_2. Those feature maps have the same shape, and φ and ω can be defined as any functions. In this model, we define φ and ω as independent GCN modules using different graphs as their inputs, which will be introduced in detail in the next section. In order to fully blend the advantages of the two GCN modules, the outputs of the first block y_1 and y_2, are then calculated to their average and half of their difference as z_1 and z_2. This invertible block can reconstruct the input from its output, where the forward pass and inverse are:

$$\begin{cases} y_1 = x_1 + \varphi(x_2) \\ y_2 = x_2 + \omega(y_1) \end{cases} \quad \begin{cases} z_1 = 0.5(y_1 + y_2) \\ z_2 = 0.5(y_2 - y_1) \end{cases} \tag{9}$$

$$\begin{cases} x_2 = y_2 - \omega(y_1) \\ x_1 = y_1 - \varphi(x_2) \end{cases} \quad \begin{cases} y_1 = z_1 - z_2 \\ y_2 = z_1 + z_2 \end{cases} \tag{10}$$

2.4. Invertible Dynamic GCN

In order to incorporate additional spatial and temporal characteristics of the brain functional connectivity network constructed by rs-fMRI data with better interpretability, we propose an invertible dynamic GCN (ID-GCN) model, which uses two different GCN as the function φ and ω in the invertible blocks to encode the functional connectivity graph and spatial connectivity graph of samples, respectively. The functional GCN, i.e., ω in the invertible block, uses the functional graph of each subject obtained by the correlation matrix. Meanwhile, the skeleton of the spatial graph is calculated directly according to the spatial distance between ROIs, and the connection weights are their correlation values. It is represented as φ for spatial GCN. The whole model includes three invertible blocks to extract explainable high dimensional features, and the inputs x_1 and x_2 of

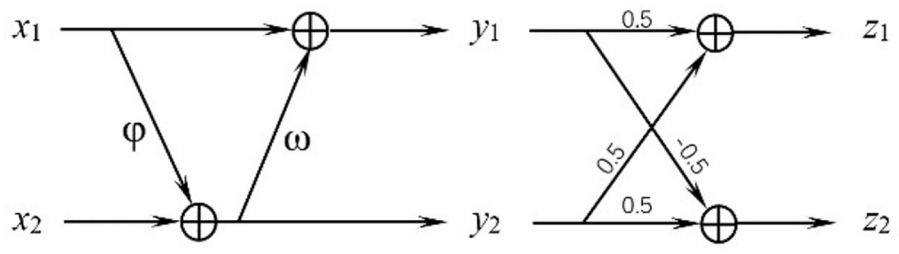

FIGURE 1 | Structure of the invertible block.

the first block are the same features that we send into the model. The proposed ID-GCN architecture for disease classification in this work is demonstrated in **Figure 2**.

To improve the computational efficiency and simplify the training process, for each node, the k connected nodes with the largest Pearson correlation coefficients in the functional graph or the smallest distance in the spatial graph are retained to construct a k-nearest graph. The correlation coefficients between each node and all other nodes are used as the sample's features which serve as input into the ID-GCN model. A fully connected layer with softmax is applied to perform the classification and the source of the collection site is included as an additional covariate. The cross-entropy is adopted in this model as loss function as:

$$L = \frac{1}{N}\sum_{i} -y_i * log(\hat{y}_i) - (1 - y_i) * log(1 - \hat{y}_i) \qquad (11)$$

where y_i is the label of the ith subject, $\hat{y}_i$ is the output of the network, and N is the number of subjects we use.

While there are usually hundreds of ROIs defined from the atlas, for a certain disease, it usually involves the changes of a portion of brain regions. Additionally, with a great individual variance of connectivity patterns, a large number of connectivity features may be easily disturbed by noise, affecting subsequent analysis and interpretation. However, reducing the number of ROIs in the input model may inevitably cause the loss of information. Therefore, rather than reducing the entire number of ROIs, we reduce the dimension of the input features of each ROI individually by selecting the M most important features for disease classification using random forest.

As our brains are a dynamic system, time-varying connectivity features have been suggested to be related to the functioning of our brain. Thus, in this model, we further utilize the dynamics of connectivity as additional features for ASD classification. The time sliding window is applied to sample the time-dependent signals and get the correlation matrix X_t of each time window. The temporal variations of dynamic connectivity are then calculated as the auxiliary feature represented as Ft, which is concatenated with other connectivity features. After the pre-selection of random forest, the reserved feature matrix $\{F_t\}$ is combined with the selected feature F of the original correlation matrix X as the final input features. The overview of the proposed model is shown in **Figure 3**.

3. EXPERIMENTS AND RESULTS

3.1. Real Dataset and Experimental Setting

We validated the proposed method on the publicly available ABIDE dataset (Martino et al., 2014), and chose 416 ASD subjects and 451 healthy controls (HC) from 13 acquisition sites. The phenotypical information of each acquisition site can refer to **Table 1**. The dataset was preprocessed with the Configurable Pipeline for the Analysis of Connectomes (C-PAC) (Sikka et al., 2013), which includes skull striping, slice timing correction, motion correction, global mean intensity normalization, nuisance signal regression, and band-pass filtering (0.01–0.1 Hz). The fMRI images were registered to the standard anatomical space (MNI152). To define brain areas, the Harvard Oxford (HO) atlas was chosen, consisting of 110 ROIs. More details of the dataset may refer to ABIDE Preprocessed.

We implemented the proposed model in a 5-fold cross-validation setting, using 80% of the data for training and 20% for testing. We set the pre-selected feature number M as 48, combined with $J = 10$ auxiliary dynamic features. Additionally, the Chebyshev polynomial order was chosen as 3, and $k = 3$ nearest nodes were selected to generate our graphs.

To test the proposed method, we compared it with other methods including siamese GCN (Ktena et al., 2018), Random Forest, SVM, and GCN, evaluating its performance improvement induced by the combination of spatial and dynamic connectivity features, and testing the effectiveness of pre-screening on the features. In these models for comparison, features input in siamese GCN is the paired subject features as implemented

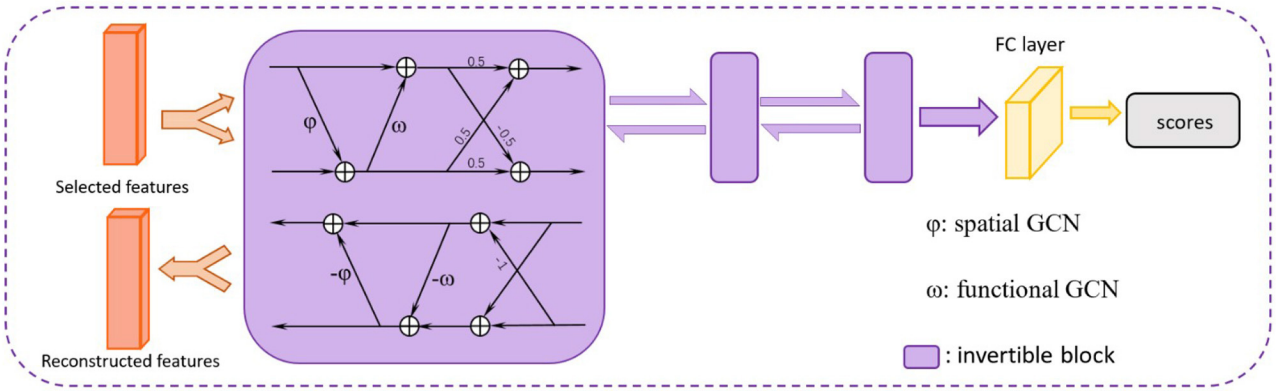

FIGURE 2 | The proposed ID-GCN architecture. The selected features are trained in three invertible blocks. A fully connected (FC) layer is finally used to obtain the output scores for ASD classification. The whole network is reversible before the FC layer, meaning that we can reconstruct the informative disease-related brain connectivity patterns by selecting important output features of the network.

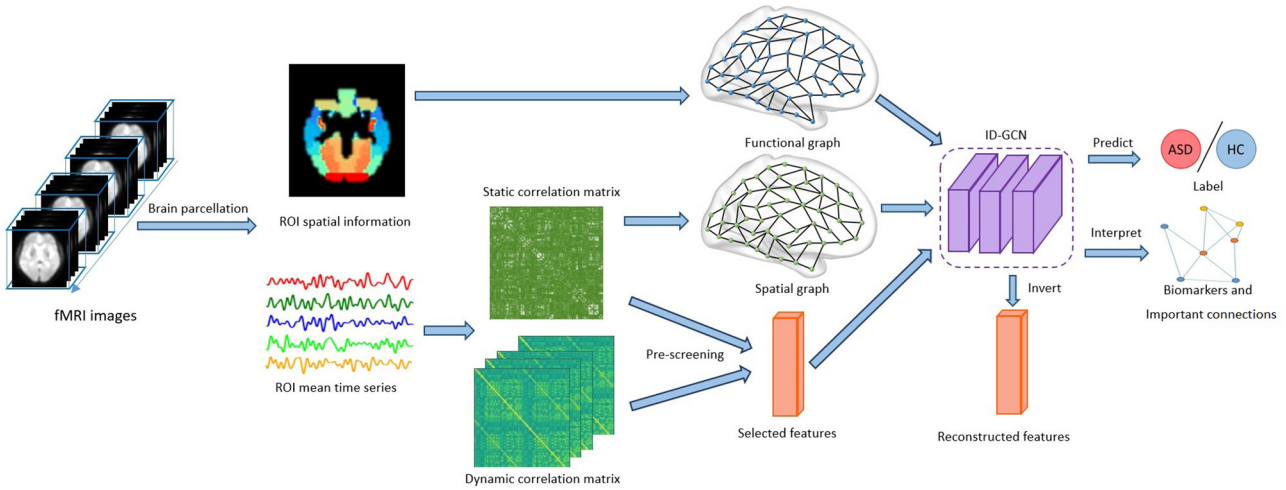

FIGURE 3 | Overview of the proposed framework. The brain connectivity features inferred from the fMRI time series and brain parcellation are fed to the model. After training using the ID-GCN model, we obtain the predictions for ASD classification, and important brain connectivity features are selected accordingly.

TABLE 1 | Phenotypical information summary of ABIDE data.

Site	ASD	HC	Gender (M/F)	Total	Age (mean±std)
PITT	30	27	49/8	57	18.9±6.8
TRINITY	24	25	49/0	49	17.2±3.6
UM_1	55	55	84/26	110	13.4±2.9
UM_2	13	22	33/2	35	16±3.3
USM	58	43	101/0	101	22.1±7.6
YALE	28	28	40/16	56	12.7±2.9
LEUVEN_1	14	15	29/0	29	22.6±3.5
LEUVEN_2	15	20	27/8	35	14.2±1.4
KKI	22	33	42/13	55	10.1±1.3
NYU	79	105	147/37	184	15.3±6.6
UCLA_1	41	32	63/10	73	13.2±2.4
UCLA_2	13	13	24/2	26	12.5±1.5
MAX_MUN	24	33	50/7	57	26.2±11.9
TOTAL	416	451	738/129	867	16.4±7.1

in the study (Ktena et al., 2018), while the other models use the whole connectivity matrix of a single subject as inputs. All the methods were evaluated in terms of accuracy, AUC value, precision, recall, and F1-score. The definitions of them are as follows:

$$Accuracy = (TP + TN)/n \tag{12}$$

$$Precision = TP/(TP + FP) \tag{13}$$

$$Recall = TP/(TP + FN) \tag{14}$$

$$F1 - score = 2 * Precision * Recall/(Precision + Recall) \tag{15}$$

where n is the total number of our subject, TP is true positive subject's number, TN is true negative, FP is false positive, FN denotes false negative, and AUC means the area under the ROC curve. We additionally performed ablation experiments to demonstrate the effects of each step of our method, including (1) GCN using the functional graph as input (GCN); (2) GCN using the spatial and functional graph in different layers (GCN adding spatial information); (3) ID-GCN with principal component analysis (PCA) for feature selection (ID-GCN with PCA); and (4) ID-GCN without dynamic features.

4. RESULTS

The classification results are shown in **Figure 4** and **Table 2**. It's noted that our proposed model, ID-GCN achieves the highest classification accuracy as 76.3%. Specifically, our model demonstrates great improvement in all the evaluation metrics compared with traditional SVM and Random Forest models and obtains 3.1% gains in accuracy compared with GCN using the same hyperparameters. Siamese GCN used paired subject features as input and generated classification results by multiplying two feature matrices from shared weight GCN. However, it's noticed that siamese GCN demonstrated worse performance on the given dataset where the paired features didn't successfully distinguish the subjects in this case.

Considering that the classification performance depends on the number of subjects, in order to have a fair comparison, we have tested our algorithm on the different number of subjects and show comparison with other state-of-the-art methods in **Table 3**. More specifically, we chose the number of subjects as 95, 459, 867, and 1,066, respectively. As they were examined on a different number of subjects, we didn't repeat their experiments but reported their datasets and results, only using same order of magnitude of subjects to run our model for better comparison. It can be seen that our results outperform other methods on the

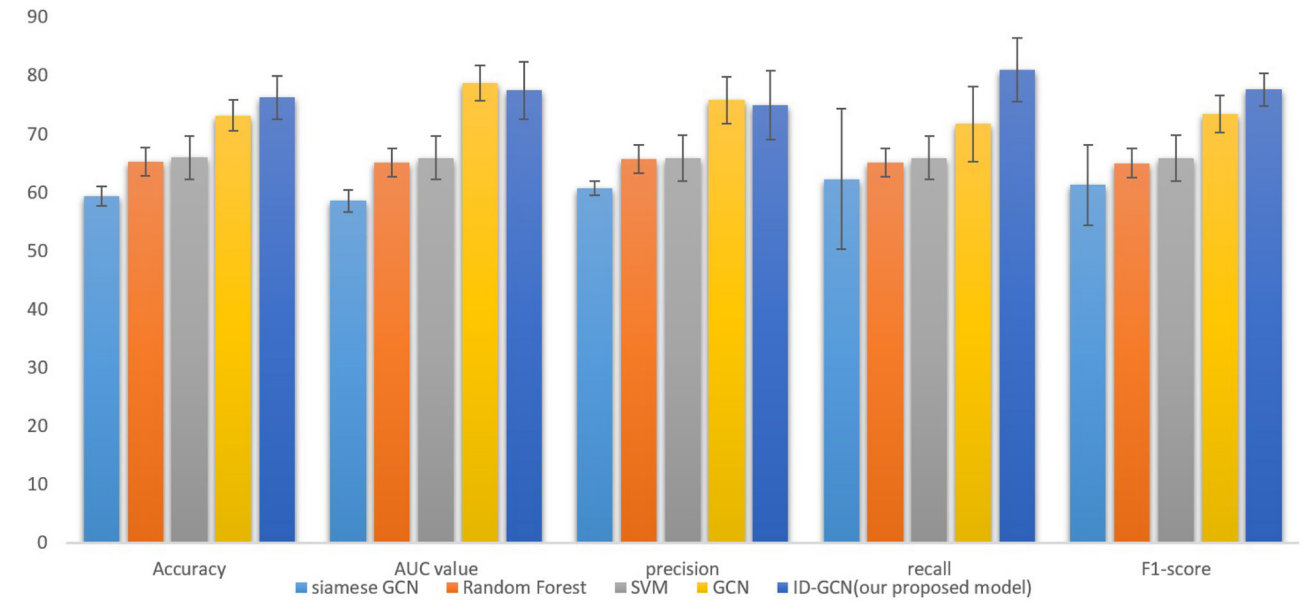

FIGURE 4 | Comparison with traditional and GCN models including siamese GCN (Ktena et al., 2018), Random Forest, SVM, and GCN.

TABLE 2 | Comparisons of different methods.

Model	Accuracy	AUC	Precision	Recall	F1-score
SVM	66.0±3.7%	65.9±3.7%	65.9±3.9%	65.9±3.7%	65.9±3.9%
Random forest	65.3±2.4%	65.1±2.4%	65.7±2.4%	65.1±2.4%	65.0±2.5%
GCN	73.2±2.7%	**78.7**±3.0%	**75.8**±4.0%	71.7±6.5%	73.4±3.2%
Siamises GCN	59.4±1.7%	58.6±1.9%	60.7±1.2%	62.3±12.0%	61.3±6.9%
ID-GCN(our model)	**76.3**±3.7%	77.5±4.9%	75±5.9%	**81.0**±5.5%	**77.6**±2.8%

TABLE 3 | Comparison with other SOTA methods.

Model	Number of subjects	Accuracy
DNN (Li et al., 2018)	95	85.3%
Combined MCNNEs (Aghdam et al., 2019)	459	70.45%
CNN-EW (Xing et al., 2018)	1096	66.88%
ASD-DiagNet (Eslami et al., 2019)	1035	70.1%
cGCN (Wang et al., 2021)	1057	70.7%
3D CNN (Thomas et al., 2020)	1162	64%
ID-GCN(our model)	95	**87.38**%
	459	**77.42**%
	867	**76.3**%
	1066	**71.44**%

same order of magnitude of data. It's worth noting that with data from different centers, the accuracy may vary. As demonstrated in **Table 3**, we can notice that more subjects do not guarantee better performance which is partially due to the great inter-center and inter-subject variability. When using 95 subjects from the same acquisition center, both (Li et al., 2018) and our method achieve high classification accuracy, and our proposed method obtains better classification performance compared with that of Li et al. (2018). Furthermore, the model performance of every single center is provided in **Table 4** that we train all the subjects and test the proposed method for each center separately. It shows that the classification accuracy varies across the centers, indicating great inter-center variability.

Additionally, the studies with multimodality data often demonstrate better performances using the same method. For example, Rakić et al. (2020) gained 85.06% of accuracy using both sMRI and fMRI features in the classification of 817 subjects. However, in this paper, we focus on the functional connectivity features. Although the proposed method has improved the classification accuracy compared with other GCN models and has interpretability, it still has several limitations. The temporal variations of brain connectivity have been utilized to represent the dynamics of brain connectivity. However, it's unable to fully delineate the time-varying connectivity. The classification accuracy of our interpretable model is limited compared with some networks without interpretability. For better performance, RNN model with temporal connectivity networks will be explored in our future work. Additionally, the biological interpretation of the biomarkers selected from our invertible network has been limited investigated. The effective

center-invariant biomarkers with sufficient biological meanings are warranted in future studies.

The results of ablation experiments are demonstrated in **Table 5**. It can be seen from the table that after adding spatial information as graph input, the accuracy of the model increased by over 1%, indicating the importance of the spatial information. As the number of connectivity features is large, great individual variation and noise may disturb the robust feature learning and degrade the classification performance. The feature selection, therefore, contributed to a significant improvement in the classification accuracy. We also evaluated other dimension reduction approach, i.e., Principal Component Analysis, for feature selection. As shown in **Table 5**, PCA led to less improvement in the classification accuracy. It may be due to the difficulty in the alignment of principal components across the subjects. Moreover, the temporal dynamics benefited the GCN model with a small accuracy gain.

In order to better understand ASD, we further identified the disease-related features by sorting the importance of each node's features extracted under the 5-fold cross-validation. The top 10% important connectivity edges were reconstructed as demonstrated in **Figure 5** and **Table 6**. It's noted that the connections between Right Pallidum and Right Inferior Frontal Gyrus, Left Frontal Orbital Cortex and Left Central Opercular Cortex, and connections involving Left Supramarginal Gyrus and Right Inferior Temporal Gyrus greatly contributed to the classification accuracy. Additionally, we evaluated the impacts of nodes by excluding each node and examining its influence on classification performance. With such lesion operation, we were able to assess the importance of each node. As shown in **Figure 6**, the highly-rated ROIs include Right Pallidum, Right Inferior Frontal Gyrus (triangle part), Right Inferior Temporal Gyrus (anterior division), Left Frontal Orbital Cortex, Left Temporal Fusiform Cortex (posterior division), and Right Temporal Occipital Fusiform Cortex, indicating their potential ROIs for ASD.

TABLE 4 | Model performance in each single center.

Site	Number of subjects	Accuracy
PITT	57	71.7±6.7%
TRINITY	49	72.0±11.7%
UM_1	110	75.5±3.6%
UM_2	35	82.9±10.7%
USM	101	84.8±7.0%
YALE	56	80.0±6.7%
LEUVEN_1	29	73.3±8.3%
LEUVEN_2	35	74.3±10.7%
KKI	55	74.5±8.9%
NYU	184	76.2±5.2%
UCLA_1	73	74.7±8.8%
UCLA_2	26	83.3±18.2%
MAX_MUN	57	66.6±11.9%
TOTAL	867	76.3±3.7%

TABLE 5 | Ablation study on the effects of different components.

Model	Accuracy
GCN	73.2%
GCN adding spatial information	74.5%
ID-GCN with PCA	74.2%
ID-GCN without dynamic features	76.1%
ID-GCN(our model)	**76.3**%

5. DISCUSSION AND CONCLUSION

The early diagnosis of ASD is a challenging task as great variations exist in the symptoms. In addition to the clinical criterion, researchers have tried to identify the effective neuroimaging biomarkers for the better diagnosis of ASD. Brain connectivity features are promising for studying ASD as

TABLE 6 | Important connectivity edges selected by feature reconstruction.

ROI1	ROI2
Right Pallidum	Right Inferior Frontal Gyrus
Left Frontal Orbital Cortex	Left Central Opercular Cortex
Left Temporal Fusiform Cortex (posterior division)	Left Heschl's Gyrus (includes H1 and H2)
Left Supramarginal Gyrus (anterior division)	Right Temporal Occipital Fusiform Cortex
Left Supramarginal Gyrus (posterior division)	Left Frontal Orbital Cortex
Right Inferior Temporal Gyrus (anterior division)	Left Supramarginal Gyrus (anterior division)
Right Inferior Temporal Gyrus (anterior division)	Left Lateral Occipital Cortex (inferior division)

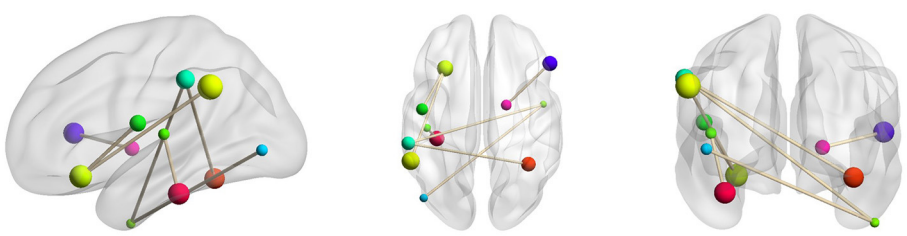

FIGURE 5 | Selected key connectivity features for ASD classification.

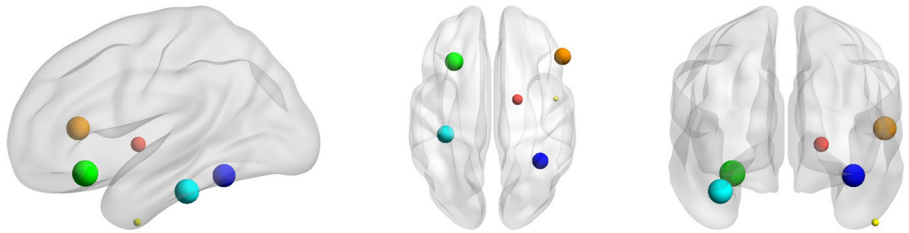

FIGURE 6 | Selected key ROIs for ASD classification, including Right Pallidum (red), Right Inferior Frontal Gyrus (triangle part) (orange), Right Inferior Temporal Gyrus (anterior division) (yellow), Left Frontal Orbital Cortex (green), Left Temporal Fusiform Cortex (posterior division) (cyan), and Right Temporal Occipital Fusiform Cortex (blue).

TABLE 7 | The classification accuracy with different k.

The value of k	2	3	4	5	6	8	10	15	20
Accuracy	73.7%	**76.3**%	75.1%	76.0%	75.0%	75.0%	75.8%	74.7%	75.3%

TABLE 8 | The classification accuracy with different M.

The value of M	10	30	48	50	70	90	110
Accuracy	72.1%	73.6%	**76.3**%	76.0%	75.7%	74.6%	73.9%

widespread connectivity changes have been observed in ASD. With various statistical and machine learning methods, we have largely expanded our understanding of the disease. However, the classification performance based on brain connectivity features is still limited, partially due to the insufficient representation ability for multi-center ASD data. It's, therefore, critical to learn the robust connectivity features for better representing the disease population. While the deep learning-based methods are promising, mos of them are designed in a black-box principle, challenging their biological interpretability.

In this study, we propose an explainable graph convolutional network, namely ID-GCN for multi-center ASD data classification and investigation by incorporating the functional, spatial and temporal information of the connectivity networks and using the invertible network to select interpretable biomarkers. The use of GCN aims to integrate the high-dimensional features of each node, and the invertible network is capable of reconstructing the extracted disease-related features back to the original connectivity graph. The proposed model contains two different GCN for brain functional connectivity and spatial connectivity, respectively. A random forest is adopted to narrow the feature space and reduce the redundancy of the data. We further integrate the dynamics of brain connectivity as important features for ASD classification. The experimental results on ABIDE dataset suggest the efficacy of our model. It is a potential classifier for large multi-center datasets despite their variations.

When classifying the ASD subjects, several connectivity features reconstructed by the model are assigned with higher importance. Those connections involve Right Pallidum, Right Inferior Frontal Gyrus, Left Frontal Orbital Cortex, Left Central Opercular Cortex, Left Temporal Fusiform Cortex, Right Temporal Occipital Fusiform Cortex, Left Supramarginal Gyrus and Right Inferior Temporal Gyrus, which are mostly consistent with the prior studies. For instance, the altered connectivity of Temporal Pole, Pallidum, and Frontal Orbital Cortex in ASD has been reported in Yerys et al. (2015); Dajani and Uddin (2016); Monk et al. (2009). In another line of studies, the changes of connectivity patterns in Fusiform Gyrus and Inferior Frontal Gyrus have been investigated for ASD subjects (Kleinhans et al., 2008; Xu et al., 2020). We additionally performed lesion analysis that sequentially removed each ROI and examined its impact on the classification accuracy. According to their contributions to the classification performance, eight ROIs including Right Superior Temporal Gyrus, Right Superior Frontal Gyrus, Right Pallidum, Right Inferior Frontal Gyrus (triangle part), Right Inferior Temporal Gyrus (anterior division), Left Frontal Orbital Cortex, Left Temporal Fusiform Cortex (posterior division), and Right Temporal Occipital Fusiform Cortex were chosen which are mostly involved in the connectivity features reconstructed by ID-GCN. It further substantiates the explainable features learned by the proposed method.

There are several parameters that need to be determined in the proposed model, and we have evaluated the impacts of different parameters on classification performance. **Table 7** demonstrates the classification accuracy as a function of the numbers of neighbors. It's observed that classification performance depends on the values of k, and when k=3, we obtained the highest classification accuracy. It indicates that there may be only a few connected areas that are most robust across the subjects. We have also chosen the number of features M using the grid search in **Table 8**, and when M=48, it achieved the best performance. If the number of M is too small or too large, the performance of the model will decline greatly.

While the proposed method is capable to identify the disease-related features and achieves a competitive classification

performance, it still has several limitations. The temporal variations of brain connectivity have been utilized to represent the dynamics of brain connectivity. However, it's unable to fully delineate the time-varying connectivity patterns, which can be further extended in our future work. The classification accuracy of our interpretable model is limited compared with some recent networks without interpretable modules. To further improve the performance, RNN models with temporal connectivity networks can be potential. Additionally, the biological interpretation of the biomarkers selected from our invertible network has been limited investigated. The effective center-invariant biomarkers with sufficient biological meanings are warranted in future studies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

YC, AL, and XF worked on the method and analyzed the data. JW interpreted the results. AL and XC supervised the project. The manuscript was drafted by YC and AL. All the authors have reviewed and revised the manuscript.

FUNDING

This study was supported in part by the National Natural Science Foundation of China (Grants 61922075, 61701158 and 22077116), and in part by the USTC Research Funds of the Double First-Class Initiative (Grants YD2100002004 and YD9110002011).

ACKNOWLEDGMENTS

The authors thank the publicly available dataset provider and the professors of the University of Science and Technology of China.

REFERENCES

Abraham, A., Milham, M. P., Martino, A. D., Craddock, R. C., Samaras, D., Thirion, B., et al. (2017). Deriving reproducible biomarkers from multi-site resting-state data: an autism-based example. *NeuroImage* 147, 736–745. doi: 10.1016/j.neuroimage.2016.10.045

Aghdam, M. A., Sharifi, A., and Pedram, M. M. (2019). Diagnosis of autism spectrum disorders in young children based on resting-state functional magnetic resonance imaging data using convolutional neural networks. *J. Digit. Imag.* 32, 899–918. doi: 10.1007/s10278-019-00196-1

Bachmann, K., Lam, A. P., Sörös, P., Kanat, M., Hoxhaj, E., Matthies, S., et al. (2018). Effects of mindfulness and psychoeducation on working memory in adult adhd: A randomised, controlled fmri study. *Behav. Res. Therapy* 106, 47–56. doi: 10.1016/j.brat.2018.05.002

Chan, H.-P., Samala, R. K., Hadjiiski, L. M., and Zhou, C. (2020). Deep learning in medical image analysis. *Adv. Exp. Med. Biol.* 1213, 3–21. doi: 10.1007/978-3-030-33128-3_1

Chandra, A., Dervenoulas, G., Politis, M., and Initiative, A. D. N. (2019). Magnetic resonance imaging in alzheimer's disease and mild cognitive impairment. *J. Neurol.* 266, 1293–1302. doi: 10.1007/s00415-018-9016-3

Dajani, D. R., and Uddin, L. Q. (2016). Local brain connectivity across development in autism spectrum disorder: a cross-sectional investigation. *Autism Res.* 9, 43–54. doi: 10.1002/aur.1494

Demirtaş, M., Falcon, C., Tucholka, A., Gispert, J. D., Molinuevo, J. L., and Deco, G. (2017). A whole-brain computational modeling approach to explain the alterations in resting-state functional connectivity during progression of alzheimer's disease. *NeuroImage Clin.* 16, 343–354. doi: 10.1016/j.nicl.2017.08.006

Dvornek, N. C., Yang, D., Ventola, P., and Duncan, J. S. (2018). "Learning generalizable recurrent neural networks from small task-fmri datasets," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Vol. 11, eds A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger (Cham: Springer International Publishing), 329–337.

Eslami, T., Mirjalili, V., Fong, A., Laird, A. R., and Saeed, F. (2019). Asd-diagnet: a hybrid learning approach for detection of autism spectrum disorder using fmri data. *Front. Neuroinformat.* 13:70. doi: 10.3389/fninf.2019.00070

Filippi, M., Elisabetta, S., Piramide, N., and Agosta, F. (2018). Functional mri in idiopathic parkinson's disease. *Int. Rev. Neurobiol.* 141, 439–467. doi: 10.1016/bs.irn.2018.08.005

Guo, X., Dominick, K. C., Minai, A. A., Li, H., Erickson, C. A., and Lu, L. J. (2017). Diagnosing autism spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method. *Front. Neurosci.* 11:460. doi: 10.3389/fnins.2017.00460

Hammond, D. K., Vandergheynst, P., and Gribonval, R. (2011). Wavelets on graphs via spectral graph theory. *Appl. Comput. Harm. Anal.* 30, 129–150. doi: 10.1016/j.acha.2010.04.005

Haweel, R., Shalaby, A., Mahmoud, A., Seada, N., Ghoniemy, S., Ghazal, M., et al. (2021). A robust dwt-cnn-based cad system for early diagnosis of autism using task-based fmri. *Med. Phys.* 48, 2315–2326. doi: 10.1002/mp.14692

Jacobsen, J.-H., Smeulders, A., and Oyallon, E. (2018). "i-revnet: deep invertible networks," in *International Conference on Learning Representations*, Vancouver, BC.

Kleinhans, N. M., Richards, T., Sterling, L., Stegbauer, K. C., Mahurin, R., Johnson, L. C., et al. (2008). Abnormal functional connectivity in autism spectrum disorders during face processing. *Brain* 131, 1000–1012. doi: 10.1093/brain/awm334

Ktena, S. I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., et al. (2018). Metric learning with spectral graph convolutions on brain connectivity networks. *NeuroImage* 169, 431–442. doi: 10.1016/j.neuroimage.2017.12.052

Li, X., Dvornek, N. C., Zhuang, J., Ventola, P., and Duncan, J. S. (2018). "Brain biomarker interpretation in asd using deep learning and fmri," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Vol. 11072 (Granada: Springer International Publishing), 206–214.

Li, Y., Liang, P., Jia, X., and Li, K. (2016). Abnormal regional homogeneity in parkinson's disease: a resting state fmri study. *Clin. Radiol.* 71, e28–e34. doi: 10.1016/j.crad.2015.10.006

Lord, C., Elsabbagh, M., Baird, G., and Veenstra-Vanderweele, J. (2018). Autism spectrum disorder. *The Lancet* 392, 508–520. doi: 10.1016/S0140-6736(18)31129-2

Martino, A. D., Yan, C.-G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* 19, 659–667. doi: 10.1038/mp.2013.78

Miller, R. L., Yaesoubi, M., and Calhoun, V. D. (2016). Cross-frequency rs-fmri network connectivity patterns manifest differently for schizophrenia patients and healthy controls. *IEEE Signal Process. Lett.* 23, 1076–1080. doi: 10.1109/LSP.2016.2585182

Monk, C. S., Peltier, S. J., Wiggins, J. L., Weng, S.-J., Carrasco, M., Risi, S., et al. (2009). Abnormalities of intrinsic functional connectivity in autism spectrum disorders. *NeuroImage* 47, 764–772. doi: 10.1016/j.neuroimage.2009.04.069

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., et al. (2018). Disease prediction using graph convolutional networks: application to autism

spectrum disorder and alzheimer's disease. *Med. Image Anal.* 48, 117–130. doi: 10.1016/j.media.2018.06.001

Rakić, M., Cabezas, M., Kushibar, K., Oliver, A., and Lladó, X. (2020). Improving the detection of autism spectrum disorder by combining structural and functional mri information. *NeuroImage Clin.* 25:102181. doi: 10.1016/j.nicl.2020.102181

Sikka, S., Cheung, B., Khanuja, R., Ghosh, S., Gan Yan, C., Li, Q., et al. (2013). Towards automated analysis of connectomes: the configurable pipeline for the analysis of connectomes (c-pac). *Front. Neuroinf.* 7. doi: 10.3389/conf.fninf.2013.09.00042

Subbaraju, V., Suresh, M. B., Sundaram, S., and Narasimhan, S. (2017). Identifying differences in brain activities and an accurate detection of autism spectrum disorder using resting state functional-magnetic resonance imaging : a spatial filtering approach. *Med. Image Anal.* 35, 375–389. doi: 10.1016/j.media.2016.08.003

Thomas, R. M., Gallo, S., Cerliani, L., Zhutovsky, P., El-Gazzar, A., and van Wingen, G. (2020). Classifying autism spectrum disorder using the temporal statistics of resting-state functional mri data with 3d convolutional neural networks. *Front. Psychiatry* 11:440. doi: 10.3389/fpsyt.2020.00440

Wang, L., Li, K., and Hu, X. P. (2021). Graph convolutional network for fmri analysis based on connectivity neighborhood. *Netw. Neurosci. (Cambridge, Mass.)* 5, 83–95. doi: 10.1162/netn_a_00171

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

Xu, J., Wang, C., Xu, Z., Li, T., Chen, F., Chen, K., et al. (2020). Specific functional connectivity patterns of middle temporal gyrus subregions in children and adults with autism spectrum disorder. *Autism Res.* 13, 410–422. doi: 10.1002/aur.2239

Yerys, B. E., Gordon, E. M., Abrams, D. N., Satterthwaite, T. D., Weinblatt, R., Jankowski, K. F., et al. (2015). Default mode network segregation and social deficits in autism spectrum disorder: evidence from non-medicated children. *NeuroImage Clin.* 9, 223–232. doi: 10.1016/j.nicl.2015.07.018

Yin, W., Mostafa, S., and Wu, F.-X. (2021). Diagnosis of autism spectrum disorder based on functional brain networks with deep learning. *J. Comput. Biol.* 28, 146–165. doi: 10.1089/cmb.2020.0252

Zhang, Z., Peng, P., and Zhang, D. (2020). Executive function in high-functioning autism spectrum disorder: a meta-analysis of fmri studies. *J. Autism Develop. Disor.* 50, 4022–4038. doi: 10.1007/s10803-020-04461-z

Zhuang, J., Dvornek, N. C., Li, X., Ventola, P., and Duncan, J. S. (2019). "Invertible network for classification and biomarker selection for asd," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Vol. 11766 (Shenzhen), 700–708.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 21 February 2022
doi: 10.3389/fnins.2021.823550

Aberrant Functional Connectivity of Sensorimotor Network and Its Relationship With Executive Dysfunction in Bipolar Disorder Type I

Wenjing Zhu[1,2†], Wenxin Tang[1†], Yan Liang[1], Xiaoying Jiang[1], Yi Li[1], Zhiyu Chen[1] and Cheng Zhu[1*]*

[1] Hangzhou Seventh People's Hospital, Hangzhou, China, [2] Department of Neurology, School of Medicine, Affiliated ZhongDa Hospital, Institution of Neuropsychiatry, Southeast University, Nanjing, China

OPEN ACCESS

Edited by:
Yu Zhang,
Lehigh University, United States

Reviewed by:
Daniele Corbo,
University of Brescia, Italy
Jiu Chen,
Nanjing Medical University, China

***Correspondence:**
Zhiyu Chen
hzczy536@163.com
Cheng Zhu
zhucheng003042@163.com

†These authors share first authorship

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *27 November 2021*
Accepted: *20 December 2021*
Published: *21 February 2022*

Citation:
Zhu W, Tang W, Liang Y, Jiang X, Li Y, Chen Z and Zhu C (2022) Aberrant Functional Connectivity of Sensorimotor Network and Its Relationship With Executive Dysfunction in Bipolar Disorder Type I. Front. Neurosci. 15:823550. doi: 10.3389/fnins.2021.823550

Background: The key pathophysiological mechanism of executive dysfunction in patients with bipolar disorder type I (BD-I) is still unclear. Previous studies have demonstrated that it may be related to the disbalance of the sensory motor network (SMN).

Objective: This study was designed to explore the aberrant functional connectivity (FC) of SMN in BD-I patients and its potential associations with executive dysfunction.

Methods: Eighteen BD-I patients and 20 healthy controls (HCs) underwent resting-state fMRI scans. The intranetwork and internetwork functional connectivities of SMN were extracted by independent component analysis (ICA). Clinical symptoms were assessed by the Bech–Rafaelsen Mania Rating Scale (BRMS) and Positive and Negative Syndrome Scale (PANSS). Executive function was measured by digit span tasks and a verbal fluency test. Finally, linear regression and correlation analyses were applied to measure the potential associations between clinical symptoms, intranetwork and internetwork functional connectivities, and executive function performance.

Results: (1) Patients with BD-I showed increased connectivity in the right paracentral lobule and the right postcentral gyrus within the SMN, and the increased connectivity value was positively correlated with the BRMS score ($P < 0.05$) but negatively correlated with digit span forward scores ($P < 0.05$). (2) Compared with HC, the connectivity value increased between the SMN and dorsal attention network (DAN) ($P < 0.01$) and between the default mode network (DMN) and DAN ($P < 0.05$) but decreased between the DAN and auditory network (AN) ($P < 0.05$) and between the SMN and DMN ($P < 0.01$) in patients with BD-I. (3) Digit span forward scores and education of all participants were negatively correlated with FC between SMN and DAN. Age of all subjects was positively correlated with FC between SMN and DMN.

Conclusion: Our findings suggest that the sensorimotor network of BD-I has abnormal functional connections within and between networks, and the abnormal FC value correlated with clinical symptoms and executive function, which provide new information for exploring the neural physiopathology of executive dysfunction in BD-I patients.

Keywords: bipolar disorder, executive function, resting-state fMRI, sensorimotor network, functional connectivity

INTRODUCTION

Bipolar disorder (BD) is a type of mood disorder characterized by the core feature of recurrence of mania (BD type I, BD-I) or hypomania (BD type II, BD-II) and depressive episodes with high disability and high burden (Hahn et al., 2014). Most of the findings show that BD, especially bipolar disorder type I (BD-I), is associated with deficits in cognitive functions, particularly in executive function, which correlated with the ability to integrate various skills to prepare for and execute complex behaviors (Ozonoff et al., 2004; Elshahawi et al., 2011; Liu et al., 2011; Peters et al., 2014; Drakopoulos et al., 2020). Interestingly, even modest executive dysfunction can lead to noticeable disturbance of behaviors, including deficiency in planning, organization, problem solving, and decision making, and it can persist throughout the course of the disease (Sole et al., 2016; Garcia-Laredo et al., 2021). However, the neural mechanism of executive dysfunction in patients with BD-I is still largely unclear.

Resting-state fMRI (rs-fMRI) is the spontaneous regulation of BOLD signals in the brain under the non-task state, which can reflect the process of promoting the integration of internal and external environmental neural signals. rs-fMRI is widely used in the research of a variety of mental and nervous system diseases (Dong et al., 2016; Breukelaar et al., 2020; Dimick et al., 2020; Waller et al., 2021). Many studies indicate that the brain is a complex system consisting of multiple functional networks subserving different functions, which consist of several brain regions with similar patterns of signal change over the course of rs-fMRI (Damoiseaux et al., 2006; De Luca et al., 2006; Maalouf et al., 2010; Power et al., 2011). The results of functional connectivity between and within functional networks could improve our understanding of the large-scale functional organization in executive dysfunction of patients with BD. For example, Amy Peters and his team investigated the cognitive-affective task-oriented engagement of the cognitive control network (CCN) and default mode network (DMN) may support real-world cognitive function in BD (Liu et al., 2015). Peters et al. (2020) probed the network of major depressive disorder and BD, and the results showed that memory impairment displays a central role in the cognitive impairment of patients with unipolar depression, whereas executive dysfunction appears to be more central in bipolar depression. Kristen K. Ellard investigated that impaired functional connectivity between the anterior insula and the inferior parietal lobule of the executive control network (ECN) distinguishes patients with bipolar depression from those with unipolar depression and healthy control (HC) subjects (Galimberti et al., 2020). However, there have been few studies focusing on the brain networks of patients with BD-I, and also few studies detailing how brain network abnormalities participate in the executive dysfunction of this type of patients.

An increasing number of research have shown that the DMN and sensory motor network (SMN) have a major role in emotional and cognitive processing, with the sensorimotor networks having become a focus of research on brain networks in BD in recent years (Ellard et al., 2018; Kebets et al., 2019; Magioncalda et al., 2020). The relevant research suggested that abnormality activities in SMN are the basis of emotion processing dysfunction and executive dysfunction in patients with BD. Meanwhile, the interaction between SMN and other brain networks including DMN has been shown to be dysfunctional (Martino et al., 2020). Nevertheless, how intranetwork and internetwork functional connectivities changed in the SMN of patients with BD-I, which in turn impacted human behaviors, especially executive dysfunction, has yet to be determined.

The aim of the present study is to measure the aberrant intranetwork and internetwork functional connectivities of SMN in patients with BD-I using the independent component analysis (ICA). Simultaneously, we also evaluated the patients' clinical symptoms and executive function. Finally, we measured the potential associations between functional connectivity, clinical symptoms, and executive function in both the BD-I group and normal group. We hypothesized that the patients of BD-I have abnormal functional connections within SMN and between SMN with other networks and that the abnormal FC correlates with clinical symptoms and executive function at the same time. The functional network connectivity of SMN was involved in the executive dysfunction of BD-I patients.

MATERIALS AND METHODS

Participants and Clinical Assessment

This study was consistent with the Declaration of Helsinki and was approved by the Hangzhou Seventh People's Hospital Ethics Committee, and written informed consent was obtained from all participants after they had been given a complete description of the study.

Nineteen subjects with BD-I (10 males and 9 females) were recruited from the outpatient or inpatient department of the Hangzhou Seventh People's Hospital; all subjects were medicated (with drugs including lithium carbonate, valproate, and second-generation antipsychotics). Diagnosis of BD-I was in accordance with the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV). The diagnosis was conducted based on structured clinical interviews by two independent psychiatrists who had more than 10 years of clinical experience. The exclusion

criteria included (1) age below 15 or above 65 years; (2) current or previous diagnosis of substance dependence, schizoaffective disorder, or schizophrenia; (3) history of brain injury, previous loss of consciousness greater than 10 min, or self-reported serious medical conditions; (4) electroconvulsive therapy history during the last 6 months; (5) any contraindications of MRI; (6) too large head motion (> 2 mm in translation or 2° in rotation). One patient was excluded because of large head motion, and the remaining 18 patients were enrolled in the research. Twenty gender-, age-, and education-matched HC volunteers (12 males and 8 females) were recruited from the community as controls *via* local advertisements. All the controls also need to accept a standard medical examination by psychiatrists to rule out the presence of current or past psychiatric illness, neurological illness, or head injury causing concussion, as well as any history of psychiatric illness in first-degree relatives. **Table 1** shows the demographic and characteristic data, including neuropsychological scales, clinical characteristics, drugs of the subjects.

Cognition and Clinical Assessment

The digit span task is a part of the executive function, which is adopted to evaluate attention (Chenji et al., 2016). All participants completed a digit span forward task followed by a digit span backward task. The digit span forward task required the participant to remember a series of numbers from two digits continuing to a maximum of 13 digits, which are presented orally. Then the participants try to verbally repeat the digits. There were two trials per digit series. All participants began with the first two digit series; if repeated correctly, they continued to the next one; otherwise, they performed the second trial at the same digit series. The task was terminated when the participant failed in the second trial. The span is defined as the maximum number of digits repeated by the participant. The digit span backward task followed the same procedure, except the order of the digits verbally repeated by the participants was reversed. During the verbal fluency test (VFT), participants were required to say as many words as possible describing an animal or a vegetable within 1 min. When the participants correctly described the animal or vegetable, 1 point was awarded. Each participant was assessed with standardized clinical instruments, including the Bech–Rafaelsen Mania Rating Scale (BRMS) and the Positive and Negative Syndrome Scale (PANSS), which were used to evaluate clinical symptoms for all patients during the 7-day period prior to the scan.

TABLE 1 | Demographic and behavioral characteristic of participants.

	BD-I	HC	[b]*T*-value/χ^2	*P*-value
Gender (male/female)[a]	10/8	12/8	0.08	0.78
Age	30.83 ± 10.28	33.30 ± 11.16	−0.71	0.49
Educational years	15.12 ± 3.22	15.56 ± 3.75	−0.51	0.67
Age of onset	22.56 ± 8.99			
Duration of illness (years)	11.45 ± 9.55			
Number of hospitalizations	5.56 ± 4.94			
Lithium (n)	8			
Anticonvulsant (n)	12			
Antipsychotic (n)	18			
Antipsychotic medication, day/mg[c,d]	605.56 ± 207.16			
BRMS	25.06 ± 6.25			
PANSS	60.11 ± 9.61			
Digit span forward	6.67 ± 1.65	8.20 ± 1.01	−3.42	0.002
Digit span backward	4.56 ± 0.78	6.90 ± 1.83	−5.22	0.001
VFT	18.72 ± 4.71	22.55 ± 5.57	−0.91	0.020
FD value	0.14 ± 0.09	0.11 ± 0.23	2.39	0.282

BD-I, bipolar disorder type I; HC, healthy control; BRMS, Bech–Rafaelsen Mania Rating Scale; PANSS, Positive and Negative Syndrome Scale; VFT, verbal fluency test; FD, framewise displacement, used to evaluate head motion during scanning.
[a]Data are presented as mean ± standard deviation except gender.
[b]Comparisons were performed with a chi-square test for the variable of gender and independent-samples t-tests for other variables.
[c]All participants were taking atypical antipsychotics.
[d]Chlorpromazine equivalent doses were calculated.

MRI Data Acquisition

Brain structural and resting-state functional images of all subjects were acquired from the Hangzhou Seventh People's Hospital. All participants were instructed before scanning to keep their eyes closed, relax, move as little as possible, think of nothing in particular, and not fall asleep during the scans. Resting-state MRI scans were conducted under a 1.5-T MRI scanner (Signa HDxt 1.5T, GE Healthcare, Buckinghamshire, United Kingdom) composed of 180 echo-planar imaging volumes with the following parameters: TR = 2,000 ms; TE = 40 ms; flip angle = 85°; matrix size = 64 × 64; field of view = 240 × 240 mm; slice thickness = 3 mm; and 28 continuous slices. The acquisition time is 6 min. A T1-weighted structural image was also acquired for each patient to further elucidate and discard gross radiological alterations (TR = 9.5 ms; TE = 3.1 ms; flip angle = 20°; field of view = 240 mm × 240 mm; slice thickness = 1.2 mm).

Functional MRI Data Processing

Functional MRI data were preprocessed using the Statistical Parametric Mapping software (SPM12)[1] and Data Processing and Analysis for Brain Imaging (DPABI)[2] (Groth-Marnat and Baker, 2003). The data preprocessing includes the following steps:

(1) Discarding of the first five volumes to achieve equilibrium and a steady state.
(2) Slice timing correction.
(3) Realignment (head motion parameters were computed by estimating the translation in each direction and the angular rotation on each axis for each volume; we required the translational or rotational motion parameters to be less than 2 mm or 2°; the framewise displacement (FD), which indexes the volume-to-volume changes in head position was also calculated).

[1]http://www.fil.ion.ucl.ac.uk/spm
[2]http://rfmri.org/dpabi

(4) These images were then spatially normalized by the following steps: individual structural images were co-registered with the mean functional image, firstly. Then the transformed structural images were segmented and normalized to the Montreal Neurological Institute (MNI) space using a high-level non-linear warping algorithm which uses the exponentiated Lie algebra (DARTEL) technique to acquire the diffeomorphic anatomical registration (Yan et al., 2016). Finally, each functional volume was spatially normalized to the MNI space using the deformation parameters estimated during the above step and resampled into a 3-mm cubic voxel.
(5) Spatial smoothing with a Gaussian kernel of 6 mm × 6 mm × 6 mm.

Independent Component Analysis

This research applied the group ICA tool named GIFT (version 3.0b)[3] to perform group spatial ICA. We analyzed a mean of 16 components for each subject by the minimum description length criteria. Spatial ICA decomposes the participant data into linear mixtures of spatially independent components that exhibit a unique time course profile. Two data reduction steps can achieve this goal: Firstly, applying principal component analysis to reduce the subject-specific data into 24 principal components; and secondly, concatenating reduced data of all participants across time and using the infomax algorithm to decompose the data into 16 independent components. To ensure estimation stability, the infomax algorithm was repeated 20 times in ICASSO,[4] and the most central run was selected and analyzed further. Finally, the GICA back-reconstruction approach was used to find the individual time courses and spatial maps of the participants.

We applied several independent components that had peak activations in gray matter; showed low spatial overlap with heartbeat, motion, and susceptibility artifacts; and exhibited low frequency power primarily to identified functional networks; then we used the methods of visual inspection and spatial network template to acquire the target brain network. In the step of visual inspection, the SMN includes the somatosensory and motor cortices (Ashburner, 2007). Finally, this research procedure resulted in six functional networks out of the 16 independent components obtained (**Figure 1**): DMN, right frontoparietal network (rFPN), dorsal attention network (DAN), SMN, auditory network (AN), and visual network (VN).

The following were performed in the intranetwork functional connectivity calculation. First, a one-sample *t*-test was carried out to determine anatomical regions within SMN at a threshold of $P < 0.05$, FEW corrected. Second, a two-sample *t*-test was performed for SMN masked by the regions identified in one-sample *t*-tests with age and gender as a nuisance covariate to identify the significantly different regions between HC and BD subjects (FEW-corrected $P < 0.001$).

[3] mialab.mrn.org/software/gift/

[4] http://research.ics.tkk.fi/ica/icasso/

In the internetwork functional connectivity calculation, the value of function connectivity was estimated by Pearson correlation coefficients between pairs of time courses of the functional networks, resulting in a symmetric 9 × 9 correlation matrix for each participant. Finally, we used Fisher's transformation to transform correlations to *z*-scores, which can improve the normality. Intranetwork connectivity was examined *via* the spatial maps, indexing the contribution of the time course to each voxel comprising a given component.

Statistical Analysis

The statistical descriptive analyses of demographic and behavioral data were conducted using the SPSS 17.0 software package. We performed multiple regression analyses to investigate the relationships between intranetwork and internetwork functional connectivity of SMN with age, education years, and behavioral performance (e.g., BRMS, PANSS, and cognitive function). Significance was determined by $p < 0.05$ (two-tailed), with no correction.

RESULTS

Demographic and Behavioral Performance

Table 1 describes the demographic, behavioral performance, and head motion data of 18 patients with BD-I and 20 HCs. There was no significant difference between two groups in terms of age, gender, and head motion. Compared with the HCs, patients with BD-I presented greater scores in BRMS and PANSS and lower scores in digit span and VFT.

Connectivity Within Sensory Motor Network

Based on previous studies and the highest ranked correlation values calculated by the spatial network template in this study, components 8 and 13 were found to match with the SMN. The other seven components were separately identified as the DMN, rFPN, DAN, AN, and VN. **Figure 1** shows the results of identified spatial network maps.

A two-sample *t*-test was performed when analyzing the within-network connectivity for SMN, and we found the brain areas with significant differences in the network between the two groups. Significantly increased within-network connectivity was found in the right paracentral lobule (voxel size = 26; peak coordinate = 8, −32, 68; $T = 4.86$, $P < 0.001$) and the right postcentral gyrus (voxel size = 20; peak coordinate = 41, −26, 53; $T = 3.33$, $P < 0.001$) in the SMN of the BD-I group (**Figure 2A**), and there were no clusters of reduced connectivity in BD-I patients compared to HCs.

The increased within-network connectivity value in the SMN was positively correlated with the BRMS score ($P < 0.05$) and negatively correlated with the digit span forward score ($P < 0.05$).

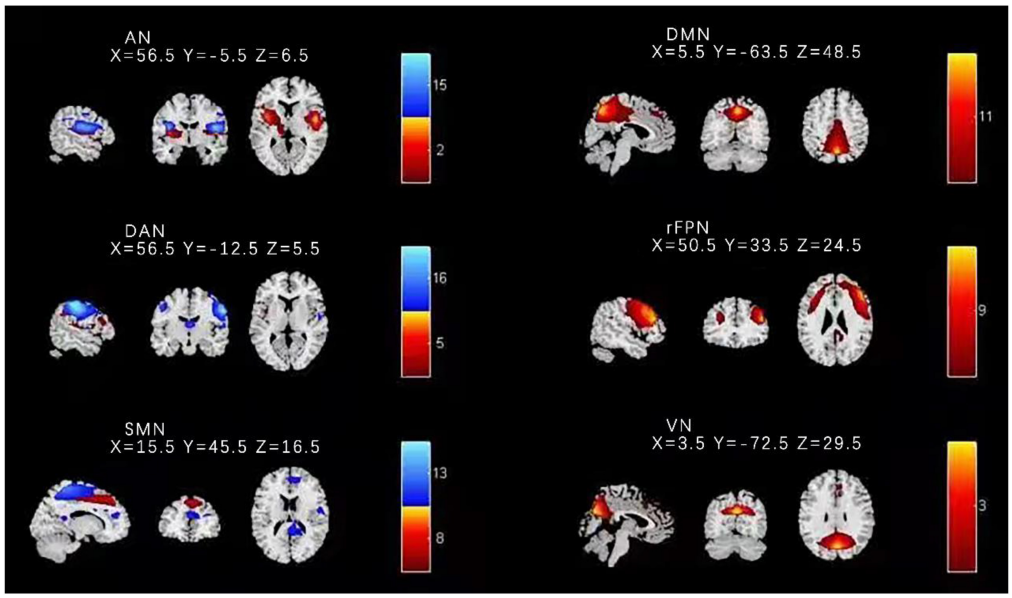

FIGURE 1 | Spatial distribution of the nine independent components for all HC and BD-I subjects. Components 8 and 13 were found to match with the SMN. DMN, default mode network; rFPN, right frontoparietal network; DAN, dorsal attention network; SMN, sensorimotor network; AN, auditory network; VN, visual network.

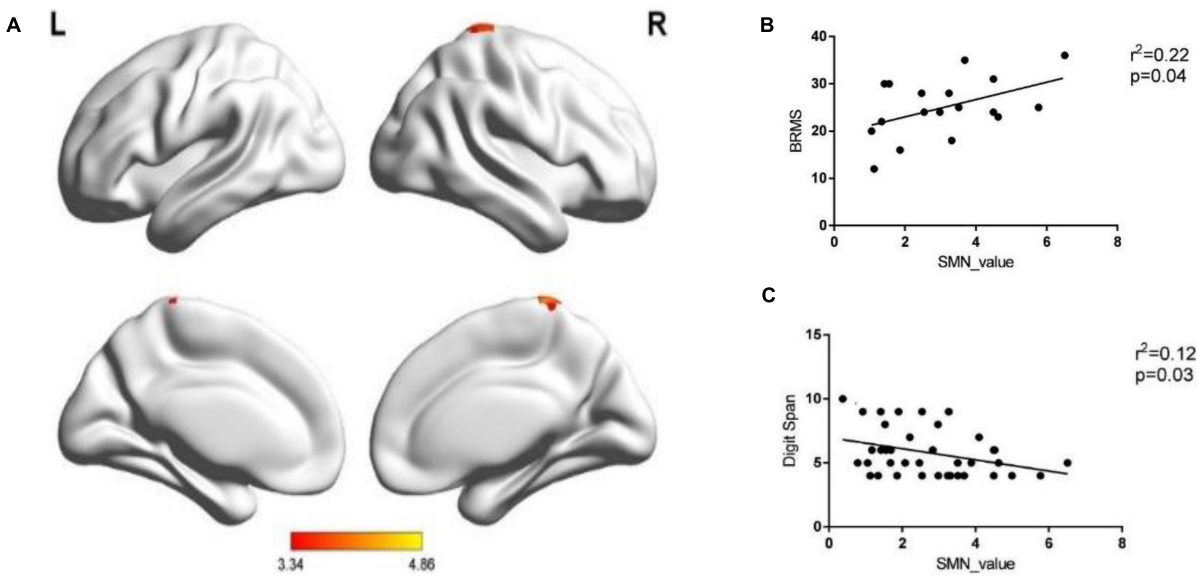

FIGURE 2 | The within-network connectivity of SMN. **(A)** Patients BD-I showed increased within-network connectivity in the right paracentral lobule and the right postcentral gyrus in the SMN compared with HCs. **(B)** The correlation between the within-network connectivity value of SMN in patients and the BRMS score. **(C)** The correlation between the within-network connectivity value of SMN in all subjects and digit span forward scores.

There was no significant correlation in the rest (all $P > 0.05$, see **Figures 2B,C**).

Connectivity Between Sensory Motor Network and Other Networks

The results of connectivity differences between networks in BD-I patients vs. HCs are shown in **Figure 3A**. Both positive and negative internetwork functional connectivities were observed. As observed from **Figures 3B,C**, compared with those in HCs, the functional connectivity value increased between SMN and DAN and decreased between SMN and DMN in BD-I subjects ($P < 0.01$, FDR corrected). Moreover, as observed from **Figures 4A–C**, digit span forward scores and education years of all participants were negatively correlated with FC between SMN and DAN ($P < 0.05$). The age of all subjects was positively correlated with FC between SMN and DMN ($P < 0.05$).

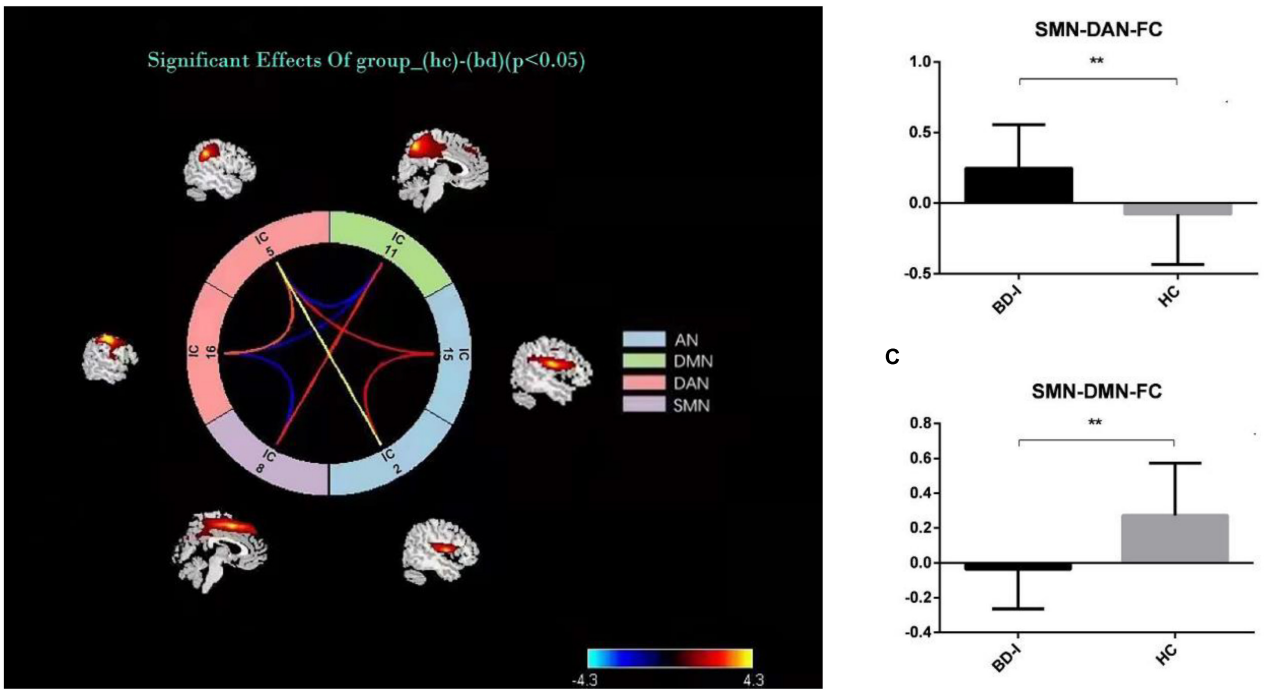

FIGURE 3 | (A) Internetwork functional connectivity in BD-I and HCs. **(B)** Comparison of functional connectivity between SMN and DAN in the two groups. **(C)** Comparison of functional connectivity between SMN and DMN in the two groups. *$p < 0.05$; **$p < 0.01$.

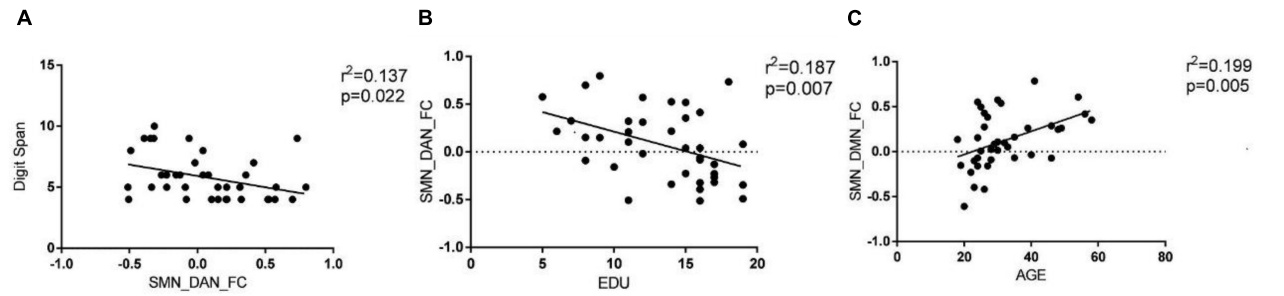

FIGURE 4 | Pearson correlation analysis of internetwork functional connectivity value and clinical information. **(A)** Digit span forward scores of all participants were negatively correlated with FC between SMN and DAN ($P < 0.05$). **(B)** Education years of all participants were negatively correlated with FC between SMN and DAN ($P < 0.05$). **(C)** Age of all subjects was positively correlated with FC between SMN and DMN ($P < 0.05$).

DISCUSSION

Main Findings

Our main findings are as follows. (1) compared with HC, patients with BD-I showed increased within-network connectivity in the right paracentral lobule and the right postcentral gyrus in the SMN, and the connectivity value was positively correlated with BRMS score but negatively correlated with digit span forward scores. (2) The between-network connectivity of BD-I increased between SMN and DAN and between DMN and DAN but decreased between DAN and AN and between SMN and DMN when compared to HC. (3) Digit span forward scores and education of all participants were negatively correlated with FC between SMN and DAN. Age of all subjects was positively correlated with FC between SMN and DMN. Aberrant intranetwork and internetwork functional connectivities of sensorimotor network may be the neural mechanism of interaction between clinical symptoms and executive dysfunction in patients of BD-I.

Executive Dysfunction of Bipolar Disorder Type-I

The two main impaired elements in individuals with BD are emotional processing and cognitive functions (Wessa and Linke,

2009; Wang C. et al., 2020). Failing in using cognitive functions to regulate and maintain emotional states can lead to a dysfunction in their emotional processing and emotion regulation (Torrent et al., 2006). The present study applied the digit span tasks and VFT to investigate executive function. The result shows that valuations of digit span tasks and VFT of BD-I patients were lower than those of HCs at baseline, and it is consistent with previous related researches (Phillips and Vieta, 2007; Kosger et al., 2015). However, the digit span tasks and VFT are just parts of the executive function, and we need more research with multiple tests to support our results.

Intranetwork Connectivity in Bipolar Disorder Type-I

Our research found increased within-network connectivity in the right paracentral lobule and the right postcentral gyrus in the SMN, and the connectivity value was positively correlated with the BRMS score. The results agreed with those of Erol et al. (2013) and Khadka et al. (2013). Kahdka reported that BD-I patients with psychosis episodes showed increased FC in the superior frontal gyrus and medial frontal gyrus in the SMN compared with HCs. Ishida et al. (2017) and Kazemi et al. (2018) found within-network connectivity in the right premotor region in the SMN was positively correlated with the total score of YMRS for BD subjects. As known from previous studies, the SMN mainly regulates sensory and motor functions, and the abnormal connection of this network function corresponds well to the patients' clinical symptoms. The anterior part of the paracentral lobule is located in the primary motor cortex, which regulates voluntary movement of the body. The concentration of norepinephrine in patients with BD-I is increased, accompanied by abnormalities of dopamine (DA) and γ-aminobutyric acid (GABA), which leads to the increased function of the primary motor cortex. The central posterior gyrus belongs to the primary sensory cortex and receives various sensory stimuli. The patients' right central posterior gyrus functional connection is abnormal, which may be related to the patients' increased motor function and compensation of adjacent sensory cortex. More importantly, different from previous SMN studies, we found that the increased connectivity of SMN was negatively correlated with digit span forward scores. In the pre-language stage of individual development, the occurrence of cognition mainly depends on sense and kinesthesia. Even if the individual gradually relies on language for learning and communication with development, sensorimotor ability is still a basic ability for the individual to recognize external things (Supekar et al., 2019). Furthermore, patients with BD-I often suffer from the inability of the brain to filter and screen irrelevant information, and the inhibitory function is impaired, leading to excessive information flooding into the consciousness and then various abnormal sensory experiences, which ultimately affect cognitive function. All these researches show that the SMN plays an important role in emotional dis-regulation and cognitive dysfunction of BD-I patients.

Internetwork Connectivity in Bipolar Disorder Type-I

The results of the second part of this study show that the between-network connectivity value of BD-I increased between SMN and DAN but decreased between SMN and DMN when compared to HC. The brain is a complex brain network, which is composed of multiple functionally interacting sub-networks (Mochizuki-Kawai et al., 2004). Previous studies have shown that the SMN is the core network that is vulnerable to dysfunction in emotional functions, emotion recognition, and cognitive functions of BD (Wood et al., 2016; Wang J. et al., 2020). So advanced CCNs related to sensorimotor functions, such as the DMN, ECN, and DAN, may also experience functional impairment or compensation, the so-called functional reorganization (Davis et al., 2017). Wang et al. (2014) found that the functions of the medial prefrontal lobe and the cingulate gyrus of DMN are related to the planning, preparation, and execution of exercise. Damage to this area will affect the intentional purpose and rough planning of the motor cortex during exercise. Roelcke et al. (1997) revealed that an imbalance was observed in the DMN/SMN activity of bipolar patients, and a low ratio of DMN/SMN activity was reported in the manic phase while the opposite happened in the depression phase. Our results are consistent with these conclusions and also show that age was positively correlated with FC between SMN and DMN, which suggests that different age groups of BD-I patients have different degrees of imbalance in DMN/SMN (Fox et al., 2006; Chen et al., 2020; Manza et al., 2020). The relationship between DMN and SMN was considered as a diagnostic marker for BD. The DAN mainly includes the bilateral parietal internal sulcus and the joint cortical area of the superior frontal gyrus and the central anterior gyrus. It is related to the active attention process in the task and is obviously activated in tasks where attention cues such as target, position, and time appear. The upper leaflet of DAN is responsible for receiving the input of visual information and plays a role of visual control in motor tasks. Patients with BD-I have the characteristics of increased active attention and shifting attention with the environment. The functional connection between DAN and SMN is enhanced, and intriguingly, it is negatively correlated with digit span forward scores and education years of all participants, which may be consistent with the clinical features of increased volitional activity in patients with BD-I and also suggests that high years of education may be a protective factor for the disease. The results of the study confirm that the patients' proprioceptive perception and visual information are impaired by the top-down and bottom-up two-way adjustment and control effects during the completion of motor coordination, which also affects the patients' executive function accordingly. The two-way regulation and control function of DAN/SMN gives us a new understanding of the imbalance between DAN and SMN networks.

CONCLUSION

In conclusion, we got valuable information about aberrant intranetwork and internetwork functional connectivities of SMN in patients with BD-I, compared with HCs. Furthermore,

we found a significant relationship between the abnormal intranetwork and internetwork functional connectivity values and clinical symptoms and executive function, which provides new information for exploring the neural physiopathology of executive dysfunction in BD-I patients.

Limitations

Some limitations in this study should be addressed. First, our study enrolled a small-sized sample. Larger samples in the future are needed to confirm current findings. Second, all patients included in this study were administrated with drugs, which made the interpretation of results complex and difficult, but it is also a high risk in the image acquisition of untreated patients with manic episodes. After the sample size is expanded, interfering factors such as the duration of illness, the number of years of therapy, and the number of hospitalizations can be analyzed by comparing patient groups. Third, this study only included BD-I patients during manic episodes but not depressive episodes, which will make the attribution of executive dysfunction in BD-I patients incomplete.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Hangzhou Seventh People's Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

CZ and ZC supervised the present study. WZ and WT performed the analysis and wrote the manuscript. YaL, XJ, and YiL helped to collect the data. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the funding from the Key Medical and Health Science and Technology Projects of Hangzhou (2015Z07), the Social Development Fund of Science and Technology Bureau of Hangzhou (20191203B120), the General Research Project of Zhejiang Provincial Department of Health (2019KY516), the Youth Innovative Talent Support Program Fund of Zhejiang Provincial Department of Health (2022493976), and Project for Hangzhou Medical Disciplines of Excellence and Key Project for Hangzhou Medical Disciplines.

REFERENCES

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage* 38, 95–113. doi: 10.1016/j.neuroimage.2007.07.007

Breukelaar, I. A., Erlinger, M., Harris, A., Boyce, P., Hazell, P., Grieve, S. M., et al. (2020). Investigating the neural basis of cognitive control dysfunction in mood disorders. *Bipolar Disord.* 22, 286–295. doi: 10.1111/bdi.12844

Chen, Y. L., Tu, P. C., Huang, T. H., Bai, Y. M., Su, T. P., Chen, M. H., et al. (2020). Using Minimal-Redundant and Maximal-Relevant Whole-Brain Functional Connectivity to Classify Bipolar Disorder. *Front. Neurosci.* 14:563368. doi: 10.3389/fnins.2020.563368

Chenji, S., Jha, S., Lee, D., Brown, M., Seres, P., Mah, D., et al. (2016). Investigating default mode and sensorimotor network connectivity in amyotrophic lateral sclerosis. *PLoS One* 11:e0157443. doi: 10.1371/journal.pone.0157443

Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., et al. (2006). Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci.* 103, 13848–13853. doi: 10.1073/pnas.0601417103

Davis, J. D., Winkielman, P., and Coulson, S. (2017). Sensorimotor simulation and emotion processing: impairing facial action increases semantic retrieval demands. *Cogn. Affect. Behav. Neurosci.* 17, 652–664. doi: 10.3758/s13415-017-0503-2

De Luca, M., Beckmann, C. F., De Stefano, N., Matthews, P. M., and Smith, S. M. (2006). fMRI resting state networks define distinct modes of long-distance interactions in the human brain. *Neuroimage* 29, 1359–1367. doi: 10.1016/j.neuroimage.2005.08.035

Dimick, M., Hird, M., MacIntosh, B. J., and Goldstein, B. (2020). Resting State Functional Connectivity and Suicidality in Adolescent Bipolar Disorder. *Biol. Psychiatry* 87:S438.

Dong, C., Weijia, G., Qing, J., Weifang, C., Ronfeng, Q., Yongxin, G., et al. (2016). Abnormal Resting-State Regional Homogeneity Relates to Cognitive Dysfunction in Manic Bipolar Disorder Adolescents: an fMRI Study. *Med. Imag. Health Inform.* 7, 1673–1678.

Drakopoulos, J., Sparding, T., Clements, C., Palsson, E., and Landen, M. (2020). Executive functioning but not IQ or illness severity predicts occupational status in bipolar disorder. *Int. J. Bipolar Disord.* 8:7. doi: 10.1186/s40345-019-0168-6

Ellard, K. K., Zimmerman, J. P., Kaur, N., Van Dijk, K. R. A., Roffman, J. L., Nierenberg, A. A., et al. (2018). Functional Connectivity Between Anterior Insula and Key Nodes of Frontoparietal Executive Control and Salience Networks Distinguish Bipolar Depression From Unipolar Depression and Healthy Control Subjects. *Biol. Psychiatry Cogn. Neurosci. Neuroimag.* 3, 473–484. doi: 10.1016/j.bpsc.2018.01.013

Elshahawi, H. H., Essawi, H., Rabie, M. A., Mansour, M., Beshry, Z. A., and Mansour, A. N. (2011). Cognitive functions among euthymic bipolar I patients after a single manic episode versus recurrent episodes. *J. Affect. Disord.* 130, 180–191. doi: 10.1016/j.jad.2010.10.027

Erol, A., Kosger, F., Putgul, G., and Ersoy, B. (2013). Ventral prefrontal executive function impairment as a potential endophenotypic marker for bipolar disorder. *Nord. J. Psychiatry* 68, 18–23. doi: 10.3109/08039488.2012.756062

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., and Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc. Natl. Acad. Sci.* 103, 10046–10051. doi: 10.1073/pnas.0604187103

Galimberti, C., Bosi, M. F., Caricasole, V., Zanello, R., Dell'Osso, B., and Vigano, C. A. (2020). Using network analysis to explore cognitive domains in patients with unipolar versus bipolar depression: a prospective naturalistic study. *CNS Spectr.* 25, 380–391. doi: 10.1017/S1092852919000968

Garcia-Laredo, E., Castellanos, M. A., Badaya, E., Paul, N., Yubero, R., Maestu, F., et al. (2021). Executive Functions Influence on Memory Process in Patients with Paranoid Schizophrenia and Bipolar Disorders with and without Psychotic Symptoms. A Pilot Study. *Span. J. Psychol.* 24:e40. doi: 10.1017/SJP.2021.38

Groth-Marnat, G., and Baker, S. (2003). Digit span as a measure of everyday attention: a study of ecological validity. *Percept. Mot. Skills* 97, 1209–1218. doi: 10.2466/pms.2003.97.3f.1209

Hahn, C., Lim, H. K., and Lee, C. U. (2014). Neuroimaging findings in late-onset schizophrenia and bipolar disorder. *J. Geriatr. Psychiatry Neurol.* 27, 56–62. doi: 10.1177/0891988713516544

Ishida, T., Donishi, T., Iwatani, J., Yamada, S., Takahashi, S., Ukai, S., et al. (2017). Interhemispheric disconnectivity in the sensorimotor network in bipolar disorder revealed by functional connectivity and diffusion tensor imaging analysis. *Heliyon* 3:e00335. doi: 10.1016/j.heliyon.2017.e00335

Kazemi, R., Rostami, R., Khomami, S., Baghdadi, G., Rezaei, M., Hata, M., et al. (2018). Bilateral Transcranial Magnetic Stimulation on DLPFC Changes Resting State Networks and Cognitive Function in Patients With Bipolar Depression. *Front. Hum. Neurosci.* 12:356. doi: 10.3389/fnhum.2018.00356

Kebets, V., Holmes, A. J., Orban, C., Tang, S., Li, J., Sun, N., et al. (2019). Somatosensory-Motor Dysconnectivity Spans Multiple Transdiagnostic Dimensions of Psychopathology. *Biol. Psychiatry* 86, 779–791. doi: 10.1016/j.biopsych.2019.06.013

Khadka, S., Meda, S. A., Stevens, M. C., Glahn, D. C., Calhoun, V. D., Sweeney, J. A., et al. (2013). Is aberrant functional connectivity a psychosis endophenotype: a resting state functional magnetic resonance imaging study. *Biol. Psychiatry* 74, 458–466. doi: 10.1016/j.biopsych.2013.04.024

Kosger, F., Essizoglu, A., Baltacioglu, M., Ulkgun, N., and Yenilmez, C. (2015). Executive function in parents of patients with familial versus sporadic bipolar disorder. *Compr. Psychiatry* 61, 36–41. doi: 10.1016/j.comppsych.2015.05.013

Liu, F., Guo, W., Fouche, J. P., Wang, Y., Wang, W., Ding, J., et al. (2015). Multivariate classification of social anxiety disorder using whole brain functional connectivity. *Brain Struct. Funct.* 220, 101–115. doi: 10.1007/s00429-013-0641-4

Liu, Y. M., Tsai, S. Y., Fleck, D. E., and Strakowski, S. M. (2011). Cross-cultural comparisons on Wisconsin Card Sorting Test performance in euthymic patients with bipolar disorder. *Psychiatry Res.* 189, 469–471. doi: 10.1016/j.psychres.2011.05.038

Maalouf, F. T., Klein, C., Clark, L., Sahakian, B. J., Labarbara, E. J., Versace, A., et al. (2010). Impaired sustained attention and executive dysfunction: bipolar disorder versus depression-specific markers of affective disorders. *Neuropsychologia* 48, 1862–1868. doi: 10.1016/j.neuropsychologia.2010.02.015

Magioncalda, P., Martino, M., Conio, B., Lee, H. C., Ku, H. L., Chen, C. J., et al. (2020). Intrinsic brain activity of subcortical-cortical sensorimotor system and psychomotor alterations in schizophrenia and bipolar disorder: a preliminary study. *Schizophr. Res.* 218, 157–165. doi: 10.1016/j.schres.2020.01.009

Manza, P., Wiers, C. E., Shokri-Kojori, E., Kroll, D., Feldman, D., Schwandt, M., et al. (2020). Brain Network Segregation and Glucose Energy Utilization: relevance for Age-Related Differences in Cognitive Function. *Cereb. Cortex* 30, 5930–5942. doi: 10.1093/cercor/bhaa167

Martino, M., Magioncalda, P., Conio, B., Capobianco, L., Russo, D., Adavastro, G., et al. (2020). Abnormal Functional Relationship of Sensorimotor Network With Neurotransmitter-Related Nuclei via Subcortical-Cortical Loops in Manic and Depressive Phases of Bipolar Disorder. *Schizophr. Bull.* 46, 163–174. doi: 10.1093/schbul/sbz035

Mochizuki-Kawai, H., Kawamura, M., Hasegawa, Y., Mochizuki, S., Oeda, R., Yamanaka, K., et al. (2004). Deficits in long-term retention of learned motor skills in patients with cortical or subcortical degeneration. *Neuropsychologia* 42, 1858–1863. doi: 10.1016/j.neuropsychologia.2004.03.012

Ozonoff, S., Cook, I., Coon, H., Dawson, G., Joseph, R., Klin, A., et al. (2004). Performance on Cambridge Neuropsychological Test automated battery subtests sensitive to frontal lobe function in people with Autistic Disorder: evidence from the Collaborative Programs of Excellence in Autism Network. *J. Autism. Dev. Disord.* 34, 139–150. doi: 10.1023/b:jadd.0000022605.81989.cc

Peters, A. T., Peckham, A. D., Stange, J. P., Sylvia, L. G., Hansen, N. S., Salcedo, S., et al. (2014). Correlates of real world executive dysfunction in bipolar I disorder. *J. Psychiatr. Res.* 53, 87–93. doi: 10.1016/j.jpsychires.2014.02.018

Peters, A., Chou, T., Testo, A., Wong, K. S., Ellard, K., Sylvia, L., et al. (2020). Executive Dysfunction in Bipolar Disorder: relationship to Default Mode and Cognitive Control Networks During Cognitive-Affective Interference. *Biol. Psychiatry.* 9:S286.

Phillips, M. L., and Vieta, E. (2007). Identifying functional neuroimaging biomarkers of bipolar disorder: toward DSM-V. *Schizophr. Bull.* 33, 893–904. doi: 10.1093/schbul/sbm060

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., et al. (2011). Functional network organization of the human brain. *Neuron* 72, 665–678.

Roelcke, U., Curt, A., Otte, A., Missimer, J., Maguire, R. P., Dietz, V., et al. (1997). Influence of spinal cord injury on cerebral sensorimotor systems: a PET study. *J. Neurol. Neurosurg. Psychiatry* 62, 61–65. doi: 10.1136/jnnp.62.1.61

Sole, B., Vieta, E., and Martinez-Aran, A. (2016). Thinking ahead: executive dysfunction in bipolar disorder. *Eur. Neuropsychopharmacol.* 26, 1348–1349. doi: 10.1016/j.euroneuro.2016.05.008

Supekar, K., Cai, W., Krishnadas, R., Palaniyappan, L., and Menon, V. (2019). Dysregulated Brain Dynamics in a Triple-Network Saliency Model of Schizophrenia and Its Relation to Psychosis. *Biol. Psychiatry.* 85, 60–69. doi: 10.1016/j.biopsych.2018.07.020

Torrent, C., Martínez-Arán, A., Daban, C., Sánchez-Moreno, J., Comes, M., Goikolea, J. M., et al. (2006). Cognitive impairment in bipolar II disorder. *Br. J. Psychiatry* 189, 254–259. doi: 10.1192/bjp.bp.105.017269

Waller, J., Miao, T., Ikedionwu, I., and Lin, K. K. (2021). Reviewing applications of structural and functional MRI for bipolar disorder. *JPN J. Radiol.* 39, 414–423. doi: 10.1007/s11604-020-01074-5

Wang, C., Cai, H., Sun, X., Si, L., Zhang, M., Xu, Y., et al. (2020). Large-Scale Internetwork Functional Connectivity Mediates the Relationship between Serum Triglyceride and Working Memory in Young Adulthood. *Neural Plast.* 2020:8894868. doi: 10.1155/2020/8894868

Wang, C., Qin, W., Zhang, J., Tian, T., Li, Y., Meng, L., et al. (2014). Altered functional organization within and between resting- state networks in chronic subcortical infarction. *J. Cereb. Blood Flow Metab.* 34, 597–605. doi: 10.1038/jcbfm.2013.238

Wang, J., Wang, Y., Wu, X., Huang, H., Jia, Y., Zhong, S., et al. (2020). Shared and specific functional connectivity alterations in unmedicated bipolar and major depressive disorders based on the triple-network model. *Brain Imaging Behav.* 14, 186–199. doi: 10.1007/s11682-018-9978-x

Wessa, M., and Linke, J. (2009). Emotional processing in bipolar disorder: behavioural and neuroimaging findings. *Int. Rev. Psychiatry* 21, 357–367. doi: 10.1080/09540260902962156

Wood, A., Rychlowska, M., Korb, S., and Niedenthal, P. (2016). Fashioning the face: sensorimotor simulation contributes to facial expression recognition. *Trends Cogn. Sci.* 20, 227–240. doi: 10.1016/j.tics.2015.12.010

Yan, C. G., Wang, X. D., Zuo, X. N., and Zang, Y. F. (2016). DPABI: data processing & analysis for (resting-state) brain imaging. *Neuroinformatics* 14, 339–351. doi: 10.1007/s12021-016-9299-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 09 March 2022
doi: 10.3389/fnins.2022.838347

Unrevealing Reliable Cortical Parcellation of Individual Brains Using Resting-State Functional Magnetic Resonance Imaging and Masked Graph Convolutions

*Wenyuan Qiu[1], Liang Ma[2,3], Tianzi Jiang[2,3] and Yu Zhang[1]**

[1] Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou, China, [2] Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, Beijing, China, [3] National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China

Brain parcellation helps to understand the structural and functional organization of the cerebral cortex. Resting-state functional magnetic resonance imaging (fMRI) and connectivity analysis provide useful information to delineate individual brain parcels *in vivo*. We proposed an individualized cortical parcellation based on graph neural networks (GNN) to learn the reliable functional characteristics of each brain parcel on a large fMRI dataset and to infer the areal probability of each vertex on unseen subjects. A subject-specific confidence mask was implemented in the GNN model to account for the tradeoff between the topographic alignment across subjects and functional homogeneity of brain parcels on individual brains. The individualized brain parcellation achieved better functional homogeneity at rest and during cognitive tasks compared with the group-registered atlas (p-values < 0.05). In addition, highly reliable and replicable parcellation maps were generated on multiple sessions of the same subject (intrasubject similarity = 0.89), while notable variations in the topographic organization were captured across subjects (intersubject similarity = 0.81). Moreover, the intersubject variability of brain parcellation indicated large variations in the association cortices while keeping a stable parcellation on the primary cortex. Such topographic variability was strongly associated with the functional connectivity variability, significantly predicted cognitive behaviors, and generally followed the myelination, cytoarchitecture, and functional organization of the human brain. This study provides new avenues to the precise individualized mapping of the cortical areas through deep learning and shows high potentials in the personalized localization diagnosis and treatment of neurological disorders.

Keywords: functional connectivity, cortical parcellation, intersubject variability, topographic variability, resting-state fMRI (rfMRI), test–retest reliability, graph neural network

OPEN ACCESS

Edited by:
Feng Liu,
Stevens Institute of Technology, United States

Reviewed by:
Guihong Wan,
Harvard Medical School, United States
Defu Yang,
Hangzhou Dianzi University, China

****Correspondence:***
Yu Zhang
yuzhang2bic@gmail.com

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *17 December 2021*
Accepted: *24 January 2022*
Published: *09 March 2022*

Citation:
Qiu W, Ma L, Jiang T and Zhang Y (2022) Unrevealing Reliable Cortical Parcellation of Individual Brains Using Resting-State Functional Magnetic Resonance Imaging and Masked Graph Convolutions. Front. Neurosci. 16:838347. doi: 10.3389/fnins.2022.838347

INTRODUCTION

Brain atlas has been an important tool to understand the neural basis of human cognition and to study the functional organization of the human brain (Ungerleider and Desimone, 1986; Felleman and Van Essen, 1991; Amunts and Zilles, 2015). Neuroanatomists have built a variety of brain atlases to depict cyto-, myelo-, and receptor architectures using postmortem brains

(Brodmann, 1909; Zilles and Amunts, 2010; Amunts et al., 2020). Recent advances in noninvasive neuroimaging techniques, such as functional magnetic resonance imaging (fMRI), provide an opportunity to explore the functional organization of the living human brain. There has been a rich and fast-growing literature on the functional brain parcellation using either spontaneous low-frequency fluctuations of fMRI activity or the aggregation of activation maps across different cognitive tasks (Blumensath et al., 2013; Eickhoff et al., 2015; Eickhoff et al., 2018a,b). The majority of current approaches focused on the group representative functional mapping of the cerebral cortex, which may provide useful insights into the intrinsic organizational principles of the human brain (Buckner et al., 2013; Wig, 2017), but ignore the variability of individual brains in areal size, location, spatial arrangement, and connectivity patterns (Mueller et al., 2013; Zuo and Xing, 2014). The precise mapping of individualized functional areas is a critical step toward better understanding of the structural–functional relationship of the human brain that underlying cognition and behavior (Wang et al., 2015; Kong et al., 2019, 2021) as well as for personalized localization diagnosis and treatment of neurological disorders (Mueller et al., 2015; Wang et al., 2020).

Traditional individualized mapping of brain atlas has relied on the linear and nonlinear registration based on the structural images in the volume space or cortical surfaces. Modern machine learning algorithms provide analytic tools to align cortical areas using neuroimaging data across multiple modalities, including myelin maps and functional localizers (Robinson et al., 2018), as well as anatomical (Ma et al., 2021) and functional connectivity fingerprints (Wang et al., 2015). As one of the most commonly used features for individualized brain mapping, functional connectivity has been shown to reveal individual-specific topographic organization that better predicted cognitive functions and behaviors (Wang et al., 2015; Cui et al., 2019; Kong et al., 2019; Li et al., 2019). However, the reliability and validity of such topographic variability has been one major concern considering that the fMRI signals are highly contaminated by noises of various physiological processes and head motions. By explicitly separating actual intersubject variability from noise components (evaluated by multiple sessions of the same subject), studies have shown that the individualized parcellation not only exhibited better functional homogeneity at rest and during cognitive tasks (Kong et al., 2021), but also captured reliable and inheritable variability in the topographic organization of the human brain (Anderson et al., 2021), demonstrated by the genetic effects of topographic variability. Yet, this multisession hierarchical Bayesian model (MS-HBM) used a global concentration parameter to model the heterogeneity of functional connectivity for different brain parcels, and resulted in similar levels of topographic variability and heritability among the primary and association cortices by treating each area equally, which is in congruence with the well-known sensory-fugal gradient in the myelination, cytoarchitecture, and functional organization of human brain.

In this study, we proposed a masked graph neural network (GNN) architecture to learn the reliable functional characteristics of each brain parcel using fMRI data from a large population and to apply such information to infer the areal probability of each vertex on unseen subjects. Specifically, we constructed a vertex-level brain graph for each subject and embed the whole-brain functional connectivity profiles as signals on the graph. Then we used high-order graph convolution to integrate the local connectivity context of each vertex such that the fluctuations in functional connectivity profiles were evaluated among a small neighboring area in the cortical surface rather than on each vertex individually, largely suppressing the noise effects from individual fMRI runs. Besides, we trained hundreds of graph convolutional kernels at each graph convolution layer to encode the variational organizational principles among cortical areas. Moreover, we implemented subject-specific confidence masks in the GNN model to maintain a consistent global topographic organization among subjects while preserving the intersubject variability in brain parcellation especially for vertices around the areal borders. Compared with the baseline approaches including the group-registered atlas and MS-HBM, our model generated highly consistent and replicable parcellation maps on individual brains, along with better functional homogeneity at rest and during cognitive tasks. Moreover, the topographic variability generally followed a sensory-fugal gradient from primary and unimodal areas to heteromodal areas, with high variations in the association cortices while keeping a stable parcellation on the primary cortex. More importantly, the topographic variability was strongly associated with individual variability in functional connectivity profiles and cognitive behaviors.

MATERIALS AND METHODS

Data Acquisition and Preprocessing

We used two independent datasets acquired from the Human Connectome Project (HCP) dataset, consisting of T1-weighted (T1w) data, resting-state functional MRI (rs-fMRI), as well as task-fMRI data for each subject. The individualized parcellation model was first trained and evaluated on a large dataset consisting of 1,022 subjects acquired from the Human Connectome Project S1200 release[1]. We then evaluated the test–retest reliability of the model on the second dataset, consisting of 44 subjects acquired from the HCP test–retest datasets. Whole-brain echo-planar imaging (EPI) acquisitions were acquired with a 32-channel head coil on a modified 3T Siemens Skyra with TR = 720 ms, 2.0-mm isotropic voxels, using a multiband sequence. Each subject underwent two fMRI sessions on separate days, including two runs of 14-min resting-state and seven task fMRI scans (we only used fMRI data with the left to right (LR) phase encoding in the current study). The task-fMRI database includes seven cognitive domains, which are emotion, gambling, language, motor, relational, social, and working memory. The detailed description of data collection and task paradigms can be found in Barch et al. (2013).

We used the minimal preprocessed fMRI data from the HCP pipelines[2]: (1) fMRIVolume pipeline generates "minimally

[1]https://db.humanconnectome.org/data/projects/HCP_1200

[2]https://github.com/Washington-University/HCPpipelines

preprocessed" 4D time series (i.e., ".nii.gz" file) that includes gradient unwarping, motion correction, fieldmap-based EPI distortion correction, brain-boundary-based registration of EPI to structural T1-weighted scan, nonlinear (FNIRT) registration into MNI152 space, and grand-mean intensity normalization. (2) fMRISurface pipeline projects fMRI data from the cortical gray matter ribbon onto the individual brain surface (fs_LR32K surface space) and then onto template surface meshes (i.e., "dtseries.nii" file), followed by surface-based smoothing using a geodesic Gaussian algorithm. Additional preprocessing steps were applied on rs-fMRI data before the calculation of functional connectivity, including regressing out the signals from white matter and csf, and the bandpass temporal filtering on frequencies between 0.01 and 0.1 Hz. Further details on fMRI data acquisition and preprocessing can be found in Glasser et al. (2013) and Barch et al. (2013).

Construction of Individual Brain Graph

The preprocessed fMRI data were mapped onto the standard surface template with 32k vertices per hemisphere. After removing confounding vertices on the medial surfaces, the cortical mask consists of 59,412 cortical vertices. Then an adjacency matrix A was generated from the surface mesh files, with $A_{ij} = 1$ indicating that the two vertices i and j are shared in a common triangle in the cortical mesh. Since all subjects have already been registered onto the standard surface template during data preprocessing, the adjacency matrix A was also shared across all subjects. As such, a binary brain graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ was constructed for each individual brain, with the node $\mathcal{V}$ indicating each vertex in the cortical mesh, and the edge $\mathcal{E}$ specified by the adjacency matrix indicating whether two nodes are connected or not. The resulting brain graph is sparsely connected and highly localized in space, with each vertex only connecting with two to six nearest vertices on average.

Functional Connectivity Profile as Graph Signals

For each vertex in the cortical mask, we calculated its functional connectivity profile by calculating Pearson correlations on preprocessed fMRI signals and treated it as a feature vector on each node of the graph. In order to save the computational cost and complexity, we did not use the vertex-wise functional connectivity maps but instead calculated the connectivity fingerprints evaluated on hundreds of functional region of interest (ROIs) from a group atlas, e.g., Schaefer400 (Schaefer et al., 2018). The connectivity fingerprint of each cortical vertex x represents the probability of assigning the seed vertex to the same label of each functional ROI. These connectivity profiles were then concatenated and embedded in the individual brain graph as graph $X \in \mathbb{R}^{N=F}$, where N indicates the number of cortical vertices, and F indicates the number of features in the connectivity profiles.

ChebNet Convolution on the Brain Graph

After defining the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with signals $X \in \mathbb{R}^{N=F}$ for each subject, a GNN architecture was applied on the combined graph data $\widetilde{\mathcal{G}} = (\mathcal{V}, \mathcal{E}, X)$ with the aim of integrating the context information of functional connectivity profiles at each vertex from its spatial neighbors by using graph convolutions. Graph convolution relies on the graph Laplacian, which is a smoothing operator characterizing the magnitude of signal changes between adjacent nodes. The normalized graph Laplacian is defined as:

$$L = I - D^{-1/2} A D^{-1/2} \tag{1}$$

where D is a diagonal matrix of node degrees, I is the identity matrix, and A is the adjacent matrix of the graph. The eigen decomposition of Laplacian matrix is defined as $L = \mathrm{U} \triangle \mathrm{U}^T$, where $\mathrm{U} = (u_0, u_1, \cdots u_{N-1})$ is the matrix of Laplacian eigenvectors and is also called graph Fourier modes, and $\triangle = \mathrm{diag}\,(\lambda_0, \lambda_1, \cdots \lambda_{N-1})$ is a diagonal matrix of eigenvalues, specifying the frequency of the graph modes. The convolution between the graph signals $X \in \mathbb{R}^{N=F}$ and a graph filter $g_\theta \in \mathbb{R}^{N=F}$ based on graph $\mathcal{G}$, is defined as their element-wise Hadamard product in the spectral domain:

$$x *_{\mathcal{G}} g_\theta = U\left(U^T g_\theta\right) \odot \left(U^T x\right) = U G_\theta U^T x \tag{2}$$

where $G_\theta = diag\left(U^T g_\theta\right)$ and θ indicate a parametric model for graph convolution g_θ, $U^T x$ projects the graph signals onto the full spectrum of graph modes. To avoid calculating the spectral decomposition of the graph Laplacian, ChebNet convolution (Defferrard et al., 2016) uses a truncated expansion of the Chebyshev polynomials, which are defined recursively by:

$$\mathrm{T}_k(x) = 2x\mathrm{T}_{k-1}(x) - \mathrm{T}_{k-2}(x), \ \mathrm{T}_0(x) = 1, \mathrm{T}_1(x) = x \tag{3}$$

Consequently, the ChebNet graph convolution is defined as:

$$x *_{\mathcal{G}} g_\theta = \sum_{k=0}^{K} \theta_k \mathrm{T}_k\left(\tilde{L}\right) x \tag{4}$$

where $\tilde{L} = 2L/\lambda_{max} - I$ is a normalized version of graph Laplacian with λ_{max} being the largest eigenvalue, θ_k is the model parameter to be learned at each order of the Chebychev polynomials. It has been proven that the ChebNet graph convolution was naturally K-localized in space by taking up to Kth order Chebyshev polynomials (Defferrard et al., 2016), which means that each ChebNet convolutional layer integrates the context of brain activity within a K-step neighborhood. We found that high-order graph convolutions might introduce over smoothing issues and result in decreased functional homogeneity in fMRI signals (**Supplementary Figure 3**). In this study, we used the third-order graph convolutions in our GNN architecture.

Masked Semi-Supervised Graph Convolutional Neural Network for Individualized Cortical Parcellation

The GNN model takes the constructed brain graph $\widetilde{\mathcal{G}} = (\mathcal{V}, \mathcal{E}, X)$ as inputs, where $\mathcal{V}$ is the set of 32k vertices in the cortical surface, $\mathcal{E}$ is the set of edges indicating whether two vertices share a common triangle in the surface, and $X \in \mathbb{R}^{N=F}$ is the set of feature vectors indicating the functional connectivity profiles of each vertex. A series of third-order graph convolution were then applied on the graph signals, with 64 kernels in the

first graph convolutional layer and 201 kernels in the second layer. The learned graph representations of the last layer were transformed to a 201-dimensional probability vector using the SoftMax function. The loss function of the proposed model was defined as follows:

$$Loss = \sum_{n} \sum_{v} w_{n,v} \sum_{k} y_{n,v,k} \log(\mathrm{p}_{v,k}) \quad (5)$$

We used the K–L divergence to compute the difference between the group prior $y_{\mathrm{n,v,k}}$ and the predicted probability $p_{\mathrm{v,k}}$ at each vertex v for each region k. The weight of uncertainty $w_{\mathrm{n,v}}$ was evaluated for each vertex, inferred by a subject-specific confidence mask. The confidence mask was generated as follows (**Supplementary Figure 1**): (1) initial parcellation: assigning each vertex in the cortical mask to the corresponding parcel with the highest similarity in functional connectivity profiles; (2) group-level alignment: excluding vertices in the initial parcellation with different parcel labels as the group atlas; (3) subject-level alignment: overlapping parcellation maps across all available sessions of a single subject. The resulting confidence map contains around half of the cortical mask (56% of vertices) that contributed to the final loss function in all subjects. The benefits of using the above loss function include: (1) high contributions from the vertices near the center of brain parcels, preserving a consistent global topographic organization across subjects; and (2) small contributions from the vertices around the areal borders, retaining the intersubject variability to some degree by introducing mismatching labels across subjects.

The proposed pipeline of individualized brain parcellation (as shown in **Supplementary Figure 5**) was trained on 50 subjects with two sessions for 100 epochs with the batch size set to 1 (processing one subject at a time), evaluated on all other subjects in the HCP S1200 dataset as well as the test–retest dataset. To avoid overfitting, an early stopping of 10 epochs was used, along with an additional l_2 regularizations of 0.0005 on model weights and a dropout rate of 0.5 on each graph convolutional layer. The best model over 100 training epochs was saved and further evaluated on the independent test sets from HCP S1200 and test–retest datasets.

Comparison With Alternative Machine Learning Approaches

Many approaches of individualized brain parcellation have been proposed in the literature, for instance, by using an iterative clustering of fMRI signals (Wang et al., 2015) or hierarchical inference through a multilevel Bayesian model (Kong et al., 2021). We included two individualized methods as baseline approaches in the current study. The first approach aligned the group atlas into individual brains using the multimodal alignment protocol (Robinson et al., 2018), which utilized myelin maps, resting state network maps, and visuotopic maps to align cortical areas. The second approach modified the individual mapping of brain atlas using a multisession hierarchical Bayesian model (Kong et al., 2019, 2021), by explicitly modeling the variability in functional connectivity at the levels of intra- and intersubject. The preprocessed HCP fMRI dataset has already included the copies of the first approach (i.e., the MSMALL version), for which we compared the functional homogeneity between our individualized brain parcellation and group-registered brain atlas (see section "Resting State Functional Homogeneity of Brain Parcels"). For fair comparisons with the second approach, we reran the MS-HBM approach on the same group of subjects along with the same preprocessing steps as in our model and compared the distribution of topographic variability between the two approaches (see section "The Intersubject Reliability and Variability of Individual Properties"). We chose the Schaefer400 atlas (Schaefer et al., 2018) as the referenced group atlas for all approaches and used the author-suggested model parameters for rerunning the MS-HBM approach (Kong et al., 2021), including priors of group spatial (100), markov random field (MRF) smoothing (50), and gradient-based spatial localization (50).

The Intersubject Reliability and Variability of Individual Properties

The reliability of individual parcellation and its intersubject variability were evaluated on both HCP S1200 and test–retest datasets. Each subject underwent two (HCP S1200 dataset) or four (test–retest dataset) fMRI sessions. The reliability of individual parcellation was evaluated by the averaged Dice coefficients among all possible pairs. The Dice coefficient was first evaluated on each brain parcel using the equation $(2 \times \mathrm{A} \cap \mathrm{B})/(\mathrm{A} + \mathrm{B})$, where A and B indicate two different parcellation schemes, and then averaged across the whole cortex. The effect size of the intersubject variability was measured by Cohen's d, representing the standardized difference between the mean values of two distributions, defined as follows:

$$\text{Cohen's } d = \frac{\mu_{inter} - \mu_{intra}}{\sqrt{\sigma_{inter}^2 + \sigma_{intra}^2}} \quad (6)$$

where μ_{inter} and σ_{inter} represent the mean and standard deviation of the variabilities between each pair of subjects, while μ_{intra} and σ_{intra} represent the mean and standard deviation of the variabilities between different fMRI scans of the same subject. We used different indices to measure the variability of parcellation maps and connectivity profiles. Specifically, the variability in brain parcellation (i.e., areal topographic variability) was measured by the Dice coefficient between two parcellation maps. The variability of connectivity profiles (i.e., functional connectivity variability) was measured by the correlation coefficients of the functional connectivity profiles between two fMRI runs.

For further validation of the biological basis of the intersubject variability in brain parcellation, additional association analyses were conducted for the areal topographic variability, including the variability of connectivity profiles, distribution of T1w/T2w myelin ratio, as well as the sensory-fugal map of laminar differentiation. As a quantitative measure of the myelin content of cerebral cortex, the myelin ratio map was defined as the ratio of T1w and T2w structural images on each subject and then averaged across all subjects on the HCP S1200 release (Glasser and Van Essen, 2011). The cortical myelin map was then

mapped onto the chosen group atlas by averaging the T1w/T2w ratio within each ROI. The laminar differentiation map identifies four different cortical types based on their cytoarchitectonic organization, namely, paralimbic, heteromodal, unimodal, and idiotypic areas (Mesulam, 1998). A cortical mapping of laminar differentiation was generated by assigning each ROI in the chosen group atlas to one of the four types.

Resting State Functional Homogeneity of Brain Parcels

Functional homogeneity in resting-state fMRI was defined as the averaged Pearson's correlations of fMRI signals between all pairs of vertices within each parcel, adjusted for the parcel size (Kong et al., 2019, 2021). Higher functional homogeneity indicates similar brain dynamics of fMRI activity within the same parcel. The functional homogeneity was evaluated on the test set from the HCP S1200 (in a total of 928 subjects, after excluding the training data, subjects included in the test–retest data, and some abnormal data with missing functional scans) for both group-registered parcellation and masked semi-supervised graph convolutional neural network (MSGCN) individualized brain parcellation. A two-sided paired *t*-test was applied to test the significant differences of functional homogeneity between the two approaches.

Regional Variability and Homogeneity in Task Activation

We chose one representative task contrast from each of the seven cognitive domains: right-hand movement for the motor task, two back conditions on tool images for working memory, math vs. story for language processing, faces vs. shapes for emotional processing, theory of mind vs. random motion for social cognition, reward for the gambling task, and relational processing vs. pattern match for relational processing. For each task contrast, we evaluated the regional variability and homogeneity in the task activations from the test subjects of the HCP S1200 dataset. Regional variability in task activation was evaluated by calculating the standard deviation of brain activation within each parcel by using the beta maps derived from the generalized linear model (GLM) analysis. Lower task variability indicates small variations of task activations within each brain parcel and large variations along the areal boundaries. Regional homogeneity in task activation was evaluated by the mean activation strength in each brain parcel. Higher task homogeneity indicates better functional alignment in the task activations across different subjects.

Prediction Cognitive Behaviors Using Individualized Parcellation

To further validate that individualized parcellation improves the intersubject functional alignment in brain organization and captures meaningful aspects of human cognition, we performed another experiment to predict cognitive behaviors using the intersubject variability in brain parcellation. A kernel regression method was used to predict the behavioral score of the test subject based on the assumption that similar topography in brain parcellation induced similar performance in behavior, defined as follows:

$$y \approx \sum_i Dice(l, l_i) y_i$$

where y_i represents the behavioral score of the i-th subject in the training set, l_i represents the parcellation map of the subject i, Dice(l, l_i) represents the Dice coefficient of parcellation maps between the test subject and the subject i, and y represents the behavioral score of the test subject. An additional l_2-regularization term was used to prevent overfitting issues, with the regularization parameter determined by a fivefold cross-validation procedure.

For each of the 58 cognitive behaviors, a prediction model was trained and evaluated on 928 individualized brain parcellations. Specifically, we trained the prediction model on 200 subjects and evaluated it on the rest of the 728 subjects. First, we used a five-fold cross-validation strategy to determine the optimal prediction model (including model parameters and the l_2-regularization parameter). The model was then used to predict the behavioral scores on the test subjects. Finally, the performance of the prediction model was evaluated by calculating the Pearson correlation of predicted and measured behavioral scores.

RESULTS

Masked Semi-Supervised Graph Convolutional Neural Network Model for Individualized Brain Parcellation

The MSGCN model (as shown in **Supplementary Figure 5**) was evaluated using 1,022 subjects from HCP S1200 dataset with two fMRI sessions for each subject, among which 40 subjects were randomly chosen for model training, 10 subjects for validation, and the rest of the datasets used for model testing. During model training, a vertex-level brain graph was first constructed from individual T1-weighted brain images, indicating the spatial adjacency between any two vertices in the cortical surface. The functional connectivity profiles of each vertex were then embedded in the brain graph as graph signals. Finally, the areal probability of each vertex was inferred by using a two-layer graph convolution architecture with a subject-specific confidence mask. As a result, the MSGCN model extracted reliable functional characteristics of each brain parcel inferred from a large population and generalized over data of unseen subjects by revealing reliable parcellation maps for each of the test subjects.

Topographic Organization of Individual Parcellation and Its Test–Retest Reliability

The cortical parcellation maps on individual brains followed a similar global topographic organization as the group atlas (e.g., Schafer400 atlas), indicating a relatively high similarity [dice = 0.847 $\pm$ 0.013 (mean $\pm$ std)] among all fMRI sessions of 1,022 HCP subjects. First, a stable brain parcellation scheme

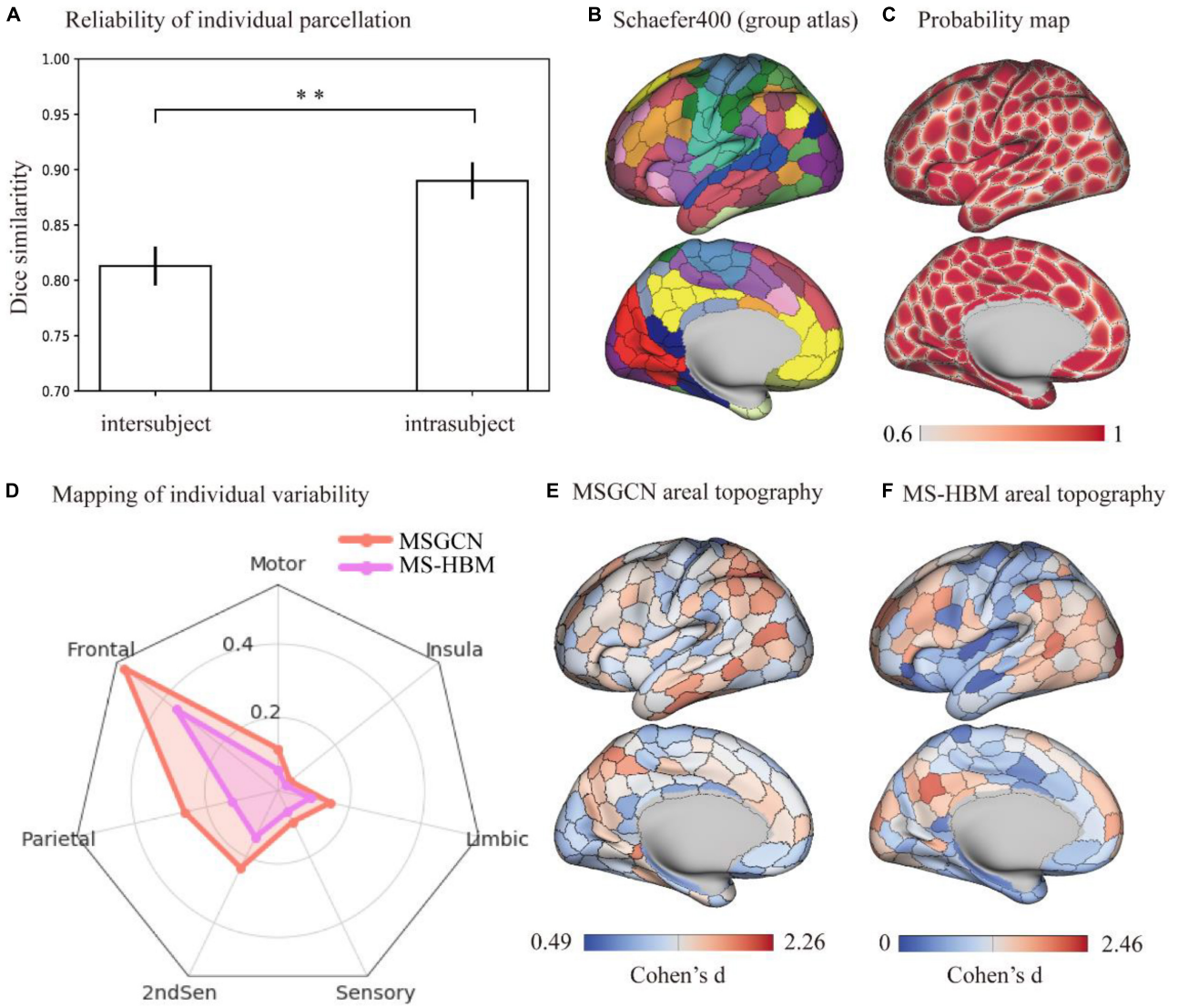

FIGURE 1 | Topographic organization of individual parcellation and its reliability measured on the HCP S1200 dataset. **(A)** Reliability of individual parcellation. The masked semi-supervised graph convolutional neural network (MSGCN) individualized parcellation showed significantly higher similarity for intrasubject (0.889 ± 0.025) than intersubject (0.810 ± 0.021), detected by a paired *t*-test ($p < 0.001$). **(B)** Cortical mapping of the group atlas. We used a similar color scheme as the Yeo-7 networks for the Scaefer400 atlas, while the areal borders were delineated by the gray line. **(C)** Probability map of individualized parcellation maps. The parcellation maps on all subjects were summarized into a population probability map, indicating the probability of assigning the vertex to the same parcel among the HCP population. A population threshold of 60% was applied to the probability map. **(D)** Distribution of intersubject variability in brain parcellation among functional networks. We used Cohen's *d* to evaluate the effect size of variability in individualized parcellation generated by MSGCN (orange line) and multisession hierarchical Bayesian model (MS-HBM) (purple line), both of which showed large variations in the association cortex including "Ffontal," "parietal," and "2ndSen" regions and low variability in the motor and sensory cortices. However, the MSGCN model detected higher variability in high-order cognitive areas, especially in the frontal and parietal regions. **(E)** Cortical mapping of intersubject variability for the MSGCN model. **(F)** Cortical mapping of intersubject variability for the MS-HBM. **p-value < 0.01.

was revealed such that the majority of vertices showed consistent parcel labels across subjects, with over 75% of cortical vertices showing a high population probability >0.6. We observed near-perfect alignment at the center of brain parcels along with notable variability around the areal borders (**Figure 1C**). Second, the reliability of individual parcellation, measured by the intra- and intersubject similarity (**Figure 1A**), showed significantly higher consistency among multiple sessions of the same subject (dice = 0.889 ± 0.025) than between different subjects (dice = 0.810 ± 0.021), as revealed by a paired *t*-test ($p < 0.001$). Moreover, the intersubject variability in brain parcellation (i.e., areal topographic variability), evaluated by the Cohen's *d* effect size among individualized parcellation maps, was not uniformly distributed in the cerebral cortex, but rather followed the functional organizational principles (**Figure 1E**). After mapping the areal topographic variability onto the seven functional networks (Ito, 2018), we found consistent parcellation schemes for the motor and sensory

cortices (i.e., low variability) and divergent brain parcels for the association cortices (i.e., high variability) including frontal, parietal, and temporal areas (**Figure 1D**). Compared with another individualized parcellation approach, e.g., the MS-HBM, our MSGCN model revealed higher areal topographic variably, in general (**Supplementary Figure 4**), with a similar level of areal topographic variability in the primary cortex but detecting much higher variability in the secondary and association cortices (**Figure 1D**).

The reliability of the MSGCN parcellation was further validated on the test–retest datasets, by revealing reliable cortical parcellation on the same subject and detecting notable variations in the areal topographic organization between subjects. We also found much higher similarity in the parcellation maps among multiple sessions of the same subject compared with similarity between different subjects (dice = 0.9122 ± 0.011 and 0.8370 ± 0.012, respectively). The cortical parcellation maps on two exemplar subjects are shown in **Figure 2**. Individual-specific areal topography was revealed, for instance, in the inferior parietal lobe (the marked region in **Figure 2**), indicating a similar topographic pattern across the four test–retest sessions of the same subject but largely different areal topography between subjects. Our findings suggest that the MSGCN individualized parcellation detects reliable individual differences in the areal topography.

Interpretability of Intersubject Variability in Brain Parcellation

The intersubject variability of MSGCN parcellation was biologically meaningful and followed the myelination and cytoarchitectonic organizational principles of the human brain. We observed a significant association between the areal topographic variability and the variability of connectivity profiles ($r = 0.42$, $p < 0.001$, **Figure 3B**), both of which showed high variations in the heteromodal and unimodal regions, along with low variability in the idiotypic and paralimibic regions, regions specified by an independent atlas of laminar differentiation (**Figure 3D**), generally following a sensory-fugal gradient from sensory–motor and unimodal areas to heteromodal areas (Mesulam, 1998). By contrast, although exhibiting a strong association with the functional connectivity variability as well ($r = 0.30$, $p < 0.001$, **Figure 3B**), the MS-HBM parcellation showed a very different distribution of areal topographic variability that weakly aligned with the laminar differentiation map. For instance, much higher variability was detected in the idiotypic and unimodal regions rather than in the heteromodal regions (**Figure 3D**). Moreover, the areal topographic variability of the MSGCN parcellation was significantly associated with the T1w/T2w myelin ratios as well ($r = -0.27$, $p < 0.001$), indicating low variability in the primary motor and visual cortex, which are heavily myelinated, and high variability in the association cortices, which are more lightly myelinated (Glasser and Van Essen, 2011). Such association with the myelination organization was missing in the MS-HBM parcellation ($r = 0.005$, $p = 0.09$). Our results suggest that the intersubject variability revealed by the MSGCN model follows the global distribution of myelo- and cytoarchitecture, as well as the variability in functional brain organization.

Improved Functional Homogeneity With Reduced Task Variability

The functional homogeneity at rest measures the internal functional consistency of brain parcels. As shown in **Figure 4A**, the global functional homogeneity on the validation dataset was gradually improved during the model training process. Besides, the individualized brain parcellation on the unseen test subjects also exhibited higher functional homogeneity than the initial parcellation derived from the group atlas (see **Supplementary Figure 1** for an example), as detected by a paired t-test ($p = 0.0006$). The averaged functional homogeneity of the MSGCN parcellation was 0.137 ± 0.001 (mean ± se), evaluated on all 928 test subjects from the HCP S1200 dataset, with a 4% improvement at the whole-brain level compared with the group-registered atlas (**Figure 4B**).

On the other hand, the regional variability in task activations (task variability) measures the functional alignment between the intrinsic brain organization at rest and task-evoked brain activation during cognitive tasks. Our results showed that the MSGCN parcellation captured more homogenously distributed task activations. Overall, MSGCN parcellation showed better functional alignment at the whole-brain level for the seven tasks in HCP data, namely, language (math–story), emotion (faces–shapes), gambling (reward), relational (rel–match), social (tom–random), motor (rh-avg), and working memory (2BK-tool) tasks, with significantly reduced task variability compared with the group atlas [False Discovery Rate (FDR) corrected p-value < 0.05, as shown in **Supplementary Table 1**]. For instance, the changes in the subject-specific activation map of language task followed the areal borders identified by the MSGCN parcellation (**Figure 5B**), e.g., lower regional variability and higher homogeneity in the inferior parietal regions (**Figure 5C**). Compared with the group-registered Schaefer400 atlas, the MSGCN parcellation showed smaller variability in task activation within the detected region ($p = 0.02$) along with higher functional consistency across subjects ($p = 0.04$). Moreover, the MSGCN parcellation detected lower variability in task activations for all seven tasks at the whole-brain level (**Figure 5A**). Our findings indicate that the MSGCN parcellation reveals a better functional alignment across subjects in both resting state and task activation.

Prediction Cognitive Behaviors Using Masked Semi-Supervised Graph Convolutional Neural Network Brain Parcellation

For each of the 58 cognitive behaviors, we trained a prediction model based on the parcellation maps of 200 subjects and evaluated the model on the rest of the 728 subjects. The models achieved significant predictions (p-value < 0.05) on 25 behavioral scores (as shown in **Figure 6A**), including motor (Strength_Unadj, Endurance_Unadj), cognition (PicVocab_Unadj), language (ReadEng_Unadj), and others (see

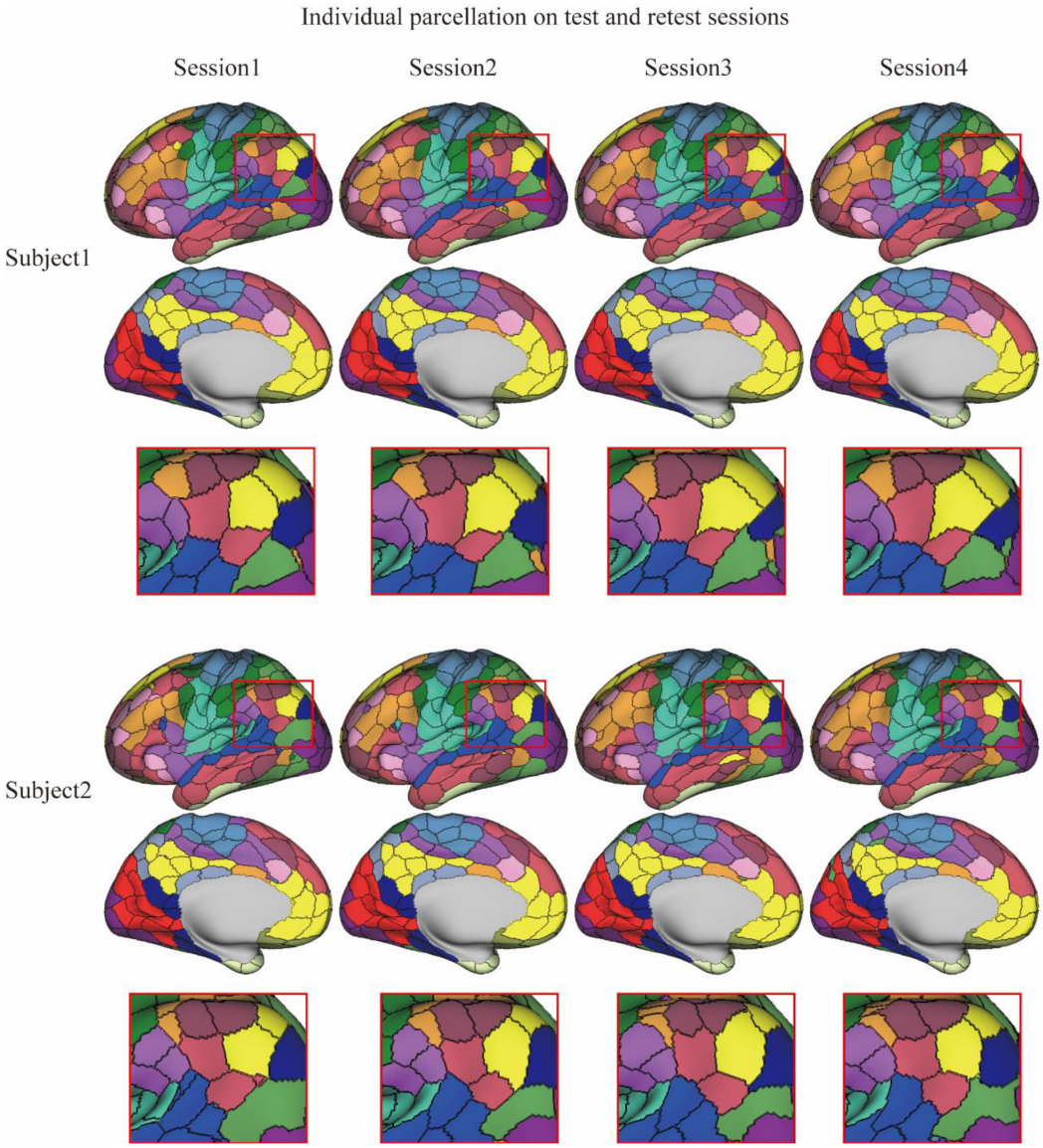

FIGURE 2 | Individualized brain parcellation on two exemplar subjects and four test–retest sessions. The inferior parietal lobe (zoomed-in areas within the red rectangle) showed a similar topographic pattern across the four sessions of each subject, but detected significantly different areal topography between two subjects.

Supplementary Table 2 for a full list of 25 behavioral measures). For instance, we found significant associations between predicted and measured behavioral scores for motor (Strength_Unadj, $r = 0.398$, $p = 7e\text{-}30$) and cognition (PicVocab_Unadj, $r = 0.1979$, $p = 5e\text{-}8$), as shown in **Figures 6B,C**. These findings indicated that the individualized parcellation maps captured meaningful aspects of individual variability in brain topography and human cognition.

DISCUSSION

In this study, we propose an individualized cortical parcellation method that projects the group atlas onto individual brains by taking into account the variations in brain topography and functional connectivity. The proposed MSGCN parcellation generated highly consistent parcellation maps on multiple sessions of the same subject (intrasubject similarity = 0.89) while capturing reasonable topographic variations between subjects (intersubject similarity = 0.81). Compared with other baseline approaches including the group-registered atlas and MS-HBM, our method generated more homogeneous parcels on individual brains that strongly aligned with the intrinsic brain organization at rest and task-evoked brain activation of cognitive tasks. Moreover, the MSGCN parcellation revealed higher intersubject variability in the association cortices while keeping a stable parcellation on the primary cortex, indicating a sensory-fugal gradient from primary and unimodal areas to heteromodal

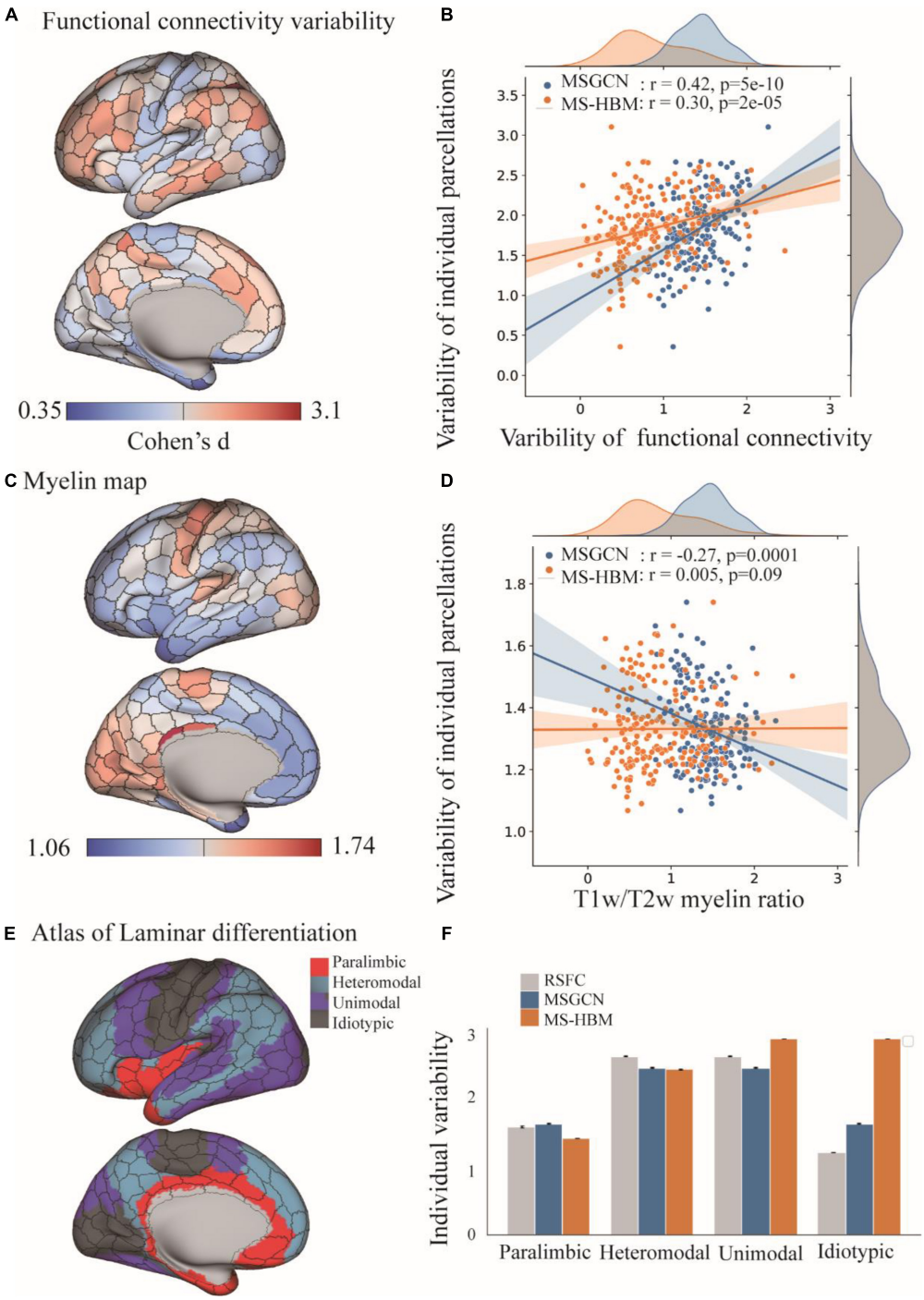

FIGURE 3 | Distribution of intersubject variability in the masked semi-supervised graph convolutional neural network (MSGCN) parcellation and its association with functional, myelination, and cytoarchitecture organization. **(A)** Cortical mapping of intersubject variability in the functional connectivity profiles, with the red colors indicating high variability among subjects and blue indicating low variability. **(B)** Associations of the areal topographic variability with the variability of functional connectivity profiles. We found a strong association in the MSGCN model ($r = 0.42$, $p < 0.001$), which was much higher than multisession hierarchical Bayesian model (MS-HBM) ($r = 0.30$, $p < 0.001$). **(C)** Cortical mapping of the T1w/T2w myelin ratio map, with the red colors indicating high myelination content in the areas. **(D)** Associations of the areal topographic variability with the distribution of the myelin ratio map. We found a significant negative association in the MSGCN model ($r = -0.27$, $p < 0.001$), but not in MS-HBM ($r = 0.005$, $p = 0.09$). **(E)** Cortical mapping of laminar differentiation, with different colors representing one of the four cortical types, namely, paralimbic, heteromodal, unimodal, and idiotypic areas. **(F)** Distribution of the intersubject variability in both functional connectivity profiles and individual parcellation. The intersubject variability was evaluated by using Cohen's *d*. Both functional connectivity profiles and MSGCN individualized parcellation showed relatively higher variability in the heteromodal and unimodal areas than the paralimbic and idiotypic areas. By contrast, the MS-HBM parcellation identified much higher variability in the idiotypic areas.

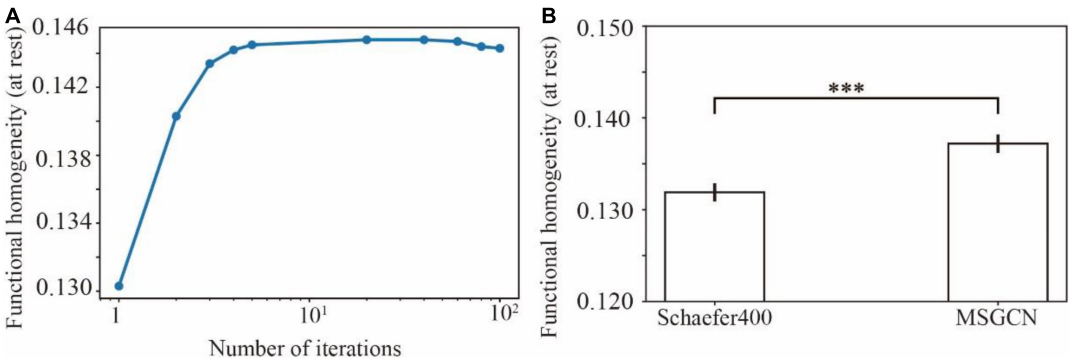

FIGURE 4 | Functional homogeneity of individualized brain parcels in the masked semi-supervised graph convolutional neural network (MSGCN) parcellation. **(A)** The changes in functional homogeneity were evaluated in the validation dataset during the training processes. **(B)** Significantly higher functional homogeneity was detected in the MSGCN model (0.137 ± 0.001) than the group-registered Schaefer400 atlas (0.132 ± 0.001), as revealed by a paired *t*-test ($p = 0.0006$). ***p-value < 0.001.

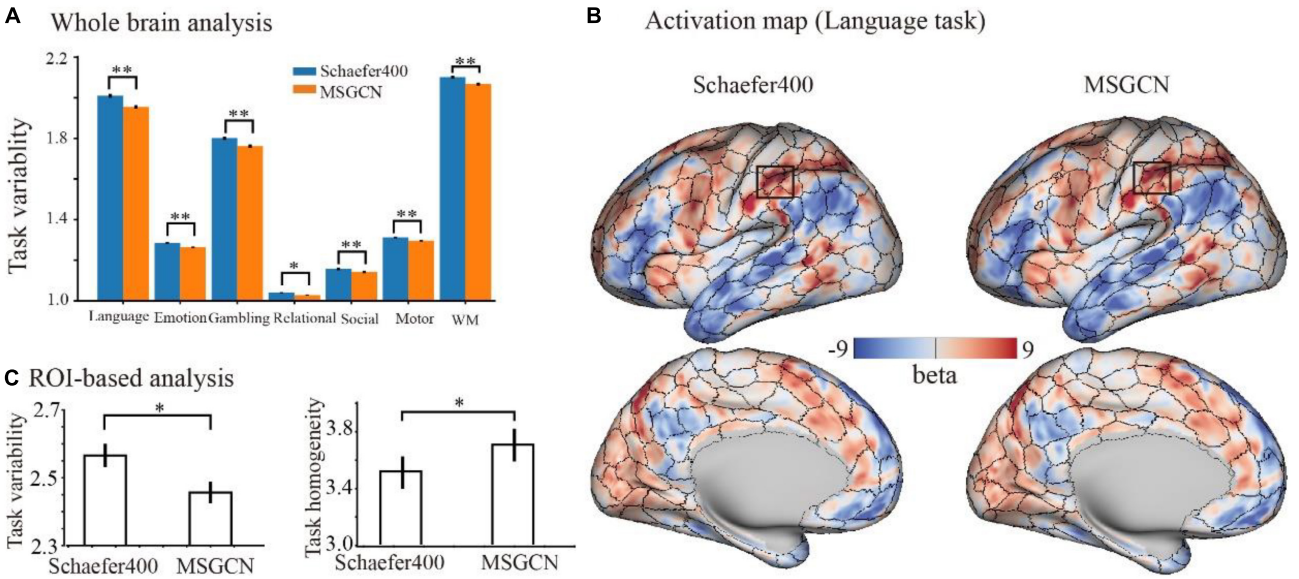

FIGURE 5 | Functional alignment of the masked semi-supervised graph convolutional neural network (MSGCN) parcellation with task activations in the Human Connectome Project (HCP) tasks. **(A)** Task variability of MSGCN parcellation and Schaefer400 atlas on the seven tasks. We observed significantly lower (FDR corrected p-values < 0.01) regional variability in task activations by using the MSGCN parcellation (in orange) compared with the Schaefer400 atlas (in blue). **(B)** Representative activation map of the language task on a single subject, with the areal borders identified by the group-registered Schaefer400 atlas and MSGCN individualized brain parcellation, respectively. The visual assessment suggested that the fluctuations in the subject-specific task activation map went along the areal borders identified by the MSGCN parcellation but not the Schaefer400 atlas. **(C)** Task variability and homogeneity of the rectangular area marked in panel **(B)**. Quantitative comparisons suggested significantly lower variability and higher homogeneity of task activation in the detected region by using the MSGCN parcellation compared with the Schaefer400 atlas, as detected by paired *t*-tests (p-value = 0.02 and 0.04, respectively). *p-value < 0.05; **p-value < 0.01.

areas. Such topographic variability in individualized parcellation strongly associated with the variability of functional connectivity profiles and cognitive behaviors, and generally followed the myelination, cytoarchitecture, and functional organization of the human brain.

Individualized brain parcellation has played a more and more important part in neuroscience research and clinical studies, which not only better predicted human cognition, behaviors, personality, and emotion (Kong et al., 2021), but also captured reliable and inheritable variability in the topographic organization of the human brain (Anderson et al., 2021), as well as more precise diagnosis and treatment of neurological disorders (Mueller et al., 2015; Wang et al., 2020). Yet, due to the inevitable contamination of fMRI signals by physiological noises and head motions, the traditional individualized parcellation approaches usually suffer from low generalizability on new datasets and low consistency among repeated scans. To address this issue, we applied a high-order graph convolution along with a multilayer deep learning architecture in this study. As a generalization of the conventional convolutional operations

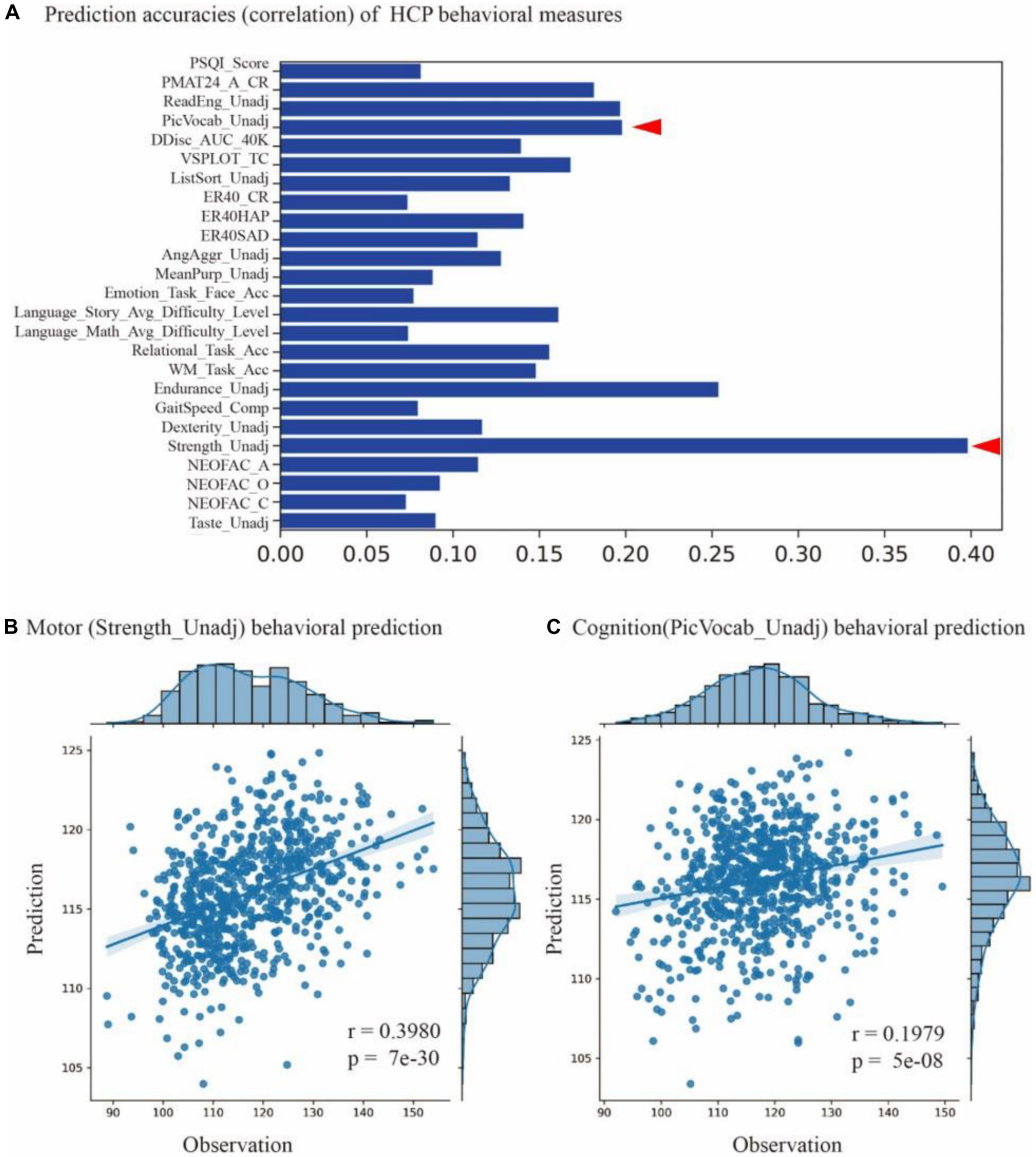

FIGURE 6 | Prediction cognitive behaviors using masked semi-supervised graph convolutional neural network (MSGCN) individualized brain parcellation. **(A)** Prediction accuracies on 25 behavioral scores measured by the Pearson correlation of predicted and measured behavioral scores, with accuracy ranging from 0.07 to 0.4 (p-values < 0.05). **(B)** Significant associations between the predicted and measured behavioral scores for motor (Strength_Unadj, $r = 0.398$, $p = 7e\text{-}30$) and **(C)** cognition (PicVocab_Unadj, $r = 0.1979$, $p = 5e\text{-}8$), as indicated by red triangles in **(A)**.

onto nongrid structures, graph convolution applies a series of low-frequency filters on the graph modes, also known as using multiple smoothing kernels on the cortical surface, and detects low fluctuations of functional connectivity along the vertex-level brain graph. The smoothing effect was controlled by the order of ChebNet graph convolution. Using this graph convolution architecture, the model generated very stable parcellations on a large population with over 75% of cortical vertices following the global topographic organization by assigning them to the same parcels among different subjects. Besides, highly consistent parcellation maps (**Figures 1**, **2**) were generated on the multiple sessions of the same subject (dice = 0.89), along with lower consistency between subjects (dice = 0.81). Yet, high-order kernels and deep architectures may introduce over-smoothing issues in GNN, which tends to generate identical parcellation maps on all subjects. We have also observed such over-smoothing effect in our MSGCN model such that the functional homogeneity of individualized brain parcels gradually reduced when using higher-order graph convolutions ($K > 3$ in **Supplementary Figure 3**). Considering the tradeoff between the intersubject variability and intrasubject reliability in individualized brain parcellation, we used the third-order graph convolution along with two layers in our model, which not only revealed high consistency among test–retest sessions

but also captured notable variability of brain parcellation between subjects.

The separation of intersubject variability from the randomly appearing noise components have also been considered in previous individualized parcellation models, for instance, using a hierarchical Bayesian model to quantify the variability of functional connectivity at multiple levels (Kong et al., 2021). However, by treating each area equally through a global concentration parameter on all regions, the MS-HBM approach revealed similar levels of topographic variability and heritability among the primary and association cortices (Anderson et al., 2021), which is in congruence with the well-known sensory-fugal gradient in brain organization. To avoid this effect, we chose a data-driven approach to learn the functional characteristics of each brain parcel inferred from a large group of subjects and to encode the intra- and inter-region heterogeneity of functional connectivity through a large set of graph convolutional kernels. The detected region-specific connectivity fingerprints have been proven to be highly generalizable and reliable when inferring the areal probability on unseen subjects (**Figures 1**, **2**). The main reasons that drive this effect include (1) integrating the local connectivity context of each vertex instead of treating each vertex independently, (2) detecting the fluctuations in functional connectivity profiles within a small neighboring area in the cortical surface, and (3) indicating the areal borders on individual brains by using the gradients of function connectivity fluctuations at multiple levels. As a result, the model revealed large intersubject variability in brain parcellation, and such topographic variability was not randomly or uniformly distributed across the cerebral cortex, but rather followed the global distribution of myelination, cytoarchitecture, and functional organization of the human brain (**Figure 3**). The model demonstrated low variability in the primary and unimodal cortices that are heavily myelinated and large variations in the heteromodal and association cortices that are lightly myelinated. Similar associations with the myelination and cytoarchitecture organization have also been reported in both brain anatomy and function (Huntenburg et al., 2018; Demirta et al., 2019), suggesting a sensory-fugal gradient in the individual developmental and evolutionary expansion of the cerebral cortex (Glasser and Van Essen, 2011).

There are two main goals for the individualized brain parcellation, including (1) functional homogeneity of brain parcels on individual brains, and (2) consistent topographic organization across multiple sessions and different subjects. The tradeoff between these two goals was addressed by using a semi-supervised learning framework with subject-specific confidence masks. Specifically, during model training, a subject-specific confidence mask was used for the guidance of topographic alignment across subjects, indicating the true labels for a small portion of cortical vertices, i.e., the labels extracted from the group atlas (see **Supplementary Figure 1** for an example). These true labels were then used to learn the association between brain topography and functional connectivity profiles and to predict the parcellation of unlabeled data in the training subjects as well as for unseen test subjects. Similar to previous approaches (Wang et al., 2015), we started from an initial parcellation (see **Supplementary Figure 1** for an example) that had the highest functional homogeneity by grouping cortical vertices according to their functional connectivity profiles to ensure the functional homogeneity of brain parcels (**Supplementary Figure 2**). Additional modifications on this parcellation map, i.e., excluding vertices that were misaligned across sessions of the same subject or with the group atlas, introduced the important features of topographic alignment in individualized parcellation. By implementing the resulting confidence map with the semi-supervised learning, the model captured homogenous parcels on individual subjects that also followed the global topographic organization of the group atlas. It is worth noting that the MSGCN parcellation not only generated the full parcellation of half-labeled training subjects but also made predictions over unseen test subjects where no labeled data were included. The generalizability of the model barely impacts by the size of training set such that labeling all subjects from the dataset, i.e., both training and testing data were drawn from the same sets of subjects, only achieved 2% of improvement on the functional homogeneity but showed much lower intra- and intersubject reliability (**Supplementary Figure 2**).

The MSGCN individualized brain parcellation not only generated replicable parcels on individual brains but also captured meaningful individual variability in brain topography and human cognition. The topographic variability generally followed a sensory-fugal gradient from primary and unimodal areas to heteromodal areas, with high variations in the association cortices while keeping a stable parcellation on the primary cortex. Such topographic variability strongly associated with the variability of functional connectivity profiles, and generally followed the myelination, cytoarchitecture, and functional organization of the human brain. More importantly, the topographic variability was highly predictive to individual variability of cognitive behaviors ranging from motor to language to cognition. However, not all behavioral scores showed a strong association with the predicted scores (only 25 out of 58 behaviors with p-value < 0.05). This is probably due to the implementation of the Dice coefficient as the kernel function in the prediction model, which was a global measure of similarity in brain parcellation and had limited power to detect the variability in specific brain regions and networks. Other kernel functions, for instance, resting-state functional connectivity and regional morphological statistics, as well as other prediction models, could be explored in the future.

LIMITATIONS

In the current study, we used the Schaefer cortical parcellation with 400 regions as the referenced group atlas. Yet, the proposed model is not limited to a specific atlas or specific resolutions per se, but rather easily generalized to other parcellation schemes including functional, anatomical, or multimodal atlases. It is worth noting that, in order to balance between the internal homogeneity in connectivity profiles and consistent topographic organization across subjects, the optimal setting of the graph convolution architecture should be tested when applying to new

datasets and atlases, including the order of graph convolution, the number of convolutional layers, as well as the confidence masks.

CONCLUSION

We proposed a masked semi-supervised GNN model for individualized brain parcellation taking into account the homogeneity of functional connectivity profiles, alignment of topographic organization across subjects, as well as the reliability of test–retest data on individual brains. Compared with other individualized approaches, the MSGCN parcellation generated more homogenous brain parcels at rest and during cognitive tasks. The model captured high topographic variability that was mainly distributed in the associated cortices while keeping a stable parcellation in the primary and unimodal areas, and generally followed the myelination, cytoarchitecture, and functional organization of the human brain. Moreover, the topographic variability strongly associated with the functional connectivity variability and significantly predicted a series of cognitive behaviors ranging from motor to language to cognition. This study provides new avenues for precise mapping of cortical areas onto individual brains, and shows potential applications in locating personalized functional areas in the diagnosis and treatment of neurological disorders.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Connectome Project (HCP). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YZ, WQ, and TJ: conceptualization. WQ, LM, and YZ: methodology. WQ and YZ: visualization and writing—original draft, review, and editing. WQ, YZ, TJ, and LM: investigation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by the Science and Technology Innovation 2030-Brain Science and Brain-Inspired Intelligence Project (Grant No. 2021ZD0200201) and the Major Scientific Project of Zhejiang Lab (Nos. 2020ND8AD02 and 2021ND0PI01).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.838347/full#supplementary-material

REFERENCES

Amunts, K., and Zilles, K. (2015). Architectonic mapping of the human brain beyond Brodmann. *Neuron* 88, 1086–1107. doi: 10.1016/j.neuron.2015.12.001

Amunts, K., Mohlberg, H., Bludau, S., and Zilles, K. (2020). Julich-Brain: a 3D probabilistic atlas of the human brain's cytoarchitecture. *Science* 369, 988–992. doi: 10.1126/science.abb4588

Anderson, K. M., Ge, T., Kong, R., Patrick, L. M., and Holmes, A. J. (2021). Heritability of individualized cortical network topography. *Proc. Natl. Acad. Sci.* U.S.A. 118:e2016271118. doi: 10.1073/pnas.2016271118

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., et al. (2013). Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* 80, 169–189. doi: 10.1016/j.neuroimage.2013.05.033

Blumensath, T., Jbabdi, S., Glasser, M. F., Van Essen, D. C., Ugurbil, K., Behrens, T. E., et al. (2013). Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. *Neuroimage* 76, 313–324. doi: 10.1016/j.neuroimage.2013.03.024

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in Ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig: Barth.

Buckner, R. L., Krienen, F. M., and Yeo, B. T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. *Nat. Neurosci.* 16, 832–837. doi: 10.1038/nn.3423

Cui, Z., Li, H., Xia, C. H., Larsen, B., Adebimpe, A., Baum, G. L., et al. (2019). Individual variation in control network topography supports executive function in youth. *bioRxiv* [Preprints] 694489, doi: 10.1016/j.neuron.2020.01.029

Defferrard, M. E. L., Bresson, X., and Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. Neural Inf. Process. Syst*. 29, 3844–3852.

Demirta, M., Burt, J. B., Helmer, M., Ji, J. L., Adkinson, B. D., Glasser, M. F., et al. (2019). Hierarchical heterogeneity across human cortex shapes large-scale neural dynamics. *Neuron* 101, 1181–1194. doi: 10.1016/j.neuron.2019.01.017

Eickhoff, S. B., Constable, R. T., and Yeo, B. T. (2018a). Topographic organization of the cerebral cortex and brain cartography. *Neuroimage* 170, 332–347. doi: 10.1016/j.neuroimage.2017.02.018

Eickhoff, S. B., Yeo, B. T., and Genon, S. (2018b). Imaging-based parcellations of the human brain. *Nat. Rev. Neurosci.* 19, 672–686. doi: 10.1038/s41583-018-0071-7

Eickhoff, S. B., Thirion, B., Varoquaux, G. E. L., and Bzdok, D. (2015). Connectivity-based parcellation: critique and implications. *Hum. Brain Mapp.* 36, 4771–4792. doi: 10.1002/hbm.22933

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1-a

Glasser, M. F., and Van Essen, D. C. (2011). Mapping human cortical areas in vivo based on myelin content as revealed by T1-and T2-weighted MRI. *J. Neurosci.* 31, 11597–11616. doi: 10.1523/JNEUROSCI.2180-11.2011

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., et al. (2013). The minimal preprocessing pipelines for the human connectome project. *Neuroimage* 80, 105–124. doi: 10.1016/j.neuroimage.2013.04.127

Huntenburg, J. M., Bazin, P., and Margulies, D. S. (2018). Large-scale gradients in human cortical organization. *Trends Cogn. Sci.* 22, 21–31. doi: 10.1016/j.tics.2017.11.002

Ito, J. (2018). *Syllable Theory in Prosodic Phonology*. London: Routledge.

Kong, R., Li, J., Orban, C., Sabuncu, M. R., Liu, H., Schaefer, A., et al. (2019). Spatial topography of individual-specific cortical networks predicts human cognition, personality, and emotion. *Cereb. Cortex* 29, 2533–2551.

Kong, R., Yang, Q., Gordon, E., Xue, A., Yan, X., Orban, C., et al. (2021). Individual-Specific Areal-Level parcellations improve functional connectivity prediction of behavior. *Cereb. Cortex* 31, 4477–4500. doi: 10.1093/cercor/bhab101

Li, M., Wang, D., Ren, J., Langs, G., Stoecklein, S., Brennan, B. P., et al. (2019). Performing group-level functional image analyses based on homologous functional regions mapped in individuals. *PLoS Biol.* 17:e2007032. doi: 10.1371/journal.pbio.2007032

Ma, L., Zhang, Y., Zhang, H., Cheng, L., Zhuo, J., Shi, W., et al. (2021). BAI-Net: individualized Human Cerebral Cartography using Graph Convolutional Network. *bioRxiv* [Preprint]. doi: 10.1101/2021.07.15.452577

Mesulam, M. (1998). From sensation to cognition. *Brain* 121, 1013–1052.

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T., Sepulcre, J., Sabuncu, M. R., et al. (2013). Individual variability in functional connectivity architecture of the human brain. *Neuron* 77, 586–595.

Mueller, S., Wang, D., Pan, R., Holt, D. J., and Liu, H. (2015). Abnormalities in hemispheric specialization of caudate nucleus connectivity in schizophrenia. *JAMA Psychiat.* 72, 552–560. doi: 10.1001/jamapsychiatry.2014.3176

Robinson, E. C., Garcia, K., Glasser, M. F., Chen, Z., Coalson, T. S., Makropoulos, A., et al. (2018). Multimodal surface matching with higher-order smoothness constraints. *Neuroimage* 167, 453–465. doi: 10.1016/j.neuroimage.2017.10.037

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X., Holmes, A. J., et al. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* 28, 3095–3114. doi: 10.1093/cercor/bhx179

Ungerleider, L. G., and Desimone, R. (1986). Cortical connections of visual area MT in the macaque. *J. Comp. Neurol.* 248, 190–222. doi: 10.1002/cne.902480204

Wang, D., Buckner, R. L., Fox, M. D., Holt, D. J., Holmes, A. J., Stoecklein, S., et al. (2015). Parcellating cortical functional networks in individuals. *Nat. Neurosci.* 18, 1853–1860. doi: 10.1038/nn.4164

Wang, D., Tian, Y., Li, M., Dahmani, L., Wei, Q., Bai, T., et al. (2020). Functional connectivity underpinnings of electroconvulsive therapy-induced memory impairments in patients with depression. *Neuropsychopharmacology* 45, 1579–1587. doi: 10.1038/s41386-020-0711-2

Wig, G. S. (2017). Segregated systems of human brain networks. *Trends Cogn. Sci.* 21, 981–996. doi: 10.1016/j.tics.2017.09.006

Zilles, K., and Amunts, K. (2010). Centenary of Brodmann's map—conception and fate. *Nat. Rev. Neurosci.* 11, 139–145. doi: 10.1038/nrn2776

Zuo, X., and Xing, X. (2014). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci. Biobehav. Rev.*. 45, 100–118. doi: 10.1016/j.neubiorev.2014.05.009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 10 March 2022
doi: 10.3389/fnins.2022.836259

Uncovering Cortical Units of Processing From Multi-Layered Connectomes

Kristoffer Jon Albers[1], Matthew G. Liptrot[1], Karen Sandø Ambrosen[1], Rasmus Røge[1], Tue Herlau[1], Kasper Winther Andersen[2], Hartwig R. Siebner[2,3,4], Lars Kai Hansen[1], Tim B. Dyrby[1,2], Kristoffer H. Madsen[1,2], Mikkel N. Schmidt[1] and Morten Mørup[1]**

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Lyngby, Denmark, [2] Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital - Amager and Hvidovre, Copenhagen, Denmark, [3] Department of Neurology, Copenhagen University Hospital Bispebjerg and Frederiksberg, Copenhagen, Denmark, [4] Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

Modern diffusion and functional magnetic resonance imaging (dMRI/fMRI) provide non-invasive high-resolution images from which multi-layered networks of whole-brain structural and functional connectivity can be derived. Unfortunately, the lack of observed correspondence between the connectivity profiles of the two modalities challenges the understanding of the relationship between the functional and structural connectome. Rather than focusing on correspondence at the level of connections we presently investigate correspondence in terms of modular organization according to shared canonical processing units. We use a stochastic block-model (SBM) as a data-driven approach for clustering high-resolution multi-layer whole-brain connectivity networks and use prediction to quantify the extent to which a given clustering accounts for the connectome within a modality. The employed SBM assumes a single underlying parcellation exists across modalities whilst permitting each modality to possess an independent connectivity structure between parcels thereby imposing concurrent functional and structural units but different structural and functional connectivity profiles. We contrast the joint processing units to their modality specific counterparts and find that even though data-driven structural and functional parcellations exhibit substantial differences, attributed to modality specific biases, the joint model is able to achieve a consensus representation that well accounts for both the functional and structural connectome providing improved representations of functional connectivity compared to using functional data alone. This implies that a representation persists in the consensus model that is shared by the individual modalities. We find additional support for this viewpoint when the anatomical correspondence between modalities is removed from the joint modeling. The resultant drop in predictive performance is in general substantial, confirming that the anatomical correspondence of processing units is indeed present between the two modalities. Our findings illustrate how multi-modal integration admits consensus representations well-characterizing each individual modality despite their biases and points to the importance of multi-layered connectomes as providing supplementary information regarding the brain's canonical processing units.

Keywords: multi-layered connectomes, dMRI, fMRI, stochastic block model, brain parcellation

OPEN ACCESS

Edited by:
Jordi Solé-Casals, Universitat de Vic—Universitat Central de Catalunya, Spain

Reviewed by:
Théodore Papadopoulo, Research Centre Inria Sophia Antipolis Méditerranée, France
Feng Liu, Stevens Institute of Technology, United States

****Correspondence:***
Kristoffer Jon Albers
kjal@dtu.dk
Morten Mørup
mmor@dtu.dk

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *15 December 2021*
Accepted: *09 February 2022*
Published: *10 March 2022*

Citation:
Albers KJ, Liptrot MG, Ambrosen KS, Røge R, Herlau T, Andersen KW, Siebner HR, Hansen LK, Dyrby TB, Madsen KH, Schmidt MN and Mørup M (2022) Uncovering Cortical Units of Processing From Multi-Layered Connectomes. Front. Neurosci. 16:836259. doi: 10.3389/fnins.2022.836259

1. INTRODUCTION

The prominent approach of viewing the organization of the brain at the macro scale needs to reconcile two fundamental aspects: while the cortex is *segregated* into specialized neuronal regions, the cognitive functions emerge from *integration* of these regions by coordinated activation (Tononi et al., 1994). Network science provides a convenient way to model complex relational systems, such as the behavior of the human brain, which does not emerge solely from the properties of the individual units, but from the complex interactions between these. Here, both aspects of brain organization can be summarized as networks, reflecting the *structural* and *functional* connectivity respectively, thereby permitting network science to provide the statistical foundation and methodology for investigating and quantifying the organization of brain connectivity (Bullmore and Sporns, 2009; Van Den Heuvel and Pol, 2010). Recent proposals aim at jointly modeling multiple modalities of brain connectivity using multi-layer networks (Battiston et al., 2017; Buldú and Porter, 2017; De Domenico, 2017), where the connections from different modalities are encoded within different layers, sharing the same network nodes (Betzel and Bassett, 2016), see also Vaiana and Muldoon (2020) for a recent review. Such multi-layer investigations allow neuroscience to integrate the complementary aspects of structural and functional data. However, the implications of multimodal integration, the extent to which it is interpretable, and the correspondence between the modalities remain unclear (Battiston et al., 2017; De Domenico, 2017).

Direct comparisons of structural and functional connectivity derived from diffusion and functional magnetic resonance imaging (dMRI/fMRI) have shown that structure to some degree reflects function (Koch et al., 2002; Greicius et al., 2009; Sporns, 2014). This suggests that a relationship does exist between the two modalities, indicated by measures of network properties, e.g., functional connectivity networks exhibiting various small-world attributes (Achard et al., 2006), which could be reflected by an evolutionarily-sound and economically-efficient structure (Bullmore and Sporns, 2009). However, the time scales of structural and functional connectivity derived from MRI are orders of magnitude apart. As such, the blood-oxygen-level-dependent (BOLD) hemodynamics quantified by fMRI are in the order of seconds with observed responses to stimuli delayed by at least a second and peaking after 4–8 s (Kim and Bandettini, 2012). Structural connections on the other hand operate in the order of milliseconds (Innocenti et al., 2014) which can thus not be directly probed by fMRI. Notably, the low temporal resolution of fMRI can be overcome by other functional neuroimaging methods such as electroencephalography (Deslauriers-Gauthier et al., 2019) but at the cost of low spatial resolution. At the whole-brain scale, previous studies suggest that functional connectivity quantified by fMRI to some extent emerges from the structural organization (Greicius et al., 2009; Sporns, 2014; Becker et al., 2015), but BOLD derived functional connectivity has also been observed between cortical regions that are not directly anatomically connected (Koch et al., 2002; Vincent et al., 2007; Skudlarski et al., 2008; Honey et al., 2009). In particular, stronger prevalence of functional connections linking right and left hemispheres have been observed (Koch et al., 2002; Vincent et al., 2007; Skudlarski et al., 2008). Additionally, various neurological disorders have been shown to cause alterations in both functional and structural connectivity (Fornito and Bullmore, 2012; Tost et al., 2012; Kaiser, 2013; van Dellen et al., 2013), though the extent to which any relation between functional and structural connectivity affects brain disease still needs further investigation (Vega Pons et al., 2016). Thus, although BOLD functional connectivity to some extent has been found to correlate with the strength of the direct anatomical connections as quantified by the number of streamlines between regions (Honey et al., 2009; Hermundstad et al., 2013), structural and BOLD functional connectivity operate on vastly different time-scales. As a consequence, the direct structural connections are not found to be very predictive of functional connections but moreso when integrating multiple steps in the structural connectome (Røge et al., 2017).

Existing attempts at jointly modeling functional and structural connectomes have primarily focused on how structure can inform function (Hinne et al., 2014) or function enhance recovery of structural connections (Chu et al., 2018). In Zhang et al. (2021), canonical correlation analysis (CCA) was used to identify optimal projections maximizing the correlation between structural and functional connections, and in Becker et al. (2018) spectral methods were used to relate structural connections and paths along the structural graph to functional connectivity. Structural and functional connectomes have further been jointly modeled using independent component analysis (ICA) combining structural and functional connections as features for the ICA (Amico and Goñi, 2018). Recently, deep learning autoencoders (Banka et al., 2020) and graph neural networks (see also Bessadok et al., 2021 for an overview) have been proposed for multimodal integration providing non-linear mappings between structural and functional connectivity (Li et al., 2021) as well as joint learning of connectivity fingerprints predictive of phenotypic traits (Filip et al., 2020; Dsouza et al., 2021). In the context of connectivity based parcellations the most prominent approaches have been to use k-means, hierarchical, or spectral clustering (Eickhoff et al., 2015; Reuter et al., 2020) to parcellate functional and structural connectivity. Whereas these frameworks can provide joint parcellations as a post processing step, joint parcellations using generic heterogeneous data clustering tools (Liu et al., 2020) based on Gabasova et al. (2017) have also been considered.

In this article, we approach the assessment of concurrence between functional connectivity (FC) and structural connectivity (SC) of high-resolution multi-layered connectomes, *hypothesizing that correspondence occurs not at the level of connectivity but organization of latent processing units*. We therefore assume that both the structural and functional connectome express different connectional fingerprints of the brain's canonical processing units. This assumption is formalized in the multi-layered network by assuming that measurements from differing modalities (i.e., network-layers) originate from the same processing units (i.e., group of network nodes) but with substantial differences in how the connectivity profiles

between these processing units are expressed across modalities. As a result, even though the elicited networks of SC and FC are different, we hypothesize that they both reflect the same underlying organization, that would emerge if neurobiological *atoms* of the cortex were to form aggregated regions which are shared across both modalities (Eickhoff et al., 2017). In particular, focusing on the consensus representation obtained by combining modalities may better reveal these regions, by providing a representation that is less polluted by modality specific biases.

Figure 1 illustrates the conceptual steps of the data-driven approach for exploring the organization of the brain, based on network modeling of structural and functional neuroimaging data. We presently consider data obtained from the publicly released Human Connectome Project (HCP) (Feinberg et al., 2010; Moeller et al., 2010; Setsompop et al., 2012; Xu et al., 2012; Van Essen et al., 2013) database, from which we generated *in vivo* whole-brain resting state FC and SC networks for a large population of 250 healthy subjects. The networks were inferred from functional and diffusion recordings in the full image resolution supported by modern MRI. The graphs for each subject were binarized and thresholded at one percent density, ignoring the sub-cortical voxels, and hence each contain 59,412 vertices. These graphs were then randomly split into five populations of 50 subjects, aggregated into single functional and structural networks for each group, and finally binarized to one percent link-density. **Figure 2** illustrates how the profiles of structural and functional connectivity substantially differ both in terms of strength within anatomical regions and in their whole brain connectional fingerprint, whilst relatively exhibiting limited variation within each modality across populations of 50 subjects. In particular, functional connectivity shows a high degree of inter-hemispheric connections when compared to the structural connectivity that is mainly ipsi-lateral. Apart from time-scale differences, the lack of inter-hemispheric structural connections can be attributed to limitations of current tractography methods (Maier-Hein et al., 2017). As such, the average area under curve (AUC) of the receiver operator characteristic directly predicting the connectivity of one group of subjects from another group of subjects (i.e., considering the total number of 0–0 matches, 0–1 matches, 1–0 matches, and 1–1 matches; Ambrosen et al., 2014; Røge et al., 2017) across the FC graphs is 0.901 whereas it is 0.935 for the SC graphs and 0.618 predicting FC from SC for the same group of subjects.

This lack of concurrence at the modality-specific connectivity level both within and between parcels does not rule out concurrence at the level of the underlying processing units. If the processing units resolved by both modalities are in perfect agreement, the inter-population variability of these units within each modality would be comparable to their inter-modality variability within populations. However, observed differences in network properties can be due to differing sources, such as noise in the data and measurement procedure including scanning parameters (Ambrosen et al., 2020), as well as inherent differences in the signals measured by the modalities including time scales as discussed above. For example, fMRI is known to suffer from motion artifacts (Diedrichsen and Shadmehr, 2005), whereas diffusion MRI is known to exhibit biases such as preference of tractography methods to terminate at gyral crowns (Schilling et al., 2018). We thus expect that modality specific biases are present and that they will drive parcellations in disagreeing directions. To investigate this we provide both a qualitative characterization as well as a quantitative predictive assessment of the differences of data driven structural and functional parcellations and contrast this to the consensus representation obtained by joint modeling of the structural and functional connectome.

We use a stochastic block model (SBM) (Nowicki and Snijders, 2001) which allows us to infer a single parcellation based upon multiple networks (see **Figure 3**). An SBM type of framework has previously been used for functional (Mørup et al., 2010; Andersen et al., 2014; Baldassano et al., 2015) and structural parcellation (Ambrosen et al., 2014; Baldassano et al., 2015) as well as joint modeling of functional and structural connectivity in low resolution (116 network nodes; Andersen et al., 2012a). We provide statistical evaluation of the predictive performance of the inferred parcellations following a similar framework to the one proposed in Albers et al. (2021) (see **Figure 4**). We compare the results of joint modeling with the comprehensive HCP_MMP1.0 atlas which is constructed using multiple modalities, including neuroanatomy (Glasser et al., 2016). We further contrast the results to a non-trivial (block permuted) null hypothesis of *non-correspondence* between the structural and functional regions. We exploit that the HCP vertex order is spatially contiguous and that a simple permutation in which the non-predicted modality is permuted according to a parcellation learned on the modality thereby preserves spatial contiguous network blocks with similar size distribution in both modalities while breaking any anatomical correspondence. In summary, we test the hypothesis that structural and functional connectomes derived from dMRI and fMRI support shared canonical processing units by:

(A) Characterizing differences of modality specific and multi-modal data-driven parcellations.
(B) Contrasting the predictive performance of parcellations trained on the same modality, different modality, or trained jointly using both modalities.
(C) Contrasting the predictive performance to the performance obtained when permuting the connectome of one modality thereby enforcing non-correspondence.

The novelty of this work lies both in the characterization of the multimodal concurrence identified in the high-resolution HCP data, despite the substantial differences in the observed connectivity profiles, and in the application of a quantitative predictive framework to assessing the validity of canonical processing units.

2. MATERIALS AND METHODS

2.1. Data

Magnetic Resonance Imaging (MRI) techniques provide non-invasive means from which functional and structural connectivity networks can be constructed. Structural

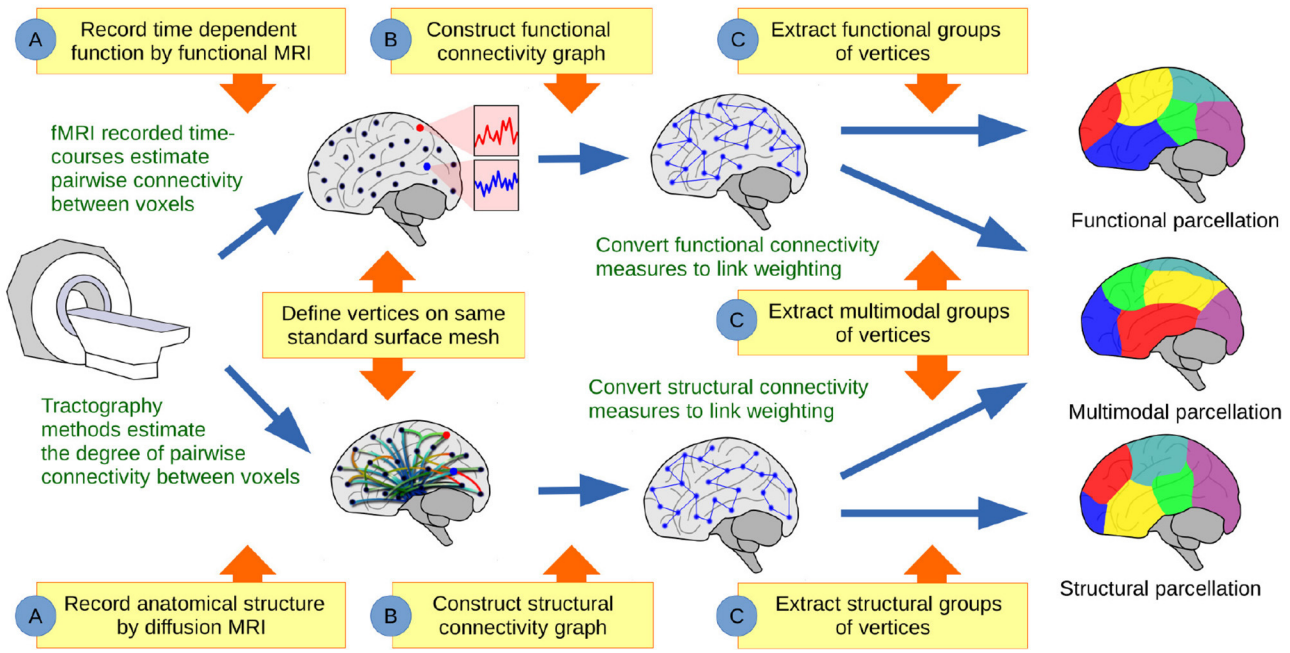

FIGURE 1 | Concept of data-driven multi-layer network modeling. **(A)** Based on functional and structural MRI, graphs of structural and functional connectivity are generated, **(B)** such that vertices are defined on the same standard surface mesh. **(C)** From these graphs data-driven parcellations can be inferred and compared, when modeling structure or function either individually or jointly.

connectivity can be derived from diffusion MRI (Gong et al., 2009) by tracking white matter streamlines across the cortex such that structural networks are obtained based on the anatomy of the brain. Functional MRI captures images of functional whole brain connectivity by indirectly measuring the time-dependent neural activity within small regions of the brain (i.e., voxels) by monitoring the blood oxygenation level dependent (BOLD) response (Ogawa et al., 1990). Networks of functional connectivity can be obtained, for instance as mapped by the correlated activation of brain regions (Bullmore and Sporns, 2009).

Networks of functional and structural brain connectivity were obtained using independent high-resolution data from the Human Connectome Project (HCP) (Feinberg et al., 2010; Moeller et al., 2010; Setsompop et al., 2012; Xu et al., 2012; Van Essen et al., 2013) database available from the MGH-USC Human Connectome Project (HCP) database (https://ida.loni.usc.edu/login.jsp). Ignoring the sub-cortical information, the networks contained 59,412 vertices covering the neocortex. We split the 250 subjects into populations, such that we obtained five non-overlapping groups of 50 subjects. For each group we created a single functional and structural training network based on the group average. The MRI data for all the subjects in each population were aggregated before the networks were constructed and thresholded in order to obtain a single functional and structural network representative for the group.

The fMRI networks were estimated from the preprocessed and structurally denoised ICA-FIX cleaned version of the resting state fMRI data, for further reference see Smith et al. (2013), Griffanti et al. (2014), and Salimi-Khorshidi et al. (2014). We formed the networks by averaging the Pearson correlation matrix estimated from the two sessions using both the left-right and right-left phase encoding directions for each subject (i.e., averaging four correlograms per subject each estimated from 1,200 time frames).

The structural connectivity networks were derived from the dMRI data preprocessed using the HCP pipeline (Glasser et al., 2013). The fiber orientation estimation was done using FSL's BedpostX for multi-shell data (Jbabdi et al., 2012) and the networks were constructed by performing probabilistic tractography using FSL's Probtrackx2 (Behrens et al., 2003, 2007) run in "matrix3" mode. One thousand streamlines were initiated in each white matter voxel, and a resulting streamline was kept if it reached two vertices of the white matter surface, resulting in weighted graphs of streamline counts between vertices. The adjacency matrices for all subjects in the group were added and binarized by thresholding the graph at 1% density keeping only the strongest links.

2.2. The HCP_MMP1.0 Atlas

To ground results, we contrasted the performance obtained using stochastic block modeling to the performance using a prominent existing parcellation, i.e., the HCP_MMP1.0 (Glasser et al., 2016) atlas. The HCP_MMP1.0 atlas (Glasser et al., 2016)

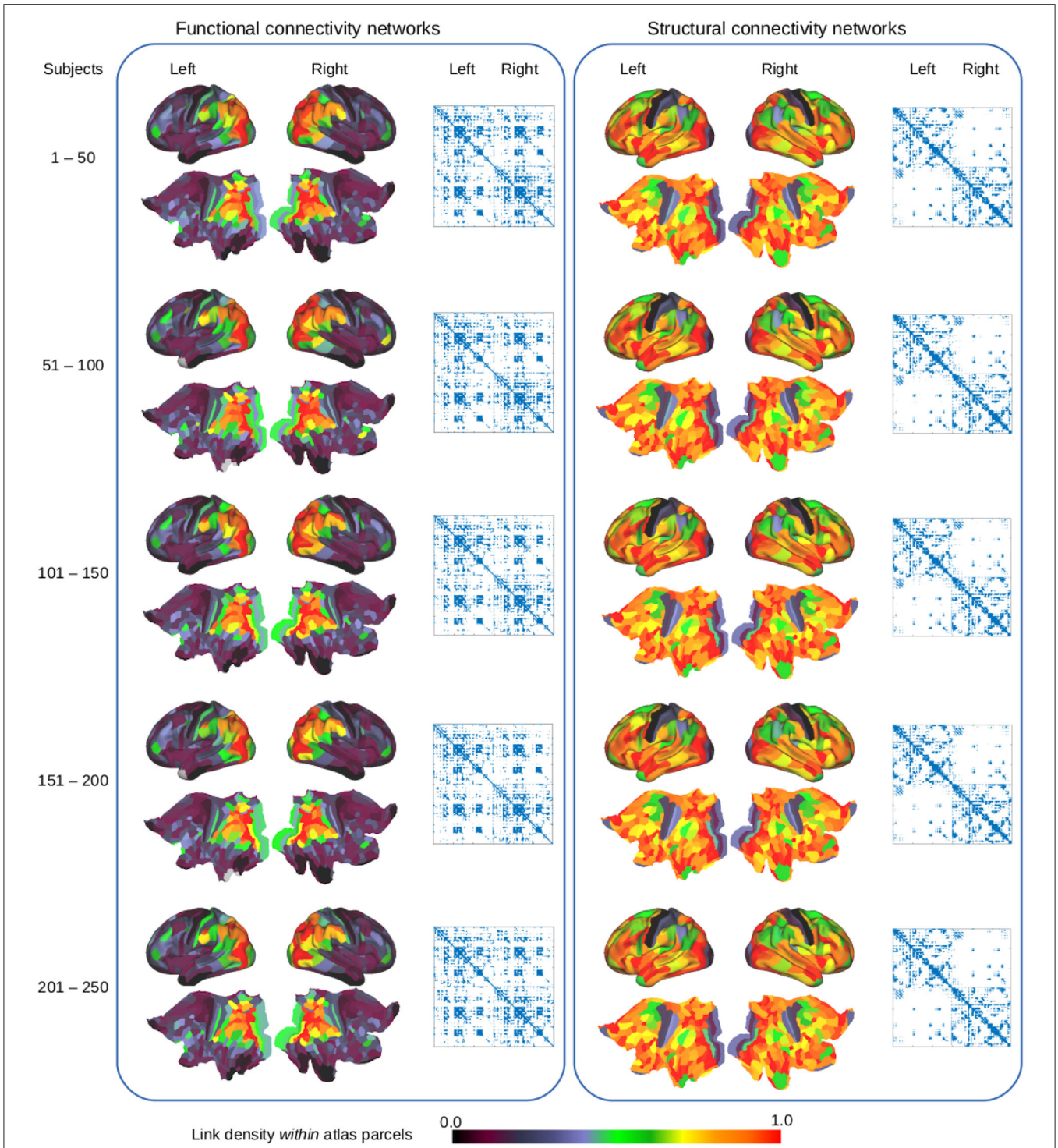

FIGURE 2 | Examination of functional connectivity **(left)** and structural connectivity **(right)** using the HCP_MMP1.0 atlas parcellation (Glasser et al., 2016). The cortical surface and flatmaps show the link-density within each of the 360 atlas parcels whereas the adjacency matrices outline the whole brain functional and structural connectivity graphs, based on the average of different populations of 50 subjects.

is based on multi-modal MRI data from the HCP and describes a total of 360 parcels split equally across both hemispheres. It was created in a combined data-driven and manual approach to obtain a single parcellation of cortical regions, based on multiple neurobiological properties including both functional information and brain anatomy obtained from 210 healthy

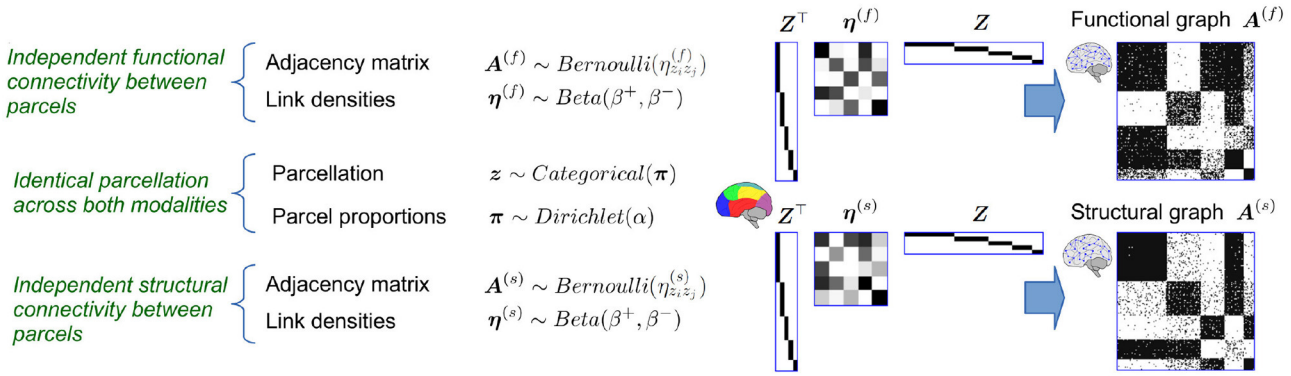

FIGURE 3 | The Stochastic Block Model (SBM) is a generative model capable of discovering a single group-structure from multiple complex networks, i.e., functional connectome $\boldsymbol{A}^{(f)}$ and structural connectome $\boldsymbol{A}^{(s)}$. Based on such a shared parcellation $\boldsymbol{z}$, the model assumes links are independently generated from a Bernoulli distribution such that the probability of observing a link between any two vertices only depends on the modality specific probability of observing a link between the two parcels that the vertices belong to given by the inter parcel (off diagonal elements) and intra parcel (diagonal elements) of the link density matrices $\boldsymbol{\eta}^{(f)}$ and $\boldsymbol{\eta}^{(s)}$ for the functional and structural connectomes, respectively. Under this assumption, the model hence allows a single parcellation to be inferred from multiple networks while accounting for differences in connectivity profiles. Let $\boldsymbol{Z}$ denote a matrix of the clustering $\boldsymbol{z}$ using a one-hot encoding. Notably, the SBM can be considered a lossy compressed representation of the connectomes such that $\boldsymbol{A}^{(f)} \approx \boldsymbol{Z}^\top \boldsymbol{\eta}^{(f)} \boldsymbol{Z}$ and $\boldsymbol{A}^{(s)} \approx \boldsymbol{Z}^\top \boldsymbol{\eta}^{(s)} \boldsymbol{Z}$.

subjects. We have previously found this atlas to perform relatively well when predicting single subject structural and functional connectivity networks and therefore include it presently as a baseline (Albers et al., 2021).

2.3. Joint Integration Using the Stochastic Block Model

Both in terms of its structural organization and functional activity the brain can be studied as a network. One approach of quantifying the latent structure in connectivity networks is to partition the nodes into groups that share a similar connectivity pattern within the network. The stochastic block model (SBM) (Nowicki and Snijders, 2001) is a data-driven Bayesian clustering approach, which, coupled with Markov Chain Monte Carlo (MCMC) sampling, has proven a valid tool for clustering and investigating structure in complex networks (Zhu et al., 2008; Schmidt and Mørup, 2013). Notably, a non-parametric SBM modeling framework [denoted the infinite relational model (IRM)] (Kemp et al., 2006; Xu et al., 2006) has previously been used for the separate modeling of functional (Mørup et al., 2010; Andersen et al., 2012b, 2014) and structural connectivity (Ambrosen et al., 2013, 2014) whereas joint modeling of structural and functional connectivity has been considered in Andersen et al. (2012a). Notably, the approach of Andersen et al. (2012a) was based on low resolution networks of 116 nodes defined by the AAL atlas (Tzourio-Mazoyer et al., 2002) with the ability to impose shared and individual segregated units of the two modalities.

The stochastic block model (SBM) (Nowicki and Snijders, 2001) partitions network nodes into clusters with similar connectivity patterns. For modeling symmetric binary networks, the model can be defined by the following generative process, where m is used to index modality:

$$\text{Links in network:} \quad \boldsymbol{A}^{(m)}_{ij} \sim \text{Bernoulli}(\eta^{(m)}_{z_i z_j}), \tag{1}$$

$$\text{Cluster-link densities:} \quad \eta^{(m)}_{\ell h} \sim \text{Beta}(\beta^+, \beta^-), \tag{2}$$

$$\text{Clustering:} \quad z_i \sim \text{Categorical}(\boldsymbol{\pi}), \tag{3}$$

$$\text{Cluster proportions:} \quad \boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{\alpha}). \tag{4}$$

The probability of observing a link between two nodes i and j in the network follows a Bernoulli distribution only depending on the probability of observing links between the clusters z_i and z_j that the nodes belong to. The probability of observing links between two clusters is considered independent given the assignment to clusters and follows a Beta distribution. Finally, the nodes are partitioned into K clusters, and the cluster proportions follow a Dirichlet distribution.

The stochastic block model when used for multimodal integration is outlined in **Figure 3**. The observed functional and structural connectomes $\boldsymbol{A}^{(f)}$ and $\boldsymbol{A}^{(s)}$ are assumed to be generated according to a shared functional and structural parcellation $\boldsymbol{z}$ such that $z_i = \ell$ indicates that vertex i belongs to parcel ℓ. Whereas, the parcellation is shared, the manner in which the different regions integrate is assumed to be modality specific and parameterized respectively for the functional and structural connectomes by $\eta^{(f)}_{\ell h}$ and $\eta^{(s)}_{\ell h}$ providing the extent (i.e., the probability) that nodes in parcel ℓ connect to nodes in parcel h. As a result, the observed connectomes can have substantially different within and between parcel connectivity structures $\boldsymbol{\eta}^{(f)}$ and $\boldsymbol{\eta}^{(s)}$ while being defined in terms of the same underlying units of processing $\boldsymbol{z}$. As parcels may differ in size, $\boldsymbol{\pi}$ is used to account for size-heterogeneity.

Due to the conjugacy between the Dirichlet and Categorical distribution, $\boldsymbol{\pi}$ can be analytically marginalized (see Schmidt and

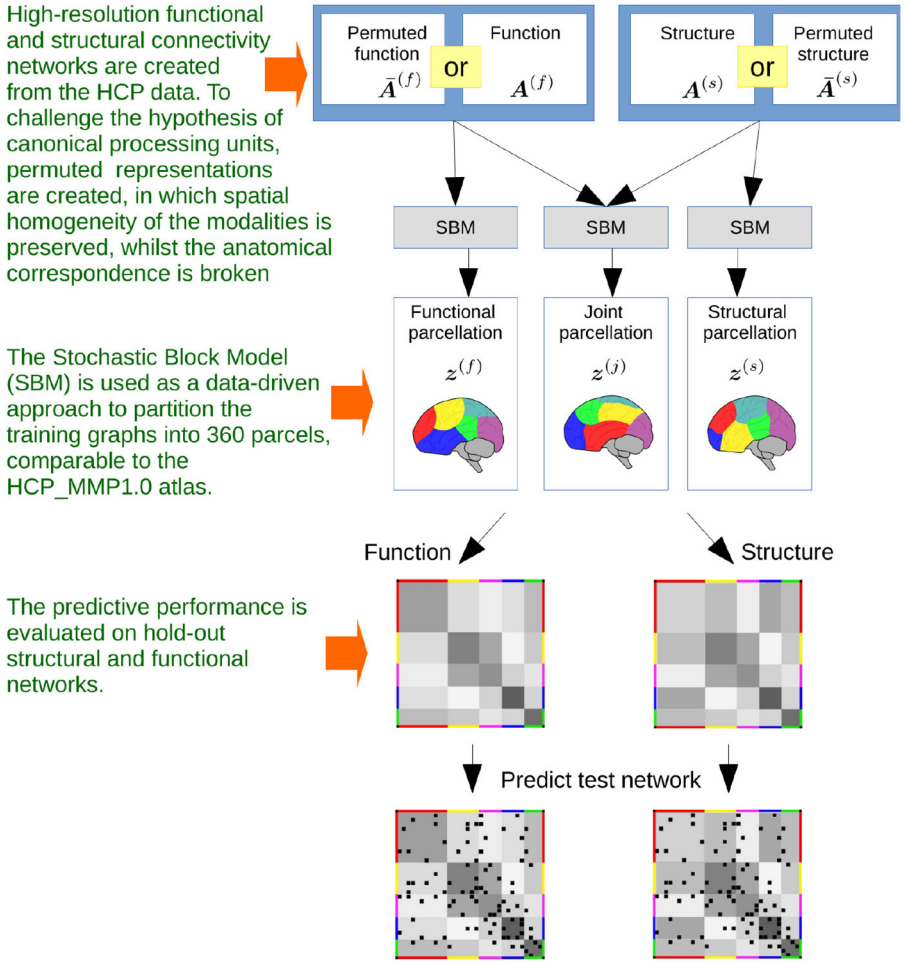

FIGURE 4 | The steps that defines the flow of the investigations. Based on high-resolution functional and structural MRI obtained from publicly released data of the Human Connectome Project, independent networks of structural and functional connectivity are generated. The networks are based on population averages, resulting in a total of five networks for each modality, based on five populations of 50 subjects. The networks are binarized by thresholding at 1% link density. Stochastic Blockmodeling is utilized to infer data-driven parcellations, based on modeling structure or function individually or jointly modeling both modalities for the five populations. The benefits of multimodal integration are hence evaluated by comparing the performance of predicting hold-out population networks using inferred single and multimodal parcellations, contrasted with that of using networks that have been spatially permuted whilst preserving the size of the parcels.

Mørup, 2013 for details). By imposing an equal concentration parameter for all K clusters $\boldsymbol{\alpha} = \frac{\alpha}{K}\mathbf{1}_{K\times 1}$ the following effective prior for the clustering can be obtained:

$$P(\boldsymbol{z}|\alpha) = \frac{\Gamma(\alpha)}{\Gamma(\alpha+N)}\prod_{l=1}^{K}\frac{\Gamma(\frac{\alpha}{K}+n_k)}{\Gamma(\frac{\alpha}{K})}, \tag{5}$$

where N is the number of nodes, n_k is the number of nodes in cluster k, and $\Gamma(x)$ is the gamma function. Notably, we use the SBM to obtain a *single* parcellation based on either a single network from one modality (either functional or structural) or two networks, one for each modality when jointly modeling structure and function. Let $\boldsymbol{A}$ represent the set of M networks, containing either $M = 1$ or $M = 2$ modalities. The beta prior is conjugate to the Bernoulli likelihood, which allows us to obtain the following joint distribution as $\boldsymbol{\eta}$ can be analytically marginalized:

$$P(\boldsymbol{A},\boldsymbol{z}|\beta^+,\beta^-,\alpha) = P(\boldsymbol{z}|\alpha)\prod_{m}^{M}\prod_{\ell\leq h}\frac{B(N_{\ell h}^{(m)+}+\beta^+, N_{\ell h}^{(m)-}+\beta^-)}{B(\beta^+,\beta^-)}, \tag{6}$$

where $N_{\ell h}^{(m)+} = \sum_{1\leq i<j\leq N}\delta_{z_i,\ell}\delta_{z_j,h}A_{ij}^{(m)}$ and $N_{\ell h}^{(m)-} = \sum_{1\leq i<j\leq N}\delta_{z_i,\ell}\delta_{z_j,h}(1-A_{ij}^{(m)})$ respectively represent the number of links and non-links between cluster ℓ and h according to network $\boldsymbol{A}^{(m)}$, while $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is the beta function.

2.4. Model Inference

We infer the model parameters using a Markov Chain Monte Carlo (MCMC) procedure. The parcellation is inferred by Gibbs sampling, where the assignment z_i for each node i in turn is

processed, based on the posterior distribution for the assignment of i to each of the K clusters ℓ. Using Bayes' theorem this can be obtained from Equation 6, where $\boldsymbol{z}^{\setminus i} = (z_j)_{j\neq i}$ are the cluster assignments for all nodes ignoring node i:

$$P(z_i = \ell | \boldsymbol{A}, \boldsymbol{z}^{\setminus i}, \beta^+, \beta^-, \alpha) = \frac{P(\boldsymbol{A}, \boldsymbol{z}^{\setminus i}, z_i = \ell | \beta^+, \beta^-, \alpha)}{\sum_{h=1}^{K} P(\boldsymbol{A}, \boldsymbol{z}^{\setminus i}, z_i = h | \beta^+, \beta^-, \alpha)}. \tag{7}$$

For inferring the hyper-parameters, β^+, β^-, and α, we use a simple Metropolis-Hastings procedure, where new proposals are drawn from a Gaussian distribution centered at the current parameter value with variance 1.

For all experiments the model parameters are inferred by sampling 100 iterations of the following sampling procedure: $\boldsymbol{z}$ is updated by one complete Gibbs sweep over all nodes followed by 1,000 MH-proposals for updating each hyper-parameter β^+, β^-, α. Due to the size of the networks, it is not computationally feasible to reach convergence. However, the Gibbs sampler quickly reach a stable cluster assignment with high posterior likelihood which is treated as a point estimate of the parameters (Albers et al., 2013). We hence treat the last sampled state as the inferred parameters. All experiments are performed with $K = 360$ clusters which limits SBM to the same complexity as the HCP_MMP1.0 atlas. For the HCP_ MMP1.0 the hyperparameters, β^+ and β^-, are inferred for each of the training networks, using the Metropolis-Hastings procedure with the parcellation fixed to the HCP atlas. Code for the SBM modeling framework is provided at brainconnectivity.compute.dtu.dk.

2.5. Predictive Performance

To assess and compare the quality of parcellations we use the predictive framework established in Ambrosen et al. (2013) and Albers et al. (2021). The quality of a parcellation is evaluated by how well it can be used to predict unseen held-out networks.

2.5.1. Predictive Likelihood

Let $\boldsymbol{A}^{(m)train}$ denote the training network and $\boldsymbol{A}^{(m)test}$ the network used for evaluating the learned parcellation $\boldsymbol{z}$ and how these segregated units integrate in terms of their intra and inter connectivity densities defined by the matrix $\boldsymbol{\eta}^{(m)}$. The expected predictive log-likelihood is then given by:

$$\begin{aligned}\langle \log p(\boldsymbol{A}^{(m),test} | \boldsymbol{A}^{(m),train}, \boldsymbol{z}, \beta^+, \beta^-, \alpha) \rangle_{p(\boldsymbol{\eta}^{(m)} | \boldsymbol{A}^{(m),train}, \boldsymbol{z})} = \\ \sum_{i>j} A_{ij}^{(m),test} \langle + \log(\eta_{z_i z_j}^{(m)}) \rangle_{p(\boldsymbol{\eta}^{(m)} | \boldsymbol{A}^{(m),train}, \boldsymbol{z})} \\ (1 - A_{ij}^{(m),test}) \langle \log(1 - \eta_{z_i z_j}) \rangle_{p(\boldsymbol{\eta}^{(m)} | \boldsymbol{A}^{(m),train}, \boldsymbol{z})},\end{aligned} \tag{8}$$

where the expectations are given with respect to the distribution $p(\boldsymbol{\eta}^{(m)} | \boldsymbol{A}^{(m),train}, \boldsymbol{z})$ as:

$$\begin{aligned}\langle \log(\eta_{\ell h}^{(m)}) \rangle_{p(\boldsymbol{\eta}^{(m)} | \boldsymbol{A}^{(m),train}, \boldsymbol{z})} &= \psi(N_{\ell h}^{(m)+} + \beta^+) \\ &\quad - \psi(N_{\ell h}^{(m)+} + N_{\ell h}^{(m)-} + \beta^+ + \beta^-)\end{aligned} \tag{9}$$

$$\begin{aligned}\langle \log(1 - \eta_{\ell h}^{(m)}) \rangle_{p(\boldsymbol{\eta}^{(m)} | \boldsymbol{A}^{(m),train}, \boldsymbol{z})} &= \psi(N_{\ell h}^{(m)-} + \beta^-) \\ &\quad - \psi(N_{\ell h}^{(m)+} + N_{\ell h}^{(m)-} + \beta^+ + \beta^-),\end{aligned} \tag{10}$$

with ψ being the digamma function $\psi(x) = \frac{d}{dx} \log[\Gamma(x)]$. In Albers et al. (2021), the log of the expected predictive likelihood was also considered but found to provide similar performance to the expected predictive log-likelihood and therefore not included herein.

2.5.2. Area Under Curve

An alternative measure is to describe how the probabilities of generating links inferred by SBM from the training network can be used to separate between links and non-links in the test network. We quantify this performance by the area under curve (AUC) of the Receiver Operator Characteristics curve (ROC) (Clauset et al., 2008), scored by the expected link probability between clusters as observed from the training graph

$$\langle \eta_{z_i z_j}^{(m)} \rangle = \frac{N_{z_i z_j}^{(m)+} + \beta^+}{N_{z_i z_j}^{(m)+} + N_{z_i z_j}^{(m)-} + \beta^+ + \beta^-}. \tag{11}$$

These scores are then evaluated in terms of how well they, for the corresponding entries $A_{ij}^{(m),test}$ in the test graph, are able to separate links (considered the positive class) from non-links (considered the negative class). Let $R_{ij}^{(m),test} = \langle \eta_{z_i z_j}^{(m)} \rangle$ be the reconstructed test connectome for modality m and $\mathrm{vec}_{\mathrm{U}}(\boldsymbol{B})$ return the upper triangular part of the matrix $\boldsymbol{B}^{I\times I}$ as the vector $\boldsymbol{b}^{I\cdot(I-1)/2\times 1}$. The score vector $\hat{\boldsymbol{y}} = \mathrm{vec}_{\mathrm{U}}[\boldsymbol{R}^{(m),test}]$ and true labels $\boldsymbol{y} = \mathrm{vec}_{\mathrm{U}}[\boldsymbol{A}^{(m),test}]$ can then be considered as inputs to the standard receiver operator characteristic function for calculating the area under curve (AUC).

2.6. Parcellation Comparison by Mutual Information

The similarity of different parcellations can be quantified using Mutual Information (MI). This constitutes a permutation invariant measure for the shared clustering information between two parcellations $\boldsymbol{z}$ and $\boldsymbol{z}'$ given by:

$$\mathrm{MI}(\boldsymbol{z}, \boldsymbol{z}') = \sum_{cc'} P(c, c') \log\left(\frac{P(c, c')}{P(c)P(c')}\right), \tag{12}$$

where $P(c) = \sum_{c'}^{K} P(c, c')$ is the probability of observing a node in cluster c while $P(c, c') = \frac{1}{N} \sum_{i=1}^{N} \delta_{z_i, c} \delta_{z'_i, c'}$ is the probability of jointly observing a node in cluster c in $\boldsymbol{z}$ and a node in cluster c' in $\boldsymbol{z}'$. We use the normalized mutual information (NMI) to get a value between zero and one:

$$\mathrm{NMI}(\boldsymbol{z}, \boldsymbol{z}') = \frac{2\ \mathrm{MI}(\boldsymbol{z}, \boldsymbol{z}')}{\mathrm{MI}(\boldsymbol{z}, \boldsymbol{z}) + \mathrm{MI}(\boldsymbol{z}', \boldsymbol{z}')}, \tag{13}$$

such that a value of one indicates that the parcellations are identical.

2.7. Blocked Permutation Procedure

To probe the correspondence of the extracted structural and functional units and their joint integration we contrast the performance to a null hypothesis assuming no correspondence. To achieve this, we use a permutation procedure that accounts for size distribution and to some extent for spatial contiguity, while upholding the assumption that the parcels do not correspond in the two modalities. The permutation procedure re-organizes all vertices of the non-predicted modality according to a clustering structure learned on the non-predicted modality in which clusters are ordered in random order. (That is, when predicting the functional connectome the vertices of the structural connectome are re-organized and vice versa). Thereby, the non-predicted modality is ordered in terms of modality specific units such that the vertices of these units correspond to different spatial contiguous regions defined through the HCP vertex traversal order in the predicted modality. The permutation procedure is illustrated in **Figure 5**. This procedure preserves size distribution and spatial contiguity as defined by the HCP vertex traversal order but does not account for anatomy nor spatial shape.

3. RESULTS

3.1. Data-Driven Parcellations

Figure 6 shows flatmap representations of the inferred parcellations (based on the extraction of 360 parcels as used in the HCP_ MMP1.0 atlas), for each of the five 50-subject populations. There is a clear similarity between the parcellations of various training populations within each modality, while there are clear characteristic differences across modalities. The functional parcellations (left column) show a high density of small, elongated parcels seemingly located in the posterior cortex, while the majority of vertices are located in few very large parcels covering the remainder of the cortex. In contrast, the structural parcellations (middle column) show spatial compactness with the majority of clusters being of a similar size. The multimodal parcellations (right column) appear to inherit features from both modalities, showing a variance of cluster sizes and shapes. Compared to the functional parcellations, the posterior cortex is segregated into fewer parcels, though still many more than in the structural parcellations. The parcels in the rest of the cortex have also inherited the spatial compactness of the structural parcellations. Although they are slightly larger, they represent a finer segregation of the cortex. The multimodal parcellations are therefore more detailed in the posterior cortex than the structural parcellations and more detailed in the rest of the cortex than the functional parcellations.

Figure 6 (lower panel) shows histograms for the distribution of cluster sizes averaged across the parcellations for all five 50-subject populations. For the functional parcellations most parcels are very small (<100 vertices) while the majority of nodes are located in extremely large parcels (larger than 1,000 nodes). In contrast, the structural populations are homogeneous, such that the majority of nodes are located in medium sized clusters (between 100 and 1,000 parcels) which is also the most common parcel size. Once again, the multimodal parcellations seem to inherit features from both modalities. Compared to the functional populations, the cluster size distribution is shifted toward larger parcels, with the concurrent removal of the few extremely large parcels, such that the majority of vertices are now located in the medium sized parcels. Furthermore, the panel shows the extent to which parcels are common to both hemispheres. This is shown both as the number of bilateral parcels and as an index, representing how evenly the nodes of the individual parcels are split across hemispheres. This laterality index for a parcel is computed as $\frac{max(N_{left}, N_{right})}{N_{left}+N_{right}}$, where N_{left} and N_{right} is the number of nodes within the parcel, that, respectively belongs to the left and right hemisphere. An average laterality index of 0.5 would indicate that all parcels are equally split across the two hemispheres, while an average index of 1 would indicate that all parcels are unilateral. The functional parcellations are significantly more bilateral (273 parcels out of 360) than both the structural (6 parcels) and multimodal (63 parcels). The average laterality index further indicates that the individual functional parcels tend to be bilateral, whilst this is uncommon for both structural and multimodal parcels.

Figure 7 indicates the similarity of the inferred parcellations, as measured by Normalized Mutual Information (NMI), between and within modalities (see also section 2). Functional parcellations are inherently noisy, as evidenced by their mutual information being far lower than those of the structural parcellations. Furthermore, the multimodal parcellations are not penalized by the noise of the functional data, as they retain an NMI almost on par with the NMI within structural parcellations. The figure further indicates that the multimodal parcellations are more in agreement with the structural parcellations than with the functional, though the functional parcellations are more in agreement with the multimodal parcellation than with the structural. This implies that the multimodal model has determined a consensus to which both the functional and structural parcellations are more in agreement than they are with each other.

3.2. Predictive Performance

Figure 8 shows the results of predicting functional and structural hold-out networks using the following inferred parcellations:

- Data-driven parcellations for a single modality, inferred from either the same modality as the hold-out networks, the other modality, or the permuted version of the other modality (enforcing non-correspondence).
- Data-driven parcellations for the multimodal model, inferred from both modalities where the non-predicted modality is considered both with and without permutation (enforcing non-correspondence).
- The fixed multi-modal HCP_ MMP1.0 atlas.

Both the area under curve of the receiver operator characteristic (AUC) and expected predictive likelihood scores (for details

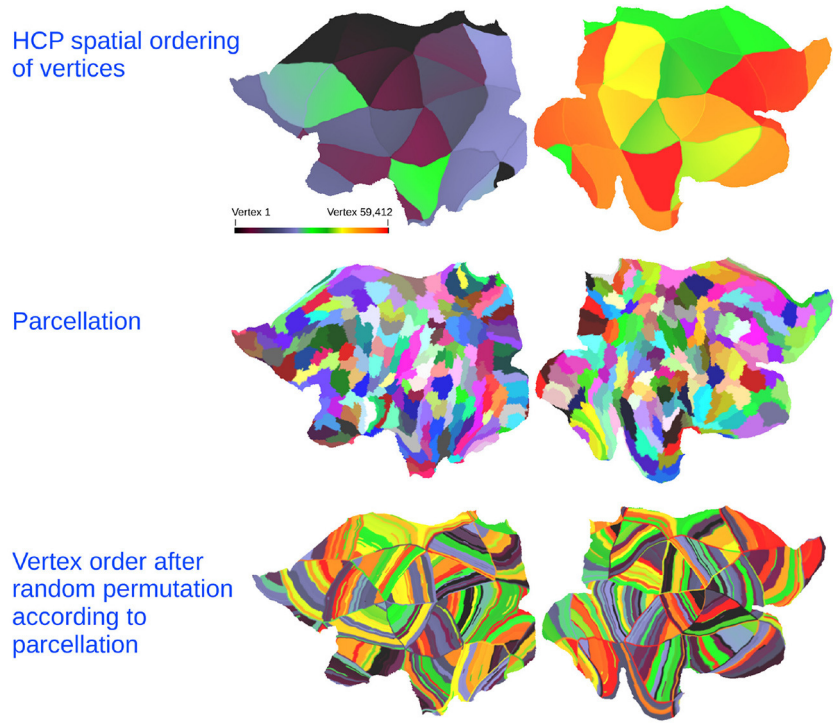

FIGURE 5 | (Top) Flatmap of each vertex in the HCP data color-coded according to the vertex traversal order of the network adjacency matrices. **(Middle)** Example of parcellation structure learned from the non-predicted modality. **(Bottom)** The random permutation of the non-predicted modality obtained by re-ordering the vertex traversal according to the learned parcellation structure in which the clusters are ordered in random order. Color-code indicates the original vertex position.

see section 2) are individually computed and averaged for the parcellations inferred for the five 50-subject populations.

Predicting both modalities, the figure shows that predicting the same modality as that from which the parcellation was inferred provides good predictions according to both predictive metrics. While structural parcellations predict function on par with the HCP atlas, functional parcellations are comparably poor predictors of structure. However, the consensus parcellations of joint modeling provide good predictions of both modalities, though slightly better for predicting function than structure.

As expected, breaking the consensus in the joint model, by using a permuted modality, decreases the predictive performance, but not to the same extent as training on the permuted data alone. In predicting functional networks, structural parcellations perform well, and far better than parcellations inferred from permuted structure. Functional parcellations on the other hand are very poor predictors of structural networks, to such a degree that parcellations inferred from the permuted functional networks actually predict better.

4. DISCUSSION

Herein we have introduced a method for quantifying multimodal integration, which assumes that structure and function are independent realizations of the same underlying processing units. While a lot of effort has recently focused on developing multilayered networks of brain connectivity, data-driven quantifications of these remains challenging and mostly limited to comparisons between modalities (Koch et al., 2002; Vincent et al., 2007; Skudlarski et al., 2008; Greicius et al., 2009; Honey et al., 2009; Sporns, 2014; Becker et al., 2015; Røge et al., 2017). By representing SC and FC graphs over the same set of network nodes, the framework proposed utilizes a stochastic blockmodel (SBM) to obtain a unified parcellation while admitting modality-specific connectivity profiles. A benefit of the considered approach, integrating connectomes using the stochastic block model is that it naturally accounts for modality specific connectivity structure between the extracted units of processing while enforcing consistency across modalities in terms of the extracted parcels thereby providing a simple computationally tractable data-driven approach. As such, the SBM can be considered a data driven approach to network compression in terms of shared learned parcels $\boldsymbol{z}$ and how the connectivity structure between these parcels form modality specific networks at the level of parcels with link densities (i.e., connectivity strengths) $\boldsymbol{\eta}^{(m)}$. This facilitates direct interpretation of large connectomes at the level of connectivity structure between the extracted parcels. As opposed to other connectivity based clustering approaches such as k-means, hierarchical, and spectral clustering that are generic clustering procedures designed for feature data (Eickhoff et al., 2015; Gabasova et al., 2017; Liu et al., 2020; Reuter et al., 2020) a benefit of

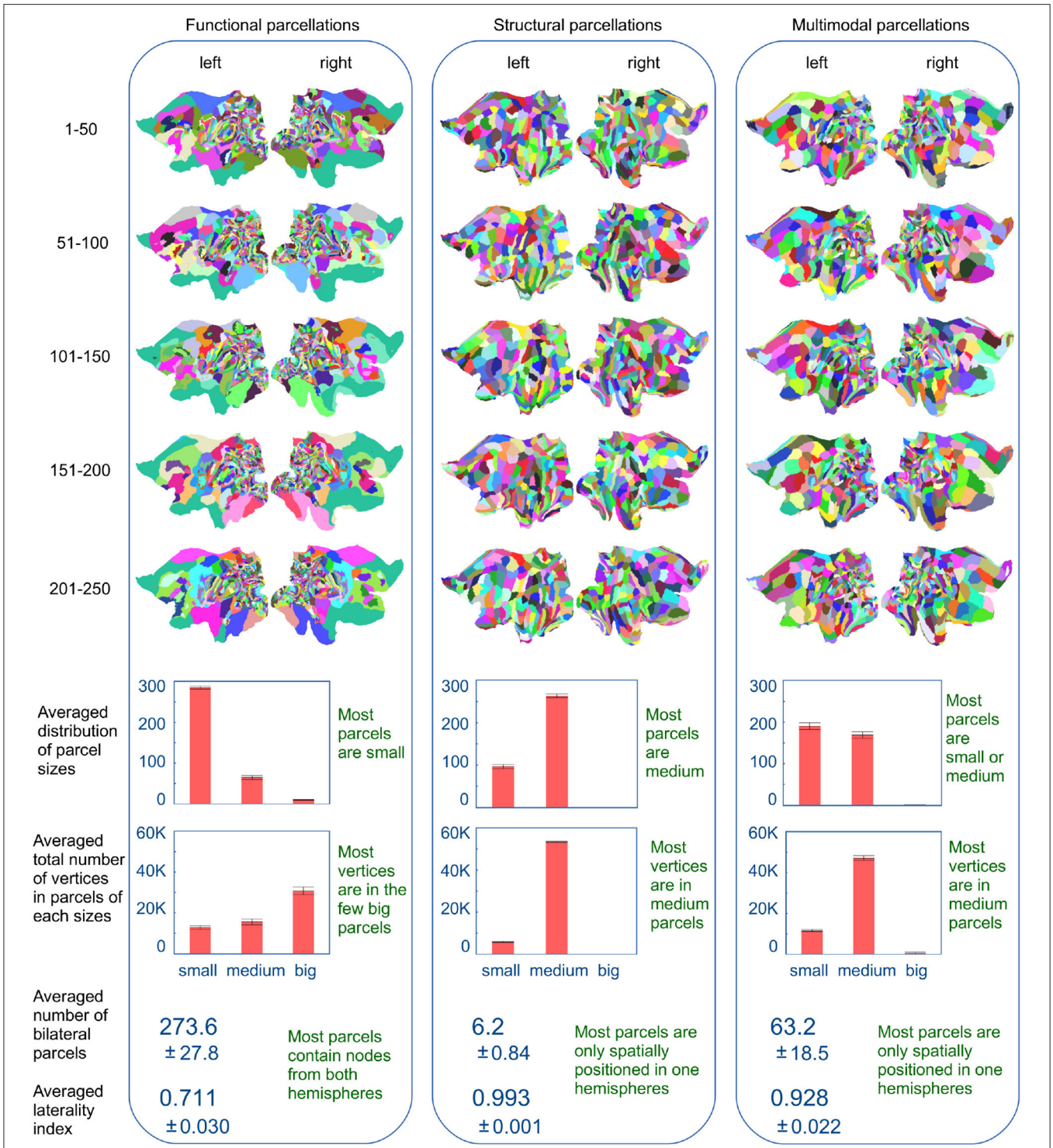

FIGURE 6 | Flatmaps of the inferred parcellations for the five populations of 50 subjects, using the different modalities. Parcels are separated into three groups based on their size: small (<100 vertices), medium (between 100 and 1,000 vertices) and big (more than 1,000 vertices). Also shown, for the per-modality averages of the five parcellations, is the number of parcels that contain nodes from both hemispheres and how evenly these nodes are split across the two hemispheres. Errorbars and ± indicate the standard deviation over the five parcellations.

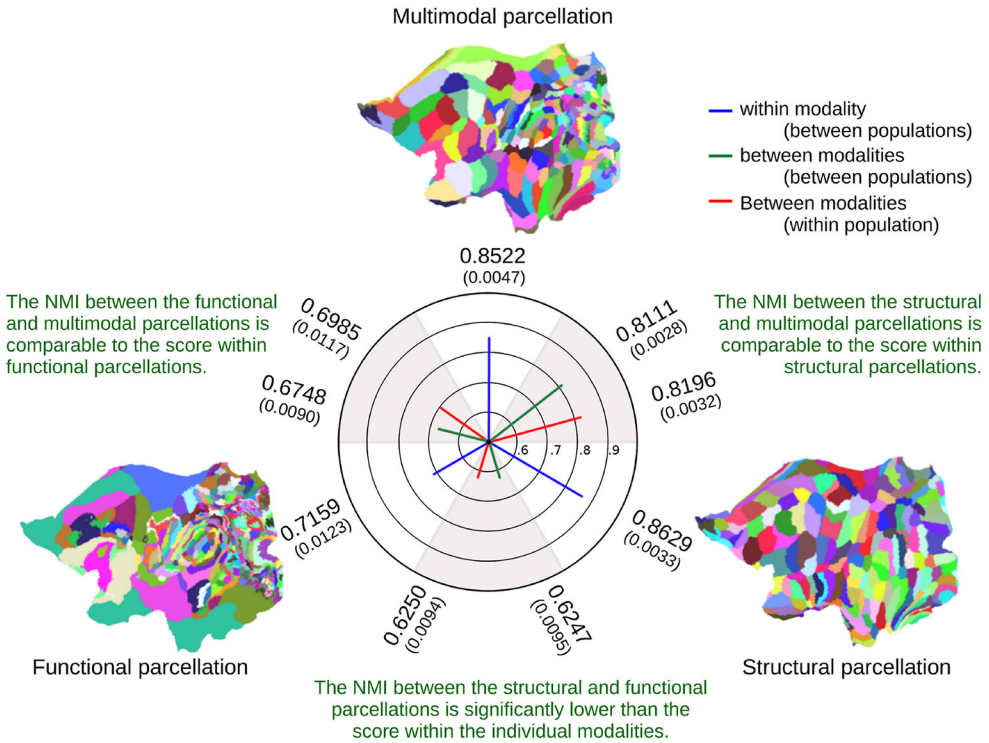

FIGURE 7 | Averaged Normalized Mutual Information (NMI) for the inferred parcellations between and within modalities, showing the average NMI from 10 comparisons of parcellations between populations and five comparisons within populations. The standard deviations are shown in parenthesis. The flatmaps illustrate one of the five parcellations for a single 50-subject population, showing only the left hemisphere.

the SBM is that it provides a statistically grounded generative model of networks. Thereby, the quantification of these unified parcellations by the SBM naturally follows the predictive framework outlined in **Figure 4** that assesses the performance of predicting independent held-out test graphs as a substitute in the absence of ground truth (Albers et al., 2021).

4.1. Qualitative Differences in Functional, Structural, and Joint Parcellations

We observe that the characteristics of the inferred parcellations heavily depend on the modality. In particular, we observe that functional parcels are much more bilateral and have a wider distribution of unit sizes. This can be attributed to the bilateral co-activation of similar functional units across both hemispheres as also observed from seed based and independent component analysis (ICA) of the HCP resting state data (Smith et al., 2013). In contrast, structural connectivity generates much more unilateral units with no parcels having more than 1,000 voxels. We attribute the unilaterality to limitations in tractography in terms of delineating inter-hemispheric structural connections and long-distance pathways (Van Essen et al., 2014; Knösche et al., 2015; Maier-Hein et al., 2017). The joint modeling also exhibits substantial differences to both the modality-specific parcellations. As such, the identified consensus representation is reduced of modality specific biases as observed from both the size distribution and laterality index being in between the structural and functional parcellations. This joint representation possesses substantial agreement with both modalities, in so much that we find that the normalized mutual information (NMI) between the joint parcellation and the functional and structural parcellations are in higher agreement than the NMI between the functional and structural parcellations. These results demonstrate that both modalities each contribute unique information about the brain's underlying organization influencing the joint representation of parcels.

4.2. Statistical Evidence for Canonical Processing Units

Predicting the organization of each modality based on a data-driven parcellation of the other we found that structural parcellations predicted almost on par with functional parcellations and much better than prediction when enforcing non-correspondence. However, for structural data we found that the functional parcellation was a poor predictor to the extent that the permutation procedure, which broke correspondence yet both preserved size distribution and enforced spatial homogeneity, was a better predictor of function. We attribute this to the parcels of the permuted functional data preserving spatial homogeneity which more favorably accounts for structure than the highly bilateral parcels extracted by function as observed in **Figure 2**. Notably, this supports the benefit of imposing spatial constraints when modeling structural connectomes, as proposed

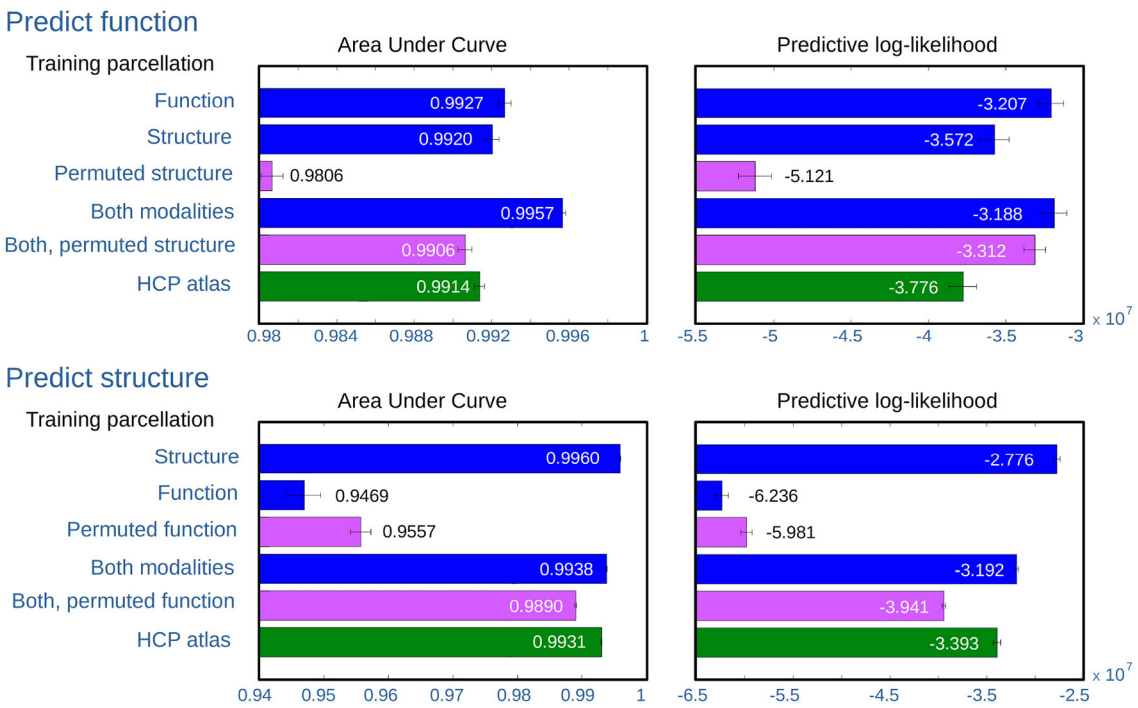

FIGURE 8 | AUC and expected predictive log-likelihood when predicting structural and functional hold-out network. The bars show the average score as obtained using the following different parcellations: (1) SBM on the same modality as the test networks, (2) SBM on the other modality, (3) SBM on the permuted networks for the other modality, (4) jointly modeling both modalities, (5) jointly modeling both modalities with the network for the other modality permuted, and (6) using the parcellation defined by the HCP atlas. For each of the five training populations, four evaluations are computed for each modality configuration, when respectively predicting from the training graph to each of the other four graphs of the same modality (providing a total of 20 predictions). The mean value is shown for each bar while the whiskers indicate the standard deviation of the mean correcting by the five independently acquired networks ($\pm$ std $/\sqrt{5}$).

in Baldassano et al. (2015) in which parcels are constrained to be spatially connected, and emphasizes that the modality-specific parcellations are heavily influenced by modality-specific biases. The consensus representation here provides a representation reduced of the modality-specific biases. Interestingly, we find that the predictions of the more noisy functional connectome are improved using the consensus representation, whereas the structural connectivity predictions are only mildly reduced when compared to using the modality-specific parcellations. In particular, integrating both modalities results in a consensus representation that has better predictive performance than permuting one modality thereby enforcing non-correspondence.

Our results points to modality-specific biases and differences in the representation of functional and structural units. We thus do not find a direct correspondence at the level of modality-specific processing units in the brain. However, we do find that imposing canonical processing units forms a useful, practical representation of structural and functional data in high resolution that well characterize both modalities. In particular, we observe that the noisier functional modality benefits from the integration of structural information. Whereas the structural network is expected to be constant, i.e., it is a static structure, the functional connectivity estimates derived from fMRI vary and are related to activations that only use some parts of the structural network. As such, functional connectomes derived from fMRI exhibit a high degree of inter- and intra-subject variability (Poldrack et al., 2015; Zuo et al., 2019; Albers et al., 2021) and the joint modeling with structural connectivity can help here by regularizing the extracted representation despite the variability of the fMRI source.

4.3. Limitations

The inter-subject alignment is based on surface morphometry, according to the standard HCP processing pipeline (Glasser et al., 2013) which could conceivably bias toward one modality more than the other. Furthermore, we arbitrarily threshold the networks at 1% density. Further studies should investigate if these findings are reproducible to change in registration methods and methods for functional and structural network construction.

We considered $K = 360$ parcels as employed by the HCP-MMP1.0 atlas (Glasser et al., 2016) enabling a direct comparison in terms of the same number of extracted units of processing. However, as we saw, the structural and functional data provide fine grained resolutions in different parts of the cortex and thus have differing preferences in terms of regions using coarse and fine grained parcels. As the joint modeling provides more data upon which parcellations can be learned it may be that better joint representations can be achieved utilizing more than the considered parcels admitting fine-grained resolutions both where it is most supported by structure and function. Future work

should explore the impact of resolution employed in terms of numbers of parcels invoked and whether the joint modeling can provide support for the use of more parcels.

In this study, we attempted to generate a random control by permuting the network for one modality. This, however, is not straightforward, as it requires a permutation that preserves the spatial homogeneity, shape, and size of the parcel along with the modality specific distribution of parcels. We approximated this using the inherent vertex adjacency ordering of the HCP data format, which results in a tendency to define contiguous yet elongated parcels. Future work should develop more advanced permutation procedures that accounts for parcel shape.

The model employed herein is based on the assumption that the connectivity profiles of the two modalities are independent, yet originating from the same underlying processing units. This assumption is a simplification of the expected true underlying organization of the brain (where at least at the neuronal level we would expect a strong structure-function relationship; Innocenti et al., 2014; Andersson et al., 2020), and is generally not expected to be the case (Eickhoff et al., 2017). It is thus widely believed that the connectivity structures are related although representing substantially different time scales. As such, structural paths have been observed predictive of functional connections (Røge et al., 2017; Becker et al., 2018) which is not accounted for by the SBM. Future work should consider more advanced non-linear modeling approaches inspired by recent deep learning approaches (Banka et al., 2020; Bessadok et al., 2021) that potentially can leverage the presently considered SBM to learn such dependencies of $\boldsymbol{\eta}^{(m)}$.

The model assumes that the modalities contribute equally and share similar properties in regards to resolution, spatial homogeneity, level of noise, and that both modalities are equally informative. Despite these constraints, our findings illustrate that the joint model allows for the identification of shared units that are useful in practice. However, even when multimodal integration allows for good predictions, care must be taken regarding the inference of the purported underlying organization that would account for such findings. This is because a perceived agreement between the modalities does not necessarily mean that the spatial extent of the brain regions, and the borders between them, are in fact located in the manner implied by the data-driven parcellation (Eickhoff et al., 2017).

The joint modeling of functional and structural connectivity extracts a consensus representation that can be reduced of modality specific biases. However, as we have no ground truth information in regards to the true optimal units of processing, the presented evaluation can be considered a qualitative assessment demonstrating concurrence beyond a rather simple block-permuted null model. Arguably, an average representation should be a better predictor than a representation based only on the complementary modality. In general, we expect the joint modeling to provide consensus representations superior to the representations provided by each modality when both modalities exhibit similar degrees of noise. However, if one modality is substantially less biased from the true (unknown) underlying representation the joint modeling may be driven undesirably by the more biased modality. In circumstances where a priori knowledge are available of the validity of the connectivity structures of the considered modalities the joint modeling can potentially be advanced to provide more emphasis to more accurate modalities.

In the present study, we considered the perhaps most simple approach to extract functional connectivity based on zero lag Pearson correlation (Bullmore and Sporns, 2009; Smith et al., 2011; Richiardi et al., 2013). Notably, it is unclear how functional connectivity is best quantified and several approaches exist, including mutual information (Bullmore and Sporns, 2009; Mørup et al., 2010; Smith et al., 2011), wavelet correlation (Achard et al., 2006), lagged correlation and partial correlation, as well as approaches quantifying directionality (see also Smith et al., 2011; Richiardi et al., 2013 for reviews). Herein we considered only positive correlation, while negative functional correlations arguably also relate to structure. Note also that the examined HCP fMRI data has a high temporal and spatial resolution, which might give a poorer signal-to-noise ratio than other protocols.

The quantified water diffusion by dMRI is a noisy and indirect measure of fiber-orientation, making structural connectivity inference based on dMRI inherently uncertain. In particular, the relative low image resolution of diffusion MRI often results in partial volume regions resulting in crossing fibers needing to be disentangled (Schilling et al., 2017; Ambrosen et al., 2020). Furthermore, as different tract systems contain different numbers of axons, and hence tract volumes, structural connectivity is expected to be volume-weighted toward the major tracts, also typically explored with tracers (Innocenti et al., 2014, 2017; Van Essen et al., 2014; Jbabdi et al., 2015). Herein we extracted the structural connectivity networks using probabilistic tractography (Behrens et al., 2003, 2007). However, tractography methods are known to suffer from systematic biases, such as a preference to terminate at gyral crowns (Van Essen et al., 2014; Schilling et al., 2018), issues characterizing multiple fiber directions in the face of limited data resolution (Jbabdi et al., 2015), and difficulties tracking long distance pathways (Van Essen et al., 2014).

5. CONCLUSION

The work presented herein is a novel approach for quantifying the relation between the function and structure of the brain and the integration of these in terms of processing units. Herein we considered joint network modeling of structural and functional connectivity data, however the proposed framework naturally extends to general multimodal modeling, including additional modalities.

Using high quality data from the Human Connectome Project, we find that shared canonical processing units cannot be discredited, despite the lack of observed correspondence between the modality-specific connectivity profiles. As such, we find that integrating both modalities allows for reasonable predictions of the individual modalities, as quantified by two separate predictive metrics remaining on par or better than using either the individual modalities or the HCP_MMP1.0

atlas. This finding supports that both modalities reflect different aspects of the same underlying processing units, which allows the joint model to infer a consensus that is a mild compromise between both modalities. At this point it is unclear whether the differences and similarities of the parcellations supported by structural and functional connectome are caused by systematic biases in the derived connectomes of the two modalities due to the current limitations extracting functional and structural connectivity networks. It will thus be interesting to re-apply the presented analysis framework as the quality of extracted functional and structural connectomes in the future improve.

The presented approach is likely to benefit studies of individuals, or populations where the data quality cannot match that of the HCP, as the integration of multiple modalities would overcome noise issues if these are more disruptive than modality specific biases. Furthermore, the similarity of inferred parcellations suggests that the consensus reduces modality-specific biases, and as such the consensus representation, if evaluated on a third modality, would likely better characterize that modality than each of the training modalities separately. If there exists an underlying truth of shared processing units, that truth may come closer to be recovered the more modalities are combined, and the presented framework provides a data-driven approach to achieve this.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. Data used in this article were taken from the Human Connectome Project and can be accessed from db.humanconnectome.org. The generated connectomes from the HCP can be provided from the authors upon request and requires acceptance of the terms of use given at: https://www.humanconnectome.org/study/hcp-young-adult/data-use-terms.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

KAl and ML contributed to conceptualization, methodology, software, formal analysis, investigation, data curation, and writing of the original draft. KAm, RR, TD, and KM contributed to conceptualization, methodology, and data curation. MS contributed to conceptualization, methodology, and software. TH, KAn, HS, and LH contributed to conceptualization, and methodology. MM contributed to conceptualization, methodology, software, formal analysis, investigation, writing of the original draft, project administration, and funding acquisition. All authors reviewed and approved the submitted version.

FUNDING

This project was funded by the Lundbeck Foundation, Grant Nr. R105-9813. HS holds a 5-year professorship in precision medicine at the Faculty of Health Sciences and Medicine, University of Copenhagen which is sponsored by the Lundbeck Foundation (Grant Nr. R186-2015-2138).

ACKNOWLEDGMENTS

The MRI data used in this work were obtained from the MGHUSC Human Connectome Project (HCP) database (https://ida.loni.usc.edu/login.jsp) in the 500 subjects release. Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) 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. A Tesla K40 general-purpose GPU, used for processing the structural graphs, was donated by NVIDIA (NVIDEA Corporation, Santa Clara, CA, USA).

REFERENCES

Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* 26, 63–72. doi: 10.1523/JNEUROSCI.3874-05.2006

Albers, K. J., Ambrosen, K. S., Liptrot, M. G., Dyrby, T. B., Schmidt, M. N., and Mørup, M. (2021). Using connectomics for predictive assessment of brain parcellations. *NeuroImage* 238, 118170. doi: 10.1016/j.neuroimage.2021.118170

Albers, K. J., Moth, A. L. A., Mørup, M., and Schmidt, M. N. (2013). "Large scale inference in the infinite relational model: gibbs sampling is not enough," in *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 1–6. doi: 10.1109/MLSP.2013.6661904

Ambrosen, K. S., Albers, K. J., Dyrby, T. B., Schmidt, M. N., and Morup, M. (2014). "Nonparametric bayesian clustering of structural whole brain connectivity in full image resolution," in *2014 International Workshop on Pattern Recognition in Neuroimaging*, 1–4. doi: 10.1109/PRNI.2014.6858507

Ambrosen, K. S., Eskildsen, S. F., Hinne, M., Krug, K., Lundell, H., Schmidt, M. N., et al. (2020). Validation of structural brain connectivity networks: the impact of scanning parameters. *NeuroImage* 204, 116207. doi: 10.1016/j.neuroimage.2019.116207

Ambrosen, K. S., Herlau, T., Dyrby, T., Schmidt, M. N., and Mørup, M. (2013). "Comparing structural brain connectivity by the infinite relational model," in *2013 International Workshop on Pattern Recognition in Neuroimaging (PRNI)*, 50–53. doi: 10.1109/PRNI.2013.22

Amico, E., and Go ni, J. (2018). Mapping hybrid functional-structural connectivity traits in the human connectome. *Netw. Neurosci.* 2, 306–322. doi: 10.1162/netn_a_00049

Andersen, K. W., Herlau, T., Mørup, M., Schmidt, M. N., Madsen, K. H., Lyksborg, M., et al. (2012a). "Joint modelling of structural and functional brain networks," in *2nd NIPS Workshop on Machine Learning and Interpretation in NeuroImaging (MLINI 2012)* (Lake Tahoe).

Andersen, K. W., Madsen, K. H., Siebner, H. R., Schmidt, M. N., Mørup, M., and Hansen, L. K. (2014). Non-parametric bayesian graph models reveal community structure in resting state fMRI. *NeuroImage* 100, 301–315. doi: 10.1016/j.neuroimage.2014.05.083

Andersen, K. W., Mørup, M., Siebner, H., Madsen, K. H., and Hansen, L. K. (2012b). "Identifying modular relations in complex brain networks," in *2012 IEEE International Workshop on Machine Learning for Signal Processing* (Santander), 1–6. doi: 10.1109/MLSP.2012.6349739

Andersson, M., Kjer, H. M., Rafael-Patino, J., Pacureanu, A., Pakkenberg, B., Thiran, J.-P., et al. (2020). Axon morphology is modulated by the local environment and impacts the noninvasive investigation of its structure-function relationship. *Proc. Natl. Acad. Sci. U.S.A.* 117, 33649–33659. doi: 10.1073/pnas.2012533117

Baldassano, C., Beck, D. M., and Fei-Fei, L. (2015). Parcellating connectivity in spatial maps. *PeerJ* 3, e784 doi: 10.7717/peerj.784

Banka, A., Buzi, I., and Rekik, I. (2020). "Multi-view brain hyperconnectome autoencoder for brain state classification," in *International Workshop on PRedictive Intelligence in MEdicine* (Lima: Springer), 101–110. doi: 10.1007/978-3-030-59354-4_10

Battiston, F., Nicosia, V., Chavez, M., and Latora, V. (2017). Multilayer motif analysis of brain networks. *Chaos* 27, 047404 doi: 10.1063/1.4979282

Becker, C. O., Pequito, S., Pappas, G. J., Miller, M. B., Grafton, S. T., Bassett, D. S., et al. (2015). Accurately predicting functional connectivity from diffusion imaging. *arXiv preprint arXiv:1512.02602*.

Becker, C. O., Pequito, S., Pappas, G. J., Miller, M. B., Grafton, S. T., Bassett, D. S., et al. (2018). Spectral mapping of brain functional connectivity from diffusion imaging. *Sci. Rep.* 8, 1–15. doi: 10.1038/s41598-017-18769-x

Behrens, T., Berg, H. J., Jbabdi, S., Rushworth, M., and Woolrich, M. (2007). Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage* 34, 144–155. doi: 10.1016/j.neuroimage.2006.09.018

Behrens, T., Woolrich, M., Jenkinson, M., Johansen-Berg, H., Nunes, R., Clare, S., et al. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* 50, 1077–1088. doi: 10.1002/mrm.10609

Bessadok, A., Mahjoub, M. A., and Rekik, I. (2021). Graph neural networks in network neuroscience. *arXiv preprint arXiv:2106.03535*.

Betzel, R. F., and Bassett, D. S. (2016). Multi-scale brain networks. *Neuroimage* 160, 73–83. doi: 10.1016/j.neuroimage.2016.11.006

Buldú, J. M., and Porter, M. A. (2017). Frequency-based brain networks: from a multiplex framework to a full multilayer description. *Netw. Neurosci.* 2, 418–441. doi: 10.1162/netn_a_00033

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198. doi: 10.1038/nrn2575

Chu, S.-H., Parhi, K. K., and Lenglet, C. (2018). Function-specific and enhanced brain structural connectivity mapping via joint modeling of diffusion and functional MRI. *Sci. Rep.* 8, 1–19. doi: 10.1038/s41598-018-23051-9

Clauset, A., Moore, C., and Newman, M. E. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature* 453, 98–101. doi: 10.1038/nature06830

De Domenico, M. (2017). Multilayer modeling and analysis of human brain networks. *Giga Sci.* 6, 1–8. doi: 10.1093/gigascience/gix004

Deslauriers-Gauthier, S., Lina, J.-M., Butler, R., Whittingstall, K., Gilbert, G., Bernier, P.-M., et al. (2019). White matter information flow mapping from diffusion MRI and EEG. *NeuroImage* 201, 116017 doi: 10.1016/j.neuroimage.2019.116017

Diedrichsen, J., and Shadmehr, R. (2005). Detecting and adjusting for artifacts in fMRI time series data. *Neuroimage* 27, 624–634. doi: 10.1016/j.neuroimage.2005.04.039

Dsouza, N. S., Nebel, M. B., Crocetti, D., Robinson, J., Mostofsky, S., and Venkataraman, A. (2021). "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* (Lübeck).

Eickhoff, S. B., Constable, R. T., and Yeo, B. T. (2017). Topographic organization of the cerebral cortex and brain cartography. *NeuroImage* 170, 332–347. doi: 10.1016/j.neuroimage.2017.02.018

Eickhoff, S. B., Thirion, B., Varoquaux, G., and Bzdok, D. (2015). Connectivity-based parcellation: critique and implications. *Hum. Brain Mapp.* 36, 4771–4792. doi: 10.1002/hbm.22933

Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., et al. (2010). Multiplexed echo planar imaging for sub-second whole brain fmri and fast diffusion imaging. *PLoS ONE* 5, e15710 doi: 10.1371/journal.pone.0015710

Filip, A.-C., Azevedo, T., Passamonti, L., Toschi, N., and Lio, P. (2020). "A novel graph attention network architecture for modeling multimodal brain connectivity," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (Montréal), 1071–1074. doi: 10.1109/EMBC44109.2020.9176613

Fornito, A., and Bullmore, E. T. (2012). Connectomic intermediate phenotypes for psychiatric disorders. *Front. Psychiatry* 3:32. doi: 10.3389/fpsyt.2012.00032

Gabasova, E., Reid, J., and Wernisch, L. (2017). Clusternomics: Integrative context-dependent clustering for heterogeneous datasets. *PLoS Comput. Biol.* 13, e1005781 doi: 10.1371/journal.pcbi.1005781

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171–178. doi: 10.1038/nature18933

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., et al. (2013). The minimal preprocessing pipelines for the human connectome project. *Neuroimage* 80, 105–124. doi: 10.1016/j.neuroimage.2013.04.127

Gong, G., He, Y., Concha, L., Lebel, C., Gross, D. W., Evans, A. C., et al. (2009). Mapping anatomical connectivity patterns of human cerebral cortex using *in vivo* diffusion tensor imaging tractography. *Cereb. Cortex* 19, 524–536. doi: 10.1093/cercor/bhn102

Greicius, M. D., Supekar, K., Menon, V., and Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cereb. Cortex* 19, 72–78. doi: 10.1093/cercor/bhn059

Griffanti, L., Salimi-Khorshidi, G., Beckmann, C. F., Auerbach, E. J., Douaud, G., Sexton, C. E., et al. (2014). ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *NeuroImage* 95, 232–247. doi: 10.1016/j.neuroimage.2014.03.034

Hermundstad, A. M., Bassett, D. S., Brown, K. S., Aminoff, E. M., Clewett, D., Freeman, S., et al. (2013). Structural foundations of resting-state and task-based functional connectivity in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 110, 6169–6174. doi: 10.1073/pnas.1219562110

Hinne, M., Ambrogioni, L., Janssen, R. J., Heskes, T., and van Gerven, M. A. (2014). Structurally-informed bayesian functional connectivity analysis. *NeuroImage* 86, 294–305. doi: 10.1016/j.neuroimage.2013.09.075

Honey, C., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.-P., Meuli, R., et al. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2035–2040. doi: 10.1073/pnas.0811168106

Innocenti, G. M., Dyrby, T. B., Andersen, K. W., Rouiller, E. M., and Caminiti, R. (2017). The crossed projection to the striatum in two species of monkey and in humans: behavioral and evolutionary significance. *Cereb. Cortex* 27, 3217–3230. doi: 10.1093/cercor/bhw161

Innocenti, G. M., Vercelli, A., and Caminiti, R. (2014). The diameter of cortical axons depends both on the area of origin and target. *Cereb. Cortex* 24, 2178–2188. doi: 10.1093/cercor/bht070

Jbabdi, S., Sotiropoulos, S. N., Haber, S. N., Van Essen, D. C., and Behrens, T. E. (2015). Measuring macroscopic brain connections *in vivo*. *Nat. Neurosci.* 18, 1546 doi: 10.1038/nn.4134

Jbabdi, S., Sotiropoulos, S. N., Savio, A. M., Gra na, M., and Behrens, T. E. (2012). Model-based analysis of multishell diffusion MR data for tractography: how to get over fitting problems. *Magn. Reson. Med.* 68, 1846–1855. doi: 10.1002/mrm.24204

Kaiser, M. (2013). The potential of the human connectome as a biomarker of brain disease. *Front. Hum. Neurosci.* 7, 484 doi: 10.3389/fnhum.2013.00484

Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., and Ueda, N. (2006). "Learning systems of concepts with an infinite relational model," in *AAAI, Vol. 3*, 5 (Boston, MA).

Kim, S.-G., and Bandettini, P. A. (2012). *Principles of BOLD Functional MRI*. Boston, MA: Springer US, 293–303. doi: 10.1007/978-1-4419-0345-7_16

Knösche, T. R., Anwander, A., Liptrot, M., and Dyrby, T. B. (2015). Validation of tractography: comparison with manganese tracing. *Hum. Brain Mapp.* 36, 4116–4134. doi: 10.1002/hbm.22902

Koch, M. A., Norris, D. G., and Hund-Georgiadis, M. (2002). An investigation of functional and anatomical connectivity using magnetic resonance imaging. *Neuroimage* 16, 241–250. doi: 10.1006/nimg.2001.1052

Li, Y., Mateos, G., and Zhang, Z. (2021). Learning to model the relationship between brain structural and functional connectomes. *arXiv preprint* arXiv:2112.09906. doi: 10.48550/arXiv.2112.09906

Liu, X., Eickhoff, S. B., Hoffstaedter, F., Genon, S., Caspers, S., Reetz, K., et al. (2020). Joint multi-modal parcellation of the human striatum: functions and clinical relevance. *Neurosci. Bull.* 36, 1123–1136. doi: 10.1007/s12264-020-00543-1

Maier-Hein, K. H., Neher, P. F., Houde, J.-C., Côté, M.-A., Garyfallidis, E., Zhong, J., et al. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nat. Commun.* 8, 1–13. doi: 10.1038/s41467-017-01285-x

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E., Strupp, J., Harel, N., et al. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magn. Reson. Med.* 63, 1144–1153. doi: 10.1002/mrm.22361

Mørup, M., Madsen, K., Dogonowski, A.-M., Siebner, H., and Hansen, L. K. (2010). "Infinite relational modeling of functional connectivity in resting state fMRI," in *Advances in Neural Information Processing Systems* (Vancouver, BC), 1750–1758.

Nowicki, K., and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *J. Am. Stat. Assoc.* 96, 1077–1087. doi: 10.1198/016214501753208735

Ogawa, S., Lee, T.-M., Kay, A. R., and Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. U.S.A.* 87, 9868–9872. doi: 10.1073/pnas.87.24.9868

Poldrack, R. A., Laumann, T. O., Koyejo, O., Gregory, B., Hover, A., Chen, M.-Y., et al. (2015). Long-term neural and physiological phenotyping of a single human. *Nat. Commun.* 6, 1–15. doi: 10.1038/ncomms9885

Reuter, N., Genon, S., Masouleh, S. K., Hoffstaedter, F., Liu, X., Kalenscher, T., et al. (2020). Cbptools: a python package for regional connectivity-based parcellation. *Brain Struct. Funct.* 225, 1261–1275. doi: 10.1007/s00429-020-02046-1

Richiardi, J., Achard, S., Bunke, H., and Van De Ville, D. (2013). Machine learning with brain graphs: predictive modeling approaches for functional imaging in systems neuroscience. *IEEE Signal Process. Mag.* 30, 58–70. doi: 10.1109/MSP.2012.2233865

Røge, R., Ambrosen, K. S., Albers, K. J., Eriksen, C. T., Liptrot, M. G., Schmidt, M. N., et al. (2017). "Whole brain functional connectivity predicted by indirect structural connections," in *2017 International Workshop on Pattern Recognition in Neuroimaging (PRNI)* (Toronto, ON), 1–4. doi: 10.1109/PRNI.2017.7981496

Salimi-Khorshidi, G., Douaud, G., Beckmann, C. F., Glasser, M. F., Griffanti, L., and Smith, S. M. (2014). Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage* 90, 449–468. doi: 10.1016/j.neuroimage.2013.11.046

Schilling, K., Gao, Y., Janve, V., Stepniewska, I., Landman, B. A., and Anderson, A. W. (2017). Can increased spatial resolution solve the crossing fiber problem for diffusion MRI? *NMR Biomed.* 30, e3787 doi: 10.1002/nbm.3787

Schilling, K., Gao, Y., Janve, V., Stepniewska, I., Landman, B. A., and Anderson, A. W. (2018). Confirmation of a gyral bias in diffusion MRI fiber tractography. *Hum. Brain Mapp.* 39, 1449–1466. doi: 10.1002/hbm.23936

Schmidt, M. N., and Mørup, M. (2013). Nonparametric Bayesian modeling of complex networks: an introduction. *Signal Process. Mag.* 30, 110–128. doi: 10.1109/MSP.2012.2235191

Setsompop, K., Gagoski, B. A., Polimeni, J. R., Witzel, T., Wedeen, V. J., and Wald, L. L. (2012). Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced G-factor penalty. *Magn. Reson. Med.* 67, 1210–1224. doi: 10.1002/mrm.23097

Skudlarski, P., Jagannathan, K., Calhoun, V. D., Hampson, M., Skudlarska, B. A., and Pearlson, G. (2008). Measuring brain connectivity: diffusion tensor imaging validates resting state temporal correlations. *Neuroimage* 43, 554–561. doi: 10.1016/j.neuroimage.2008.07.063

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., et al. (2013). Resting-state fMRI in the human connectome project. *Neuroimage* 80, 144–168. doi: 10.1016/j.neuroimage.2013.05.039

Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., et al. (2011). Network modelling methods for fMRI. *Neuroimage* 54, 875–891. doi: 10.1016/j.neuroimage.2010.08.063

Sporns, O. (2014). Contributions and challenges for network models in cognitive neuroscience. *Nat. Neurosci.* 17, 652–660. doi: 10.1038/nn.3690

Tononi, G., Sporns, O., and Edelman, G. M. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc. Natl. Acad. Sci. U.S.A.* 91, 5033–5037. doi: 10.1073/pnas.91.11.5033

Tost, H., Bilek, E., and Meyer-Lindenberg, A. (2012). Brain connectivity in psychiatric imaging genetics. *Neuroimage* 62, 2250–2260. doi: 10.1016/j.neuroimage.2011.11.007

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Vaiana, M., and Muldoon, S. F. (2020). Multilayer brain networks. *J. Nonlin. Sci.* 30, 2147–2169. doi: 10.1007/s00332-017-9436-8

van Dellen, E., Hillebrand, A., Douw, L., Heimans, J. J., Reijneveld, J. C., and Stam, C. J. (2013). Local polymorphic delta activity in cortical lesions causes global decreases in functional connectivity. *Neuroimage* 83, 524–532. doi: 10.1016/j.neuroimage.2013.06.009

Van Den Heuvel, M. P., and Pol, H. E. H. (2010). Exploring the brain network: a review on resting-state fmri functional connectivity. *Eur. Neuropsychopharmacol.* 20, 519–534. doi: 10.1016/j.euroneuro.2010.03.008

Van Essen, D. C., Jbabdi, S., Sotiropoulos, S. N., Chen, C., Dikranian, K., Coalson, T., et al. (2014). "Mapping connections in humans and non-human primates: aspirations and challenges for diffusion imaging," in *Diffusion MRI, 2nd Edn*, eds H. Johansen-Berg and T. E. J. Behrens (San Diego, CA: Elsevier), 337–358. doi: 10.1016/B978-0-12-396460-1.00016-0

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., et al. (2013). The wu-minn human connectome project: an overview. *Neuroimage* 80, 62–79. doi: 10.1016/j.neuroimage.2013.05.041

Vega Pons, S., Olivetti, E., Avesani, P., Dodero, L., Gozzi, A., and Bifone, A. (2016). Differential effects of brain disorders on structural and functional connectivity. *Front. Neurosci.* 10, 605 doi: 10.3389/fnins.2016.00605

Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., Van Essen, D. C., et al. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature* 447, 83–86. doi: 10.1038/nature05758

Xu, J., Moeller, S., Strupp, J., Auerbach, E., Chen, L., Feinberg, D., et al. (2012). "Highly accelerated whole brain imaging using aligned-blipped-controlled-aliasing multiband EPI," in *Proceedings of the 20th Annual Meeting of ISMRM, Vol. 2306* (Melbourne, VIC).

Xu, Z., Tresp, V., Yu, K., Kriegel, H.-P., et al. (2006). "Learning infinite hidden relational models," in *Uncertainity in Artificial Intelligence (UAI2006)* (Cambridge), 2.

Zhang, S., He, Z., Du, L., Zhang, Y., Yu, S., Wang, R., et al. (2021). Joint analysis of functional and structural connectomes between preterm and term infant brains via canonical correlation analysis with locality preserving projection. *Front. Neurosci.* 15, 724391 doi: 10.3389/fnins.2021.724391

Zhu, S., Yu, K., and Gong, Y. (2008). "Stochastic relational models for large-scale dyadic data using MCMC," in *Advances in Neural Information Processing Systems* (Vancouver, BC), 1993–2000. Available online at: https://papers.nips.cc/paper/2008/file/2291d2ec3b3048d1a6f86c2c4591b7e0-Paper.pdf

Zuo, X.-N., Biswal, B. B., and Poldrack, R. A. (2019). Editorial: Reliability and reproducibility in functional connectomics. *Front. Neurosci.* 13, 117 doi: 10.3389/fnins.2019.00117

Conflict of Interest: HS has received honoraria as speaker from Sanofi Genzyme, Denmark, and Novartis, Denmark, as consultant from Sanofi Genzyme, Denmark, Lophora, Denmark, and Lundbeck AS, Denmark, and as editor-in-chief (Neuroimage Clinical) and senior editor (NeuroImage) from Elsevier Publishers, Amsterdam, The Netherlands. He has received royalties as book editor from Springer Publishers, Stuttgart, Germany and from Gyldendal Publishers, Copenhagen, Denmark.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 13 April 2022
doi: 10.3389/fnins.2022.867466

A Graph Fourier Transform Based Bidirectional Long Short-Term Memory Neural Network for Electrophysiological Source Imaging

*Meng Jiao[1,2], Guihong Wan[3,4], Yaxin Guo[1], Dongqing Wang[2], Hang Liu[5], Jing Xiang[6] and Feng Liu[1]**

[1] School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, United States, [2] College of Electrical Engineering, Qingdao University, Qingdao, China, [3] Department of Dermatology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States, [4] Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States, [5] Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, United States, [6] MEG Center, Division of Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States

OPEN ACCESS

Edited by:
Tolga Cukur, Bilkent University, Turkey

Reviewed by:
Guang Ling, Wuhan University of Technology, China
Boyu Wang, Western University, Canada

***Correspondence:**
Feng Liu
fliu22@stevens.edu

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *01 February 2022*
Accepted: *14 March 2022*
Published: *13 April 2022*

Citation:
Jiao M, Wan G, Guo Y, Wang D, Liu H, Xiang J and Liu F (2022) A Graph Fourier Transform Based Bidirectional Long Short-Term Memory Neural Network for Electrophysiological Source Imaging. Front. Neurosci. 16:867466. doi: 10.3389/fnins.2022.867466

Electrophysiological source imaging (ESI) refers to the process of reconstructing underlying activated sources on the cortex given the brain signal measured by Electroencephalography (EEG) or Magnetoencephalography (MEG). Due to the ill-posed nature of ESI, solving ESI requires the design of neurophysiologically plausible regularization or priors to guarantee a unique solution. Recovering focally extended sources is more challenging, and traditionally uses a total variation regularization to promote spatial continuity of the activated sources. In this paper, we propose to use graph Fourier transform (GFT) based bidirectional long-short term memory (BiLSTM) neural network to solve the ESI problem. The GFT delineates the 3D source space into spatially high, medium and low frequency subspaces spanned by corresponding eigenvectors. The low frequency components can naturally serve as a spatially low-band pass filter to reconstruct extended areas of source activation. The BiLSTM is adopted to learn the mapping relationship between the projection of low-frequency graph space and the recorded EEG. Numerical results show the proposed GFT-BiLSTM outperforms other benchmark algorithms in synthetic data under varied signal-to-noise ratios (SNRs). Real data experiments also demonstrate its capability of localizing the epileptogenic zone of epilepsy patients with good accuracy.

Keywords: electroencephalography, source localization, inverse problem, graph Fourier transform, BiLSTM

INTRODUCTION

EEG/MEG source imaging (ESI), also known as EEG/MEG source localization, is a non-invasive neuroimaging technology that infers the location, direction, and distribution of the corresponding brain sources from the EEG or MEG data (He et al., 2018). Compared with the invasive modalities, the recording of EEG/MEG signals imposes minimum risks of blooding and inflammation of the brain (Portillo-Lara et al., 2021). Compared to other non-invasive brain imaging modalities, like

computed tomography (CT), positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS), the temporal resolution of EEG is up to a millisecond (He et al., 2018), which allows it to track the electrical activity of neurons in smaller temporal granularity (Numata et al., 2019). The study of ESI is of great significance in both neuroscience and clinical applications (Congedo and Sherlin, 2011). Accurate estimation of brain sources can not only help neuroscientists to better understand the brain mechanism (Liu et al., 2019) and the pathological characteristics of brain injury or mental disorders (da Silva, 2013), but also help doctors to identify the lesion areas of brain diseases such as epilepsy focal regions, which can contribute to the improvement of the accuracy of presurgical evaluations (Sanei and Chambers, 2013).

However, the inverse problem of ESI is highly ill-posed (Qin et al., 2017; He et al., 2018; Cui et al., 2019), and there can be infinite numbers of source configurations that explain the EEG recording since the number of EEG sensors on the scalp is far less than the number of brain sources (Liu et al., 2017; Hecker et al., 2021). Consequently, numerous methods have been proposed to solve the ESI problem by incorporating different regularizations or prior information to seek a unique solution, as further discussed in see section "Related Work." In recent years, deep learning has achieved great success in the fields of computer vision (Voulodimos et al., 2018), natural language processing (Young et al., 2018), bioinformatics (Min et al., 2017), etc., by employing its end-to-end feature extraction and representation capability (Deng and Yu, 2014; LeCun et al., 2015). Solving the inverse problems in the computer vision domain such as image reconstruction (Schlemper et al., 2017), super resolution (Dong et al., 2015), etc., has achieved great success by using a variety of artificial neural network (ANN) architectures such as the convolutional neural network (CNN) (LeCun and Bengio, 1995), the recurrent neural network (RNN) (Rodriguez et al., 1999), and the RNN with long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997).

To solve the ESI problem, deep learning frameworks have also been proposed in the past years, but with only a few existing works available. For example, Bore et al. (2021) introduced an RNN with LSTM units for spatiotemporal EEG source imaging and the proposed approach achieved good performance against the benchmark algorithms. Hecker et al. (2021) constructed a novel CNN-based structure, named ConvDip, to detect multiple sources, and this architecture is shown to outperform state-of-the-art methods. Wei et al. (2021) proposed an edge sparse basis network to learn the mapping between edge sparse source activation and recorded EEG signal.

As the source signal is defined on an irregular source space, where each source is defined as a vertex in a 3D source space (Liu et al., 2018), there exists a spatially connected graph structure among sources that have not been fully explored in the existing literature, especially with the recent advance of graph signal processing (Huang et al., 2016). In this work, we propose to employ the *spatial-temporal structure* of EEG source signal and come up with a new framework based on spatial graph Fourier transform (GFT) (Sandryhaila and Moura, 2013), and bidirectional LSTM (BiLSTM) neural network (Schuster and Paliwal, 1997), termed as GFT-BiLSTM to solve the ESI inverse problem. The main contributions of this paper are as follows:

(i) We propose to use the GFT on the 3D source space, and delineate the source space into spatially high, medium and low frequency subspaces spanned by corresponding eigenvectors, and the low frequency components naturally serve as a basis to estimate an extended areas of source activation.
(ii) By projecting the original source signal into a reduced dimensional subspace with low frequency eigenvectors, the dimension of output layer of BiLSTM can be greatly reduced.
(iii) The numerical experiments show that the proposed GFT-BiLSTM outperforms the benchmark algorithms based on area under the curve (AUC) and the localization error (LE).

RELATED WORK

Given the ill-posedness nature of ESI, traditional methods typically adopt parsimonious models to get a unique solution by introducing priors or regularizations based on the assumptions from neural physiology, brain anatomy, etc. (Scherg and Berg, 1991). The first category of ESI approaches is the equivalent current dipole (ECD) source localization (Cover et al., 2007). This method treats the neural electrical activity of the cerebral cortex as one of several ECDs. With such a constraint, the spatial location and orientation of each ECD can be optimized to best interpret the measured EEG signals. The ECD model has played a certain role in the localization of focal brain activity. However, the real brain sources can have multiple source activations (Sanei and Chambers, 2013), while the ECD method can only locate a single source point which makes it unable to reconstruct the distributed pattern of activated sources (Zumer et al., 2007). Another category of ESI methods, namely distributed source localization framework, has become more widely used in recent years. The current density distribution (CDD) model-based approach does not make any prior assumptions on the number of dipoles but divides the cerebral cortex into numerous triangular grids (Liao et al., 2012). The neural electrical activity on the brain voxels is represented by brain sources defined on the 3D mesh grid. Since the location of each source in the CDD model is fixed, the distributed source imaging only needs to solve a linear inverse problem (Astolfi et al., 2004). Over the past few decades, many distributed source imaging algorithms have been developed. The most popular ones are based on L_2 norm constraints such as the minimum norm estimate (MNE) (Koles, 1998), the dynamic statistical parametric mapping (dSPM) (Tanaka et al., 2009), the low-resolution electromagnetic tomography analysis (LORETA), and the exact LORETA (eLORETA), (Jatoi et al., 2014), etc. The computation of these methods is simple, but the resulting solutions can be overdiffuse (Ou et al., 2009). Consequently, the sparsity constraints-based source imaging algorithms have been proposed by many researchers, such as the minimum current

estimate (MCE) (Wen et al., 1998), the focal underdetermined system solver (FOCUSS) (Murray and Kreutz-Delgado, 2001), etc. Another class of methods for the ESI inverse problem is the data-driven method, which mainly includes the subspace-based classic Multiple Signal Classification (MUSIC) (Vergallo and Lay-Ekuakille, 2013) and the beamforming approaches, such as the linearly constrained minimum variance (LCMV) beamformer (Lin et al., 2008). MUSIC is the version of the Spatio-temporal approach. Multiple dipoles can be found in this technique *via* scanning potential locations through one dipole model (Mosher et al., 1992). The LCMV beamformer is a type of adaptive spatial filter that localizes activity sources by minimizing the contributions of other uncorrelated sources (Wong and Gordon, 2009). Recent developments on ESI include some interesting works such as utilizing more sophisticated edge-sparse regularization (Sohrabpour et al., 2020), or multitask framework for source localization among multiple subjects (Janati et al., 2020), or employing manifold graph structure in the EEG source space (Liu et al., 2021), source localization using multimodality of fMRI and EEG (Nguyen et al., 2018). However, the graph structure of the spatially connected sources is not fully explored in the literature, as the graph signal processing technique (Ortega et al., 2018) can have a principled way to decompose the spatial graph signal into components with different spatial frequencies. In this work, we come up with a new framework based on spatial graph Fourier transform and bidirectional LSTM (BiLSTM) neural network to efficiently solve the brain source extents reconstruction problem.

MATERIALS AND METHODS

In this section, we first give a brief introduction of the forward problem, then the spatial graph signal processing technique is explained, followed by structure of the BiLSTM neural network and finally, the GFT-BiLSTM model is introduced.

Forward Problem

The relationship between the scalp potential measured by the electrodes and the brain source distribution can be expressed as follows:

$$\boldsymbol{x}(t) = \boldsymbol{H}\boldsymbol{s}(t) + \boldsymbol{\varepsilon}(t) \tag{1}$$

where t represents the time, vector $\boldsymbol{x}(t) \in \boldsymbol{R}^{n \times 1}$ represents the EEG or MEG signal measured by n electrodes, matrix $\boldsymbol{H} \in \boldsymbol{R}^{n \times m}$ represents the lead field, vector $\boldsymbol{s}(t) \in \boldsymbol{R}^{m \times 1}$ represents the source signal generated by m brain sources, and vector $\boldsymbol{\varepsilon}(t) \in \boldsymbol{R}^{n \times 1}$ represents the additive noise from observation.

The forward model models the linear mapping between scalp potential measured by the electrodes and the brain source signal (Birot et al., 2014). The solution to the forward problem relies on the establishment of the head model, which is determined by the geometry and corresponding electrical conductivity of different head tissues such as brain, skull, scalp, etc. (Acar and Makeig, 2010). In the early days, the mainly used head models are the spherical model and the ellipsoid model (Gutiérrez et al., 2005). With the development of brain imaging technology, real head models, which can be calculated by the boundary element method (BEM), the finite elements method (FEM), and the finite difference method (FDM) are increasingly used (Akalin-Acar and Gençer, 2004). Once the head model is established, the lead field matrix can be determined.

Graph Signal Processing in the Brain Source Space

The connectivity relationship between m sources can be represented by an undirected graph, which can be defined as follows:

$$G = (V, \boldsymbol{A}) \tag{2}$$

where $V \in \boldsymbol{R}^{m \times 1}$ is a set of m nodes, $\boldsymbol{A} \in \boldsymbol{R}^{m \times m}$ is the corresponding adjacency matrix. If there is no edge connecting nodes i and j, then $a_{ij} = 0$; otherwise, $a_{ij} > 0$, and its value represents the weight of the edge between the two nodes. In addition, since G is an undirected graph, then $a_{ij} = a_{ji}$, which means the adjacency matrix $\boldsymbol{A}$ is symmetric. In the EEG source space, all the potential source locations are represented by the nodes defined on a 3D mesh as is illustrated in **Figure 1**. When the source locations i and j are neighbors on the 3D mesh, then we set $a_{ij} > 0$.

The graph signal is defined on the set of graph nodes V, which is represented by a vector, and each element represents the signal value at the corresponding node. The brain source signal $\boldsymbol{s} = [s_1, s_2, \ldots, s_m]^T$, in which each element s_i represents the signal value of the i-th source voxel, is defined on the nodes of graph G. The traditional Fourier transform calculates the projection of a function $f(t)$ on the basis function e^{-iwt}, and the projected value of a time series signal using Fourier transform $F(w)$ represents its magnitude at the basis of a specific frequency. Different from the traditional Fourier transform defined in the temporal domain, for signals defined on a graph, the eigenvectors of the Laplacian matrix $\boldsymbol{L}$ of the graph can be used as the basis vectors of the GFT, where the Laplacian matrix $\boldsymbol{L}$ can be calculated as follows:

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A} \tag{3}$$

where $\boldsymbol{L} \in \boldsymbol{R}^{m \times m}$, and $\boldsymbol{D} \in \boldsymbol{R}^{m \times m}$ is a diagonal matrix called the degree matrix, in which the diagonal elements satisfy $d_{ii} = \sum_j^m a_{ij}$, that is, the sum of elements in the i-th row of $\boldsymbol{A}$. Since $\boldsymbol{A}$ and $\boldsymbol{L}$ is real and symmetric, therefore, $\boldsymbol{L}$ can be decomposed as follows:

$$\boldsymbol{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T \tag{4}$$

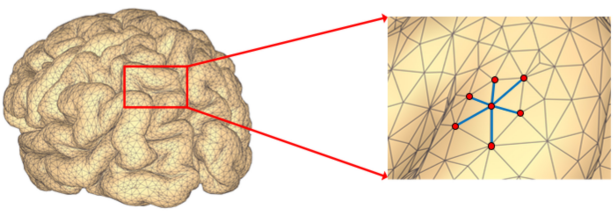

FIGURE 1 | Illustration of brain mesh and brain source extent activation.

where $\mathbf{\Lambda} \in \boldsymbol{R}^{m \times m}$ is a diagonal matrix and the diagonal elements $\lambda_i, (\ i = 1, 2, \ldots, m)$ are the eigenvalues of $\boldsymbol{L}$ and satisfy $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_m$, $\boldsymbol{U} \in \boldsymbol{R}^{m \times m}$ is the eigenvector matrix, which is also an orthogonal matrix satisfying $\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{I}$, each column in $\boldsymbol{U}$ is an eigenvector of $\boldsymbol{L}$ and corresponds to the eigenvalue in $\mathbf{\Lambda}$. With the eigenvectors of the Laplacian matrix $\boldsymbol{L}$ as the Fourier basis vectors, each eigenvector can be regarded as a graph basis with a certain frequency, and this frequency corresponds to the eigenvalue. The smaller the eigenvalue, the lower the frequency of the corresponding eigenvector, which is manifested as a small difference between the signals of adjacent nodes on the graph; on the contrary, a larger variation among neighboring signals. The value of a graph Fourier coefficient can measure the amplitude of the graph signal at different frequencies. With the eigenvectors as the Fourier basis vectors, the GFT of a given graph signal $\boldsymbol{s}$ can be defined as follows:

$$\widetilde{\boldsymbol{s}} = \boldsymbol{U}^T \boldsymbol{s} \tag{5}$$

where vector $\widetilde{\boldsymbol{s}} \in R^{m \times 1}$ is the graph Fourier coefficient. Further, the inverse graph Fourier transform (IGFT) of $\boldsymbol{s}$ can be defined as:

$$\boldsymbol{s} = \boldsymbol{U}\widetilde{\boldsymbol{s}} \tag{6}$$

The above two formulas show that a graph signal can be decomposed into components with different frequencies through the GFT, and can also be recovered through the IGFT.

To characterize the graph frequency, we introduce the following definition:

Definition 1: Graph Frequency (GF): GF, denoted as f_G, is a function of u_i which represents the total number of sign flips of u_i between any two connected nodes on G, it is defined as follows:

$$f_G(u_i) = \sum_{j=1}^{m} \sum_{p \in \Omega(j)} I(u_i(j)\, u_i(p) < 0)/2 \tag{7}$$

where $\Omega(j)$ represents all neighbors of node j, and $I(\cdot)$ is an indicator function to check whether the values of u_i at nodes j and p have a sign flip. The number of sign flips is analogous to counting the number of zero crossings of the basis signal within a given window for a time series data. We constructed the Laplacian matrices within first-order neighbors and second-order neighbors, and the associated GF spectrum is shown in **Figure 2**. It can be seen that the GF value of the eigenvector increases as the eigenvalue increases.

Similar to the counterpart in the time domain, the spatial frequency basis matrix $\boldsymbol{U}$ can be similarly decomposed into different spatial frequency bands, such as $\boldsymbol{U} = [\boldsymbol{U}_{low}, \boldsymbol{U}_{medium}, \boldsymbol{U}_{high}]$. The graph signal $\boldsymbol{s}$ can be projected into a subspace of $\boldsymbol{U}$. For example, $\widetilde{\boldsymbol{s}} = \boldsymbol{U}_{low}^T \boldsymbol{s}$ is the projection of $\boldsymbol{s}$ into a space spanned by low frequency eigenvectors. In our work, we use the spatially low frequency components as a filter to reconstruct the focally extended sources.

Bidirectional Long-Short Term Memory Neural Network

Bidirectional long-short term memory neural network is an extension of the traditional RNN (Schuster and Paliwal, 1997).

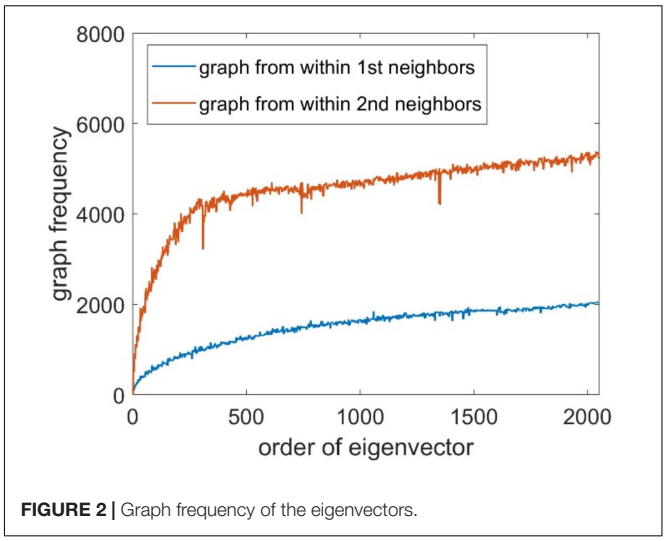

FIGURE 2 | Graph frequency of the eigenvectors.

For the time series, it is recognized that RNN can effectively estimate the information at the future moment based on the previous states. However, for the time series with a long sequence of states, the estimation performance of RNN will be greatly discounted, because the future information in a long time series usually depends on the information from distant history moments, which is the long-term dependence. However, the superior structure of the LSTM unit equips the network with the ability to solve long-term dependence. The RNN with LSTM units can filter the information by a unique structure called "gate" and store the valid information by the so-called "memory cell." The elements in a gate vector have values in the interval [0,1]. When preceding time series information arrives at the gate, it will be multiplied with the gate vector element-wise, if the element value in the gate vector is 1, then the timing information multiplied with it will be retained, and if the element value is 0, the information will be discarded after a multiplication with 0. In this way, the filtering of information propagated in the LSTM unit is achieved. The valid information obtained after filtering is then stored in the memory cell and passed on to the next moment to prevent being lost over time, thus effectively addressing the long-term dependence existing in traditional RNNs.

The structure of a standard LSTM unit is shown in **Figure 3A**, where $\boldsymbol{x}_t$, $\boldsymbol{h}_{t-1}$, and $\boldsymbol{c}_{t-1}$, respectively, represent the input sample at the current moment, the unit output and the memory state at the previous moment. The information contained in $\boldsymbol{x}_t$ and $\boldsymbol{h}_{t-1}$ is first activated by $\sigma(\cdot)$ function and got the forget gate $\boldsymbol{f}_t$, the input gate $\boldsymbol{i}_t$, and the output gate $\boldsymbol{o}_t$. At the same time, $\boldsymbol{x}_t$ and $\boldsymbol{h}_{t-1}$ are also activated by $tanh(\cdot)$ function and got a temporary state $\widetilde{\boldsymbol{c}}_t$. On the one hand, the information passed from the previous moment which contained in $\boldsymbol{c}_{t-1}$ is filtered by the forget gate $\boldsymbol{f}_t$. On the other hand, the newly input information contained in $\widetilde{\boldsymbol{c}}_t$ is filtered by the input gate $\boldsymbol{i}_t$. Then, the valid information retained by the above two filtering processes is integrated together as a new memory state $\boldsymbol{c}_t$. This newly updated memory state is passed along time to the next moment, and simultaneously, it is also filtered by

the output gate o_t. Finally, the new output of the LSTM unit h_t is obtained.

This propagation process can be formulated as follows:

$$f_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \tag{8}$$

$$i_t = \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right) \tag{9}$$

$$c_t = tanh\left(W_c\left[h_{t-1}, x_t\right] + b_c\right) \tag{10}$$

$$c_t = f_t * c_{t-1} + i_t * \widetilde{c}_t \tag{11}$$

$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{12}$$

$$h_t = o_t * \tanh\left(c_t\right) \tag{13}$$

where W_f, W_i, W_c, W_o are weight matrices; b_f, b_i, b_c, b_o are bias vectors; the symbol * stands for the element-wise multiplication.

The hidden layer of the BiLSTM neural network is composed of two layers of LSTM units that are reversely connected, and its structure is shown in **Figure 3B**. In a BiLSTM layer, the time series performs both forward propagation and backward

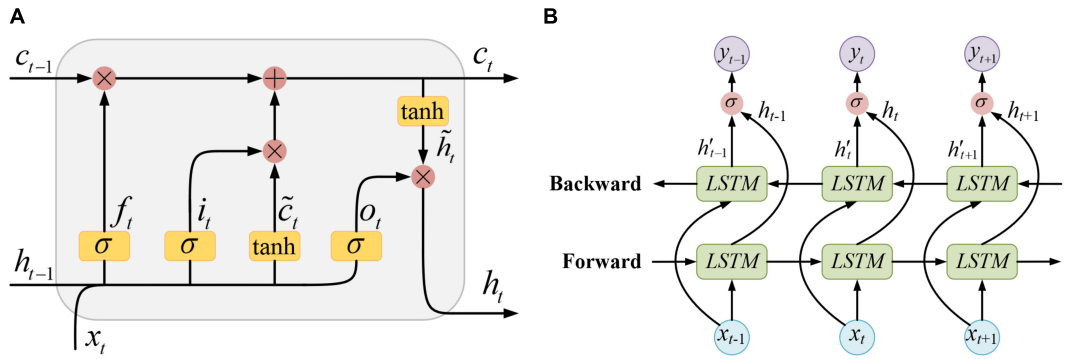

FIGURE 3 | (A) The LSTM unit, **(B)** The BiLSTM network.

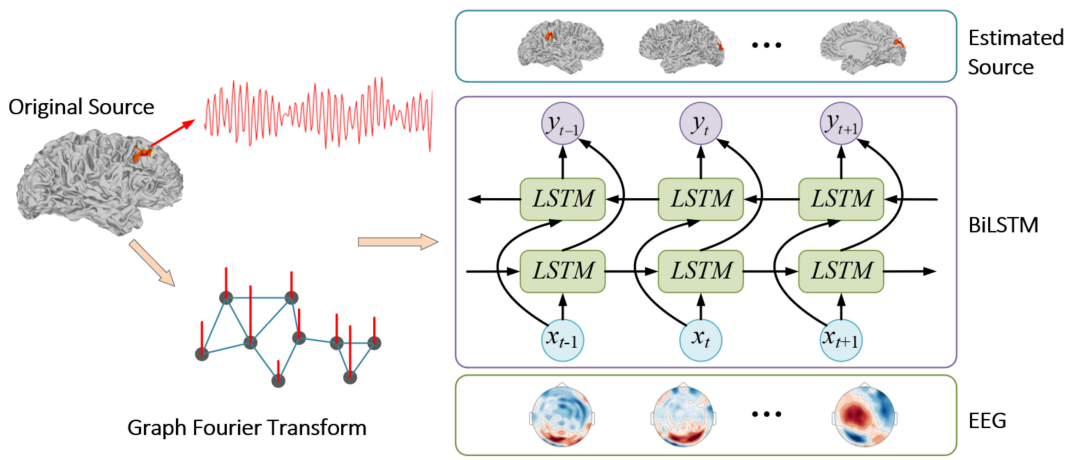

FIGURE 4 | The flowchart for the proposed method.

TABLE 1 | The evaluation metrics corresponding to different ESI inverse solutions with a single activated area.

	AUC			LE		
	SNR = 20	**SNR = 30**	**SNR = 40**	**SNR = 20**	**SNR = 30**	**SNR = 40**
GFT-BiLSTM	0.9668	0.9821	0.9844	15.5683	13.2668	13.2067
dSPM	0.7733	0.8769	0.9237	58.1761	45.4715	40.3180
MNE	0.7020	0.8192	0.8954	94.8286	69.2218	48.5848
sLORETA	0.7637	0.8784	0.9339	83.3185	46.3218	25.1723

propagation. Therefore, both the information at the previous moments and that at the future moments can be fully utilized.

The output of the BiLSTM neural network can be calculated as follows:

$$y_t = \sigma\left(W_s\left[h_t \bigoplus h_t'\right] + b_s\right) \tag{14}$$

where y_t is the final output, h_t' is the unit output of the backward propagation, W_s and b_s are the weight matrix and the bias vector of the output layer, respectively, and the symbol $\bigoplus$ stands for the vector concatenation.

Graph Fourier Transform-Bidirectional Long-Short Term Memory for Electrophysiological Source Imaging

Graph Fourier Transform-Bidirectional Long-Short Term Memory Training Procedure

Generally, the brain is divided into smaller voxels, and each voxel can be activated and regarded as a source. Therefore, when a BiLSTM network is adopted to solve the inverse problem of ESI with the recorded EEG signal as the inputs and the source signal as the outputs, the number of nodes in the output layer of the network equals to the number of sources. This will lead to a significant number of parameters in the network. In order to improve the training speed of the BiLSTM network, in this paper, we reduce its output nodes by using projected coefficients as the output dimension based on low frequency eigenvectors, given the extended source activation pattern mainly contains signal from the low frequency subspace. With the training dataset $\{x_i, s_i\}$, the training setup procedure is as follows:

Step 1: Perform the GFT on the original brain source signal $\boldsymbol{s}$ according to (5), then the Fourier coefficient $\tilde{\boldsymbol{s}}$ is obtained.

Step 2: With the eigenvectors in $\boldsymbol{U}$ as the Fourier basis vectors, and the corresponding eigenvalues in $\boldsymbol{\Lambda}$. Then, set the eigenvalue threshold as $\boldsymbol{T_f}$, and the number of eigenvalues less than $\boldsymbol{T_f}$ is $\boldsymbol{k}$.

Step 3: Take the first $\boldsymbol{k}$ columns of eigenvectors in $\boldsymbol{U}$, denoted as $\boldsymbol{U_k}$ and the first $\boldsymbol{k}$ elements in the Fourier coefficients $\tilde{\boldsymbol{s}}$ as $\tilde{\boldsymbol{s}}'$.

Step 4: Set the number of the input nodes in the BiLSTM network as $\boldsymbol{n}$, the number of the output nodes as $\boldsymbol{k}$, and the number of the BiLSTM units in the hidden layer as $\boldsymbol{l}$. Then take the EEG signal $\boldsymbol{x}$ as the input, $\tilde{\boldsymbol{s}}'$ as the output to train the BiLSTM network.

The mean square error (MSE) is chosen as the loss function:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\widetilde{s}_i - \widetilde{s}_i'\right)^2 \tag{15}$$

where N is the number of data points, $\widetilde{s}_i$ is the true values, and $\widetilde{s}_i'$ is the estimated values by the network. The Nadam optimizer is adopted during the training process.

In general, we take the projections of the brain source signal on the basis spanned by the low frequency eigenvectors instead of the

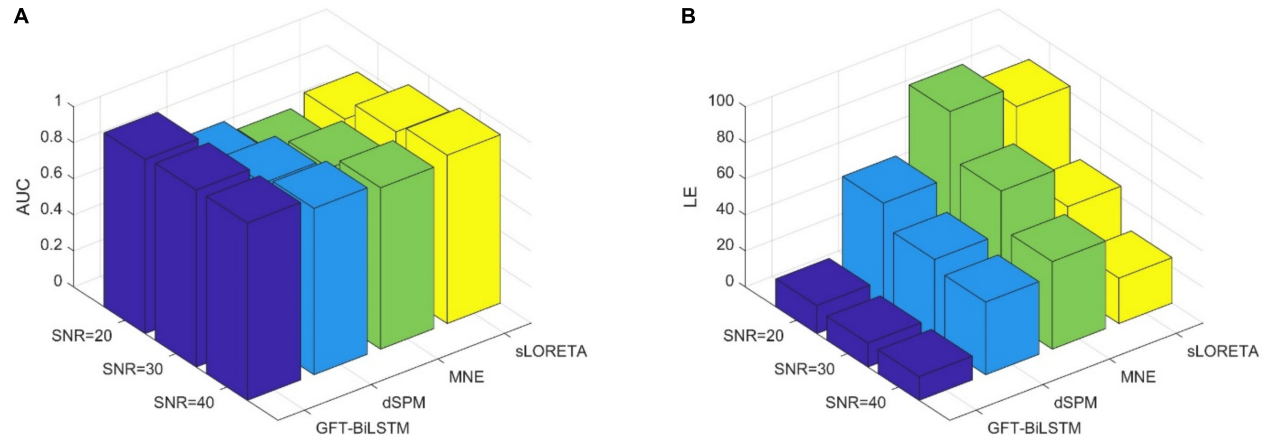

FIGURE 5 | The performance comparison of different ESI inverse solutions with a single activated area. **(A)** The comparison of AUC at different SNR levels. **(B)** The comparison of LE at different SNR levels.

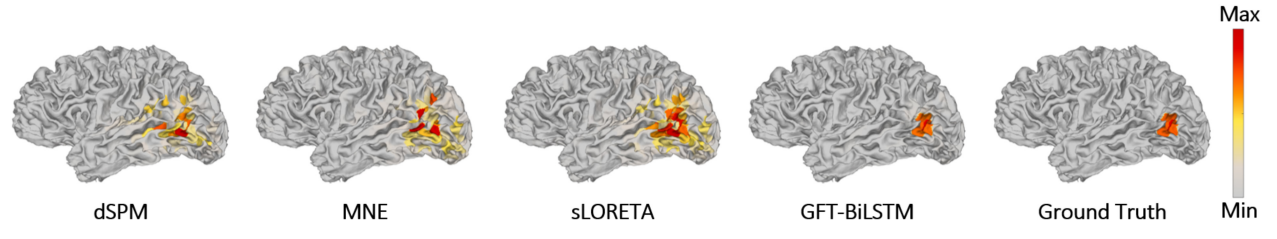

FIGURE 6 | Brain source activations estimated by different ESI algorithms with a single activated area.

brain source signal itself as the output of the BiLSTM network. By doing this, the number of output nodes in the network can be reduced from m to k, which can make the parameters in the network significantly decrease and the training speed increase. In the meanwhile, the source extent pattern is recovered better after removing the spatial high frequency noise.

Source Signal Recovery

When the network training is completed, perform the IGFT on the estimated values $\widehat{s}$ as follows:

$$\widehat{s} = U_k \widehat{s}' \tag{16}$$

where $\widehat{s}$ is the estimated source signal. The whole process is summarized in **Figure 4**.

EXPERIMENTS

In this section, the proposed GFT-BiLSTM is evaluated using both the synthetic data and the real data. The benchmark ESI algorithms, including dSPM, MNE, and sLORETA, are used for comparison.

In the simulated data, the number of brain sources is 2,052 and the number of electrodes is 128, then these source regions are activated in turn with one level neighborhood sources (sources

TABLE 2 | The evaluation metrics corresponding to different ESI inverse solutions with multiple activated areas.

	AUC			LE		
	SNR = 20	**SNR = 30**	**SNR = 40**	**SNR = 20**	**SNR = 30**	**SNR = 40**
GFT-BiLSTM	0.9602	0.9796	0.9818	11.3105	6.1145	5.6589
dSPM	0.7130	0.8214	0.8743	67.4522	52.6055	47.9739
MNE	0.6523	0.7640	0.8415	103.3403	79.0106	58.2711
sLORETA	0.7045	0.8253	0.8892	93.6530	60.1024	39.0890

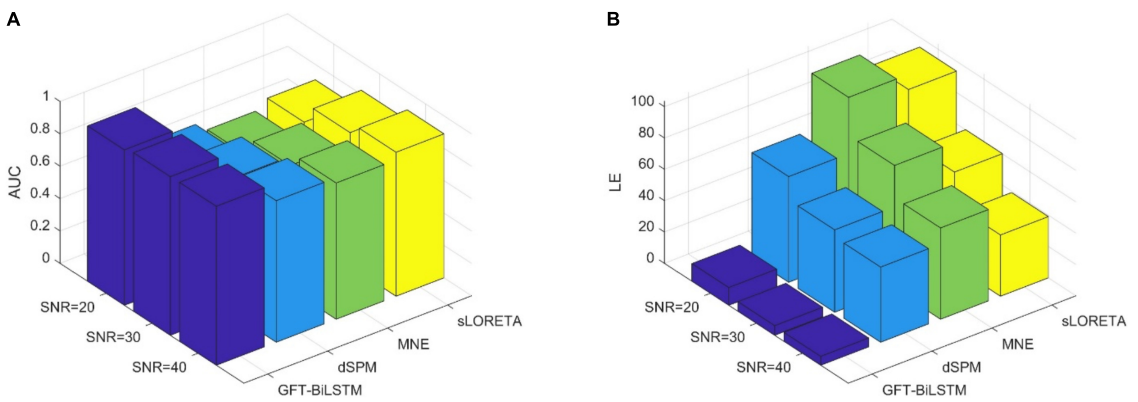

FIGURE 7 | The performance metrics comparison of different ESI inverse solutions with multiple activated areas. **(A)** The comparison of AUC at different SNR levels. **(B)** The comparison of LE at different SNR levels.

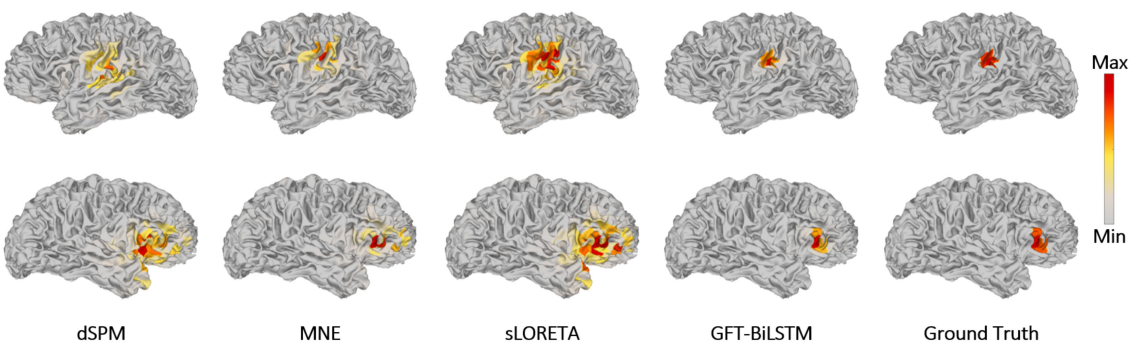

FIGURE 8 | Brain source activation reconstructed by different ESI algorithms with multiple activated areas. The upper figures correspond to the activated area in the left side of the brain, the bottom figures correspond to the activated area in the right side of the brain.

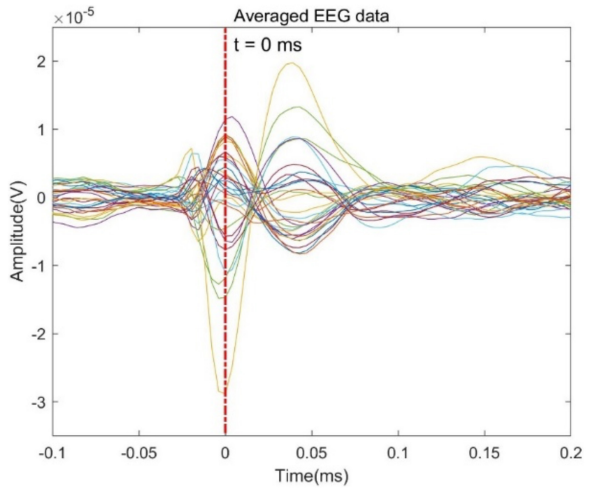

FIGURE 9 | Average EEG time series plot around the inter-ictal spike.

that are directly connected to activated source in 3D mesh) activated at the same time, and the source signal time series is generated based on the 5th-order autoregressive (AR) model (Haufe and Ewald, 2019), with 100 Hz sampling rate and 1 s of length, and the simulated source signal s is obtained. Given the lead field matrix H, by using the forward model, the EEG data $\boldsymbol{x}$ can be calculated according to Eq. (1), in which the sensor noise ε is generated based on different signal-to-noise ratio (SNR) levels (20 dB, 30 dB, and 40 dB), where SNR is defined based on the ratio of the power of signal P_{signal} to the power of noise P_{noise}, as prescribed below:

$$SNR = 10log(\frac{P_{signal}}{P_{noise}}) \tag{17}$$

The Laplacian matrix $\boldsymbol{L}$ of the brain source signal is calculated according to Eq. 3, then decomposed according to Eq. 4 to obtain the eigenvalue matrix $\boldsymbol{\Lambda}$ and the corresponding eigenvector matrix $\boldsymbol{U}$. Use the eigenvectors as the basis vectors of the GFT, then the Fourier coefficient is obtained according to Eq. 5. We set $k = 615$ as the number of eigenvectors, based on the frequency spectrum illustrate in **Figure 2**. We use the first k values of $\widetilde{\boldsymbol{s}}$ as the model output $\widetilde{\boldsymbol{s}}'$, and the EEG data is taken as the model input. The simulated data is divided into training, validation, and testing datasets according to the proportion of 70%, 15%, and 15%, respectively. The number of input nodes in the BiLSTM neural network is set to be 128, the hidden nodes is set to be 2,560, and the number of output nodes is 615. Adopt the MSE as the loss function and Nadam optimizer to train the BiLSTM neural network on the training set. After training, the testing dataset is used for model testing. The following two metrics are used as the metrics for model evaluation:

- Localization error: LE can be quantified as the distance between the true peak source point and estimated peak source point.
- Area under the curve: AUC measures the area underneath the receiver operating characteristic (ROC) curve.

Evaluation With Single Source

To render source extents activation, the adjacent sources along with a central source are activated at the same time. The signal strength of the adjacent sources is set to be lower than that of the central region. All the 2,052 potential source locations are chosen as the central source in turn to generate the scalp EEG data. In the first experiment, we test the proposed algorithm on the simulated EEG data and the true source activation pattern with one source extents activated. Apply the training and validation dataset to train and validate the proposed GFT-BiLSTM model, and then test it on the testing dataset. The performance of the GFT-BiLSTM is compared with that of dSPM, MNE, and sLORETA. The evaluation metrics of each model are shown in **Table 1** and **Figure 5**. The comparison between the ground truth source and the reconstructed sources by different algorithms is shown in **Figure 6**.

From **Table 1** and **Figure 5**, it can be seen that the proposed GFT-BiLSTM shows the better performance when compared to other methods. For different SNR levels, the AUC corresponding to GFT-BiLSTM is the highest while the LE is the lowest. The brain source distribution estimated by the proposed GFT-BiLSTM is closer to the ground truth as illustrated in **Figure 6**. The numerical result demonstrates the superiority of the GFT-BiLSTM when applied to solve the ESI inverse problem.

Evaluation With Multiple Sources

In order to study the performance of the proposed GFT-BiLSTM when there are multiple activated sources, we randomly select 2 out of 2,052 brain source locations to be activated and the

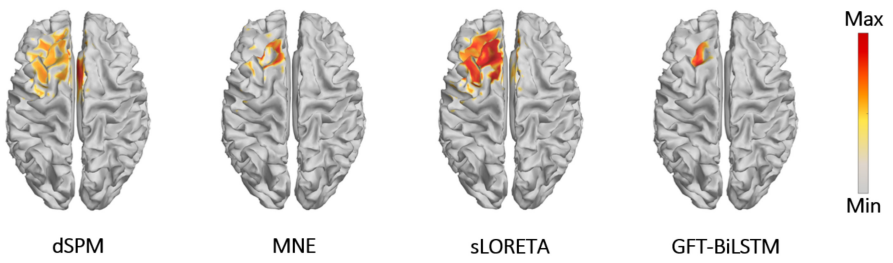

FIGURE 10 | Reconstructed sources by different ESI algorithms for epilepsy EEG data.

first level neighboring sources of these two source locations are also activated with a lower signal magnitude. Apply this simulated dataset to train and validate the GFT-BiLSTM, and then test it on the testing dataset. The performance of the GFT-BiLSTM is compared with that of dSPM, MNE, and sLORETA. The evaluation metrics of each model are shown in **Table 2** and **Figure 7**. The comparison between the real source and the estimated source is shown in **Figure 8**.

From **Table 2** and **Figure 7**, the proposed GFT-BiLSTM demonstrates better performance when compared with other methods. In addition, with the increased number of activated source, the estimation performance of all methods except the GFT-BiLSTM deteriorates significantly. This is demonstrated as an increase in LE and a decreased AUC value. There is also a situation for other method in which the localization of the central region is accurate while the localization of its adjacent regions is slightly deviated. The reason is that as the number of activated regions increases, the distribution of the sources is no longer concentrated, and it is more challenging to accurately estimate the locations of all active regions. The reduction in performance for AUC is much more pronounced for the benchmark algorithms. In contrast, the performance of the GFT-BiLSTM is more stable and robust when it comes to multiple activated sources.

Evaluation With Real Epilepsy Data

In order to further evaluate the performance of the proposed GFT-BiLSTM, we applied it on the public epilepsy EEG dataset from the Brainstorm tutorial datasets (Tadel et al., 2011). This dataset was recorded from a patient who suffered from focal epilepsy. The patient underwent invasive EEG to identify the epileptogenic area then underwent a left frontal tailored resection and was seizure-free during a 5-year follow-up period. We followed the Brainstorm tutorial to obtain the head model, and the lead field matrix. Then we calculated the average spikes (as shown in **Figure 9**) of the provided EEG measurements with 29 channels. Apply the averaged EEG data for brain source localization, and the estimated sources at peak (0 ms) from different methods are shown in **Figure 10**, as compared to other methods including dSPM, MNE, sLORETA.

It can be seen from **Figure 10** that the proposed GFT-BiLSTM provides a good reconstruction of the epileptogenic zone which was validated by the follow-up survey after resection on the left frontal region. The source area estimated by dSPM and sLORETA spans a wide range cortical areas and includes part of the right frontal lobe which is not related to the epilepsy lesion. In contrast, the source location estimated by the MNE method and the GFT-BiLSTM proposed in this paper is more accurate. However, compared between the two methods, the range of sources estimated by the GFT-BiLSTM is smaller, and the source estimated by GFT-BiLSTM shows better continuity of the spatial signal, due to the benefit of using GFT.

CONCLUSION

The inverse problem of source extents reconstruction is challenging due to its highly ill-posed nature. In this paper, we present a novel ESI framework, named GFT-BiLSTM, which is based on the delineation of spatial graph frequency using graph Fourier transform and BiLSTM, to solve the ESI problem in a more efficient and robust way. Our numerical results based on the synthetic data and real data show that the proposed GFT-BiLSTM has a superior performance compared to other benchmark methods. The future work can further explore more clinical applications using the proposed framework. A more rigorous selection of the low frequency set of eigenvectors can also be investigated.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

MJ: methodology, software, investigation, data curation, and writing – original draft. GW: methodology, conceptualization, writing, review, and editing. YG: methodology and writing – original draft. DW and HL: methodology, writing, review, and editing. JX: conceptualization, writing, review, and editing. FL: conceptualization, methodology, writing – original draft, review, editing, and supervision. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors are grateful to Andreas Schulze-Bonhage and Marcel Heers at the Epilepsy Center Freiburg for their permission to use the epilepsy dataset in this work.

REFERENCES

Acar, Z. A., and Makeig, S. (2010). Neuroelectromagnetic forward head modeling toolbox. *J. Neurosci. Methods* 190, 258–270. doi: 10.1016/j.jneumeth.2010.04.031

Akalin-Acar, Z., and Gençer, N. G. (2004). An advanced boundary element method (BEM) implementation for the forward problem of electromagnetic source imaging. *Phys. Med. Biol.* 49:5011. doi: 10.1088/0031-9155/49/21/012

Astolfi, L., Cincotti, F., Mattia, D., Salinari, S., Babiloni, C., Basilisco, A., et al. (2004). Estimation of the effective and functional human cortical connectivity with structural equation modeling and directed transfer function applied to high-resolution EEG. *Magn. Reson. Imaging* 22, 1457–1470. doi: 10.1016/j.mri.2004.10.006

Birot, G., Spinelli, L., Vulliémoz, S., Mégevand, P., Brunet, D., Seeck, M., et al. (2014). Head model and electrical source imaging: a study of 38 epileptic patients. *NeuroImage: Clin.* 5, 77–83. doi: 10.1016/j.nicl.2014.06.005

Bore, J. C., Li, P., Jiang, L., Ayedh, W. M., Chen, C., Harmah, D. J., et al. (2021). A long short-term memory network for sparse spatiotemporal EEG source imaging. *IEEE Trans. Med. Imaging* 40, 3787–3800. doi: 10.1109/TMI.2021.3097758

Congedo, M., and Sherlin, L. (2011). "EEG source analysis: methods and clinical implications," in *Neurofeedback and Neuromodulation Techniques and Applications*, eds R. Coben and J. R. Evans (Cambridge, CA: Academic Press), 25–433.

Cover, K. S., Verbunt, J. P., de Munck, J. C., and van Dijk, B. W. (2007). Fitting a single equivalent current dipole model to MEG data with exhaustive search optimization is a simple, practical and very robust method given the speed of modern computers. *Int. Cong. Ser.* 1300, 121–124. doi: 10.1016/j.ics.2007.01.026

Cui, S., Duan, L., Gong, B., Qiao, Y., Xu, F., Chen, J., et al. (2019). EEG source localization using spatio-temporal neural network. *China Commun.* 16, 131–143. doi: 10.23919/jcc.2019.07.011

da Silva, F. L. (2013). EEG and MEG: relevance to neuroscience. *Neuron* 80, 1112–1128. doi: 10.1016/j.neuron.2013.10.017

Deng, L., and Yu, D. (2014). Deep learning: methods and applications. *Found Trends Signal Process.* 7, 197–387.

Dong, C., Loy, C. C., He, K., and Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal Mac. Intell.* 38, 295–307.

Gutiérrez, D., Nehorai, A., and Preissl, H. (2005). Ellipsoidal head model for fetal magnetoencephalography: forward and inverse solutions. *Phys. Med. Biol.* 50:2141. doi: 10.1088/0031-9155/50/9/015

Haufe, S., and Ewald, A. (2019). A simulation framework for benchmarking EEG-based brain connectivity estimation methodologies. *Brain Topogr.* 32, 625–642. doi: 10.1007/s10548-016-0498-y

He, B., Sohrabpour, A., Brown, E., and Liu, Z. (2018). Electrophysiological source imaging: a noninvasive window to brain dynamics. *Ann. Rev. Biomed. Eng.* 20, 171–196. doi: 10.1146/annurev-bioeng-062117-120853

Hecker, L., Rupprecht, R., Tebartz van Elst, L., and Kornmeier, J. (2021). ConvDip: a convolutional neural network for better EEG Source Imaging. *Front. Neurosci.* 15:569918. doi: 10.3389/fnins.2021.569918

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput* 9, 1735–1780.

Huang, W., Goldsberry, L., Wymbs, N. F., Grafton, S. T., Bassett, D. S., and Ribeiro, A. (2016). Graph frequency analysis of brain signals. *IEEE J. Sel.Top. Signal Process.* 10, 1189–1203. doi: 10.1109/JSTSP.2016.2600859

Janati, H., Bazeille, T., Thirion, B., Cuturi, M., and Gramfort, A. (2020). Multi-subject MEG/EEG source imaging with sparse multi-task regression. *NeuroImage* 220:116847. doi: 10.1016/j.neuroimage.2020.116847

Jatoi, M. A., Kamel, N., Malik, A. S., and Faye, I. (2014). EEG based brain source localization comparison of sLORETA and eLORETA. *Australas. Phys Eng. Sci. Med.* 37, 713–721. doi: 10.1007/s13246-014-0308-3

Koles, Z. J. (1998). Trends in EEG source localization. *Electroencephalogr. Clin. Neurophysiol.* 106, 127–137. doi: 10.1016/s0013-4694(97)00115-6

LeCun, Y., and Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *Handbook Brain Theory Neural Netw.* 3361:1995.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444.

Liao, K., Zhu, M., Ding, L., Valette, S., Zhang, W., and Dickens, D. (2012). Sparse imaging of cortical electrical current densities via wavelet transforms. *Phys. Med. Biol.* 57:6881. doi: 10.1088/0031-9155/57/21/6881

Lin, F. H., Witzel, T., Zeffiro, T. A., and Belliveau, J. W. (2008). Linear constraint minimum variance beamformer functional magnetic resonance inverse imaging. *Neuroimage* 43, 297–311. doi: 10.1016/j.neuroimage.2008.06.038

Liu, F., Rosenberger, J., Lou, Y., Hosseini, R., Su, J., and Wang, S. (2017). Graph regularized EEG source imaging with in-class consistency and out-class discrimination. *IEEE Trans. Big Data* 3, 378–391. doi: 10.1109/tbdata.2017.2756664

Liu, F., Stephen, E. P., Prerau, M. J., and Purdon, P. L. (2019). "Sparse multi-task inverse covariance estimation for connectivity analysis in EEG source space," in *In Proceedings of the 2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, (Francisco, CA: IEEE), 299–302. doi: 10.1109/NER.2019.8717043

Liu, F., Wang, L., Lou, Y., Li, R. C., and Purdon, P. L. (2021). Probabilistic structure learning for EEG/MEG source imaging with hierarchical graph priors. *IEEE Trans. Med. Imaging* 40, 321–334. doi: 10.1109/TMI.2020.3025608

Liu, F., Wang, S., Qin, J., Lou, Y., and Rosenberger, J. (2018). Estimating latent brain sources with low-rank representation and graph regularization. *Int Conf. Brain Inform.* 11309, 304–316. doi: 10.1007/978-3-030-05587-5_29

Min, S., Lee, B., and Yoon, S. (2017). Deep learning in bioinformatics. *Brief. Bioinform.* 18, 851–869.

Mosher, J. C., Lewis, P. S., and Leahy, R. M. (1992). Multiple dipole modeling and localization from spatio-temporal MEG data. *IEEE Trans. Biomed. Eng.* 39, 541–557. doi: 10.1109/10.141192

Murray, J. F., and Kreutz-Delgado, K. (2001). "An improved FOCUSS-based learning algorithm for solving sparse linear inverse problems," in *Proceedings of the Conference Record of Thirty-Fifth Asilomar Conference on Signals, Systems and Computers (Cat. No. 01CH37256)*, Vol 1, (Grove, CA: Institute of Electrical and Electronics Engineers, IEEE), 347–351.

Nguyen, T., Potter, T., Grossman, R., and Zhang, Y. (2018). Characterization of dynamic changes of current source localization based on spatiotemporal fMRI constrained EEG source imaging. *J. Neural Eng.* 15:036017. doi: 10.1088/1741-2552/aa9fb2

Numata, T., Kiguchi, M., and Sato, H. (2019). Multiple-Time-Scale analysis of attention as revealed by EEG. NIRS, and pupil diameter signals during a free recall task: a multimodal measurement approach. *Front. Neurosci.* 13:1307. doi: 10.3389/fnins.2019.01307

Ortega, A., Frossard, P., Kovačević, J., Moura, J. M., and Vandergheynst, P. (2018). "Graph signal processing: Overview, challenges, and applications," in *Proceedings of the IEEE* 106, (Piscataway: Institute of Electrical and Electronics Engineers, IEEE), 808–828. doi: 10.1109/jproc.2018.2820126

Ou, W., Hämäläinen, M. S., and Golland, P. (2009). A distributed spatio-temporal EEG/MEG inverse solver. *NeuroImage* 44, 932–946. doi: 10.1016/j.neuroimage.2008.05.063

Portillo-Lara, R., Tahirbegi, B., Chapman, C. A., Goding, J. A., and Green, R. A. (2021). Mind the gap: State-of-the-art technologies and applications for EEG-based brain–computer interfaces. *APL Bioeng.* 5:031507. doi: 10.1063/5.0047237

Qin, J., Liu, F., Wang, S., and Rosenberger, J. (2017). "EEG source imaging based on spatial and temporal graph structures," in *Proceedings of the 2017 seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, (Montreal, QC: Institute of Electrical and Electronics Engineers, IEEE), 1–6. doi: 10.1117/1.jei.28.4.043032

Rodriguez, P., Wiles, J., and Elman, J. L. (1999). A recurrent neural network that learns to count. *Connect. Sci.* 11, 5–40. doi: 10.1080/095400999116340

Sandryhaila, A., and Moura, J. M. (2013). "Discrete signal processing on graphs: Graph Fourier transform," in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, (Vancouver: Institute of Electrical and Electronics Engineers, IEEE), 6167–6170.

Sanei, S., and Chambers, J. A. (2013). *EEG Signal Processing*. Hoboken, NJ: John Wiley & Sons.

Scherg, M., and Berg, P. (1991). Use of prior knowledge in brain electromagnetic source analysis. *Brain Topogr.* 4, 143–150. doi: 10.1007/BF01132771

Schlemper, J., Caballero, J., Hajnal, J. V., Price, A. N., and Rueckert, D. (2017). A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans. Med. Imaging* 37, 491–503. doi: 10.1109/TMI.2017.2760978

Schuster, M., and Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* 45, 2673–2681. doi: 10.1109/78.650093

Sohrabpour, A., Cai, Z., Ye, S., Brinkmann, B., Worrell, G., and He, B. (2020). Noninvasive electromagnetic source imaging of spatiotemporally distributed epileptogenic brain sources. *Nat. Commun.* 11:1946. doi: 10.1038/s41467-020-15781-0

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Computat. Intell. Neurosci.* 2011:879716. doi: 10.1155/2011/879716

Tanaka, N., Cole, A. J., von Pechmann, D., Wakeman, D. G., Hämäläinen, M. S., Liu, H., et al. (2009). Dynamic statistical parametric mapping for analyzing ictal magnetoencephalographic spikes in patients with intractable frontal lobe epilepsy. *Epilepsy Res.* 85, 279–286. doi: 10.1016/j.eplepsyres.2009.03.023

Vergallo, P., and Lay-Ekuakille, A. (2013). Brain source localization: a new method based on MUltiple SIgnal Classification algorithm and spatial sparsity of the field signal for electroencephalogram measurements. *Rev. Sci. Instrum.* 84:085117. doi: 10.1063/1.4818966

Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computat. Intell. Neurosci.* 2018, 1–13. doi: 10.1155/2018/7068349

Wei, C., Lou, K., Wang, Z., Zhao, M., Mantini, D., and Liu, Q. (2021). "Edge sparse basis network: a deep learning framework for EEG source localization," in *Proceedings of the 2021 International Joint Conference on Neural Networks (IJCNN)*, (Shenzhen: Institute of Electrical and Electronics Engineers, IEEE), 1–8.

Wen, P., He, F., and Sammut, K. (1998). "A pseudo-conductivity inhomogeneous head model for computation of EEG," in *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No. 98CH36286)*, Vol. 4, (Kong: Institute of Electrical and Electronics Engineers, IEEE), 2167–2170.

Wong, D. D., and Gordon, K. A. (2009). Beamformer suppression of cochlear implant artifacts in an electroencephalography dataset. *IEEE Trans. Biomed. Eng.* 56, 2851–2857. doi: 10.1109/TBME.2009.2029239

Young, T., Hazarika, D., Poria, S., and Cambria, E. (2018). Recent trends in deep learning based natural language processing. *IEEE Comput. Intell. Mag.* 13, 55–75. doi: 10.1109/mci.2018.2840738

Zumer, J. M., Attias, H. T., Sekihara, K., and Nagarajan, S. S. (2007). A probabilistic algorithm integrating source localization and noise suppression for MEG and EEG data. *NeuroImage* 37, 102–115. doi: 10.1016/j.neuroimage.2007.04.054

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 23 May 2022
doi: 10.3389/fnins.2022.866666

Combining Neuroimaging and Omics Datasets for Disease Classification Using Graph Neural Networks

*Yi Hao Chan, Conghao Wang, Wei Kwek Soh and Jagath C. Rajapakse**

School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

Both neuroimaging and genomics datasets are often gathered for the detection of neurodegenerative diseases. Huge dimensionalities of neuroimaging data as well as omics data pose tremendous challenge for methods integrating multiple modalities. There are few existing solutions that can combine both multi-modal imaging and multi-omics datasets to derive neurological insights. We propose a deep neural network architecture that combines both structural and functional connectome data with multi-omics data for disease classification. A graph convolution layer is used to model functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) data simultaneously to learn compact representations of the connectome. A separate set of graph convolution layers are then used to model multi-omics datasets, expressed in the form of population graphs, and combine them with latent representations of the connectome. An attention mechanism is used to fuse these outputs and provide insights on which omics data contributed most to the model's classification decision. We demonstrate our methods for Parkinson's disease (PD) classification by using datasets from the Parkinson's Progression Markers Initiative (PPMI). PD has been shown to be associated with changes in the human connectome and it is also known to be influenced by genetic factors. We combine DTI and fMRI data with multi-omics data from RNA Expression, Single Nucleotide Polymorphism (SNP), DNA Methylation and non-coding RNA experiments. A Matthew Correlation Coefficient of greater than 0.8 over many combinations of multi-modal imaging data and multi-omics data was achieved with our proposed architecture. To address the paucity of paired multi-modal imaging data and the problem of imbalanced data in the PPMI dataset, we compared the use of oversampling against using CycleGAN on structural and functional connectomes to generate missing imaging modalities. Furthermore, we performed ablation studies that offer insights into the importance of each imaging and omics modality for the prediction of PD. Analysis of the generated attention matrices revealed that DNA Methylation and SNP data were the most important omics modalities out of all the omics datasets considered. Our work motivates further research into imaging genetics and the creation of more multi-modal imaging and multi-omics datasets to study PD and other complex neurodegenerative diseases.

Keywords: attention, diffusion tensor imaging, disease classification, functional magnetic resonance imaging, Generative Adversarial Networks, graph convolutional networks, multi-omics, Parkinson's disease

OPEN ACCESS

Edited by:
Yu Zhang,
Lehigh University, United States

Reviewed by:
Geng Chen,
Northwestern Polytechnical University, China
Regina Júlia Deák-Meszlényi,
Hungarian Academy of Sciences (MTA), Hungary

****Correspondence:***
Jagath C. Rajapakse
ASJagath@ntu.edu.sg

Specialty section:
This article was submitted to Original Research Article, a section of the journal Frontiers in Neuroscience

Received: *31 January 2022*
Accepted: *06 April 2022*
Published: *23 May 2022*

Citation:
Chan YH, Wang C, Soh WK and Rajapakse JC (2022) Combining Neuroimaging and Omics Datasets for Disease Classification Using Graph Neural Networks. Front. Neurosci. 16:866666. doi: 10.3389/fnins.2022.866666

1. INTRODUCTION

Neurodegenerative diseases such as Parkinson's Disease (PD) have been shown to be associated with both brain connectivity and genetic factors. While measurements of cortical thickness from structural Magnetic Resonance Imaging (MRI) have produced contradictory findings about its utility to predict PD (Yadav et al., 2016), analysis of Diffusion Tensor Imaging (DTI) data has consistently shown that PD patients, with and without cognitive deficits, have reduced fractional anisotropy in prefrontal areas (Deng et al., 2013; Price et al., 2016). Studies on functional MRI (fMRI) data have also consistently revealed lower activity in the supplementary motor complex (Nachev et al., 2008), reduced functional connectivity in the posterior putamen (Herz et al., 2014), as well as changes in the activity levels of the dopaminergic cortico-striatal (Tessitore et al., 2019) and mesolimbic-striatal loops (Filippi et al., 2018) in PD patients.

On the genomics front, several genes (such as alpha-synuclein, LRRK2 and PARK2) and their variants, in the form of Single Nucleotide Polymorphism (SNP) data, have been associated with PD (Klein and Westenberger, 2012). However, none of them have complete penetrance and it is likely that there are multiple risk factors involved in both familial and sporadic PD (Tran et al., 2020), as well as influence from non-coding ribonucleic acid (RNA) (Majidinia et al., 2016). Thus, small non-coding RNA (sncRNA) such as micro RNA (miRNA) should be considered as well. miRNA has been associated with PD: the mitochondrial cascade hypothesis stems from miRNA dysregulation, which causes oxidative stress in neurons and ultimately lead to aggregation of alpha-synuclein and neurodegeneration (Watson et al., 2019). With sporadic PD representing a much larger proportion of PD cases as compared to familial PD, epigenetics alterations (such as DNA Methylation) could be a potential biomarker for PD (Miranda-Morales et al., 2017). Recent findings have revealed that hypo-regulation of some PD-associated genes, such as the SNCA promoter region, upregulates SNCA and leads to the formation of Lewy bodies (Wang et al., 2019).

Neuroimaging and multi-omics data capture different aspects of brain disease manifestations. Neuroimaging modalities such as DTI and fMRI capture macroscopic differences in the structure and function of healthy and diseased brains while multi-omics data zoom into a microscopic view of various molecular signatures in neurodegenerative diseases. Although these modalities have been implicated in PD, their relative importance over each other is less clear. Thus, integrating imaging and omics modalities could reveal new links between these levels of analysis and unravel the pathway of complex neurodegenerative diseases such as PD (Antonelli et al., 2019). However, methods to combine imaging and genetics data are very limited. Existing studies typically study multi-modal imaging data (Subramanian et al., 2020) and multi-omics data (Chaudhary et al., 2018; Zhang et al., 2018; Jin et al., 2021) separately, or combine one imaging modality with only one omics dataset (Kim et al., 2017; Markello et al., 2021). Notably, there have also been works that merged multi-modal imaging data with non-imaging data such as demographic features (Kazi et al., 2019a,b); as well as combining genetic data with clinico-demographic data (Nalls et al., 2015). However, none has attempted to combine both multi-modal imaging and multi-omics data.

One reason for this is due to the very large number of features involved in both imaging and omics datasets. Depending on the choice of atlases, structural and functional connectivity matrices could introduce several thousands of features, while omics datasets are even bigger, ranging from thousands in sncRNA to half a million in DNA Methylation data. Existing methods to combine both data modalities are rudimentary and often involve concatenation. This makes modeling challenging, especially because number of data samples with both imaging and omics data are very few. Models trained on such small datasets overfit easily.

To overcome these issues, we propose a deep neural network architecture that uses a combination of graph convolution layers and the attention mechanism to model multi-modal imaging and multi-omics datasets simultaneously. This is demonstrated on the Parkinson's Progression Markers Initiative (PPMI) dataset, which has a rich collection of imaging (DTI, fMRI) and omics datasets (SNP, sncRNA, miRNA, RNA sequencing and DNA Methylation). However, the number of disease classification studies based on this dataset has been limited, likely due to the very imbalanced distribution of classes (many more PD patients than controls). To alleviate the problem of imbalanced data, we propose the use of CycleGAN to generate structural and functional connectivity matrices of healthy subjects to augment the existing dataset. Existing methods for addressing class imbalance are not feasible for our problem—synthetic data generation algorithms such as SMOTE and ADASYN could generate more data but it will not be possible to associate them to a particular set of omics data sample. Under-sampling exacerbates the issue of having small datasets, while over-sampling merely duplicates the existing dataset. Given a structural connectivity matrix, CycleGAN is able to generate a functional connectivity matrix (and vice versa) such that it corresponds to the same subject and it is not just another repeated data sample in the existing dataset.

With these augmented and less imbalanced datasets, we propose an architecture named JOIN-GCLA (Joining Omics and Imaging Networks *via* Graph Convolutional Layers and Attention) to model both connectome and genomics data simultaneously. Based on our proposed algorithm, a population graph generated from both structural and functional connectivity matrices is used as the graph of the graph convolution layer. Thus, the learnt embedding of the feature vectors—which could be arbitrarily chosen—will be influenced by the multi-modal imaging data. The learnt representations are then passed into multiple graph convolution layers, each based on a graph that is built using different omics datasets. Each graph convolution layer produces its own intermediate representations and interim prediction. These are fused together *via* an attention mechanism, leading to a final decision of the disease classification problem.

Experiment results showed that the best performing model made use of both multi-modal imaging and multi-omics data. Both were crucial for the good performance—model performance fell significantly when only 1 imaging modality,

only 1 omics or when no omics dataset were used. Data augmentation was essential for the models to perform well—without it, the extreme imbalance hinders proper model training even with the use of class-weighted cost functions. JOIN-GCLA was shown to outperform existing approaches of multi-modal fusion (Long et al., 2012; Kazi et al., 2019b). Ablation studies demonstrated the importance of the initial graph convolution layer used to learn representations of the connectome data - replacing the graph convolution layer with fully-connected or convolution layers saw significant reduction in model performance. The proposed attention layer was also shown to outperform a self-attention baseline. Furthermore, JOIN-GCLA provides improved model interpretability. With a carefully designed attention mechanism, the resultant attention matrix revealed that out of the omics datasets used, DNA methylation was the most important omics data when predicting that the data sample is a healthy control, while SNP was most important when predicting PD patients.

In sum, we have made the following novel contributions in this work:

- Proposed an architecture, JOIN-GCLA, that is able to incorporate both multi-modal connectome datasets and multi-omics datasets simultaneously.
- JOIN-GCLA provides better model interpretability from the generated attention score matrix—it is able to identify which omics modalities are being focused on when predicting a certain disease class.
- Found that amongst all the multi-omics datasets used, DNA methylation and SNP are the most important omics modalities for PD classification.

2. METHODS

2.1. JOIN-GCLA Architecture

We propose a deep neural network architecture, named Joining Omics and Imaging Networks *via* Graph Convolutional Layers and Attention (JOIN-GCLA), that consists of multiple graph convolution layers and an attention mechanism to combine multi-modal imaging data and multi-omics datasets for prediction of PD. **Figure 1** illustrates the JOIN-GCLA architecture that is made up of 3 cascaded networks: the connectome encoder, omics networks, and an attention layer.

Fusion of multi-modal imaging data and multi-omics data is performed within the graph convolution layer of the connectome encoder and omics network, respectively. Thus, the inputs to the JOIN-GCLA architecture can be arbitrarily defined, depending on what is desired to be studied. In this work, we use features from the connectomes derived from each imaging modality as inputs to JOIN-GCLA. Let us assume we receive a multi-modal imaging dataset $X = \{X^m\}_{m=1}^M$ with connectivity feature matrix $X^m \in \mathbb{R}^{P\times J_m}$ where J_m is the number of connectivity features derived from each imaging modality m, obtained from P imaging scans. For the omics networks, the information from N omics data types are encoded in the graphs of N graph convolution layers. Let $O = \{O^n\}_{n=1}^N$ denote the features of omics data where N denotes the number of omics data types and $O^n \in \mathbb{R}^{P\times K_n}$ denotes the features from the n-th omics data type. K_n is the number of omics features from each omics data type n. Finally, let the set of weights, biases, output and size of the l-th layer be denoted by $W^{(l)}$, $b^{(l)}$, $H^{(l)}$ and $L^{(l)}$, respectively.

2.1.1. Population Graphs

Both the connectome encoder and the omics networks make use of graph convolution layers that decode the information encoded in population graphs where each node in a population graph represents a data sample. The connectome encoder condenses the structural and functional connectivity matrices into a small and compact vector representation. The omics networks receive the representation realized from imaging data and combine them with omics data for disease classification.

The graph of the connectome encoder is built from multiple connectome datasets derived from neuroimaging data. Formally, we define the imaging-based population graph as a population scan graph (PSG) where imaging scans are represented as nodes and the similarity between each pair of scans is calculated as the edge weight, making it a fully connected weighted graph.

Let us denote x_v^m as the connectivity features for imaging modality m from an individual v and let $A^m = \{a_{uv}^m\} \in \mathbb{R}^{P\times P}$ denote the adjacency matrix of a PSG where u and v denote two data samples. Each weight a_{uv} represents the similarity $\texttt{sim}$ between two samples:

$$a_{uv}^m = \texttt{sim}(x_u^m, x_v^m) \tag{1}$$

Similarly, we make population omics graphs (POG) from features of each omics data type. Let o_v^n represent the omics features for omics data type n from an individual v and let $B^n = \{b_{uv}^n\} \in \mathbb{R}^{P\times P}$ represent the adjacency matrix of a POG. Each weight b_{uv} represents the similarity $\texttt{sim}$ between two samples:

$$b_{uv}^n = \texttt{sim}(o_u^n, o_v^n) \tag{2}$$

The similarity measure $\texttt{sim}$ is chosen as the Pearson's correlation coefficient.

2.1.2. Connectome Encoder

The connectome encoder is made up of a linear layer and a graph convolution layer. The input to the connectome encoder is the modality-wise concatenation of connectivity feature matrices, represented by $X^c \in \mathbb{R}^{P\times J}$, where $J = \sum_{m=1}^M J_m$. A linear layer is first used to reduce data dimensionality. This is needed because the connectivity matrices were built by computing correlations between time-series from brain regions-of-interests (ROI), which produces a large number of features. In this work, since both fMRI and DTI data were involved, we used the AAL atlas which defines 116 ROIs and produces 6,670 features for each imaging modality, warranting the need for the linear layer:

$$H^{(1)} = \text{ReLU}(X^c W^{(1)} + b^{(1)}) \tag{3}$$

where ReLU denotes the ReLU activation function.

The output of the linear layer is then passed to the graph convolution layer. Additionally, the graph convolution layer takes in a PSG as the graph A. The PSG was created by setting

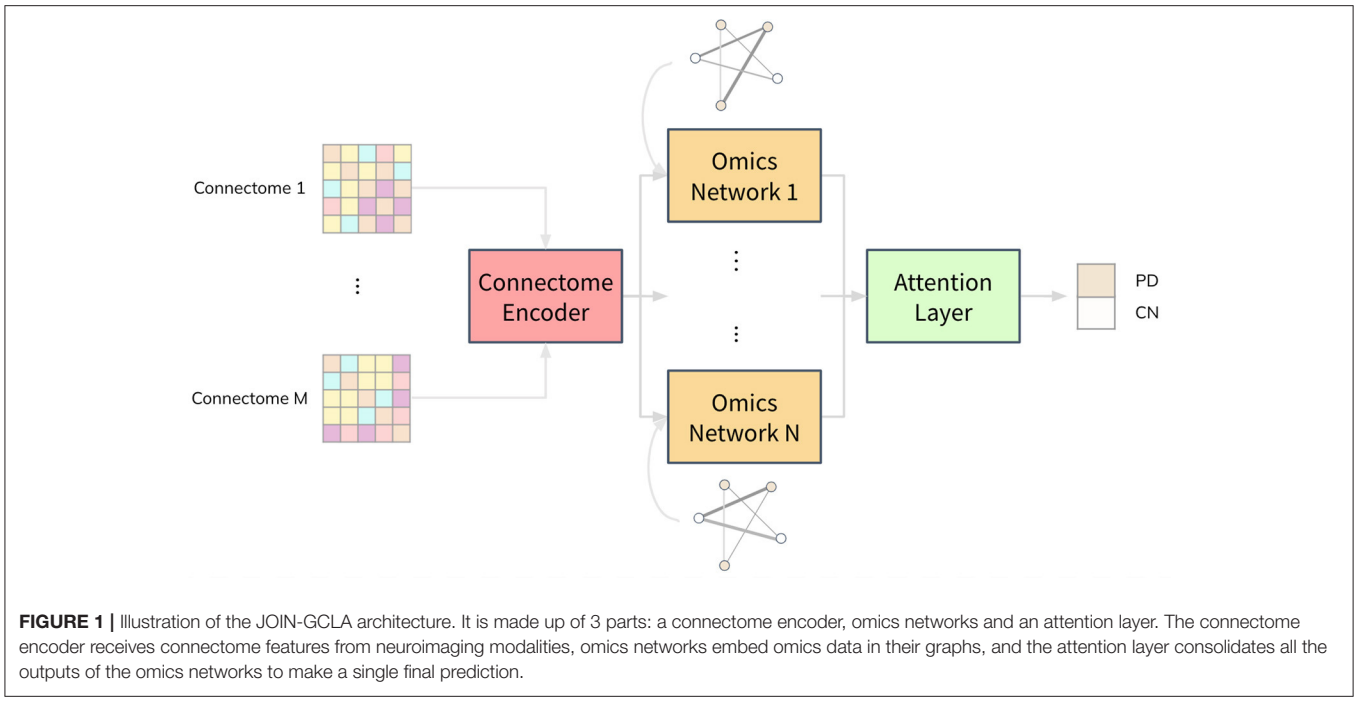

FIGURE 1 | Illustration of the JOIN-GCLA architecture. It is made up of 3 parts: a connectome encoder, omics networks and an attention layer. The connectome encoder receives connectome features from neuroimaging modalities, omics networks embed omics data in their graphs, and the attention layer consolidates all the outputs of the omics networks to make a single final prediction.

the edge weights between each pair of subjects as the Pearson's correlation of their vectorised connectivity matrices. Min-max normalization is then performed on the PSG and each element in the PSG is incremented by 1 to ensure that the minimum value is 1. When there are multiple modalities involved, let the PSG of modality m be denoted by A^m. A^m is multiplied with the existing PSG A, which is initialized as a matrix of ones. A is then used as the graph of the graph convolution layer.

Since the PSG is fully connected, the graph convolution layer should incorporate edge weights from the graph when improving the feature vector. One such layer was proposed in Kipf and Welling (2016):

$$H^{(2)} = \text{ReLU}(\hat{D}_A^{-1/2}\hat{A}\hat{D}_A^{-1/2}H^{(1)}W^{(2)}) \tag{4}$$

where $\hat{A} = A + I$ represents the PSG (of dimensions $P \times P$) with self-loops added, and $\hat{D}_A = \{\hat{d}_{vv}\}$ represents the diagonal degree matrix of A with $\hat{d}_{vv} = \sum_{u \in V} \hat{d}_{vu}$ where V is the vertex set of scans. The output of the connectome encoder, $H^{(2)}$, is subsequently used as input to each omics network.

2.1.3. Omics Networks

Each omics network is made up of a graph convolution layer and a softmax layer. Despite receiving the same output from the connectome encoder, each omics network produces different outputs because the POG used in each omics network is different. Creating the POG O involves a different procedure from Parisot et al. (2018) due to the nature of omics datasets. For example, the population graph of DNA Methylation and miRNA data have values very close to each other (as seen in **Figure 3**), which requires further scaling. This is done *via* the WGCNA algorithm (Zhang and Horvath, 2005) which re-scales the values to follow a power law distribution. Furthermore, while one subject has only one set of multi-omics data, a single subject can have multiple imaging scans. Thus, a duplication step has to be introduced to replicate the omics features when a subject has multiple imaging scans.

In short, POGs are generated by producing an adjacency matrix *via* computing the correlation between each scan's omics vector, followed by addition of self-loops, WGCNA scaling and scan duplication for subjects with more than 1 imaging scan, producing an $P \times P$ matrix. Since POGs are also always fully connected, the model proposed in Kipf and Welling (2016) can be used.

$$H^{(3)} = \hat{D}_B^{-1/2}\hat{B}\hat{D}_B^{-1/2}H^{(2)}W^{(3)} \tag{5}$$

where $\hat{B} = B + I$ represents the POG with self-loops added, $\hat{D}_B = \{\hat{d}_{vv}\}$ represents the diagonal degree matrix of B. Subsequently, the output of the graph convolution layer is passed to a linear layer with $L^{(4)}$ hidden nodes, where $L^{(4)}$ represents the number of classes for the classification task.

$$H^{(4)} = \text{ReLU}(\text{ReLU}(H^{(3)})W^{(4)} + b^{(4)}) \tag{6}$$

The above equations detail the process of generating the outputs of a single omics network. In the case where only a single omics data is available, $H^{(4)}$ can be passed to a softmax layer to produce the final prediction. Given N different sets of omics data, we will repeat these steps for each omics dataset, each producing their own omics network. Then, both $H^{(3)}$ and $H^{(4)}$ will be used by the attention layer shown in the next section.

2.1.4. Attention Layer

When multiple omics datasets are used, not all of them will be useful for the classification task. Thus, we introduce an attention layer that learns which omics network to pay more attention

to when making the final prediction. For each data sample, the attention layer will learn an attention matrix of dimensions $N \times L^{(4)}$, showing which omics network is being focused on for the classification task. It will also produce a single prediction for the disease classification task.

The attention mechanism, following the terminology in Vaswani et al. (2017), involves two components: (i) the attention weights produced from a pair of query and key matrices, and (ii) the value matrix, i.e. the term to be weighted. The latter refers to $H^{(4)}$, the logits from each omics network. Thus, let $H^{(4c)} \in \mathbb{R}^{P \times N \times L^{(4)}}$ be the concatenated logits from all omics networks. For the former, since it is desirable to arrive at an attention matrix of dimension $N \times L^{(4)}$ for better model interpretability, the query matrix is defined as $H^{(4m)} \in \mathbb{R}^{P \times L^{(4)} \times 1}$, the mean of logits from all omics networks, averaged across dimension N and transposed so that the shape of the attention matrix is correct. The key matrix is defined as $H^{(3c)} \in \mathbb{R}^{P \times N \times L^{(3)}}$, which represents the combined outputs concatenated from the graph convolution layer in each omics network. Since the last dimension of $H^{(4m)}$, $H^{(3c)}$ and $H^{(4c)}$ are different, $H^{(3c)}$ is projected *via* a projection matrix $W^{(3c)} \in \mathbb{R}^{L^{(3)} \times 1}$. Similarly, $H^{(4c)}$ is projected by $W^{(4c)} \in \mathbb{R}^{L^{(4)} \times 1}$.

Finally, the query matrix $H^{(4m)}$ and the key matrix $H^{(3c)}$ are combined to compute the attention score used to weigh the value matrix $H^{(4c)}$. In sum, this operation finds the best set of weights to weigh the output of each $H^{(4)}$ from the omics networks, producing $H^{(5)} \in \mathbb{R}^{P \times L^{(4)} \times 1}$.

$$H^{(5)} = \text{softmax}\left(H^{(4m)}\left(H^{(3c)}W^{(3c)}\right)^{T}\right)\left(H^{(4c)}W^{(4c)}\right)\Big) \quad (7)$$

2.1.5. Output Layer

$H^{(5)}$ is then passed into a softmax layer to produce the predicted class label y.

$$p(y_i = y_k|H^{(5)}) = \text{softmax}\left(H^{(5)}\right) \quad (8)$$

2.1.6. Training

Training of the JOIN-GCLA architecture is done by minimizing the error between predicted class label y and the target class label y_d *via* a weighted cross-entropy cost function J to account for data imbalance. Let $w_{y_d} = 1 - \frac{P_{y_d}}{P}$ be the weight of the class y_d, where P_{y_d} refers to the data subset that belongs to the class y_d.

$$J = E_x\{-w_{y_d} y_d \, log(y) - (1 - w_{y_d})(1 - y_d) \, log(1 - y)\} \quad (9)$$

The cost function J is minimized using an Adam optimiser. Also, during model training, dropouts are added after the graph convolution layer in both the connectome encoder and the omics networks.

3. RESULTS

3.1. Dataset and Pre-processing

Data used in this study were obtained from the Parkinson's Progressive Markers Initiative (PPMI) (Marek et al., 2018). PPMI is a clinical study that seeks to build data driven approaches for early diagnosis of PD by discovering novel biomarkers. For this study, we have utilized both imaging and genetic data downloaded from the website. **Tables 1**, **2** summarizes key demographic information and statistics of the PPMI dataset for imaging data, while **Table 3** shows the sample and feature sizes of the omics datasets. PD subjects included in this study are those who either have a pathogenic genetic variant or are newly diagnosed and have yet to commence medication for PD.

TABLE 1 | Basic statistics of subjects with DTI scans in PPMI dataset.

	Healthy control (HC)	Parkinson's disease (PD)
Number of subjects (scans)	66 (178)	154 (705)
Male/Female	43/23	98/56
Age	60.9 ± 10.6	60.8 ± 9.3

Details about the pre-processing steps are shown in the **Supplementary Materials**. In brief, after pre-processing the raw diffusion weighted imaging data to correct for motion, eddy currents and echo planar imaging distortions *via* the `dwi-preprocessing-using-t1` pipeline in Clinica (Routier et al., 2021), structural connectivity matrices were obtained by performing probabilistic tractography using the `BedpostX GPU` (Hernández et al., 2013) and `ProbtrackX GPU` (Hernandez-Fernandez et al., 2019) tool from FSL (Jenkinson et al., 2012). Since the raw connectivity matrix is not symmetric, the average of the upper and lower triangular was computed and was further log-transformed and standardized to ensure that the values follow a standard normal distribution (which will aid downstream modeling tasks). The fMRI dataset was processed using fMRIPrep (Esteban et al., 2019) and the AAL atlas was used to generate 116 regions of interests (ROI) from both the cortex and subcortex. The activation of a ROI is computed by taking the mean time series of all voxels less than 2.5 mm away from the ROI. Pearson correlation was used to obtain a symmetric matrix containing the functional connectivities between pairs of ROIs for each scan.

Most of the DTI and fMRI scans in the PPMI datasets are taken on different sessions (i.e. different days). Just relying on scans which are taken on the same day will result in a small and unusable dataset. Instead, for every DTI scan, we pair it up with fMRI scans which are taken not more than 1 year away from the date the DTI scan was performed. This produces 351 PD and 25 HC scans with paired DTI and fMRI data.

For multi-omics datasets, PPMI provides pre-processed data, with steps such as quality control and normalization performed. RNA-Seq data are given in format of Transcripts Per Million, and sncRNA and miRNA data are given in Reads Per Million (RPM) and RPM Mapped to miRNA formats. DNA Methylation (Met) and Single Nucleotide Polymorphism (SNP) data have been distilled with p-value detection. Based on the above processing, we further perform noise removal and Wilcoxon Signed Rank test to eliminate irrelevant features on the sample set required for downstream experiments. More details about the pre-processing steps can be found in the **Supplementary Materials**.

TABLE 2 | Basic statistics of subjects with fMRI scans in PPMI dataset.

	Healthy control (HC)	Parkinson's disease (PD)
Number of subjects (scans)	18 (19)	94 (194)
Male/Female	14/4	64/30
Age	61.0 ± 10.8	59.7 ± 10.2

TABLE 3 | Dataset and feature sizes of multi-omics data before and after pre-processing.

Omics data type	Dataset size	Original feature size	Processed feature size
RNAseq	226	34,569	19,728
Met	152	864,067	677,506
SNP	206	267,607	239,731
sncRNA	184	29,585	4,366
miRNA	184	2,656	748

3.2. Data Augmentation

Most multi-modal imaging and multi-omics datasets are small because not all the subjects with one imaging modality come along with other modalities. For instance, not all subjects with DTI scans will have a corresponding fMRI scans (and vice versa). This is also true for the PPMI dataset. Another major issue in the PPMI dataset is the huge class imbalance, with the number of PD subjects about 10 times larger than the number of healthy controls, as seen in **Tables 1**, **2**. To address these issues, we use CycleGAN, a type of Generative Adversarial Network (GAN) proposed by Zhu et al. (2017), to generate functional connectomes from structural connectomes of healthy subjects. GANs are generative models that can generate additional data samples with distributions similar to that of the distribution of the training dataset. CycleGAN is made up of conditional GANs, which are able to use images of one modality as latent variable so as to generate images of another modality. CycleGAN goes further to introduce a cycle consistency loss that ensures that the source and target images are consistent with each other as the network is able to both generate the target image from the source image and reconstruct the source image from the generated target image.

To train the CycleGAN architecture, functional and structural connectivity matrices, generated from preprocessed fMRI and DTI data from the Human Connectome Project (HCP) S1200 release (Glasser et al., 2013), was used as the training data and the CycleGAN model was tuned and tested using data from the Amsterdam Open MRI Collection (AOMIC) (Snoek et al., 2021). PIOP1 was used as validation set, while PIOP2 was the test set. Both HCP and AOMIC datasets are made up of brain imaging scans from healthy young adults. These were chosen, despite the age differences from PPMI, due to the large dataset sizes available (1062 for HCP, 189 for PIOP1 and 183 for PIOP2). To the best of our knowledge, no publicly available datasets with such dataset sizes exist for elderly populations. Pre-processing steps for the HCP and AOMIC datasets are similar to Section 3.1 and more details about the dataset and pre-processing steps are provided in the **Supplementary Materials**. With a trained CycleGAN model, structural connectivity matrices are passed into it to generate additional fMRI scans of healthy subjects. These are used to augment the original dataset. This results in 208 PD and 186 HC scans, a more balanced dataset (52.3% as compared to 91.6% previously). For the paired DTI-fMRI dataset, this results in 351 PD and 364 HC scans, also resulting in a more balanced dataset (53.3% as compared to 93.4%).

3.3. Hyperparameter Tuning

The huge number of possible omics and imaging data combinations makes it unfeasible to tune the model for each of them. Rather, hyperparameter tuning was performed once on the largest dataset available for the baseline model (i.e. a graph convolutional layer, without the omics networks, trained only on DTI data). We first split the dataset into non-test and test sets at a 2:1 ratio, before performing 5 fold cross-validation on the non-test split. Once the optimal parameters are found, the experiments are repeated over 10 seeds and the mean accuracies (along with standard deviation) are reported in the next sections. Importantly, synthetic data are only added to the training set–the validation and test set always uses real data only.

Parameters tuned include dropout {0.1, 0.3, 0.5}, number of hidden neurons in the graph convolution layers {2, 4, 8, 16, 32} and learning rate {0.001, 0.0005, 0.0001}. Early stopping with a patience of 20 epochs was applied during the tuning process and the largest number of epochs taken to reach the best Matthew Correlation Coefficient (MCC) score was used as the number of epochs to train the model for before applying the model on the test set. The optimal parameters are dropout of 0.1, 16 hidden neurons and learning rate of 0.001. Adam optimiser was used to train the model. This set of parameters is consistently used throughout all combinations of data modalities, with no further model tuning done for the other imaging and omics combinations. All experiments were repeated over 10 seeds.

3.4. Data Augmentation Improves Disease Classification

The PPMI dataset is heavily imbalanced. Even when the cost function is weighted by the classes, **Table 4** showed that the trained JOIN-GCLA model cannot classify well without data augmentation. While the accuracy achieved is high, that is an indication that the model is stuck at predicting the majority class (PD) and cannot predict the minority class (HC) well. **Supplementary Table S1** shows the percentage of the dataset represented by the majority class. It is evident that model performance on the original dataset is often around or even below this percentage. Additional confirmation is provided by the MCC scores, which are very low without data augmentation. With data augmentation, MCC increased significantly on most omics combinations. Thus, data augmentation helps to reduce the imbalance and it is necessary for good model performance. Analyses in subsequent sections will use this augmented dataset.

TABLE 4 | Comparison of model performance on DTI-fMRI data, with and without training set augmentation.

	No augmentation		With augmentation	
Omics	**Accuracy**	**MCC**	**Accuracy**	**MCC**
None	93.09 ± 0.03	0.00 ± 0.00	93.09 ± 0.03	0.00 ± 0.00
Met	89.59 ± 0.04	0.02 ± 0.05	90.21 ± 0.04	0.13 ± 0.25
SNP	93.18 ± 0.03	0.03 ± 0.09	93.89 ± 0.03	0.08 ± 0.25
miRNA	96.16 ± 0.00	0.00 ± 0.00	96.16 ± 0.00	0.00 ± 0.00
sncRNA	96.16 ± 0.00	0.01 ± 0.01	96.16 ± 0.00	0.00 ± 0.00
RNAseq	92.82 ± 0.03	0.00 ± 0.00	92.82 ± 0.03	0.01 ± 0.04
RNAseq-Met	81.21 ± 0.20	0.17 ± 0.33	92.52 ± 0.10	0.79 ± 0.23
RNAseq-SNP	88.96 ± 0.11	0.28 ± 0.19	84.92 ± 0.13	0.38 ± 0.29
RNAseq-miRNA	87.48 ± 0.26	0.03 ± 0.05	95.65 ± 0.02	0.02 ± 0.03
RNAseq-sncRNA	92.11 ± 0.11	0.02 ± 0.05	96.10 ± 0.01	0.08 ± 0.18
Met-SNP	85.59 ± 0.14	0.39 ± 0.34	83.99 ± 0.13	0.43 ± 0.35
Met-miRNA	90.62 ± 0.20	0.03 ± 0.05	95.72 ± 0.11	0.61 ± 0.51
Met-sncRNA	85.99 ± 0.24	0.02 ± 0.05	98.16 ± 0.02	0.43 ± 0.49
SNP-miRNA	96.84 ± 0.03	0.04 ± 0.10	100.0 ± 0.00	1.00 ± 0.00
SNP-sncRNA	85.91 ± 0.29	0.02 ± 0.06	99.78 ± 0.01	0.90 ± 0.32
miRNA-sncRNA	94.87 ± 0.04	0.06 ± 0.15	96.25 ± 0.01	0.06 ± 0.18
RNAseq-Met-SNP	89.45 ± 0.04	0.37 ± 0.25	86.46 ± 0.16	0.56 ± 0.38
RNAseq-Met-miRNA	97.13 ± 0.01	0.01 ± 0.01	97.93 ± 0.01	0.29 ± 0.45
RNAseq-Met-sncRNA	97.28 ± 0.01	0.11 ± 0.25	97.80 ± 0.02	0.32 ± 0.47
RNAseq-SNP-miRNA	88.64 ± 0.30	0.01 ± 0.02	99.63 ± 0.01	0.81 ± 0.40
RNAseq-SNP-sncRNA	97.99 ± 0.00	0.02 ± 0.03	99.77 ± 0.01	0.90 ± 0.30
RNAseq-miRNA-sncRNA	90.18 ± 0.19	0.04 ± 0.06	95.60 ± 0.02	0.01 ± 0.03
Met-SNP-miRNA	96.62 ± 0.01	0.06 ± 0.20	99.68 ± 0.01	0.91 ± 0.30
Met-SNP-sncRNA	96.94 ± 0.01	0.16 ± 0.33	100.0 ± 0.00	1.00 ± 0.00
Met-miRNA-sncRNA	90.23 ± 0.22	0.00 ± 0.02	98.70 ± 0.01	0.52 ± 0.50
SNP-miRNA-sncRNA	90.77 ± 0.23	0.01 ± 0.01	99.80 ± 0.01	0.90 ± 0.32
RNAseq-Met-SNP-miRNA	87.20 ± 0.29	0.12 ± 0.31	99.72 ± 0.01	0.90 ± 0.32
RNAseq-Met-SNP-sncRNA	85.23 ± 0.29	0.05 ± 0.10	99.42 ± 0.01	0.80 ± 0.42
RNAseq-Met-miRNA-sncRNA	96.70 ± 0.01	0.03 ± 0.06	97.08 ± 0.03	0.31 ± 0.48
RNAseq-SNP-miRNA-sncRNA	98.16 ± 0.01	0.11 ± 0.31	99.36 ± 0.01	0.70 ± 0.48
Met-SNP-miRNA-sncRNA	87.39 ± 0.29	0.08 ± 0.17	93.21 ± 0.21	0.91 ± 0.29
RNAseq-Met-SNP-miRNA-sncRNA	96.78 ± 0.01	0.19 ± 0.31	89.67 ± 0.30	0.73 ± 0.44

3.5. Effects of Incorporating Different Omics Datasets

JOIN-GCLA takes in two or more omics networks. When less than two omics datasets are available, the attention layer can be removed. Thus, in the case where one omics dataset is used, the resulting architecture has 2 graph convolution layers (1 for imaging, 1 for omics). When no omics datasets are used, the resulting architecture has 1 graph convolution layer for the multi-modal imaging data only. From **Table 4**, it is evident that almost all the models trained without omics data or only with a single omics data modality fared poorly, with MCC ranging from 0.00 to 0.13 as compared to the multi-omics models (bolded rows) with MCC ranging from 0.73 to 1.00. Furthermore, it is observed that data augmentation has greatest efficacy when multi-omics data is involved. The increase of MCC score ranged from 0.00 to 0.11 when no or one omics data was used, while the increment for multi-omics combinations ranged from 0.54 to 0.84.

3.6. Selection of the Optimal Omics Combination

In **Table 4**, results for the power set of omics combinations were shown for completeness. A principled way to arrive at the optimal combination of omics data is to perform backward elimination at the level of omics data type, based on MCC score. From the full set of omics data (RNAseq-Met-SNP-miRNA-sncRNA), $m - 1$ separate models are trained independently, each with a different subset of $m-1$ omics data types obtained by removing a different omics dataset for each model. If any of the new models produces a higher MCC score than the existing best model (initialized as the original set), it is set as the best model and the process continues recursively until it gets terminated when either no omics data is left or the current iteration of models do not perform better than the existing best model from the previous iteration. Following this procedure, Met-SNP-sncRNA was determined to be the optimal omics combination. For clearer presentation of results,

TABLE 5 | Comparison of model performance between DTI-fMRI data and fMRI data.

	DTI-fMRI		fMRI	
Omics	**Accuracy**	**MCC**	**Accuracy**	**MCC**
Model 3	100.0 ± 0.00	1.00 ± 0.00	97.15 ± 0.03	0.80 ± 0.27
Model 4	93.21 ± 0.21	0.91 ± 0.29	96.16 ± 0.04	0.71 ± 0.34
Model 5	89.67 ± 0.30	0.73 ± 0.44	97.43 ± 0.04	0.77 ± 0.41

subsequent analyses will focus on the rows in bold in **Table 4**, which represent the best models for each number of omics combinations considered in the process of backward elimination. We adopt the following notation in the tables below: Model 3 = Met-SNP-sncRNA, Model 4 = Met-SNP-miRNA-sncRNA, Model 5 = RNAseq-Met-SNP-miRNA-sncRNA.

3.7. Effect of Using Multi-Modal Imaging Data

Table 5 shows that models using multi-modal imaging data generally results in better MCC score than models trained with uni-modal imaging data[1]. In particular, Met-SNP-sncRNA is able to achieve a MCC score of 1 across all 10 seeds, but when DTI data was dropped, the MCC score reduced to 0.80 (p-value of 0.04 when performing a *t*-test to check for identical population means). Higher MCC score was also observed for Met-SNP-miRNA-sncRNA when multi-modal imaging data was involved. While the accuracies obtained when only fMRI used seems generally higher, their lower MCC suggest that the model still tends to predict the majority class. This issue is alleviated when multi-modal imaging data are used.

3.8. JOIN-GCLA Outperforms Existing Approaches for Disease Classification

To the best of our knowledge, there has been no existing work proposed to process both multi-modal imaging and multi-omics data in a single architecture. Early methods such as Long et al. (2012) extracted features from structural and functional brain images and used a support vector machine (SVM) to perform disease classification. However, such approaches do not combine omics features. Nevertheless, a comparison will be made between JOIN-GCLA and machine learning models such as SVM and logistic regression (LR) to ascertain whether JOIN-GCLA give any advantage over these models.

Tuning of the machine learning models was performed with Optuna (Akiba et al., 2019) and the models were implemented in Python using Scikit-learn. For SVM, a linear SVM was used and the regularization parameter `C` is randomly sampled from a log uniform distribution ranging between 1×10^{-5} and 1×10^{5}. For LR, besides the regularization parameter `C` (sampled from 1×10^{-3} to 1×10^{2}), the parameter `l1_ratio` is sampled from a uniform distribution ranging between 0 and 1. The best set of model parameters across 10 trials are used to train the final model. Model performance over 10 seeds is reported in **Table 6**.

[1]Data augmentation was done by using the trained CycleGAN model to generate functional connectivity matrices only as the PPMI dataset has too few fMRI scans to generate structural connectivity matrices from. Instead, the DTI dataset is used to demonstrate effects of undersampling, as shown in **Supplementary Table S7**.

TABLE 6 | Comparison between alternative fusion approaches and JOIN-GCLA.

Model	**Modality**	**Accuracy**	**MCC**
Logistic Regression	DTI	45.07 ± 5.26	–0.10 ± 0.11
Logistic Regression	fMRI	56.84 ± 3.74	0.20 ± 0.07
Logistic Regression	DTI + fMRI	58.53 ± 4.96	0.22 ± 0.10
Support Vector Machine	DTI	46.47 ± 4.96	–0.06 ± 0.12
Support Vector Machine	fMRI	45.87 ± 4.73	0.16 ± 0.11
Support Vector Machine	DTI + fMRI	37.05 ± 8.56	0.14 ± 0.10
JOIN-GCLA, Model 3	DTI + fMRI	100.0 ± 0.00	1.00 ± 0.00
JOIN-GCLA, Model 4	DTI + fMRI	93.21 ± 0.21	0.91 ± 0.29
JOIN-GCLA, Model 5	DTI + fMRI	89.67 ± 0.30	0.73 ± 0.44

TABLE 7 | Comparison between JOIN-GCLA with alternative fusion methods.

	JOIN-GCLA		Kazi et al. (2019a)	
Omics	**Accuracy**	**MCC**	**Accuracy**	**MCC**
Model 3	100.0 ± 0.00	1.00 ± 0.00	73.71 ± 0.22	0.32 ± 0.32
Model 4	93.21 ± 0.21	0.91 ± 0.29	82.14 ± 0.17	0.18 ± 0.18
Model 5	89.67 ± 0.30	0.73 ± 0.44	77.11 ± 0.19	0.35 ± 0.27

While it is evident that the JOIN-GCLA results with multi-omics data outperforms machine learning models, comparing the results in **Table 6** with the rows in **Table 4** where no omics datasets were used, deep learning models do not seem to perform better than SVM nor logistic regression models. This is true for both cases where fMRI or DTI-fMRI datasets are used. This suggest that the good model performances seen in **Table 4** are likely contributed by the addition of omics dataset and the omics networks, rather than just the use of deep learning models in the connectome encoder. While the number of test samples involved in these 3 examples (∼55) are indeed smaller than the number of test samples used when no omics data are involved (∼115), the difference in performance is unlikely to be attributed to the difference in sample sizes between the experiments. This is supported by the result from omics combination RNAseq-SNP-miRNA-sncRNA, which still has an MCC score of 0.70 with ∼ 95 test samples, much higher than what was obtained from machine learning models despite having a similar number of test samples.

More recent works related to JOIN-GCLA include architectures that combine both imaging data and demographic information in the form of population graphs (Parisot et al., 2018; Kazi et al., 2019a). However, they do not use omics datasets. The closest architecture to JOIN-GCLA is the multi-layered parallel graph convolutional network presented in Kazi et al. (2019a). In their model, separate population graphs were built based on each demographic feature used (e.g. age, gender). Each population graph was used as the graph for a different graph convolutional network (GCN). Features from MRI, fMRI and cognitive tests were used as the node vector of the GCNs. The representations learnt by the GCNs were then fused *via* a weighted sum, with the

TABLE 8 | Ablation study of the connectome encoder on DTI-fMRI dataset.

	JOIN-GCLA		Fully-connected layer		Convolution layers	
Omics	**Accuracy**	**MCC**	**Accuracy**	**MCC**	**Accuracy**	**MCC**
Model 3	100.0 ± 0.00	1.00 ± 0.00	85.82 ± 0.11	0.42 ± 0.27	95.59 ± 0.09	0.57 ± 0.46
Model 4	93.21 ± 0.21	0.91 ± 0.29	83.69 ± 0.20	0.47 ± 0.39	88.62 ± 0.23	0.26 ± 0.34
Model 5	89.67 ± 0.30	0.73 ± 0.44	89.97 ± 0.11	0.54 ± 0.36	72.86 ± 0.32	0.23 ± 0.30

weight assigned to each GCN being a parameter learnt during model training. JOIN-GCLA is different in two key aspects: (i) our connectome encoder can incorporate multiple modalities of connectome data and (ii) our proposed attention layer is used for fusing multiple views of information. In our implementation of Kazi et al. (2019a), instead of using demographic information, POGs were used as the graph for the graph convolution layers and the connectome encoder is replaced by a fully-connected layer. **Table 7** shows that JOIN-GCLA significantly outperforms their approach of modality fusion.

3.9. Effects of Graph Convolution Layer in the Connectome Encoder

The connectome encoder in JOIN-GCLA can also be compared with other deep learning approaches by replacing the graph convolution layer with alternatives such as layers in the connectome convolutional neural network proposed by Meszlényi et al. (2017), which uses customized horizontal and vertical filters of dimensions $1 \times |ROI|$ and $|ROI| \times 1$, respectively. Such a model can accept multi-modal imaging data by treating each modality as an additional channel. Alternatively, the graph convolution layer could be simply replaced with a linear layer. Such a model will take in multi-modal imaging data by flattening the original matrices into vectors and concatenating them into one large feature vector.

From **Table 8**, it can be seen that both models with the fully-connected layer and convolution layers perform rather poorly. The connectome convolution layers does not seem to aid model performance relative to the fully connected layers. Both model performances are also inferior to the results obtained by JOIN-GCLA, as shown in **Figure 1**. A limitation of the comparison made in **Table 8** is the significantly smaller number of parameters involved in the model with the convolution layers ($\sim 30,000$) as compared to the model with the fully-connected layer and JOIN-GCLA ($\sim 200,000$). In view of this, another experiment was performed where the number of parameters in the model with convolution layers was increased by increasing the number of filters (for the convolution layer in connectome encoder) and hidden nodes (for the graph convolution layer in omics networks) such that the total number of parameters is similar to the other two models. Results shown in **Supplementary Table S5** demonstrates that the larger model using convolution layers is still outperformed by JOIN-GCLA. Thus, it is evident that the proposed method to fuse multi-modal imaging data *via* PSG helps to improve model performance.

TABLE 9 | Ablation study of the attention layer on DTI-fMRI dataset.

	JOIN-GCLA		Self-attention	
Omics	**Accuracy**	**MCC**	**Accuracy**	**MCC**
Model 3	100.0 ± 0.00	1.00 ± 0.00	99.31 ± 0.01	0.80 ± 0.42
Model 4	93.21 ± 0.21	0.91 ± 0.29	98.94 ± 0.02	0.70 ± 0.48
Model 5	89.67 ± 0.30	0.73 ± 0.44	98.46 ± 0.02	0.62 ± 0.49

3.10. Effects of Different Attention Layers for Fusing Multi-View Data

Section 3.5 demonstrated the importance of using multi-omics datasets and showed how the attention mechanism improves the final disease prediction. In this section, this will be compared with alternative approaches to fuse the representations learnt from each omics network. One baseline for comparison is to use self-attention, instead of the customized formulation of the attention mechanism proposed in Section 2.1.4. **Table 9** shows that our proposed attention layer performs better than self-attention.

3.11. Model Interpretability

The performance of models with graph convolution layers is highly dependent on the graph used (Parisot et al., 2018; Cosmo et al., 2020). This warrants the need to analyse the PSG used in the connectome encoder and POGs used in the omics networks. Additionally, our proposed method to construct the attention scores allows for greater interpretability into the models decision from the weights assigned to the intermediate representations produced from the omics networks when predicting HC or PD.

3.11.1. Imaging Population Scan Network Distributions

The number of scans considered in the PSG vary according to the omics combinations used in the JOIN-GCLA model. As seen in **Figure 2**, the PSGs have similar distributions, with most values being around 2.0 with a smaller peak around 3.0. Thus, they are not likely to explain the difference in model performances when the same imaging modalities are used, as shown in **Table 4**.

3.11.2. Omics Population Graph Distributions

Figure 3 shows the distributions of POGs. These are generated by taking the lower triangular of the POG (which is symmetric) and producing kernel density plots for each omics dataset. miRNA and Met have very high values, indicating that most subjects share very similar data. While sncRNA and RNAseq has a longer left

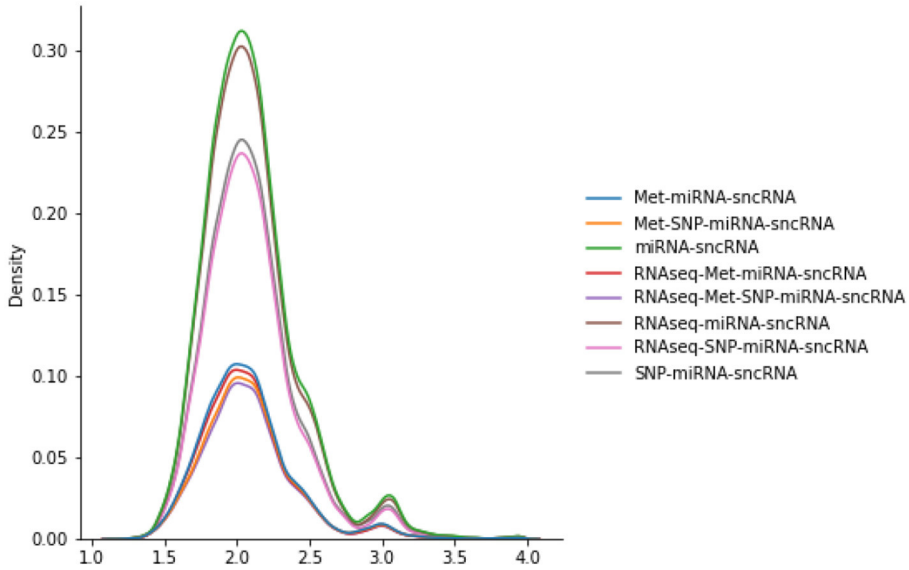

FIGURE 2 | Distributions of various PSGs for DTI-fMRI data, used in the connectome encoder.

tail, SNP has a different distribution: most of the data range from 0.2 to 0.4, indicating very little similarity between subjects. When WGCNA is applied, Met and SNP clearly have very different distributions from the rest, with a majority of values being very low (below 0.3). On the other hand, miRNA and sncRNA still have most of the values above 0.6. RNAseq has many values close to 0, but also a significant amount of values spread across the range of 0 and 1.

3.11.3. Attention Weights

JOIN-GCLA provides model interpretability in the form of attention matrices with shape $N \times L^{(4)}$. In this regard, an existing method (Kazi et al., 2019a) provides a scalar value for each view. JOIN-GCLA goes further to show which view is being focused on when predicting a certain class. **Figure 4A** shows the attention matrix for the omics combination SNP-miRNA, which was one of the omics combinations with high MCC score. SNP has a slightly higher weight in both the cases when the model predicts HC or PD. Thus, it could be inferred that the high performance of SNP-miRNA was due to the attention mechanism's focus on SNP. Similarly, **Figure 4B** shows the attention matrix for SNP-miRNA-sncRNA (i.e. sncRNA is added), which had an MCC of 0.9. While the attention scores when predicting PD (the majority class) are now equally spread, the attention scores when predicting HC was heavily weighted toward SNP.

Another set of examples is presented in **Figures 4C,D**—with both cases having an MCC of 0.9. Met has the highest weight when predicting HC, but when miRNA was added, the attention weights are slightly more distributed between Met and SNP. Also, SNP has the highest attention score when predicting PD. It could be inferred from these attention matrices that while SNP is evidently the most important omics modality when performing disease prediction, Met contributes to the high performance too especially when predicting HC.

Overall, it can be seen that when predicting PD (majority class), the attention scores tend to focus on SNP, but it could still be equally distributed. However, when predicting HC (minority), focusing on Met (or SNP, when Met is not present) helps to improve model performance. These insights, which are more detailed than (Kazi et al., 2019a), are only possible with the use of JOIN-GCLA and our proposed attention layer.

4. DISCUSSION

Overall, our results demonstrated that the combination of connectome encoder, omics networks and the customized attention layer is essential for JOIN-GCLA to work well and provide better model interpretability. From the above experiments, it is evident that our proposed architecture, JOIN-GCLA, was the best performing model. Past works have demonstrated that is it not possible to perform disease classification successfully by solely using DTI data (Prasuhn et al., 2020). Our results in **Table 6** support this finding and we went further to demonstrate that disease classification can be done well if imaging and omics datasets are used simultaneously. However, datasets with both multi-modal imaging and multi-omics data are typically small. Thus, deep neural network models have to be small. JOIN-GCLA is made as lean as possible with only 1 graph convolution layer in the connectome encoder and each omics network. The number of hidden nodes is kept small as well. In the case of JOIN-GCLA, the number of parameters, as seen in **Supplementary Table S4**, is large in this example as the flattened correlation matrix from imaging data is used as the feature vector. However, feature vectors can be any arbitrary data of interest and

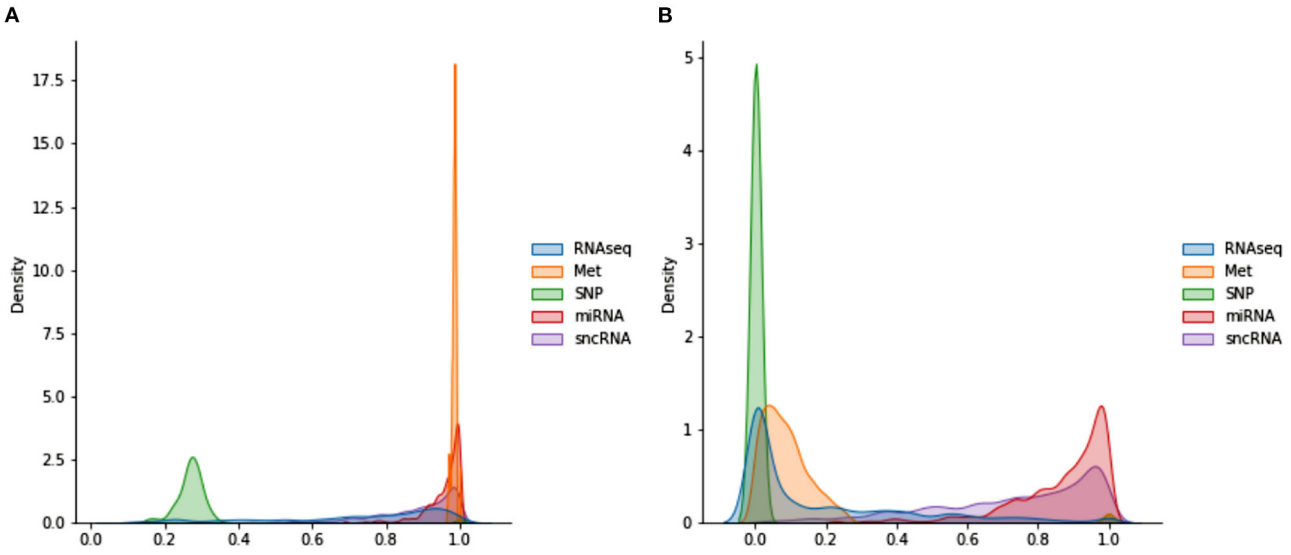

FIGURE 3 | Distributions of POGs used in omics networks **(A)** before WGCNA **(B)** after WGCNA.

thus the number of parameters could be reduced significantly, especially when dealing with smaller datasets.

It was shown in **Table 8** that PSG was essential for better model performance. While JOIN-GCLA gave the best performance, there are also other important considerations such as the scalability of the model. For instance, using convolution layers instead allows for multiple connectivity matrices to be combined without a huge increase in the number of parameters as additional modalities simply increases the number of input channels. However, this comes with the limitation that connectivity matrices of the same size have to be used (i.e. same brain atlas). JOIN-GCLA is also able to merge multiple modalities *via* PSG, but if the imaging data has to be used as feature vectors (in the form of vectorised connectivity matrices), the the number of parameters increases significantly as more modalities are included, as seen in **Supplementary Table S2**. Thus, the model with convolution layer in the connectome encoder is most suitable for small datasets where overfitting is a concern, while JOIN-GCLA is the best choice if low-dimensional feature vectors are used.

We demonstrated the feasibility of incorporating multi-omics datasets into the model *via* the use of omics networks. As seen in **Table 3**, omics datasets often have a huge number of features, even more than imaging data. Thus, it is not feasible to use the entire set of omics features as feature vectors. Instead, the use of POGs allowed information from multi-omics datasets to be included into the modeling process. A population graph built from an omics dataset is used as the input graph for the graph convolution layer and fusion between the omics data and the representations of the imaging data learnt by the connectome encoder (in the form of feature vectors) happens in this graph convolution layer. Such an approach scales up well with minimal increase of parameters, as seen in **Supplementary Tables S3, S4**. Notably, the best model performances were obtained when 3 multi-omics datasets were used: DNA Methylation, SNP and sncRNA.

The attention layer performs a key role in combining the interim predictions from each omics network and producing a final decision. Besides performing better than baseline attention methods such as self-attention as seen in **Table 9**, our proposed approach ensures that an attention matrix of shape $N \times L^{(4)}$ is generated, providing greater model interpretability as seen from **Figure 4**. This has highlighted the relative importance of SNP and DNA Methylation in distinguishing PD patients from healthy controls.

These results were only achieved after data augmentation was introduced, as shown in **Table 4**. This is largely attributed to the data imbalance that exists in the PPMI dataset, with PD scans forming the majority of the data as seen in **Supplementary Table S1**. By comparing **Table 4** with **Supplementary Table S6**, it is possible to observe the effects of gradually introducing more data augmentation to the DTI-fMRI dataset. When only 100 samples was added (majority class taking up 74% of the dataset), model performance did not change much as compared to the original baseline (with no augmented data, majority class takes up 93% of the dataset). But when the imbalance was further reduced by adding 200 samples (reducing the imbalance to 61%), model performance started to improve, but still significantly poorer than the performance obtained when all 339 samples were added to the dataset (resolving the imbalance, 53.3%). Since the best model performance was obtained when the data imbalance is resolved, it is evident that data augmentation is another key aspect needed to perform disease classification on the PPMI dataset successfully.

We have used the CycleGAN architecture for producing additional scans to be augmented to the original dataset. The main motivation of using CycleGAN is to overcome the limitations of the existing approaches for tackling data

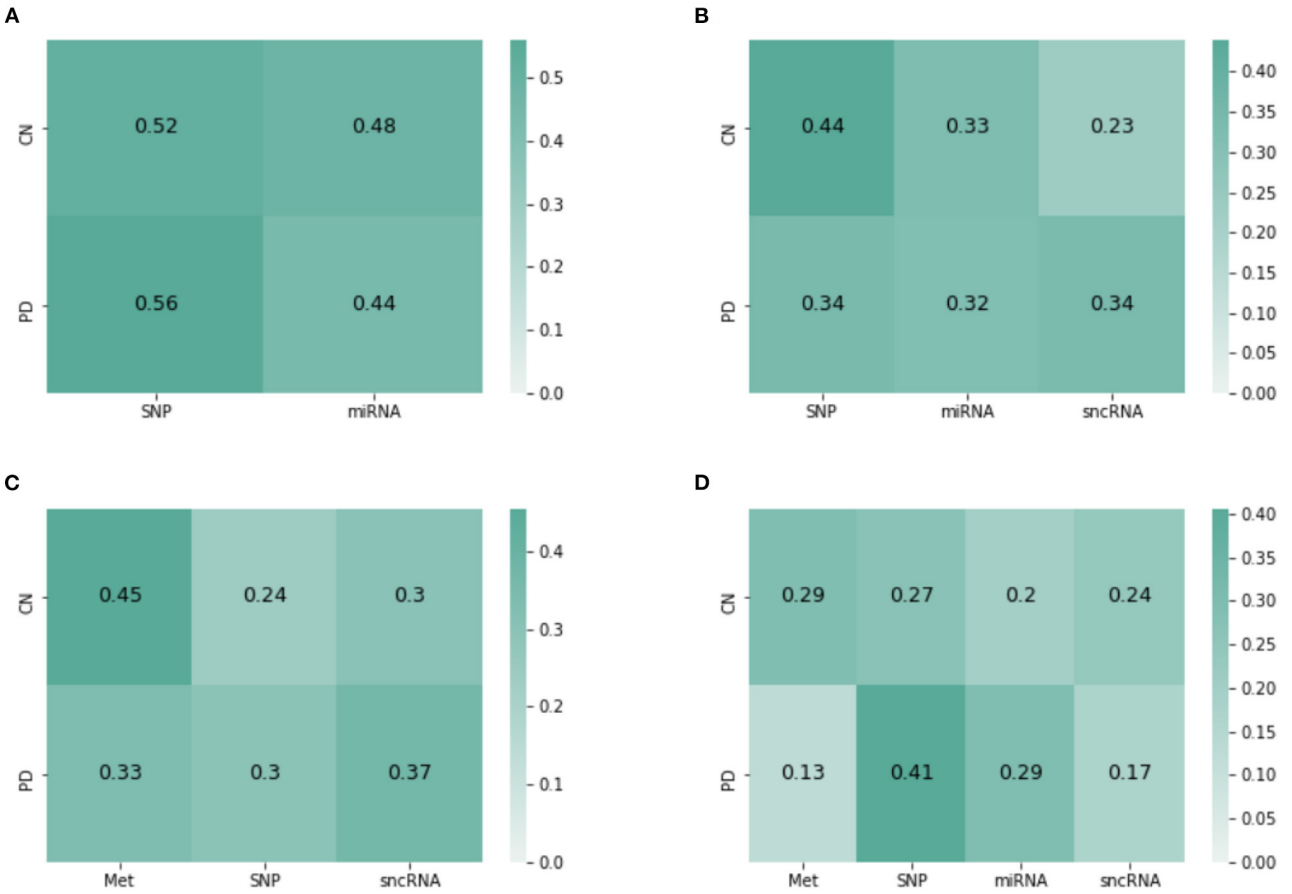

FIGURE 4 | Attention matrices from JOIN-GCLA for the omics combination of **(A)** SNP-miRNA, **(B)** SNP-miRNA-sncRNA, **(C)** Met-SNP-sncRNA **(D)** Met-SNP-miRNA-sncRNA.

imbalance. As seen in **Table 4** (without augmentation), class weighting applied to the loss function did not improve model performance at all, likely due to the extreme imbalance in the dataset. Undersampling is not a viable approach when dealing with small dataset, as demonstrated on an experiment in **Supplementary Table S7** where the DTI dataset was undersampled—while the imbalance was well addressed (as seen in **Supplementary Table S1**), model performance did not improve significantly. On the other hand, oversampling on the DTI-fMRI dataset did help to improve model performance to a level similar to what was obtained from the CycleGAN-augmented dataset (comparing **Table 4** with **Supplementary Table S8**).

While both oversampling and CycleGAN generates data that can be attributed to a specific subject (hence making it possible to link it to a genetic dataset, unlike synthetic data generation algorithms such as SMOTE and ADASYN), oversampling merely duplicates the existing dataset. CycleGAN-generated data are not just another repeated data sample in the existing dataset. However, when compared to the results obtained from oversampling, the marginal benefit introduced by the use of CycleGAN might not always justify the additional complexity added. Below, we present details on the data produced by CycleGAN to propose possible reasons for these observations.

Examples of the data generated by the CycleGAN architecture are shown in **Supplementary Figure S1**. Although generated scans have low mean squared errors (MSE) (approximately 0.03 when compared to actual functional connectivity matrices from the same pair ; approximately 0.5 for DTI), they do not have the same variability. On examining all the other generated matrices, it is evident that the synthetic connectomes have very slight differences and seem to capture patterns that exist across most scans, while missing out on more subtle variations that exist in functional connectivity matrices. These variations are visually stark (for fMRI), but it might not have been captured by the GAN as the overall numerical significance is not great (since the MSE achieved is rather low already). This issue is likely to be alleviated with the introduction of more data (Karras et al., 2020). Additionally, several architecture changes to the original CycleGAN were attempted to improve the variability of data generated, including adding Edge-to-edge (E2E) layers from Kawahara et al. (2017) as the first layer of the generator and discriminator and reducing the number of residual blocks (from 9 to 3) and number of filters (from 256 to 32) in the CycleGAN architecture. However, from the results

in **Supplementary Table S9**, the best model was still the original CycleGAN architecture.

In this study, data augmentation was limited to healthy controls as the goal was to resolve the data imbalance in the PPMI dataset. Having demonstrated the feasibility of this approach, further studies could explore the use of CycleGAN to generate connectomes for various neurodegenerative disorders. Using GANs to generate connectome datasets is at a nascent stage: a recent work used GAN to generate functional connectivity matrices for schizophrenia and major depressive disorder patients (Zhao et al., 2020). Our work has extended the application of GANs on connectome datasets to multi-modal settings and the results demonstrated that using relatively large connectome datasets (∼1,000 samples) to train CycleGAN is still not yet sufficient to significantly outperform oversampling as rather similar matrices are produced by the GAN. However, since CycleGAN is capable of learning from unpaired data, this is not an unsurmontable problem and future studies should consider using more data when training CycleGAN architectures to augment multi-modal connectome datasets. If obtaining more data is not feasible, oversampling presents a limited but effective approach for data augmentation.

5. CONCLUSION

We have proposed a new architecture, JOIN-GCLA, which is able to model multi-modal imaging data and multi-omics datasets simultaneously. Through the experiments, it has been demonstrated that the best performing data combination utilizes both multi-modal imaging data (DTI, fMRI) and multi-omics datasets (SNP, DNA Methylation and sncRNA). While several combinations of imaging and omics data led to very high model performance, this must be seen in the light of the small test dataset size available in the PPMI dataset. Our experiments on the PPMI dataset showed that JOIN-GCLA can work well, but this should be further tested on larger datasets that have both multi-modal imaging data and multi-omics datasets. Examples of such sources of data would be the Alzheimer's Disease Neuroimaging Initiative, UK Biobank and also future versions of PPMI, which has recently expanded its data collection with a few thousand more data samples to be expected by year 2023.

One possible area of future work is to perform decoding. Given a trained neural network model, it has been demonstrated that saliency scores can be computed to identify important features that contributed most to the model's decision (Gupta et al., 2021). While such an approach cannot be simply applied to JOIN-GCLA due to the attention layer, novel methods could be developed to weigh the saliency scores by the attention scores for each view. This could be explored as a follow-up work after this paper.

Another direction for further research on combining neuroimaging and omics datasets is the use of transformers. While originally proposed for natural language processing (Vaswani et al., 2017), it has been demonstrated to work on images too (Dosovitskiy et al., 2020), motivating recent works on using transformer-based architectures for multi-modal settings (Hu and Singh, 2021; Kim et al., 2021). One limitation of such models is their reliance on pre-training from large datasets (Dosovitskiy et al., 2020). Modifying transformers to work on small datasets is still an open area of research (Lee et al., 2021). This could explain the paucity of works on using transformers for neuroimaging datasets (especially on connectivity matrices). Recent works on the use of transformers utilizes raw fMRI signals (Nguyen et al., 2020; Malkiel et al., 2021). Notably, one of the key findings in Malkiel et al. (2021) is the need for pre-training for best model performance. Addressing this issue for connectome datasets could be possible with the use of larger datasets such as UK Biobank.

While this paper focuses on PD classification using multi-modal imaging data (DTI, fMRI) and multi-omics data (miRNA, DNA methylation, RNAseq, sncRNA, SNP), JOIN-GCLA can be easily extended to other diseases, omics modalities and imaging modalities too. For instance, diseases such as ADHD could benefit from the use of multi-modal imaging and multi-omics data (Klein et al., 2017) and the problem of limited multi-modal data could be addressed by using CycleGAN to generate more data. However, our results suggest that such approaches will need large amounts of data (more than 1,000 data points) to train the CycleGAN architecture.

In sum, the JOIN-GCLA architecture makes it possible to analyse multi-modal imaging data along with multi-omics datasets. Our proposed architecture alleviates the issue of high dimensionality of imaging and omics data by incorporating them in graph convolution layers in the form of PSG and POG, respectively. This enables multi-scale analysis, incorporating both macro-scale imaging data with micro-scale genomics analysis, to be conducted. The greater interpretability provided by JOIN-GCLA's attention matrices gives greater insight into the relative importance of the omics datasets taken into consideration, potentially revealing more novel insights for complex neurodegenerative diseases in future studies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: Dataset provided by PPMI can be found in https://www.ppmi-info.org/. Information about HCP S1200 are presented in https://www.humanconnectome.org/study/hcp-young-adult/document/1200-subjects-data-release and the dataset can be downloaded from https://db.humanconnectome.org/. Download links for the AOMIC dataset are provided at https://nilab-uva.github.io/AOMIC.github.io/.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was partially supported by AcRF Tier-1 grant RG116/19 and AcRF Tier-2 grant MOE T2EP20121-0003 of Ministry of Education, Singapore.

ACKNOWLEDGMENTS

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/access-dataspecimens/download-data).
For up-to-date information on the study, visit ppmi-info.org. PPMI a public-private partnership is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including 4D Pharma, AbbVie Inc., AcureX Therapeutics, Allergan, Amathus Therapeutics, Aligning Science Across Parkinson's (ASAP), Avid Radiopharmaceuticals, Bial Biotech, Biogen, BioLegend, Bristol Myers Squibb, Calico Life Sciences LLC, Celgene Corporation, DaCapo Brainscience, Denali Therapeutics, The Edmond J. Safra Foundation, Eli Lilly and Company, GE Healthcare, GlaxoSmithKline, Golub Capital, Handl Therapeutics, Insitro, Janssen Pharmaceuticals, Lundbeck, Merck & Co., Inc., Meso Scale Diagnostics, LLC, Neurocrine Biosciences, Pfizer Inc., Piramal Imaging, Prevail Therapeutics, F. HoffmannŘLa Roche Ltd and its affiliated company Genentech Inc., Sanofi Genzyme, Servier, Takeda Pharmaceutical Company, Teva Neuroscience, Inc., UCB, Vanqua Bio, Verily Life Sciences, Voyager Therapeutics, Inc. and Yumanity Therapeutics, Inc. Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) 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.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.866666/full#supplementary-material

REFERENCES

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). "Optuna: a next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Anchorage, AK), 2623–2631.

Antonelli, L., Guarracino, M. R., Maddalena, L., and Sangiovanni, M. (2019). Integrating imaging and omics data: a review. *Biomed. Signal Process. Control*. 52, 264–280. doi: 10.1016/j.bspc.2019.04.032

Chaudhary, K., Poirion, O. B., Lu, L., and Garmire, L. X. (2018). Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin. Cancer Res*. 24, 1248–1259. doi: 10.1158/1078-0432.CCR-17-0853

Cosmo, L., Kazi, A., Ahmadi, S.-A., Navab, N., and Bronstein, M. (2020). "Latent-graph learning for disease prediction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Lima: Springer), 643–653.

Deng, B., Zhang, Y., Wang, L., Peng, K., Han, L., Nie, K., et al. (2013). Diffusion tensor imaging reveals white matter changes associated with cognitive status in patients with parkinson's disease. *Am. J. Alzheimers Dis. Other Dement*. 28, 154–164. doi: 10.1177/1533317512470207

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint* arXiv:2010.11929.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., et al. (2019). fmriprep: a robust preprocessing pipeline for functional mri. *Nat. Methods* 16, 111–116. doi: 10.1038/s41592-018-0235-4

Filippi, M., Elisabetta, S., Piramide, N., and Agosta, F. (2018). Functional mri in idiopathic parkinson's disease. *Int. Rev. Neurobiol*. 141:439–467. doi: 10.1016/bs.irn.2018.08.005

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., et al. (2013). The minimal preprocessing pipelines for the human connectome project. *Neuroimage* 80, 105–124. doi: 10.1016/j.neuroimage.2013.04.127

Gupta, S., Chan, Y. H., Rajapakse, J. C., Initiative, A. D. N., et al. (2021). Obtaining leaner deep neural networks for decoding brain functional connectome in a single shot. *Neurocomputing* 453, 326–336. doi: 10.1016/j.neucom.2020.04.152

Hernández, M., Guerrero, G. D., Cecilia, J. M., García, J. M., Inuggi, A., Jbabdi, S., et al. (2013). Accelerating fibre orientation estimation from diffusion weighted magnetic resonance imaging using gpus. *PLoS ONE* 8, e61892. doi: 10.1371/journal.pone.0061892

Hernandez-Fernandez, M., Reguly, I., Jbabdi, S., Giles, M., Smith, S., and Sotiropoulos, S. N. (2019). Using gpus to accelerate computational diffusion mri: from microstructure estimation to tractography and connectomes. *Neuroimage* 188, 598–615. doi: 10.1016/j.neuroimage.2018.12.015

Herz, D. M., Eickhoff, S. B., Løkkegaard, A., and Siebner, H. R. (2014). Functional neuroimaging of motor control in parkinson's disease: a meta-analysis. *Hum. Brain Mapp*. 35, 3227–3237. doi: 10.1002/hbm.22397

Hu, R., and Singh, A. (2021). "Unit: multimodal multitask learning with a unified transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (Montreal, QC), 1439–1449.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Jin, C., Qi, S., Teng, Y., Li, C., Yao, Y., Ruan, X., et al. (2021). Integrating structural and functional interhemispheric brain connectivity of gait freezing in parkinson's disease. *Front. Neurol*. 12, 411. doi: 10.3389/fneur.2021.609866

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., and Aila, T. (2020). Training generative adversarial networks with limited data. *arXiv preprint* arXiv:2006.06676.

Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V., Grunau, R. E., et al. (2017). Brainnetcnn: convolutional neural networks for brain networks; towards predicting neurodevelopment. *Neuroimage* 146, 1038–1049. doi: 10.1016/j.neuroimage.2016.09.046

Kazi, A., Shekarforoush, S., Kortuem, K., Albarqouni, S., Navab, N., et al. (2019a). "Self-attention equipped graph convolutions for disease prediction," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)* (Venice: IEEE), 1896–1899.

Kazi, A., Shekarforoush, S., Krishna, S. A., Burwinkel, H., Vivar, G., Wiestler, B., et al. (2019b). "Graph convolution based attention model for personalized disease prediction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Shenzhen: Springer), 122–130.

Kim, M., Kim, J., Lee, S.-H., and Park, H. (2017). Imaging genetics approach to parkinson's disease and its correlation with clinical score. *Sci. Rep*. 7, 1–10. doi: 10.1038/srep46700

Kim, W., Son, B., and Kim, I. (2021). "Vilt: vision-and-language transformer without convolution or region supervision," in *International Conference on Machine Learning* (Sydney, NSW: PMLR), 5583–5594.

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

Klein, C., and Westenberger, A. (2012). Genetics of parkinson's disease. *Cold Spring Harb. Perspect. Med*. 2, a008888. doi: 10.1101/cshperspect.a008888

Klein, M., Onnink, M., van Donkelaar, M., Wolfers, T., Harich, B., Shi, Y., et al. (2017). Brain imaging genetics in adhd and beyond-mapping pathways from gene to disorder at different levels of complexity. *Neurosci. Biobehav. Rev*. 80, 115–155. doi: 10.1016/j.neubiorev.2017.01.013

Lee, S. H., Lee, S., and Song, B. C. (2021). Vision transformer for small-size datasets. *arXiv preprint* arXiv:2112.13492.

Long, D., Wang, J., Xuan, M., Gu, Q., Xu, X., Kong, D., et al. (2012). Automatic classification of early Parkinson's disease with multi-modal mr imaging. *PLoS ONE* 7, e47714. doi: 10.1371/journal.pone.0047714

Majidinia, M., Mihanfar, A., Rahbarghazi, R., Nourazarian, A., Bagca, B., and Avci, Ç. B. (2016). The roles of non-coding rnas in Parkinson's disease. *Mol. Biol. Rep.* 43, 1193–1204. doi: 10.1007/s11033-016-4054-3

Malkiel, I., Rosenman, G., Wolf, L., and Hendler, T. (2021). Pre-training and fine-tuning transformers for fmri prediction tasks. *arXiv preprint* arXiv:2112.05761.

Marek, K., Chowdhury, S., Siderowf, A., Lasch, S., Coffey, C. S., Caspell-Garcia, C., et al. (2018). The Parkinson's progression markers initiative (ppmi)-establishing a pd biomarker cohort. *Ann. Clin. Transl. Neurol.* 5, 1460–1477. doi: 10.1002/acn3.644

Markello, R. D., Shafiei, G., Tremblay, C., Postuma, R. B., Dagher, A., and Misic, B. (2021). Multimodal phenotypic axes of parkinson's disease. *NPJ Parkinsons Dis.* 7, 1–12. doi: 10.1038/s41531-020-00144-9

Meszlényi, R. J., Buza, K., and Vidnyánszky, Z. (2017). Resting state fmri functional connectivity-based classification using a convolutional neural network architecture. *Front. Neuroinform.* 11, 61. doi: 10.3389/fninf.2017.00061

Miranda-Morales, E., Meier, K., Sandoval-Carrillo, A., Salas-Pacheco, J., Vázquez-Cárdenas, P., and Arias-Carrión, O. (2017). Implications of dna methylation in parkinson's disease. *Front. Mol. Neurosci.* 10, 225. doi: 10.3389/fnmol.2017.00225

Nachev, P., Kennard, C., and Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nat. Rev. Neurosci.* 9, 856–869. doi: 10.1038/nrn2478

Nalls, M. A., McLean, C. Y., Rick, J., Eberly, S., Hutten, S. J., Gwinn, K., et al. (2015). Diagnosis of parkinson's disease on the basis of clinical and genetic classification: a population-based modelling study. *Lancet Neurol.* 14, 1002–1009. doi: 10.1016/S1474-4422(15)00178-7

Nguyen, S., Ng, B., Kaplan, A. D., and Ray, P. (2020). "Attend and decode: 4d fmri task state decoding using attention models," in *Machine Learning for Health* (Montreal, QC: PMLR), 267–279.

Parisot, S., Ktena, S. I., Ferrante, E., Lee, M., Guerrero, R., Glocker, B., et al. (2018). Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimers disease. *Med. Image Anal.* 48, 117–130. doi: 10.1016/j.media.2018.06.001

Prasuhn, J., Heldmann, M., Münte, T. F., and Brüggemann, N. (2020). A machine learning-based classification approach on parkinson's disease diffusion tensor imaging datasets. *Neurol. Res. Pract.* 2, 1–5. doi: 10.1186/s42466-020-00092-y

Price, C. C., Tanner, J., Nguyen, P. T., Schwab, N. A., Mitchell, S., Slonena, E., et al. (2016). Gray and white matter contributions to cognitive frontostriatal deficits in non-demented parkinson's disease. *PLoS ONE* 11, e0147332. doi: 10.1371/journal.pone.0147332

Routier, A., Burgos, N., Díaz, M., Bacci, M., Bottani, S., El-Rifai, O., et al. (2021). Clinica: an open-source software platform for reproducible clinical neuroscience studies. *Front. Neuroinform.* 15, 689675. doi: 10.3389/fninf.2021.689675

Snoek, L., van der Miesen, M. M., Beemsterboer, T., van der Leij, A., Eigenhuis, A., and Scholte, H. S. (2021). The amsterdam open mri collection, a set of multimodal mri datasets for individual difference analyses. *Sci. Data* 8, 1–23. doi: 10.1038/s41597-021-00870-6

Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. (2020). Multi-omics data integration, interpretation, and its application. *Bioinform. Biol. Insights*. 14, 1177932219899051. doi: 10.1177/1177932219899051

Tessitore, A., Cirillo, M., and De Micco, R. (2019). Functional connectivity signatures of parkinson's disease. *J. Parkinsons Dis.* 9, 637–652. doi: 10.3233/JPD-191592

Tran, J., Anastacio, H., and Bardy, C. (2020). Genetic predispositions of parkinson's disease revealed in patient-derived brain cells. *NPJ Parkinsons Dis.* 6, 1–18. doi: 10.1038/s41531-020-0110-8

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). "Attention is all you need," in *Advances in Neural Information Processing Systems* (Long Beach, CA), 5998–6008.

Wang, C., Chen, L., Yang, Y., Zhang, M., and Wong, G. (2019). Identification of potential blood biomarkers for parkinson's disease by gene expression and dna methylation data integration analysis. *Clin. Epigenet.* 11, 1–15. doi: 10.1186/s13148-019-0621-5

Watson, C. N., Belli, A., and Di Pietro, V. (2019). Small non-coding rnas: new class of biomarkers and potential therapeutic targets in neurodegenerative disease. *Front. Genet.* 10, 364. doi: 10.3389/fgene.2019.00364

Yadav, S. K., Kathiresan, N., Mohan, S., Vasileiou, G., Singh, A., Kaura, D., et al. (2016). Gender-based analysis of cortical thickness and structural connectivity in parkinson's disease. *J. Neurol.* 263, 2308–2318. doi: 10.1007/s00415-016-8265-2

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4, 17. doi: 10.2202/1544-6115.1128

Zhang, L., Lv, C., Jin, Y., Cheng, G., Fu, Y., Yuan, D., et al. (2018). Deep learning-based multi-omics data integration reveals two prognostic subtypes in high-risk neuroblastoma. *Front. Genet.* 9, 477. doi: 10.3389/fgene.2018.00477

Zhao, J., Huang, J., Zhi, D., Yan, W., Ma, X., Yang, X., et al. (2020). Functional network connectivity (fnc)-based generative adversarial network (gan) and its applications in classification of mental disorders. *J. Neurosci. Methods* 341, 108756. doi: 10.1016/j.jneumeth.2020.108756

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision* (Venice: IEEE), 2223–2232.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 01 July 2022
doi: 10.3389/fnins.2022.866735

Graph Empirical Mode Decomposition-Based Data Augmentation Applied to Gifted Children MRI Analysis

Xuning Chen[1], Binghua Li[1], Hao Jia[1], Fan Feng[1], Feng Duan[1], Zhe Sun[2]*, Cesar F. Caiafa[1,3] and Jordi Solé-Casals[1,4,5]**

[1] Department of Artificial Intelligence, Nankai University, Tianjin, China, [2] Computational Engineering Applications Unit, Head Office for Information Systems and Cybersecurity, RIKEN, Saitama, Japan, [3] Instituto Argentino de Radioastronomía, Consejo Nacional de Investigaciones Científicas y Técnicas – Centro Científico Tecnológico La Plata/Comisión de Investigaciones Científicas – Provincia de Buenos Aires/Universidad Nacional de La Plata, Villa Elisa, Argentina, [4] Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom, [5] Data and Signal Processing Research Group, University of Vic-Central University of Catalonia, Vic, Spain

OPEN ACCESS

Edited by:
Sergey M. Plis,
Georgia State University,
United States

Reviewed by:
Cota Navin Gupta,
Indian Institute of Technology Guwahati, India
Guang Ling,
Wuhan University of Technology, China

***Correspondence:**
Feng Duan
duanf@nankai.edu.cn
Zhe Sun
zhe.sun.vk@riken.jp
Jordi Solé-Casals
jordi.sole@uvic.cat

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 31 January 2022
Accepted: 27 May 2022
Published: 01 July 2022

Citation:
Chen X, Li B, Jia H, Feng F, Duan F, Sun Z, Caiafa CF and Solé-Casals J (2022) Graph Empirical Mode Decomposition-Based Data Augmentation Applied to Gifted Children MRI Analysis. Front. Neurosci. 16:866735. doi: 10.3389/fnins.2022.866735

Gifted children and normal controls can be distinguished by analyzing the structural connectivity (SC) extracted from MRI data. Previous studies have improved classification accuracy by extracting several features of the brain regions. However, the limited size of the database may lead to degradation when training deep neural networks as classification models. To this end, we propose to use a data augmentation method by adding artificial samples generated using graph empirical mode decomposition (GEMD). We decompose the training samples by GEMD to obtain the intrinsic mode functions (IMFs). Then, the IMFs are randomly recombined to generate the new artificial samples. After that, we use the original training samples and the new artificial samples to enlarge the training set. To evaluate the proposed method, we use a deep neural network architecture called BrainNetCNN to classify the SCs of MRI data with and without data augmentation. The results show that the data augmentation with GEMD can improve the average classification performance from 55.7 to 78%, while we get a state-of-the-art classification accuracy of 93.3% by using GEMD in some cases. Our results demonstrate that the proposed GEMD augmentation method can effectively increase the limited number of samples in the gifted children dataset, improving the classification accuracy. We also found that the classification accuracy is improved when specific features extracted from brain regions are used, achieving 93.1% for some feature selection methods.

Keywords: GEMD, MRI, gifted children, structural connectivity, BrainNetCNN

INTRODUCTION

Intelligence can be seen as the ability to recognize and understand reality and use knowledge and experience to solve problems such as memory, observation, imagination, thinking, and judgment. Gifted children are regarded to have higher intelligence and perform better in attention, language, mathematics, verbal working memory, shifting, and social problem-solving (Bucaille et al., 2022). At the same time, gifted children demonstrate high working memory capacity and more effective executive attention (Aubry et al., 2021). They also have significant differences in cognitive flexibility function and problem-solving and reasoning (Rocha et al., 2020).

Gifted children have higher intelligence and learn faster than others, probably due to differences in neurophysiology (Gross, 2006). Neurological differences mean that gifted children may experience neurodevelopmental trajectories different from normal children, leading to a greater connection of neuronal pathways (Navas-Sánchez et al., 2014). Gifted children have larger subcortical structures and more robust white matter microstructural organization between those structures in regions associated with explicit memory (Kuhn et al., 2021). They are also characterized by highly developed functional interactions between the right hemisphere and excellent cognitive control of the prefrontal cortex, enhanced frontoparietal cortex, and posterior parietal cortex (Wei et al., 2020). Ma et al. found that gifted children have network topological properties of high global efficiency and high clustering with a low wiring cost and a higher level of local connection density (Ma et al., 2017). Gifted children's structural brain network has a more integrated and versatile topology than normal children (Solé-Casals et al., 2019).

Based on previous work on the brain neuroscience of gifted children, we believe that it is significant to identify gifted children through the structure of their brains. In the past decades, many neuroscientists have tried to understand the brain mechanisms and proposed many types of neuroimaging techniques, such as magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), and diffusion tensor imaging. In recent years, deep learning algorithms have achieved good results in processing these types of signals. Abdelaziz Ismael et al. proposed an enhanced deep learning approach, residual networks, for brain cancer MRI images classification and achieved 99% accuracy (Abdelaziz Ismael et al., 2020). Sarraf et al. used convolutional neural network (CNN) architectures Lenet-5 and GoogleNet to classify fMRI data of Alzheimer's disease subjects and normal controls, and the accuracy of the test dataset reached 96.85% (Sarraf and Tofighi, 2016). The BrainNetCNN is proposed to predict clinical neurodevelopmental outcomes by brain networks (Kawahara et al., 2017). It utilizes structural brain networks' topological locality to create edge-to-edge (E2E), edge-to-node (E2N), and node-to-graph (N2G) convolutional filters, which makes it perform well on human brain data classification. Leonardsen et al. proposed that neural network is able to identify subject brain from its MRI (Leonardsen et al., 2022).

The deep learning technology is notable for its impressive performance and generalization capability, but the number of effective samples in the medical imaging dataset is usually small, leading to performance degradation. The training model needs large amount of data to avoid overfitting (Caiafa et al., 2020). However, obtaining enough MRI data is not easy. The acquisition and preprocessing of brain data are more difficult than image and voice data, for example. It is difficult to find gifted children in our daily life. The number of gifted children is small, especially those whose IQ test score is higher than 140. In this work, we use a sample of 29 children, from which the MRI was obtained. The brain was parcellated into 308 regions and from each region 7 morphometric features were extracted. Hence, we have a total of 2,156 features per subject (7 morphological features by 308 brain regions). Training a model in such a small and high-dimensional MRI dataset is complicated. Therefore, we focus on MRI data augmentation to improve model training. Data augmentation has proven to be useful in MRI, improving the accuracy of schizophrenia classification by 5% (7–8% relative improvement using augmentation) (Ulloa et al., 2015). Also, Nguyen et al. proposed a data augmentation method synthesizing a new fMRI image by performing a T1-based coregistration to another subject's brain in native space. This method was tested on antidepressant treatment response fMRI and demonstrated a 26% improvement in predicting response using augmented images (Nguyen et al., 2020). Previous work proves that increasing the amount of neuroimaging data through an appropriate data augmentation method can significantly improve the accuracy of deep learning classification.

In our MRI dataset, we propose to use a data decomposition method, graph empirical mode decomposition (GEMD) (Tremblay et al., 2014). GEMD is an adaptation to graph signals of the well-known empirical modal decomposition (EMD) (Huang et al., 1998). EMD has some variants, such as GEMD, masking EMD, ensemble-EMD (EEMD), and multivariate EMD (MEMD). Masking EMD, EEMD, and MEMD can primarily alleviate the mode mixing problem, and masking EMD and MEMD can perform spatiotemporal reconstruction of active sources (Muñoz-Gutiérrez et al., 2018). GEMD improves many aspects of the critical points of EMD, namely, extrema, interpolation, and stopping criterion (Tremblay et al., 2014). Because a parcellation of 308 brain regions is used, which can help to build a brain region connection graph, GEMD is the best choice for our work, as we will base our data augmentation on the decomposition-recombination strategy first presented in Dinarès-Ferran et al. (2018) for EEG signals. To our knowledge, this is the first time this technique has been used on MRI data. To compare the results of the proposed method, we also generate artificial samples through a more classical approach, the synthetic minority over-sampling technique (SMOTE) (Chawla et al., 2002).

In this work, the BrainNetCNN is used as a deep learning classifier. The main motivation for using a deep learning method is that the MRI data can then be fed directly into the model without the need for any feature selection method. This is an important aspect to keep in mind as feature selection methods are usually very database-dependent, and the results could drop if the database is changed. We train the BrainNetCNN for the classification task, showing that a well-trained classification model can increase the classification accuracy from 55.7 to 78% when using artificial data. Moreover, in Zhang et al. (2021), a hybrid selection method of morphological features and brain regions on the same gifted children dataset was derived. They used a completion method, simultaneous tensor decomposition, and completion (STDC), for outlier correction. After tensor completion, several feature selection methods were applied to the training set to explore which morphometric features and brain regions could perform better in the classification step. Based on their methodology, we used GEMD to generate artificial data on Zhang et al.'s work to achieve an accuracy of 93.1% on the F-score (FS), combined feature selection, and rank FS method.

The rest of the article is organized as follows. the materials and methods and the details of the experiments are introduced. Then,

the experimental results are discussed, followed bydiscussion. Finally, the conclusions are summarized.

MATERIALS AND METHODS

The overall experimental process is shown in **Figure 1**. In this section, we introduce the six parts in order. The details of the data are first described. Then, we show the brain region atlas and the morphometric features. After that, the basic algorithm principle of GEMD will be provided. Then, the data augmentation with GEMD is introduced. The following is the structural connectivity (SC) analysis, which converts MRI images into a correlation matrix. Finally, we introduce a deep learning network, the BrainNetCNN, as a classifier.

Gifted Children MRI Dataset

The MRI dataset of gifted children contains 29 healthy, right-handed male subjects without neurological diseases (Solé-Casals et al., 2019). We refer to this dataset as the UVic-gifted children dataset (UVic-GC dataset). There is no significant age difference between the two groups. Gifted children have a high IQ and outstanding performance in various tasks such as spatial, numerical, reasoning, verbal, and memory (Gras et al., 2010). The criteria for gifted group included having an IQ in the very superior range ($\geq$140). Gifted children also had a performance above the 90th percentile in three of the following aptitudes, namely, spatial, numerical, abstract reasoning, verbal reasoning, and memory. More details on the dataset can be found in Solé-Casals et al. (2019). **Table 1** summarizes the details of the dataset. Using similar procedure and scanning parameters, all participants underwent examinations in a 3 T MRI scanner (Magnetom Trio Tim, Siemens Medical Systems, Germany). The raw (anonymized) MRI data are available in the OpenNeuro repository (https://openneuro.org/datasets/ds001988).

Brain Region Atlas and Morphometric Features

In our study, the brain is divided into 308 cortical regions following previous work (Romero-Garcia et al., 2012). The parcellation atlas is based on the Desikan-Killiany Atlas (68 cortical areas). Each area defined in the Desikan-Killiany atlas is subdivided into spatially contiguous areas through a backtracking algorithm available in FreeSurfer (Desikan et al., 2006). The size of each area is approximately equal to 500 mm^2.

The original feature matrix includes seven morphological features measured in each of the 308 brain regions. **Figure 2** shows the morphological features such as gray matter volume, cortical thickness, surface area, intrinsic curvature, mean curvature, curvature index, and fold index.

Graph Empirical Mode Decomposition

Empirical modal decomposition can decompose a signal into a set of intrinsic mode functions (IMFs), each covering different frequency bands by interpolating the extremes in the time series (Huang et al., 1998). The IMFs have two characteristics, namely, (1) the number of its zero crossings must be equal or differ up to one compared to its number of extrema and (2) IMFs' upper and lower envelopes must be symmetric to zero. When all the IMFs of the original signal are extracted, the iterative process is terminated. GEMD is an adaptation of the classical EMD for graph signals (Tremblay et al., 2014). It improves many aspects of the critical points of EMD, namely, extrema, interpolation, and stopping criterion.

For the graph creation, the set of N regions is used as nodes for the graph. A weighted graph parameter δ is used to define edges in the graph. Only pairs of regions (i, j) at a distance $d_{i,j}$, shorter than δ, are connected by an edge, with weight $w_{i,j} = \exp(-d^2_{i,j}/2\delta^2)$. The distance $d_{i,j}$ is the Euclidean distance in the features space. In that case, we get a graph $G = (N, E)$, where E is the set of edges. The adjacency matrix A, which contains all the weights $w_{i,j}$ connecting the nodes, is also needed. We use the 3D coordinate points of 308 brain regions to calculate the adjacency matrix for the 308 brain regions graph.

For the definition of local extrema, a node n will be a local minimum (or maximum) if for all its neighbors in G, $x(n) < x(m)$ [or $x(n) > x(m)$, where $x(n)$ and $x(m)$ represent the value of one of the features in the n_{th} and m_{th} brain regions]. Once the extremes have been obtained, the graph signal is interpolated to get the lower and upper graph envelopes needed to derive the IMFs.

To maintain the hypothesis-free nature of the classical EMD method, interpolation is regarded as a discrete partial differential equation on the graph (Grady and Schwartz, 2003). As the envelope is a slowly changing component, the interpolation signal s needs to minimize the total graph change, $s^{'}Ls$, where L is the graph Laplacian matrix under the constraint that the graph signal value of the known vertex remains unchanged. Let K denote the set of vertices of the known graph signal, and U denote the set of unknown vertices. Then, to calculate the new, interpolated, graphical signals, we need to solve the following equation minimize $s^{'}Ls$ subject to:

$$s(K) = x(K) \tag{1}$$

Through simple rearrangement of vertices, s can be rewritten as $s^{'} = [s^{'}_K \; s^{'}_U]$ in its equivalent vector expression, where s_K and s_U are the vector representations of $s(K)$ and $s(U)$, and the rearranged Laplacian matrix $L = \begin{bmatrix} L_K & R \\ R' & L_U \end{bmatrix}$. Finally, the graph interpolation is a Dirichlet problem on the graph, and its solution depends on the following linear equation (Kalaganis et al., 2020):

$$L_U S_U = -R s_k \tag{2}$$

We refer the reader to Grady and Schwartz (2003) and Kalaganis et al. (2020) for a detailed explanation of the graph interpolation method. With the mentioned elements, the sifting process can be modified easily. We set the parameter of the stopping criterion, which was defined in Tremblay's work (Tremblay et al., 2014), as follows: stop the loop (steps 4–8 in the following algorithm) as soon as the energy of the average envelope z (computed in the step 6) is lower than the energy of the analyzed signal x_i divided by 1,000.

After defining the graph extremum and interpolation process, the classic EMD algorithm can easily be extended to graph

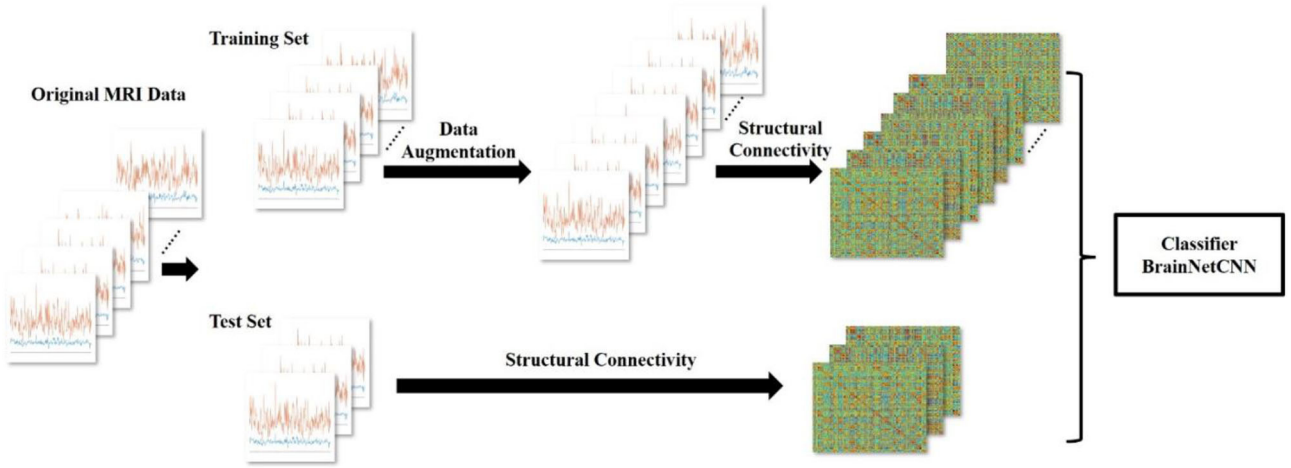

FIGURE 1 | Overview of the process for the data augmentation of the gifted children MRI dataset using GEMD. From left to right, the seven curves representing the morphometric features in the 308 regions of the brain, for each subject (original MRI data). Then, the data are split into training and test sets; the training set is augmented using GEMD; the structural connectivity of the data is calculated and used to feed the deep learning model. Finally, the structural connectivity is also derived for the samples of the test set to demonstrate the capability of the classifier.

TABLE 1 | Membership information of gifted children MRI dataset.

Group	Gifted group	Control group
Average IQ	148.80 ± 2.93	122.71 ± 3.89
Average age	12.03 ± 0.54	12.53 ± 0.77
Sample size	15	14

signals. The process of data decomposition with GEMD is shown in **Figure 3**. The GEMD algorithm (Tremblay et al., 2014) is defined as follows:

- Step 1: Create the adjacency matrix A for the graph G;
- Step 2: Initialize $m = x_i$;
- Step 3: While m does not meet the stopping criterion, repeat step 4 to step 8;
- Step 4: Detect the local extreme of m ;
- Step 5: Interpolate the upper and lower extremes of m and get the envelope$e_{\max}$and $e_{\min}$;
- Step 6: Calculate the average envelope $z = \frac{e_{\min} + e_{\max}}{2}$;
- Step 7: Subtract the average envelope from the signal: $m = m - z$;
- Step 8: Set $d_{i+1} = m$ and $x_{i+1} = x_i - m$;
- Step 9: If m meets the stopping criteria: stop the decomposition and terminate, return stored IMFs, and get [Mathtype-mtef1-eqn-3.mtf].

Data Augmentation

The MRI dataset contains $P = 29$ subjects. Therefore, the training set can be regarded as a three-dimensional tensor $T \in R^{BxFxP}$ (B: number of brain regions; F: number of features; P: number of subjects). If the number of subjects in the training set is too small, the model will tend to be overfitted. To overcome the overfitting problem in the UVic-GC dataset, we propose to increase the training set through GEMD. The data augmentation procedure is based on a decomposition-recombination strategy, originally proposed in Dinarès-Ferran et al. (2018), and first used in a deep learning context in Zhang et al. (2019). The data augmentation process is shown in **Figure 4**. This method has the following steps:

- *Step 1: Data decomposition.*
- Create the adjacency matrix A for the graph G. In our work, A is obtained by calculating the Euclidean distance among the 308 regions.
- Organize the MRI data of all subjects and get the concatenated tensor $T \in R^{BxFxP}$.
- Decompose T with GEMD and get $T_{IMF} \in R^{M \times B \times F \times P}$, where M is the total number of IMFs ($M = 5$ in our experiments).
- *Step 2: Artificial data generation.*
- Randomly select M subjects from one of the groups (gifted group or control group).
- Take one IMF from each subject so that you end up with one IMF from each category (IMF_1 to IMF_5), i.e., each subject contributes with one IMF to create the new artificial data. The artificial data of the n_{th} feature is the sum of the M IMFs.

Structural Connectivity Analysis

After creating artificial samples, we use the original subjects and the artificial to perform the classification. For that purpose, we calculate the SC between features in all the regions. The SC matrix (one matrix per sample, i.e., original subjects and artificial subjects created *via* GEMD) will be used later as the input data for the deep learning classification system. SC represents the data correlation between two brain regions (Qi et al., 2019). Pearson's correlation or coherence is usually used to compute the correlation. We use Pearson's correlation and z-score to obtain

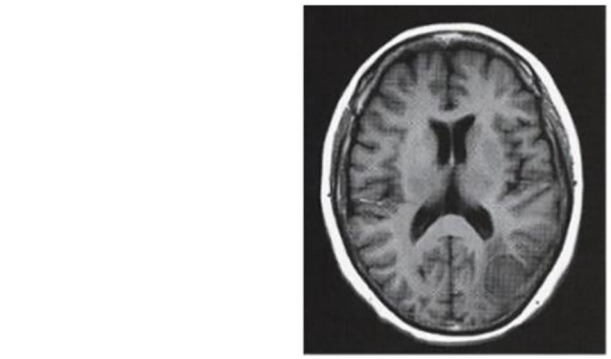

FIGURE 2 | Morphometric features extraction pipeline.

the SC in this work. We correlate the seven values (morphometric features) of one region with the seven values (morphometric features) of another region. We perform these correlations for all possible pairs, obtaining a 308 × 308 matrix per subject. Assuming two brain region data x and y, Pearson's correlation (Kotu and Deshpande, 2019) between x and y can be expressed as follows:

$$c\left(x,y\right)=\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \tag{3}$$

where S_{xy} is the covariance of x and y, which is defined as,

$$S_{xy}=\sum_{i=1}^{n}\left(x_i-x\right)\left(y_i-y\right) \tag{4}$$

S_{xx} and S_{yy} can be calculated as the variance of x and y, respectively. After we get the Pearson's correlation of MRI data, z-score is used to standardize it. Finally, a three-dimensional tensor of dimensions 29 × 308 × 308 is obtained.

This procedure was introduced by Seidlitz et al. (2018) to estimate the inter-regional correlation of multiple MRI features in a single subject instead of estimating the inter-regional correlation of a single feature measured in multiple subjects (which is done with the structural covariance analysis). Therefore, we end up with an SC matrix per subject.

Neural Network Classifier

As the BrainNetCNN (Kawahara et al., 2017) outperforms lots of other methods on structural brain networks datasets, we choose it as a neural network classifier in our experiments. There are three kinds of convolutional filters in BrainNetCNN, namely, E2E, E2N, and N2G filters. They leverage the topological locality of

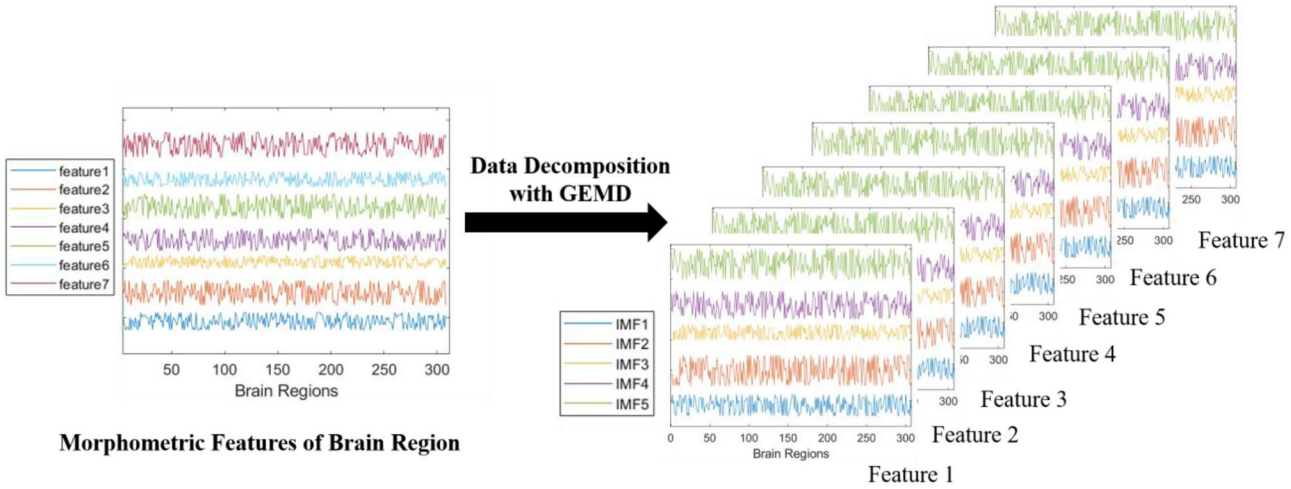

FIGURE 3 | The progress of data decomposition with GEMD. GEMD can decompose feature series (different color lines on the left) into IMFs simultaneously. Here, we use seven morphometric features from MRI data as a decomposition example. Every feature is decomposed into four or five IMFs (different color lines on the right). If only four IMFs are decomposed from raw data, zero-padding will be used to have a total of five IMFs in all the decompositions, so that the data augmentation can proceed successfully.

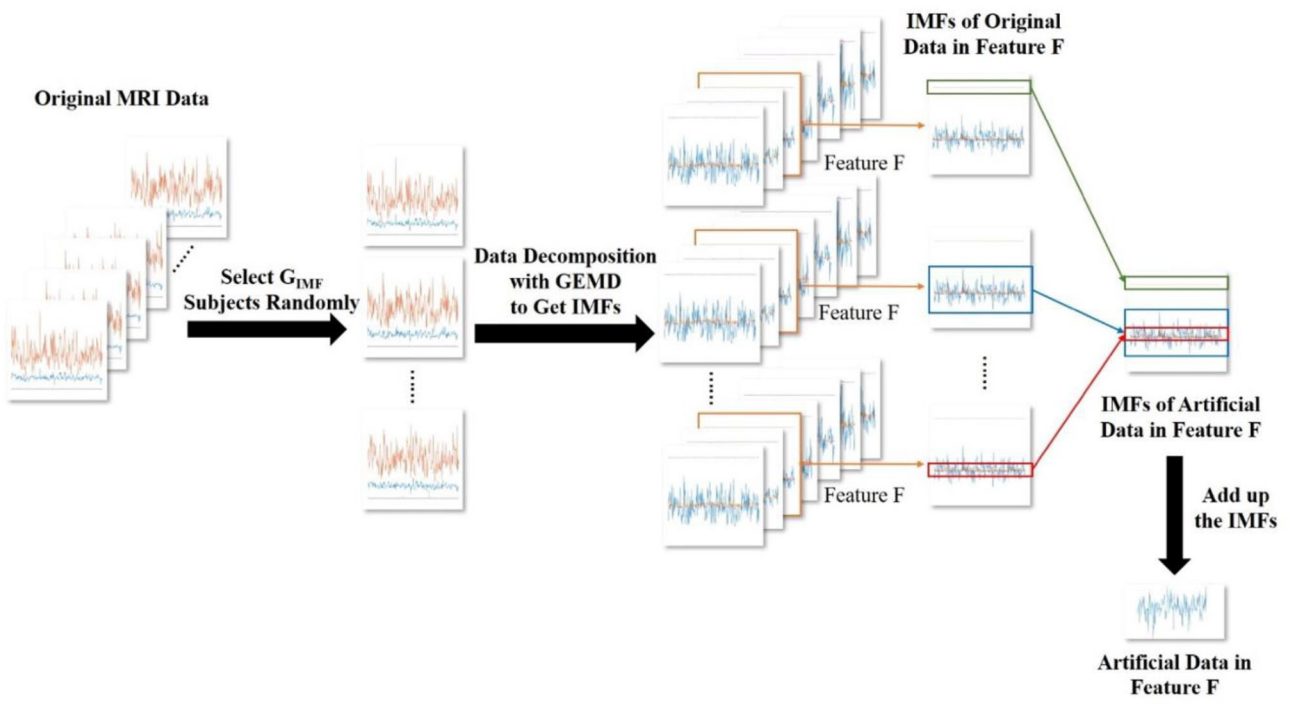

FIGURE 4 | The generation of artificial MRI data. Here, we generate the artificial data in feature F as an example. We randomly select *M* subjects from the original MRI data. Then, we obtain the IMFs, which are decomposed with GEMD. The IMFs decomposed form feature F of the *M* randomly selected subjects are recombined. Then, they are added up to obtain the artificial data of the feature F.

structural brain networks. E2E filter convolves the brain network adjacency matrix and weights edges of adjacent brain regions. E2N filter assigns each brain region a weighted sum of its edges. N2G assigns a single response based on all the weighted nodes. These three filters consist of convolution kernels: kernel $c_1 \in R^{1\times D}$, $c_2 \in R^{D\times 1}$. The kernel of the E2E filter is $c_{E2E} = c_1 + c_2 \in R^{D\times D}$. D is the number of nodes in a graph, which corresponds to the number of brain regions in this work. The kernels of the

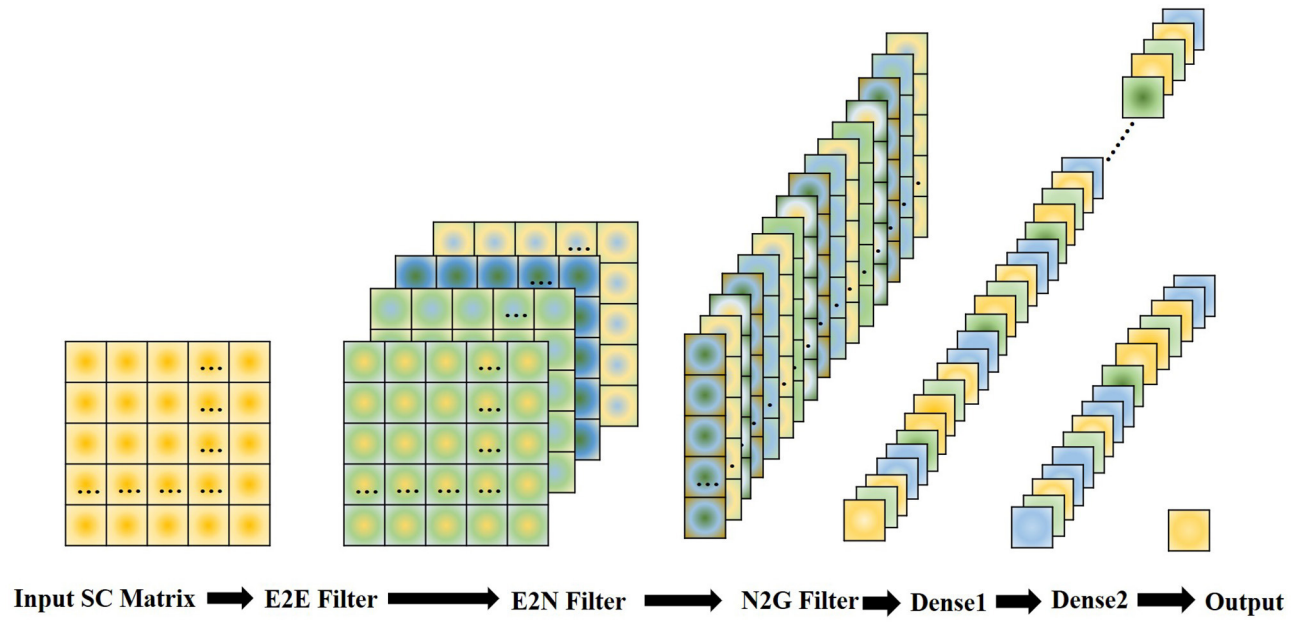

FIGURE 5 | Structure of the BrainNetCNN network.

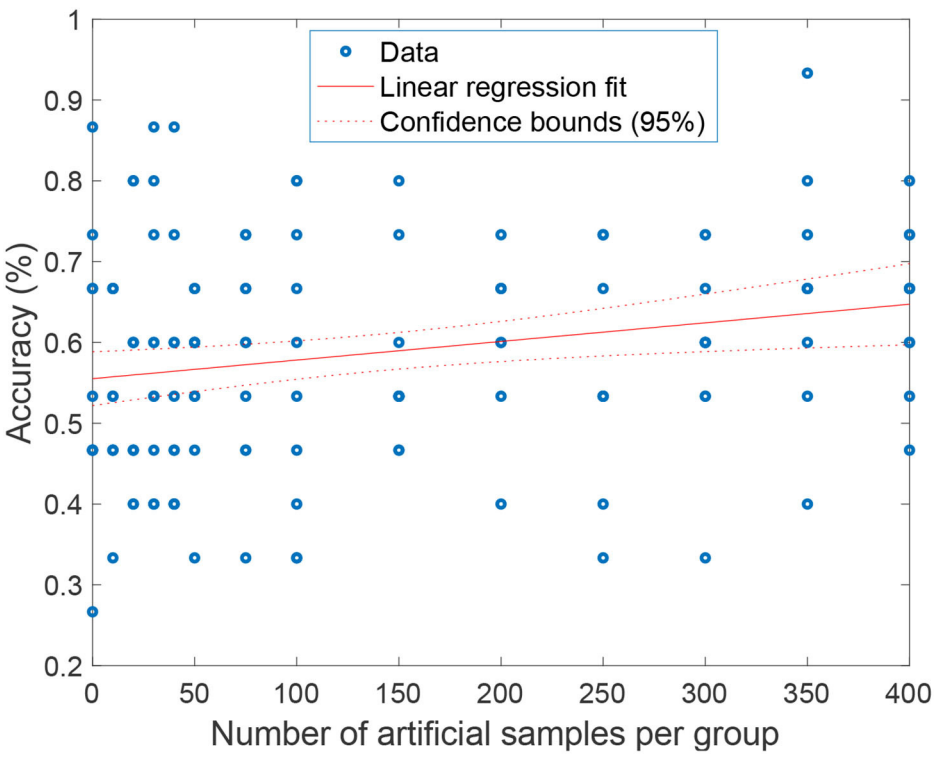

FIGURE 6 | The mean of accuracy and standard deviation with GEMD augmentation for the different number of artificial samples.

E2N filter and N2G filter are $c_{E2N} = c_1$, $c_{N2G} = c_2$. In our experiment, the structure of the BrainNetCNN can be simply expressed as Input (308 × 308 SC matrix) -> E2E (4 channels) -> relu -> E2N (16 channels) -> relu -> N2G (32 channels) -> dense1 (16 channels) -> dense2 (1 channels). This structure is shown in **Figure 5**. We use the adaptive moment estimation

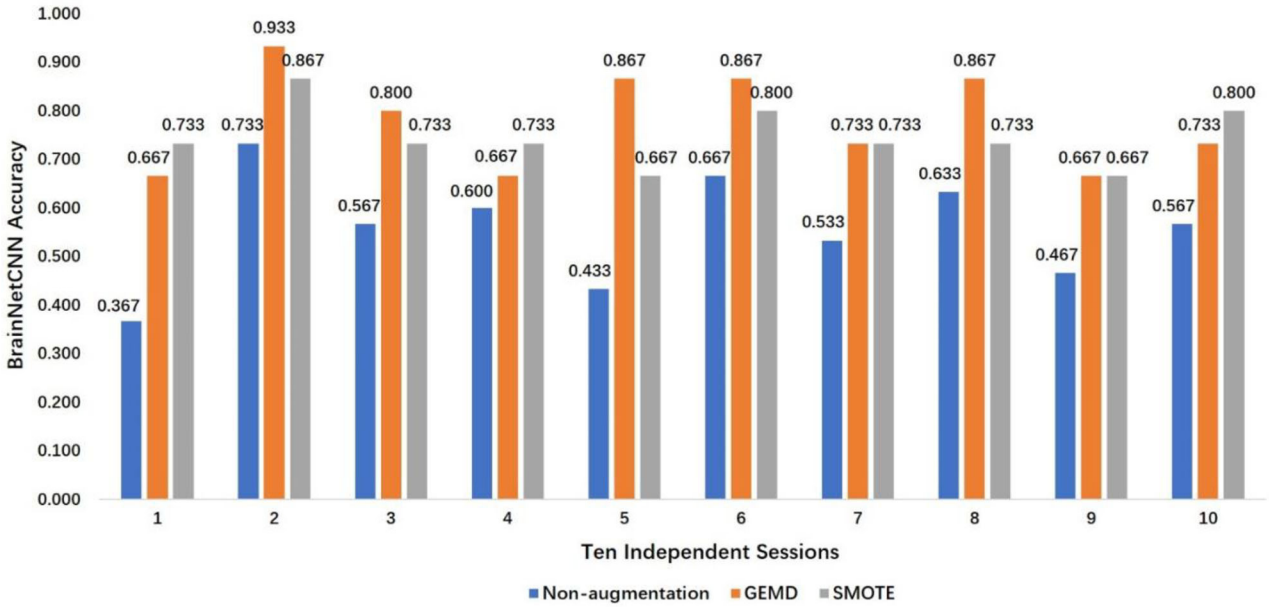

FIGURE 7 | Best accuracy using GEMD, SMOTE, and non-augmented case in 10 independent sessions.

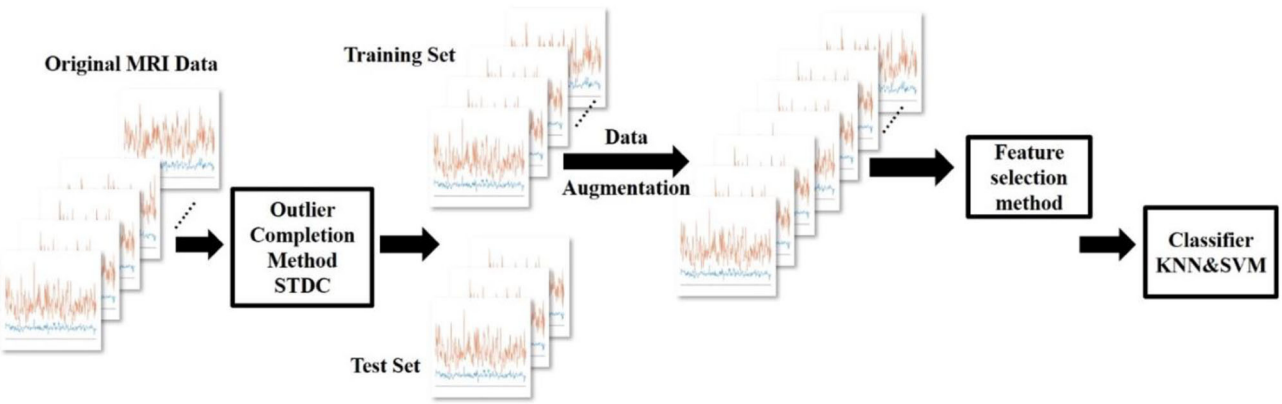

FIGURE 8 | The data augmentation process on the UVic-GC dataset with GEMD and feature selection experiment.

(Adam) optimizer, with learning rate $lr = 0.001$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$. The network is trained using 300 epochs, and the batch size is 32. Considering the size of the dataset, we applied 10-fold cross-validation and repeated the experiment 10 times to get the average accuracy.

RESULTS

GEMD Performance on BrainNetCNN

We want to prove that the data augmentation with GEMD can improve the performance of the BrainNetCNN in the classification of the UVic-GC dataset. Therefore, we randomly selected 14 subjects (7 from the gifted group and 7 from the control group) as the original MRI data for the training set. The training set also contains artificial MRI data generated through GEMD from the original data of this training set. The rest of the subjects are used as the test set, containing 15 subjects.

Aiming to study how the number of artificial subjects affects the performance in the training set, we increase the number of artificial samples from 0 to 400 for each group. For each session, the original MRI data are split into the training set and test set. The training set is used to generate the required number of artificial samples. The model is then trained using the original training set and the artificial samples generated from it, and finally the model is tested with the remaining test set. This process is repeated 10 times for each number of artificial samples to get the final average accuracy.

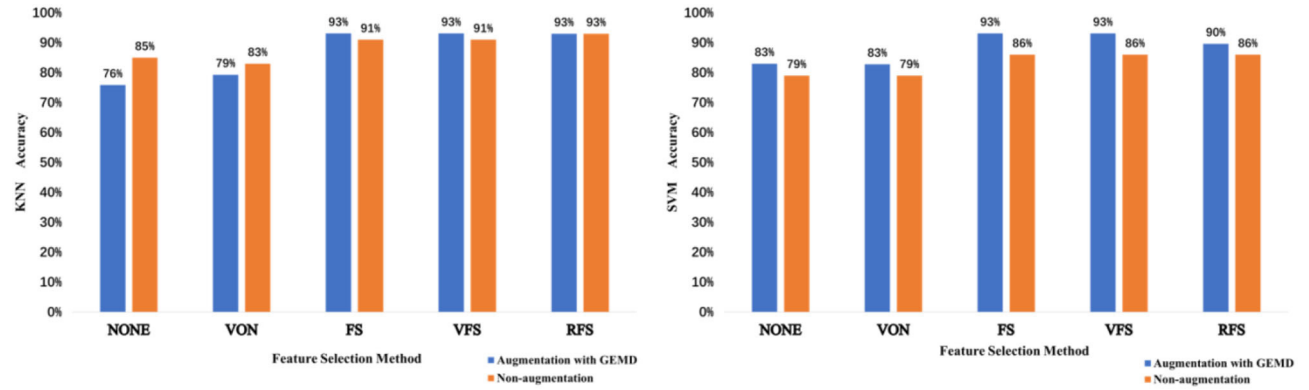

FIGURE 9 | The KNN (left) and SVM (right) accuracies obtained depending on the feature selection method used for the non-augmentation case and augmentation with GEMD case.

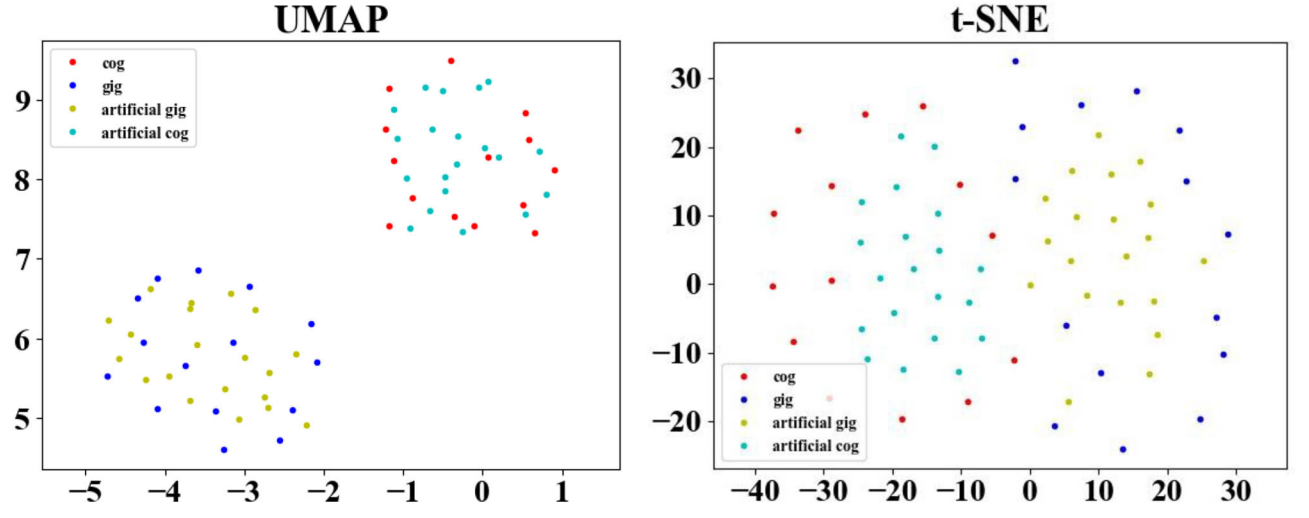

FIGURE 10 | Feature map of the artificial gifted group (artificial gig), artificial control group (artificial cog), original gifted group (gig), and original control group (cog) plotted with UMAP.

In **Figure 6**, we show the classification accuracy for a different number of artificial samples. As can be seen, the performance of the BrainNetCNN can be improved when adding artificial samples, from 10 artificial samples to 400 artificial samples. The improvement increases when the number of artificial samples increases. Fitting a linear regression model gives us an idea of the expected improvement when adding artificial samples. The model shows a positive trend of gradient $x_1 = 0.00023035$, with a p-value < 0.01. This means that we should expect a 2.3% increase in the accuracy per 100 artificial samples added. The BrainNetCNN has the best mean accuracy performance at 66.7% when the number of artificial samples is 350, while without GEMD, the mean accuracy is only 56%. This is an increase of 10.7%, slightly better than the 8% predicted by the linear model.

The SMOTE is also used. The results are compared and depicted in **Figure 7**, which presents the best accuracy with GEMD, SMOTE, and non-augmented cases (baseline) in 10 different sessions. The accuracy is always improved, with respect of the non-augmented case, when GEMD and SMOTE are used. This emphasizes the importance of having more data to train the model. Specifically, GEMD shows higher classification accuracy than SMOTE in sessions 2, 3, 5, 6, and 8; while SMOTE has better performance in sessions 1, 4, and 10. In sessions 7 and 9, the accuracies of both GEMD and SMOTE methods are almost the same. In addition, a classification performance of 93.3% is obtained in session 2 by using GEMD, which is the best result obtained with this database so far. The average of the 10 sessions' best accuracy using GEMD achieves 78%, which is better than using SMOTE (74.7%) and the baseline case (55.7%).

GEMD Performance on Feature Selection Methods

In this section, we evaluate the performance of the GEMD using the procedures described in Zhang et al. (2021). In summary,

Zhang et al. proposed an outlier correction on the morphometric features based on the STDC algorithm (Chen et al., 2013) and explored several feature selection methods to classify MRI data from controls and gifted groups. These methods were applied to the UVic-GC dataset with outstanding performance.

According to Zhang et al. (2021), the NONE feature selection method used all the features in the raw feature matrix. The VON feature selection method used only the regions belonging to types 2 and 3 of the von Economo atlas (van den Heuvel et al., 2015), which corresponds to the associative areas of the brain. Choosing the top highest features selected with a threshold, from all the morphometric features and brain regions, was defined as the FS feature selection method. The rank F-score (RFS) method is a variation of the previous one in which, for each region, the FS values are sorted by descending order, where the morphometric features with the highest FS value are the selected ones. Finally, the combination of VON and FS will lead the VFS feature selection method, in which only type 2 and 3 regions are considered when calculating the FS value for morphometric features. Two traditional machine learning methods, KNN and SVM, were used as classifiers (Zhang et al., 2021), with leave-one-out as a cross-validation strategy.

The process of this experiment is shown in **Figure 8**. First, we use the outlier completion method STDC to compute the missing entries from the estimated latent factors. Then, we enlarge the training set of original MRI data by using GEMD. After that, we use feature selection methods NONE, VON, FS, VFS, and RFS to select different features. Finally, the model is trained by KNN and SVM for classification.

From **Figure 9**, we observe that using data augmentation with GEMD generally improves the performance of feature selection methods. For the SVN case (**Figure 9, right**), the GEMD method always improves the accuracy regardless of the feature selection method, while for the KNN case (**Figure 9, left**) only in two cases the accuracy is lower using artificial data. Note that for both KNN and SVM the classification accuracy reaches 93% using FS and VFS, which is the best result with this database, to the best of our knowledge.

DISCUSSION

In our study, we have used GEMD to enlarge the UVic-GC dataset. The motivation for exploring a data augmentation strategy is 2 fold. First, the UVic-GC dataset is small. Second, there are many parameters in the deep neural network that need to be learned from the data. Therefore, overfitting could appear due to insufficient amount of data.

We propose the GEMD augmentation method to solve the problems mentioned above in this work. We analyze the GEMD augmentation result in three aspects, namely, the influence of the number of artificial subjects, the classification accuracy between non-augmentation and augmentation, and the feature selection method used.

It can be seen from **Figure 6** that the accuracy shows an upward trend with the increase in the amount of artificial data. When the number of artificial data reaches 350, the classification accuracy achieves the maximum. Note that the result may vary considerably from experiment to experiment. This is due to the non-convergence of the BrainNetCNN and the random factor added when selecting the data for each experiment. Prettier but unfair results could be shown by discarding the non-convergent experiments, for example, but we show the full set of results to point out these potential problems.

To clearly illustrate the distribution of the artificial data generated by GEMD, **Figure 10** depicts the original SC matrices, named the original gifted group (gig) and original control group (cog), and 20 artificial SC matrices of the artificial gifted group (artificial gig) and artificial control group (artificial cog). This figure uses Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) and Distributed Stochastic Neighborhood Embedding (van der Maaten and Hinton, 2008) for dimensionality reduction. It can be seen that the artificial data of each group are projected around the original data of the corresponding group, which is a way of showing that the artificial data are meaningful, i.e., the data generated by GEMD are consistent with the distribution of the original data. Furthermore, the two classes (control and gifted) in the two figures can be accurately separated. There is no obvious overlap between the two groups, explaining why the linear classifiers (SVM and KNN) combined with feature selection methods perform very well.

Even if our proposed method can augment the dataset so that the artificial data help improve the classification accuracy, we must highlight that the results of the BrainNetCNN are not stable. This is due to two main factors, the non-convergence of the model and the overfitting that appears despite the amount of artificial data generated. This is the main drawback of the proposed method. We are now investigating it and other possible neural network models with fewer parameters to improve the classification results when using a small number of original MRI subjects in the training dataset and artificial data generated with them. **Figure 10** shows that the artificial data created using the GEMD method are consistent with the original (real) data, which encourages us to use this method and improve the classification model.

CONCLUSIONS

Medical data such as MRI are difficult to obtain, and gifted children are rare in our society. Identifying gifted children from a small set of MRI data is not easy. At the same time, deep neural networks require a large amount of data to improve their performance. They cannot exert their full performance when the dataset is too small. In that case, our work provides a feasible solution by data augmentation. We use the UVic-GC dataset and artificial data generated by GEMD to train the BrainNetCNN neural network. This avoids using a feature selection method as we feed the model directly with the SC data. The results show that GEMD has a significant effect that improves the performance of the classifier. Furthermore, the GEMD data augmentation method can be extended to other similar small datasets. Our future work will focus on the application of GEMD on multisite MRI data, such as the Human Connectome Project data. Due to different scanner settings, parameters, and operators, the distribution of MRI data collected in various regions is different. We expect to be able to adjust the distribution of other datasets by domain adaptation. In that case, we can predict the classification

results of multiple MRI datasets using the trained model after augmentation with GEMD.

DATA AVAILABILITY STATEMENT

The raw anonymized MRI data is available in the OpenNeuro repository: https://openneuro.org/datasets/ds001988. The code for replication is available at: https://github.com/CynthiaChern/GEMD-Based-Data-Augmentation-for-Gifted-Children-MRI-Dataset-Analysis GraphEMD code is available at https://github.com/fkalaganis/graph_emd.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board (IRB00003099) of the University of Barcelona (Catalonia). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

ZS, CC, and JS-C: conceptualization. FD, ZS, CC, and JS-C: methodology and supervision. XC, BL, HJ, and FF: formal analysis and investigation. XC: writing. XC, BL, HJ, FF, FD, ZS, CC, and JS-C: writing—review and editing. FD, CC, and JS-C: funding acquisition. FD and JS-C: resources. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported in part by the National Natural Science Foundation of China (Key Program) under Grant No. 11932013 and in part by the Tianjin Natural Science Foundation for Distinguished Young Scholars under Grant No. 18JCJQJC46100. JS-C's work was partially based upon work from COST Action CA18106, supported by COST (European Cooperation in Science and Technology) and the University of Vic – Central University of Catalonia (R0947). CC's work was partially supported by grants PICT 2017-3208, PICT 2020-SERIEA-00457, UBACYT 20020190200305BA, and UBACYT 20020170100192BA (Argentina).

ACKNOWLEDGMENTS

We are grateful to Dr. Fotis Kalagnis for providing the GraphEMD code and for fruitful discussions about it. We also thank the reviewers for their suggestions and criticisms that have helped us to improve the paper.

REFERENCES

Abdelaziz Ismael, S. A., Mohammed, A., and Hefny, H. (2020). An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artif. Intell. Med.* 102, 101779. doi: 10.1016/j.artmed.2019.101779

Aubry, A., Gonthier, C., and Bourdin, B. (2021). Explaining the high working memory capacity of gifted children: contributions of processing skills and executive control. *Acta Psychol.* 218, 103358. doi: 10.1016/j.actpsy.2021.103358

Bucaille, A., Jarry, C., Allard, J., Brochard, S., Peudenier, S., and Roy, A. (2022). Neuropsychological profile of intellectually gifted children: a systematic review. *J. Int. Neuropsychol. Soc.* 28, 424–440. doi: 10.1017/S1355617721000515

Caiafa, C. F., Solé-Casals, J., Marti-Puig, P., Zhe, S., and Tanaka, T. (2020). Decomposition methods for machine learning with small, incomplete or noisy datasets. *Appl. Sci.* 10, 8481. doi: 10.3390/app10238481

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi: 10.1613/jair.953

Chen, Y.-L., Hsu, C.-T., and Liao, H.-Y. M. (2013). Simultaneous tensor decomposition and completion using factor priors. *IEEE Transac. Pattern Anal. Mach. Intell.* 36, 577–591. doi: 10.1109/TPAMI.2013.164

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021

Dinarès-Ferran, J., Ortner, R., Guger, C., and Solé-Casals, J. (2018). A new method to generate artificial frames using the empirical mode decomposition for an EEG-based motor imagery BCI. *Front. Neurosci.* 12, 308. doi: 10.3389/fnins.2018.00308

Grady, L. J., and Schwartz, E. L. (2003). *Anisotropic Interpolation on Graphs: The Combinatorial Dirichlet Problem*. Boston, MA: Citeseer.

Gras, R. M. L., Bordoy, M., Ballesta, G. J., and Berna, J. C. (2010). Creativity, intelectual abilities and response styles: Implications for academic performance in the secondary school.[Creatividad, aptitudes intelectuales y estilos de respuesta: implicaciones para el rendimiento académico en secundaria]. *Ann. Psychol.* 26, 212–219. Available online at: https://doi.org/10.6018/analesps

Gross, M. U. M. (2006). Exceptionally gifted children: long-term outcomes of academic acceleration and nonacceleration. *J. Educ. Gifted* 29, 404–429. doi: 10.4219/jeg-2006-247

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., et al. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London.* 454, 903–995. doi: 10.1098/rspa.1998.0193

Kalaganis, F. P., Laskaris, N. A., Chatzilari, E., Nikolopoulos, S., and Kompatsiaris, I. (2020). A Data Augmentation Scheme for Geometric Deep Learning in Personalized Brain–Computer Interfaces. *IEEE Access* 8, 162218–162229. doi: 10.1109/ACCESS.2020.3021580

Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V., Grunau, R. E., et al. (2017). BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage.* 146, 1038–1049. doi: 10.1016/j.neuroimage.2016.09.046

Kotu, V., and Deshpande, B. (2019). "Chapter 4 - Classification," in *Data Science (Second Edition)*, eds. V. Kotu and B. Deshpande. *Morgan Kaufmann)* p. 65–163. doi: 10.1016/B978-0-12-814761-0.00004-6

Kuhn, T., Blades, R., Gottlieb, L., Knudsen, K., Ashdown, C., Martin-Harris, L., et al. (2021). Neuroanatomical differences in the memory systems of intellectual giftedness and typical development. *Brain Behav.* 11, e2348. doi: 10.1002/brb3.2348

Leonardsen, E. H., Peng, H., Kaufmann, T., Agartz, I., Andreassen, O. A., Celius, E. G., et al. (2022). Deep neural networks learn general and clinically relevant representations of the ageing brain. *NeuroImage* 256, 119210. doi: 10.1016/j.neuroimage.2022.119210

Ma, J., Kang, H. J., Kim, J. Y., Jeong, H. S., Im, J. J., Namgung, E., et al. (2017). Network attributes underlying intellectual giftedness in the developing brain. *Sci. Rep.* 7, 11321. doi: 10.1038/s41598-017-11593-3

Muñoz-Gutiérrez, P. A., Giraldo, E., Bueno-López, M., and Molinas, M. (2018). Localization of active brain sources from EEG signals using empirical

mode decomposition: a comparative study. *Front. Integr. Neurosci.* 12, 55. doi: 10.3389/fnint.2018.00055

Navas-Sánchez, F. J., Alemán-Gómez, Y., Sánchez-Gonzalez, J., Guzmán-De-Villoria, J. A., Franco, C., Robles, O., et al. (2014). White matter microstructure correlates of mathematical giftedness and intelligence quotient. *Hum. Brain Map.* 35, 2619–2631. doi: 10.1002/hbm.22355

Nguyen, K., Chin Fatt, C., Treacher, A., Mellema, C., Trivedi, M., and Montillo, A. (2020). *Anatomically Informed Data Augmentation for Functional MRI With Applications to Deep Learning.* (Houston, TX: SPIE). doi: 10.1117/12.2548630

Qi, P., Ru, H., Gao, L., Zhang, X., Zhou, T., Tian, Y., et al. (2019). Neural Mechanisms of Mental Fatigue Revisited: New Insights from the Brain Connectome. *Engineering* 5, 276–286. doi: 10.1016/j.eng.2018.11.025

Rocha, A., Almeida, L., and Perales, R. G. (2020). Comparison of gifted and non-gifted students' executive functions and high capabilities. *J. Educ. Gifted Young Sci.* 8, 1397–1409. doi: 10.17478/jegys.808796

Romero-Garcia, R., Atienza, M., Clemmensen, L. H., and Cantero, J. L. (2012). Effects of network resolution on topological properties of human neocortex. *NeuroImage.* 59, 3522–3532. doi: 10.1016/j.neuroimage.2011.10.086

Sarraf, S., and Tofighi, G. (2016). Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks. *arXiv preprint arXiv:1603.08631*.

Seidlitz, J., Váša, F., Shinn, M., Romero-Garcia, R., Whitaker, K. J., Vértes, P. E., and NSPN Consortium (2018). Morphometric similarity networks detect microscale cortical organization and predict inter-individual cognitive variation. *Neuron.* 97, 231–247 doi: 10.1016/j.neuron.2017.11.039

Solé-Casals, J., Serra-Grabulosa, J. M., Romero-Garcia, R., Vilaseca, G., Adan, A., Vilaró, N., et al. (2019). Structural brain network of gifted children has a more integrated and versatile topology. *Brain Struct. Function* 224, 2373–2383. doi: 10.1007/s00429-019-01914-9

Tremblay, N., Borgnat, P., and Flandrin, P. (2014). "Graph Empirical Mode Decomposition", in: *2014 22nd European Signal Processing Conference (EUSIPCO)* (Lisbon: IEEE), p. 2350–2354.

Ulloa, A., Plis, S., Erhardt, E., and Calhoun, V. (2015). "Synthetic structural magnetic resonance image generator improves deep learning prediction of schizophrenia", in: *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)* (Boston, MA: IEEE), p. 1–6. doi: 10.1109/MLSP.2015.7324379

van den Heuvel, M. P., Scholtens, L. H., Barrett, L. F., Hilgetag, C. C., and de Reus, M. A. (2015). Bridging cytoarchitectonics and connectomics in human cerebral cortex. *J. Neurosci.* 35, 13943–13948. doi: 10.1523/JNEUROSCI.2630-15.2015

van der Maaten, L., and Hinton, G. (2008). Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605. Available online at: https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

Wei, M., Wang, Q., Jiang, X., Guo, Y., Fan, H., Wang, H., et al. (2020). Directed connectivity analysis of the brain network in mathematically gifted adolescents. *Comput. Intell. Neurosci.* 2020:10. doi: 10.1155/2020/4209321

Zhang, J., Feng, F., Han, T., Duan, F., Sun, Z., Caiafa, C. F., et al. (2021). A hybrid method to select morphometric features using tensor completion and F-score rank for gifted children identification. *Sci. China Technol. Sci.* 64, 1863–1871. doi: 10.1007/s11431-020-1876-3

Zhang, Z., Duan, F., Solé-Casals, J., Dinares-Ferran, J., Cichocki, A., Yang, Z., et al. (2019). A novel deep learning approach with data augmentation to classify motor imagery signals. *IEEE Access.* 7, 15945–15954. doi: 10.1109/ACCESS.2019.2895133

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Neuroscience

METHODS
published: 29 July 2022
doi: 10.3389/fnins.2022.866734

Systematic Fusion of Multi-Source Cognitive Networks With Graph Learning - A Study on Fronto-Parietal Network

Xiaofei Zhang[1,2,3,4†], *Yang Yang*[5†], *Hongzhi Kuai*[3,4,6], *Jianhui Chen*[1,3,4], *Jiajin Huang*[1,3,4], *Peipeng Liang*[7*] and *Ning Zhong*[1,3,4,6*]

[1] Faculty of Information Technology, Beijing University of Technology, Beijing, China, [2] School of Computer, Jiangsu University of Science and Technology, Zhenjiang, China, [3] International WIC Institute, Beijing University of Technology, Beijing, China, [4] Beijing International Collaboration Base on Brain Informatics and Wisdom Services, Beijing, China, [5] Department of Psychology, Beijing Forest University, Beijing, China, [6] Department of Life Science and Informatics, Maebashi Institute of Technology, Maebashi, Japan, [7] School of Psychology and Beijing Key Laboratory of Learning and Cognition, Capital Normal University, Beijing, China

OPEN ACCESS

Edited by:
Feng Liu,
Stevens Institute of Technology,
United States

Reviewed by:
Xinglong Ju,
University of Oklahoma, United States
Linh Ho Manh,
University of Texas at Arlington,
United States

***Correspondence:**
Peipeng Liang
ppliang@cnu.edu.cn
Ning Zhong
zhong@maebashi-it.ac.jp

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: 31 January 2022
Accepted: 22 June 2022
Published: 29 July 2022

Citation:
Zhang X, Yang Y, Kuai H, Chen J, Huang J, Liang P and Zhong N (2022) Systematic Fusion of Multi-Source Cognitive Networks With Graph Learning - A Study on Fronto-Parietal Network. Front. Neurosci. 16:866734. doi: 10.3389/fnins.2022.866734

Cognitive tasks induce fluctuations in the functional connectivity between brain regions which constitute cognitive networks in the human brain. Although several cognitive networks have been identified, consensus still cannot be achieved on the precise borders and distribution of involved brain regions for each network, due to the multifarious use of diverse brain atlases in different studies. To address the problem, the current study proposed a novel approach to generate a fused cognitive network with the optimal performance in discriminating cognitive states by using graph learning, following the synthesization of one cognitive network defined by different brain atlases, and the construction of a hierarchical framework comprised of one main version and other supplementary versions of the specific cognitive network. As a result, the proposed method demonstrated better results compared with other machine learning methods for recognizing cognitive states, which was revealed by analyzing an fMRI dataset related to the mental arithmetic task. Our findings suggest that the fused cognitive network provides the potential to develop new mind decoding approaches.

Keywords: systematic fusion, brain atlas, cognitive network, fronto-parietal network, fMRI

1. INTRODUCTION

Cognitive functions of the human brain rely on neuronal activities, as well as the intra-neural networks and inter-neural networks. Modern neuroimaging technologies, such as functional magnetic resonance imaging (fMRI), have provided effective approaches to revealing the patterns of the neural network, also known as cognitive network, during the cognitive processes. Being a vital control-type cognitive network, the fronto-parietal network (FPN) occupying brain regions across the lateral prefrontal cortex to the posterior parietal cortex, plays a critical role in imposing cognitive control on a variety of tasks by initiating and deploying executive control abilities. As a result, it has always been in a flexible state full of dynamic changes while other processing-type cognitive networks are deemed to be more comparatively modular and static (Dosenbach et al., 2008). Previous studies on structural and functional neuroimaging have reached a general

consensus that the FPN is responsible for intelligence, integrated with the cognitive functions including perception, attention, memory, language, and planning (Colom et al., 2010). Many fMRI and PET studies on attention, working memory, and episodic memory retrieval have reported the frequent detection of the FPN's activity. Moreover, the activation of the FPN was observed in some fMRI studies for conscious visual perception (Naghavi and Nyberg, 2005), and the FPN in the Theta band was found to take a vital role in a mentally demanding arithmetic task (Mizuhara and Yamaguchi, 2007). Consequently, the exploration of the FPN can help to provide a more comprehensive and precise understanding of the intelligence and cognitive abilities of the human brain.

Given that the human brain is a precisely interconnected system network, graph theory has increasingly proved to be a popular tool for the analysis of human MRI data (Fornito, 2016). By adopting graph analysis, it is found that the local and global integrity of the FPN, the cingulo-opercular network (CON), and other control-type cognitive networks, are significantly positively associated with cognitive abilities. It suggests that greater network efficiency supports better cognitive ability, evidenced by the similar performance in healthy participants and patients with schizophrenia (Sheffield et al., 2015). Under the resting state, the functional connectivity between the critical regions of the FPN is identified to be linked to the cognitive performance of patients with glioma as well as their cognitive outcome after the surgery treatment (Lang et al., 2017). The FPN and its subregions can change the functional connectivity with nodes of other cognitive networks on the different goals of cognitive tasks. In addition, the functional connectivity pattern of the FPN can indicate its involvement in specific tasks, and facilitate the novel tasks' learning in the form of a transferable code (Zanto and Gazzaley, 2013). Therefore, the FPN can be regarded as a defined control network, whose partial function is to interact with and change other cognitive networks (Marek and Dosenbach, 2018).

The cognitive networks in the human brain are often defined on the basis of anatomical or functional brain atlases. Automated Anatomical Labeling (AAL) is a commonly used anatomy-based structural brain atlas (Tzourio-Mazoyer et al., 2002). On the basis of AAL, the FPN and its default mode network (DMN) can be structurally defined, where the FPN has six regions in the frontal lobe and four regions in the parietal lobe (Oliver et al., 2019). The FPN can also be defined with some functional brain atlases. For instance, Dosenbach-160 (Dosenbach et al., 2010) is a human brain atlas consisting of 160 regions of interest (ROIs), where each ROI is uniquely assigned to one of its six cognitive networks. Power-264 (Power et al., 2011) is a human brain atlas composed of 264 ROIs, of which 236 ROIs are uniquely assigned to the given part of its 13 cognitive networks yet the remaining 28 ROIs belong to which part of the cognitive networks remains uncertain. There are 21 ROIs and 28 ROIs identified in the FPNs of Dosenbach-160 and Power-264, respectively. Willard-499 (Richiardi et al., 2015) is a voxel-defined brain atlas, and its 142 regions can be identified in one of its 14 cognitive networks while the cognitive networks to which the remaining 357 regions belong are unknown. Willard-499 does not make an explicit definition of the FPN, yet it defines two executive control networks (ECNs), namely, the left ECN and the right ECN located in the two hemispheres. Given that FPN and ECN are conceptually equivalent (Seeley et al., 2007; Vincent et al., 2008), the FPN can be obtained by tailoring the ECN of Willard-499. Moreover, some researchers defined a small brain atlas manually for their own research. For example, the brain atlas Gao-32 (Gao and Lin, 2012) consists of 32 ROIs, where each ROI is uniquely assigned to one of its five cognitive networks. The FPN of Gao-32 contains nine ROIs in the frontal lobe, parietal lobe, and insula. Despite the conceptual consistency in cognitive neuroscience, the above mentioned frameworks of FPN differ in their structures and distributions. The differences between the multiple FPNs bring challenges to exploring the processing mechanisms of FPNs in human brain cognitive tasks.

The various versions of the FPN definitions result in the fact that any exploration of the FPN from just a single perspective may only result in one sided outcome. Given the complexity of brain science, brain informatics (Zhong et al., 2011) claimed the importance and necessity of a thorough exploration for the research of human information processing system (HIPS). Thus, a systematic fusion of multi-source FPNs may provide an effective way to address this problem by providing a comprehensive and systematic investigation. This article aims to propose a graph-learning-based method for fusing multi-source cognitive networks and tends to evaluate it with its application to the fusion of the FPNs from multiple brain atlases. Three steps are involved to achieve this method as shown in **Figure 1**. The first step is to study and adjust the FPNs defined in the multiple brain atlases, so as to ensure the consistent boundaries between the regions contained in all the FPNs and to spatially synthesize the multi-source FPNs for the realization of a combined FPN. The second step is to analyze the functional connectivity of the combined FPN under a specific cognitive task and to calculate the graph properties of each independent FPN in the combined FPN. The single FPN with the optimal performance in discriminating the graph properties under different cognitive states is chosen as the main FPN while other FPNs are accepted as the supplementary FPNs. In the last step, the main cognitive network is adopted as the initial candidate fused FPN, into which the ROIs in all the supplementary FPNs are added one by one according to their nodal index. The iteratively fused FPN composed of the main FPN and the added ROIs has the optimal index and, thus, is set as the final fused FPN. The experiment materials are the fMRI data generated by mental arithmetic task (Yang et al., 2017) and the resulting fused FPN will be evaluated by comparing it with other classic machine learning methods.

2. METHODOLOGY

2.1. Synthesizing Multi-Source Cognitive Networks

Closely related to the core cognitive functions of the human brain, the FPN has been widely discussed in the research literature on brain atlases. However, no agreement has been reached on the definition of the FPN and its relative descriptions, and some obvious differences or even contradictions still exist

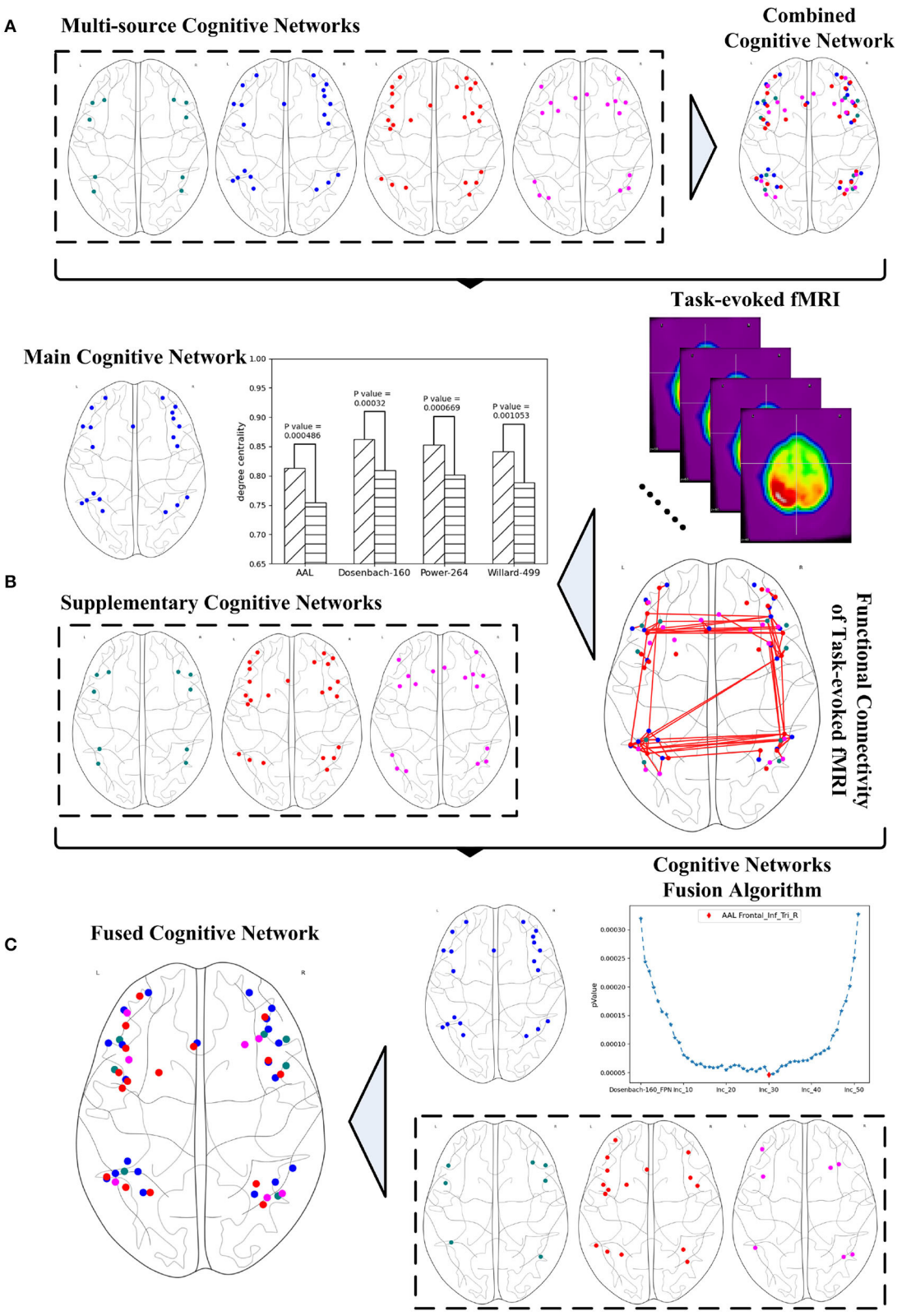

FIGURE 1 | Systematic fusion of multi-source cognitive networks, **(A)** combined cognitive network synthesized from multiple brain atlases, **(B)** main and supplementary cognitive networks obtained from graph analysis, **(C)** fused cognitive network computed from cognitive networks fusion algorithm.

between different brain atlases. To better explore the fusion of cognitive networks congruent with the FPN, this section will first analyze several typical brain atlases that define the FPN, and then clip some FPNs to obtain a reasonable combined FPN. Given the FPN is just one specific type of cognitive network in the human brain, it is necessary to generalize the definitions of cognitive network initially.

Definition 1. *CN* is the concept of a cognitive network that specializes in specific cognitive functions of the human brain, and $(CN)^I$ refer to the set of all the concrete cognitive network instances of *CN*. CN_{inst}, the element in the set of $(CN)^I$, is one instance of *CN* and constituted by a collection of ROIs, as shown in Equation 1.

$$CN_{inst} = \{ROI_1, ROI_2, ..., ROI_{|CN_{inst}|}\}, CN_{inst} \in (CN)^I \quad (1)$$

where $|CN_{inst}|$ is the ROIs' number of CN_{inst}.

The ROIs in different instances of *CN* vary from different reference brain atlases. If a specific *CN* has the instances of $CN_1, CN_2, ..., CN_{|(CN)^I|}$, the union of all the instances constitutes the combined cognitive network of *CN*, defined and shown in Equation 2.

Definition 2. A combined cognitive network noted as *CCN* and formulated in Equation 2 is the set of ROIs from all the instances of the same specific *CN*. $|(CN)^I|$ is the instances' number of the specific *CN*.

$$CCN = \bigcup_{i=1}^{|(CN)^I|} CN_i, CN_i \in (CN)^I \quad (2)$$

It should be noted that ROIs with equal Montreal Neurological Institute (MNI) coordinates may come from different instances of *CN*. Given that these ROIs originate from different brain atlases or research, the ROIs with the same MNI coordinates are still considered to be different. That is, $CN_i \bigcap CN_j = \emptyset, \forall CN_i, \forall CN_j \in (CN)^I, i \neq j$. With regards to the different ROIs from the same instance of *CN*, they do not have the same MNI coordinates.

To synthesize a combined cognitive network with the FPNs as a specific case, the investigation of the ROIs contained in each instance of the FPN must be initially conducted. The brain atlas Gao-32 defines an FPN with nine ROIs, as shown in **Figure 2A**, of which seven red ROIs belong to the frontal lobe or parietal lobe, while the two blue ROIs belong to the insula. The present study does not define the insula as the portion of the FPN. Previous emotion-controlling studies on the exploration of the emotion-regulating strategies failed to prove the possible connections between insula-active and FPN-active regions (Li et al., 2021), which indicates that the insula may not be a part of the FPN on the edge. Suppose these results are correct, there are only seven ROIs left in the FPN of Gao-32 after the removal of the insula ROIs. Moreover, very scant follow-up studies were conducted to make a further validation of Gao-32, thus the FPN of Gao-32 is not on the consideration list for the present study. The brain atlas Power-264 defines an FPN with 25 ROIs, as shown in **Figure 2B**, in which 24 red ROIs belong to the frontal lobe or parietal lobe, while the only blue ROI belongs to the temporal lobe. Although previous studies have indicated that many cognitive networks vary among subjects, a great level of overlap can be identified from the FPN of multiple subjects in the internal parietal sulcus, ventral inferior temporal gyrus, and lateral prefrontal cortex (Marek and Dosenbach, 2018). Given that one ROI in the FPN of Power-264 belongs to the middle temporal gyrus in the temporal lobe, it is removed from the definition of the FPN with the Power-264 in the present study. The brain atlas Willard-499 defines an ECN containing 24 regions, as shown in **Figure 2C**, 17 of which belong to the frontal lobe or parietal lobe. In addition, there is one specific region in both limbic and temporal lobes and five particular regions in the cerebellum. Considering Willard-499 is defined with numbers for different regions rather than with specific names, only the regions located in the frontal lobe or parietal lobe in the ECN of Willard-499 are retained, so as to reach the congruity between the Power-264 clipping strategy and the lobes involved in the FPN of other brain atlases. A customized Willard-499 FPN is finally obtained as shown in **Figure 2D**.

Partial brain atlases defined their FPNs containing regions or ROIs that only belong to the frontal lobe or parietal lobe, resulting in the congruity between conceptual boundaries when synthesizing multi-source FPNs. The FPN defined on the basis of AAL contains 10 regions (Oliver et al., 2019), all of which belong to the frontal lobe or parietal lobe, as shown in **Figure 3A**. The FPN obtained from the central ROI of each region is shown in **Figure 3B**. Dosenbach-160 defines an FPN with 21 ROIs, and all ROIs belongs to the frontal lobe or parietal lobe exclusively, as shown in **Figure 3C**. The original FPN of Power-264 is clipped to obtain a pruned FPN with 24 ROIs, as shown in **Figure 3D**. An FPN with 17 regions is obtained by clipping the original Willard-499 ECN, and the ROIs in the new FPN of Willard-499 are shown in **Figure 3E**.

In the present study, the original or clipped FPNs with reference to AAL, Dosenbach-160, Power-264, and Willard-499 are confined to the frontal lobe and parietal lobe, providing the congruent lobe boundaries to synthesize multi-source FPNs. The combined FPN contains 72 ROIs, located in the frontal lobe or parietal lobe, and their distributions in the brain are shown in **Table 1**. Of all the forty-seven frontal ROIs in the combined FPN, six come from AAL, 13 from Dosenbach-160, 17 from Power-264, and 11 from Willard-499. On top of that, four ROIs from AAL, eight ROIs from Dosenbach-160, seven ROIs from Power-264, and six ROIs from Willard-499 constitute the 25 parietal ROIs in the combined FPN.

Taking the *FPN* as a specific concept of the cognitive network as an example, if the four FPNs mentioned above make up all the instances of the *FPN*, the definition of the instances set of the *FPN* can be shown in Equation 3.

$$(FPN)^I = \{FPN_a, FPN_d, FPN_p, FPN_w\} \quad (3)$$

where FPN_a, FPN_d, FPN_p, and FPN_w represent the FPNs in the brain atlas of AAL, Dosenbach-160, Power-264, and Willard-499, respectively.

To arrive at a comprehensive examination of all the FPNs defined in these human brain atlases, all the contained ROIs need

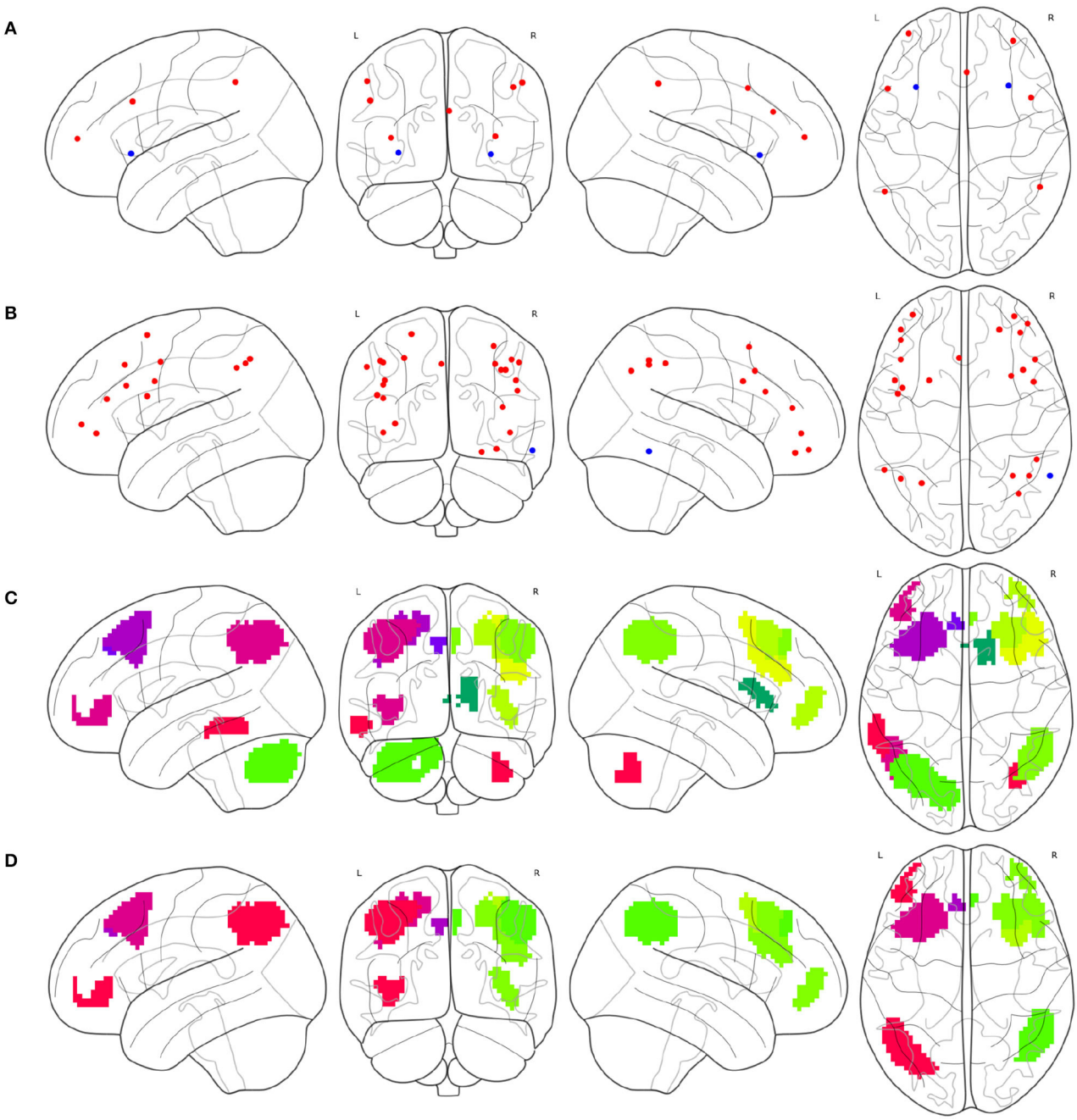

FIGURE 2 | Definition issues of the FPNs in some brain atlases, **(A)** the ROIs in the FPN of Gao-32, where the blue ROIs belong to the insula, **(B)** the ROIs in the FPN of Power-264, where the only blue ROI belongs to the temporal lobe, **(C)** the regions in the ECN of Willard-499, where some of the regions belong to limbic lobe, temporal lobe, or cerebellum, and **(D)** the frontal and parietal regions in the ECN of Willard-499.

to be synthesized. And the definition of the combined FPN is shown in Equation 4.

$$CombinedFPN = FPN_a \bigcup FPN_d \bigcup FPN_p \bigcup FPN_w \quad (4)$$

Compared with other FPN instances, more ROIs in *CombinedFPN* mean more various dimensions of the correlation matrices for functional connectivity analysis, resulting in the difficulty in interpreting the final results of the *CombinedFPN*. Consequently, the following section will discuss how to set a given FPN instance as the main FPN in the *CombinedFPN*, and then choose some ROIs with higher priority from the remaining supplementary FPNs to construct a fused FPN so as to ensure a reasonable range of the dimensions, as well as the optimal

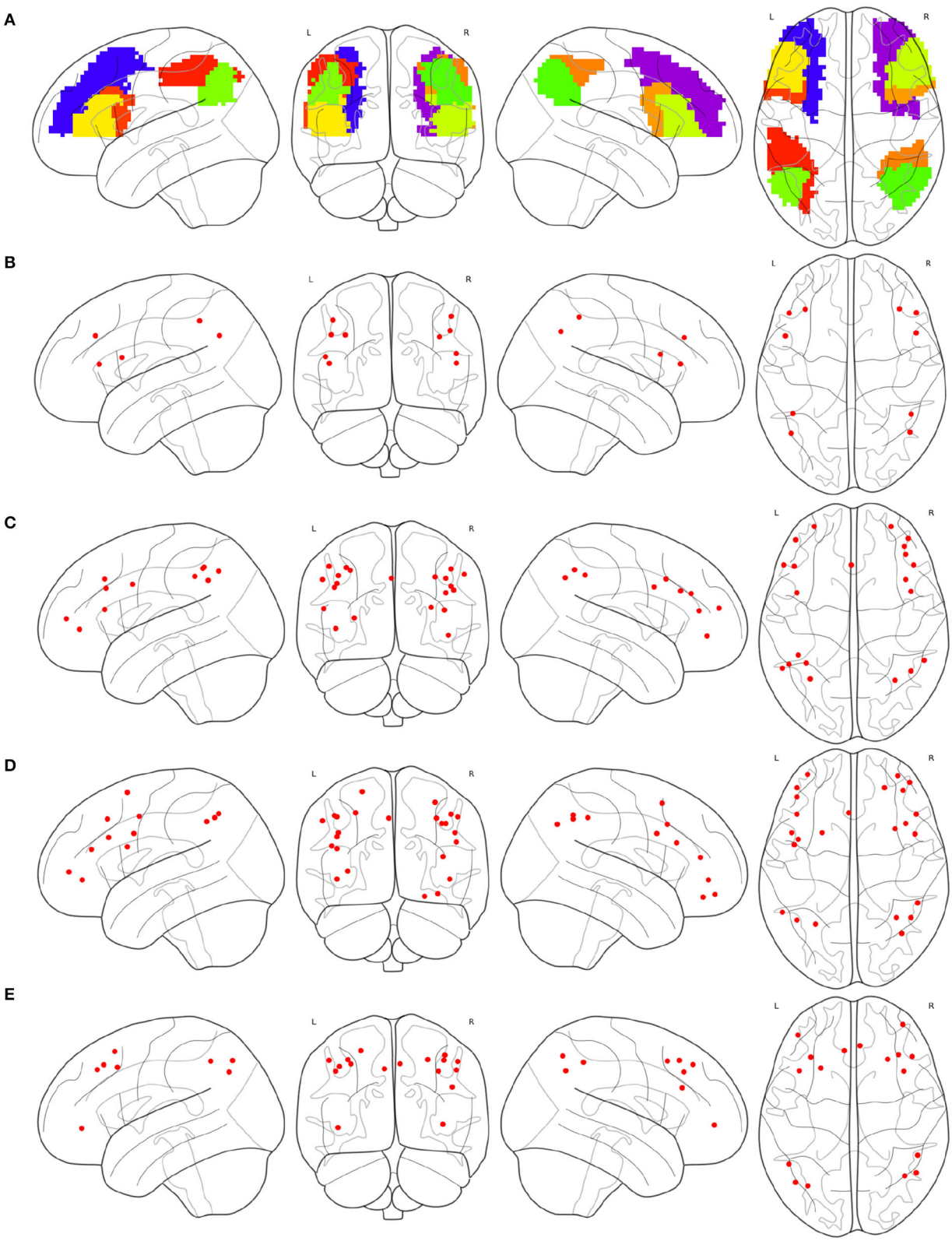

FIGURE 3 | Multi-source FPNs for synthesizing the combined FPN, **(A)** the regions belong to the FPN of AAL, **(B)** 10 ROIs of the FPN in AAL, **(C)** 21 ROIs of the FPN in Dosenbach-160, **(D)** 24 ROIs of the FPN in Power-264, and **(E)** 17 ROIs of the FPN in Willard-499.

TABLE 1 | Combined fronto-parietal network.

Cognitive network	Number of ROIs					
	Frontal	Parietal	Occipital	Temporal	Limbic	Cerebellum
FPN of AAL	6	4	0	0	0	0
FPN of Dosenbach-160	13	8	0	0	0	0
FPN of Power-264	17	7	0	1	0	0
ECN of Willard-499	11	6	0	1	1	5
Combined FPN	47	25	0	0	0	0

TABLE 2 | Fused fronto-parietal network.

Cognitive network	Number of ROIs	
	Frontal	Parietal
FPN of AAL	5	2
FPN of Dosenbach-160	13	8
FPN of Power-264	11	5
ECN of Willard-499	4	3
Fused FPN	33	18

performance in discriminating the graph properties of the FPN under different cognitive states.

2.2. Selecting Main Cognitive Network

Although the synthesis of multi-source cognitive networks can examine more cortical regions relevant to the cognitive task, such an operation will contribute to the increase of computational load during data analysis, as well as the worse performances in discriminating the graph properties of the FPN under different cognitive states. Alternatively, the concentration on one specific instance of the cognitive network may help to provide a consensus to reach a better interpretation of the analyzed results. Consequently, there is a need to select one instance of *CN* as the main cognitive network and set the remaining instances of *CN* as the supplementary cognitive networks.

Definition 3. The main cognitive network noted as *MCN* and formulated in Equation 5 is the instance of *CN* and has the optimal performance, compared with any other instances of *CN*, in discriminating the graph properties under different cognitive states for the relative task.

$$MCN \in (CN)^I, P(MCN) \geq P(CN_i), CN_i \in (CN)^I, CN_i \neq MCN \tag{5}$$

where *P* stands for the performance, usually set as the *P*-value, in discriminating the graph properties of the FPN under different cognitive states. The *P-Value* of each instance of *CN* is given in **Algorithm 1**. The remaining instances of *CN* excluding *MCN* are the supplementary cognitive networks.

Algorithm 1: Main cognitive network selection.

Data: *MatricesConditions*
Input: $(CN)^I$, *Conditions*, *Subjects*
Output: *MainROIs*, *SuppROIs*
begin
  // Stage 1
  subnetP-Values = ∅
  for *candidateMain* ∈ $(CN)^I$ **do**
    MatricesIndices = *InitMatrix*(|*Conditions*|, |*Subjects*|)
    for *c* ∈ *range*(|*Conditions*|) **do**
      for *s* ∈ *range*(|*Subjects*|) **do**
        index = *GraphProperty*($MatricesConditions_{c,s}$, *candidateMain*)
        $MatricesIndices_{c,s}$ = *index*
    P-Value = *StatisticalTest*(*MatricesIndices*)
    add *P-Value* to *subnetP-Values*
  MainIndex = *indexOfMin*(*subnetP-Values*)
  MainROIs = $(CN)^I_{MainIndex}$

  // Stage 2
  UnsortedSuppROIs = *Union*[$(CN)^I$] − *MainROIs*
  nodalP-Values = ∅
  for *roi* ∈ *UnsortedSuppROIs* **do**
    MatricesIndices = *InitMatrix*(|*Conditions*|, |*Subjects*|)
    for *c* ∈ *range*(|*Conditions*|) **do**
      for *s* ∈ *range*(|*Subjects*|) **do**
        index = *GraphProperty*($MatricesConditions_{c,s}$, {*roi*})
        $MatricesIndices_{c,s}$ = *index*
    P-Value = *StatisticalTest*(*MatricesIndices*)
    add *P-Value* to *nodalP-Values*
  SuppROIs = *sort*(*UnsortedSuppROIs*, *nodalP-Values*)

  return *MainROIs*, *SuppROIs*

Definition 4. The supplementary cognitive network noted as *SCN* and formulated in Equation 6 refers to any instances of *CN* excluding *MCN*.

$$SCN \in (CN)^I - \{MCN\} \tag{6}$$

The selection procedure of *MCN* and *SCN* is depicted in **Algorithm 1**, with the manipulated data *MatricesConditions*. The input parameters include $(CN)^I$, *Conditions*, and *Subjects*, and the output parameters include *MainROIs* and *SuppROIs*. $(CN)^I$ is the set of all instances of *CN*, while *Conditions* and *Subjects* are arrays containing the conditions and subjects of the cognitive task respectively. The returned *MainROIs* are a non-priority list of ROIs from *MCN*, and *SuppROIs* are a prior list of ROIs from all *SCNs*. **Algorithm 1** can be divided into two stages as follows:

Stage 1: Select the cognitive network in $(CN)^I$ with the optimal performance as the *MCN*, and append the ROIs of *MCN* into a non-priority list as *MainROIs*.

Stage 2: Merge the ROIs of all *SCNs* in $(CN)^I$ into a non-priority list as *UnsortedSuppROIs*, and reorder the ROIs in *UnsortedSuppROIs* into a priority list as *SuppROIs* according to their nodal graph property.

The data *MatricesConditions* is a four-dimensional matrix, where the first dimension refers to the conditions of the cognitive task, and the dimension size stands for the number of conditions. For instance, if the mental arithmetic task only involve the conditions of addition and subtraction, the dimension size will be set as two. The subjects recruited for the cognitive task stand for the second dimension, thus this dimension size is measured by the number of subjects. In the present study, twenty-one subjects participated in the present mental arithmetic task, and thus, the subjects' dimension size is 21. The third and fourth dimensions of

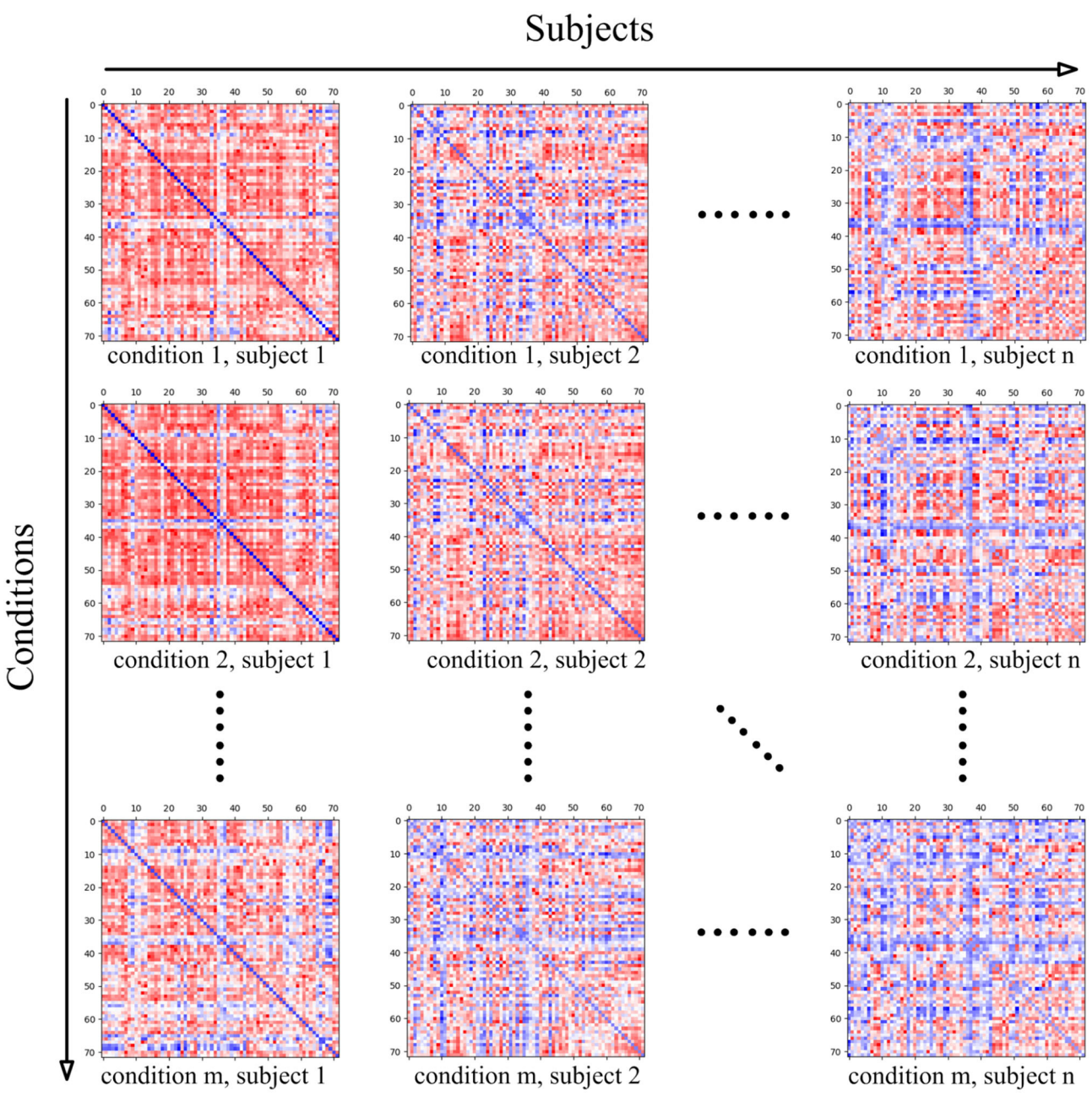

FIGURE 4 | The four-dimensional matrix *MatricesConditions*, where the first dimension represents the conditions in the cognitive task, and the second dimension represents the subject. When the condition and subject are set, the obtained two-dimensional matrix stands for the brain functional connectivity matrix of the subject under the current cognitive task condition, and the shape of the two-dimensional matrix is determined by the ROIs' number of the adopted cognitive network.

the matrix both represent the nodes contained in the CCN. Since the combined FPN here contains a total of 72 nodes, the sizes of the two dimensions are both 72 in the current case.

The intermediate two-dimensional matrix $MatricesConditions_{c,s}$ at both stages is the sub-matrix of *MatricesConditions* where *c* represents the index of the cognitive task condition in *Conditions* and *s* represents the index of the subject in *Subjects*. $MatricesConditions_{c,s}$ basically stands for the functional connectivity matrix of the subject $Subjects_s$ under the cognitive task condition $Conditions_c$. Since the functional connectivity matrix of $MatricesConditions_{c,s}$ is constructed on the basis of *CCN*, the shape of $MatricesConditions_{c,s}$ is $|CCN| \times |CCN|$. The relationship between *MatricesConditions* and its sub-matrix $MatricesConditions_{c,s}$ is shown in **Figure 4**.

The functional connectivity matrix of $MatricesConditions_{c,s}$ for each subject under specific cognitive conditions has been given, yet what actually needs to be calculated is the graph property of the candidate's main cognitive network, noted as *candidateMain* in the top loop at Stage 1, as well as the nodal graph property of the ROI, noted as *roi* in the top loop at Stage 2. Since *candidateMain* is a proper subset of *CCN*, and *roi* is an element of *CCN*, the graph property calculation implemented by the function of *GraphProperty* is based on $MatricesConditions_{c,s}$ as the first parameter. The second parameter of *GraphProperty* is of great significance in that its setting size can determine whether the calculating processing is targeted at graph property or nodal graph property. The metric choice of graph property includes degree centrality, clustering coefficient, and network efficiency. In the present study, degree centrality is set as the metric for its popular application in discriminating the graph properties under different cognitive states.

The intermediate two-dimensional matrix *MatricesIndices* at the two stages is used for storing the performance of each candidate's main cognitive network in $(CN)^I$ or the performance of each ROI in *UnsortedSuppROIs*. The shape of *MatricesIndices* is $|Conditions| \times |Subjects|$, and $MatricesIndices_{c,s}$ is the cell in the *cth* row and the *sth* column to store the performance corresponding to subject $Subjects_s$ under cognitive condition $Conditions_c$.

The function *StatisticalTest* at the two stages is used to conduct the statistical test between the rows of *MatricesIndices*, and return the *P-Value* which represents the performance in discriminating the graph properties under different cognitive states. The row number, namely the number of cognitive conditions, of *MatricesIndices* needs to be considered in the choice of a specific test function. If the row number of *MatricesIndices* is two, the statistical analysis of the *t*-test or the χ^2 test can be adopted, yet if the row number is greater than two, the statistical analysis of variance needs to be utilized.

When the *P-Value* of each *candidateMain* at Stage 1 is obtained, the cognitive network with the minimum *P-Value* is selected as the main cognitive network, and the ROIs in the main cognitive network are appended into the non-priority list *MainROIs*. Similarly, after the acquisition of *P-Value* of each *roi* from *UnsortedSuppROIs* at Stage 2, the ROIs in *UnsortedSuppROIs* are reordered and appended into the prior list *SuppROIs*. Finally, the *MainROIs* and *SuppROIs* are returned by the algorithm.

Algorithm 2: Cognitive networks fusion.

Data: *MatricesConditions*
Input: *Conditions*, *Subjects*, *MainROIs*, *SuppROIs*
Output: *FusedROIs*, *FusedP-Value*
begin
  // Stage 1
  $P\text{-}Values = \emptyset$
  $iterFusedROIs = MainROIs$
  for $roi \in SuppROIs$ **do**
    add *roi* to *iterFusedROIs*
    $MatricesIndices = InitMatrix(|Conditions|, |Subjects|)$
    for $c \in range(|Conditions|)$ **do**
      for $s \in range(|Subjects|)$ **do**
        $index = GraphProperty(MatricesConditions_{c,s}, iterFusedROIs)$
        $MatricesIndices_{c,s} = index$
    $P\text{-}Value = StatisticalTest(MatricesIndices)$
    add *P-Value* to *P-Values*

  // Stage 2
  $FusedIndex = indexOfMin(P\text{-}Values)$
  $FusedP\text{-}Value = P\text{-}Values_{fusedIndex}$
  $FusedROIs = MainROIs$
  for $i = 0; i \leq fusedIndex; i++$ **do**
    add $SuppROIs_i$ to *FusedROIs*

  return *FusedROIs*, *FusedP-Value*

2.3. Searching a Fused Cognitive Network

When *MCN*, consisting of the ROIs in *MainROIs*, is selected and set as the main cognitive network, it will be utilized as the base for fusing the ROIs, which are stored in the priority list *SuppROIs*, from all the supplementary cognitive networks.

Definition 5. Fused cognitive network, noted as *FCN* and formulated in Equation 7, is the union of *MCN* and the set, noted as $Sub(SuppROIs, fusedIndex)$, constituted by the first number of *fusedIndex* ROIs in *SuppROIs*, and has the optimal performance in discriminating the graph properties under different cognitive states for the relative task.

$$FCN = MCN \bigcup Sub(SuppROIs, fusedIndex) \tag{7}$$

The second parameter in *Sub* operation determines how many ROIs are chosen from the beginning of *SuppROIs* to constitute the set, and its value ranges from 0 to $|SuppROIs|$. $Sub(SuppROIs, 0)$ means an empty set, and $Sub(SuppROIs, |SuppROIs|)$ means the set containing all the ROIs in *SuppROIs*. Since *FCN* has the optimal performance, it can be inferred that $P(MCN \bigcup Sub(SuppROIs, fusedIndex)) \geq P(MCN \bigcup Sub(SuppROIs, i))$, where *P* is the same as the one defined in Equation 5, and $i \neq fusedIndex$.

Therefore, the essence of searching the *FCN* is to find the value of *fusedIndex* and to integrate the ROIs in $Sub(SuppROIs, fusedIndex)$ into *MCN* to construct the *FCN*. The search process of the *fusedIndex* is implemented

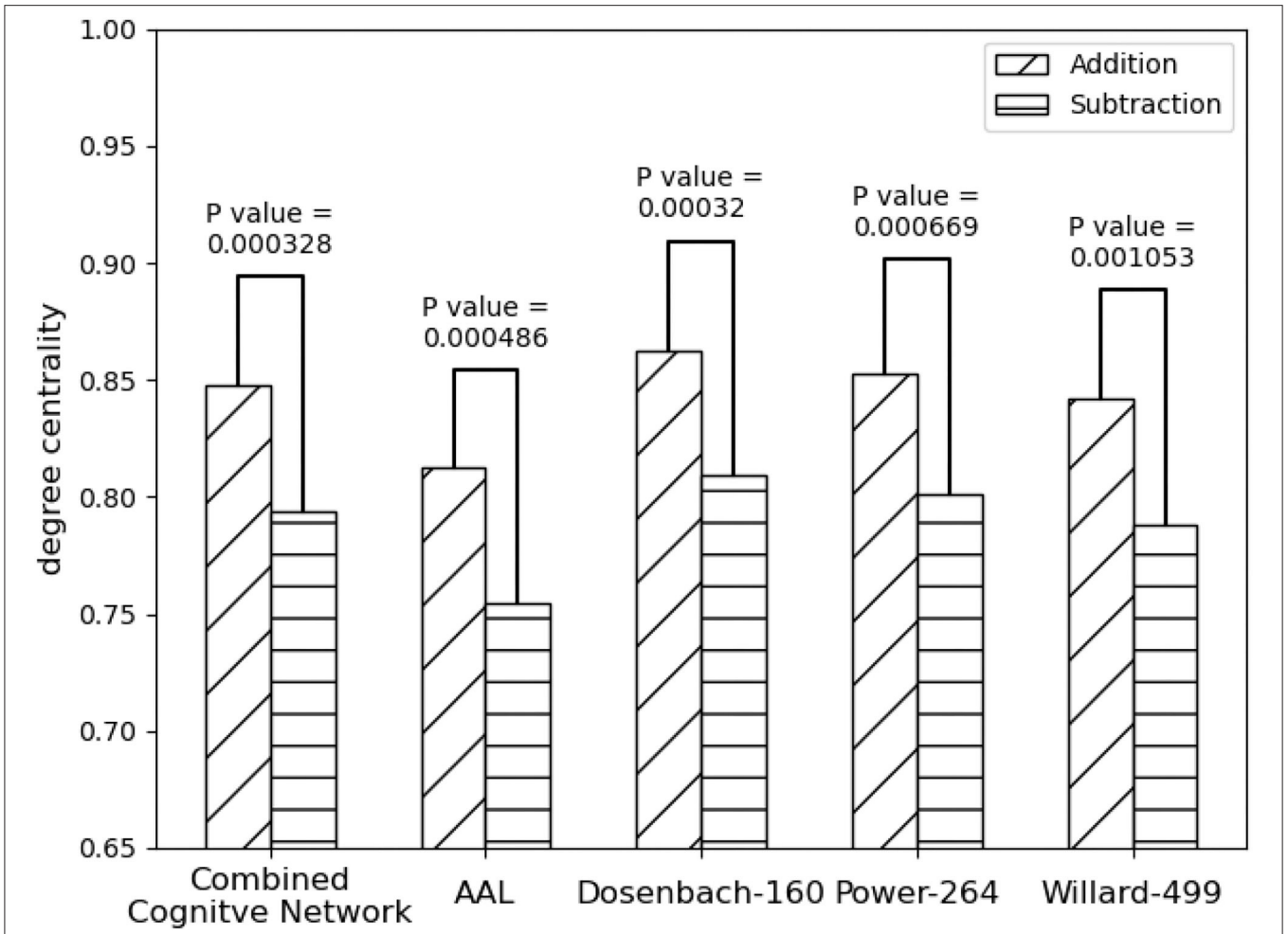

FIGURE 5 | Statistical comparison results of the FPNs' topological property (degree centrality) between the addition and subtraction cognitive states. The FPN of Dosenbach-160 has yielded the sharp differences with the biggest statistical significance between the two cognitive states of mental arithmetic, followed by the FPNs of CCN (Combined Cognitive Network), AAL, Power-264, and Willard-499.

by **Algorithm 2**, which can be divided into two stages as follows:

Stage 1: Initialize the candidate fused cognitive network, represented as the list *iterFusedROIs*, with the ROIs in the non-priority list *MainROIs*. Then, add the ROIs in *SuppROIs* into the candidate fused cognitive network iteratively and append the performance of the current candidate fused cognitive network into the list of *P-Values*.

Stage 2: Select the minimum *P-Value* in *P-Values* and its corresponding position, namely *FusedIndex*, in the list. The ROI at the position of *FusedIndex* and all its leading ROIs in *SuppROIs*, together with the ROIs in *MainROIs*, constitute the final fused cognitive network *FCN* with the optimal performance in discriminating the graph properties under different cognitive states.

MatricesConditions, *Conditions*, and *Subjects* are the same as the ones in **Algorithm 1** and will be reused. The returned *MainROIs* and *SuppROIs* by **Algorithm 1** are set as the input parameters here. The output results, namely *FusedROIs* and *FusedP-Value*, are the generated list of ROIs in the optimal fused cognitive network *FCN* and its performance, respectively.

Despite the optimal performance in the entire iterative searching process, further comparisons between the fused cognitive network and the results obtained from other typical machine learning methods are necessity.

3. EXPERIMENTS AND RESULTS

3.1. Experiment

The fMRI data comes from the mental arithmetic task of simple addition and subtraction (Yang et al., 2017) designed by the Web Intelligence Consortium (WIC). The goal of this cognitive task is to study the regularity of brain neural activity in simple arithmetic operations. Twenty-one subjects (12 males, 9 females) with no statistically significant differences were recruited. Before the experiment, each subject was made clear about the possible natural responses during the task process, and all subjects signed an informed consent form. After obtaining

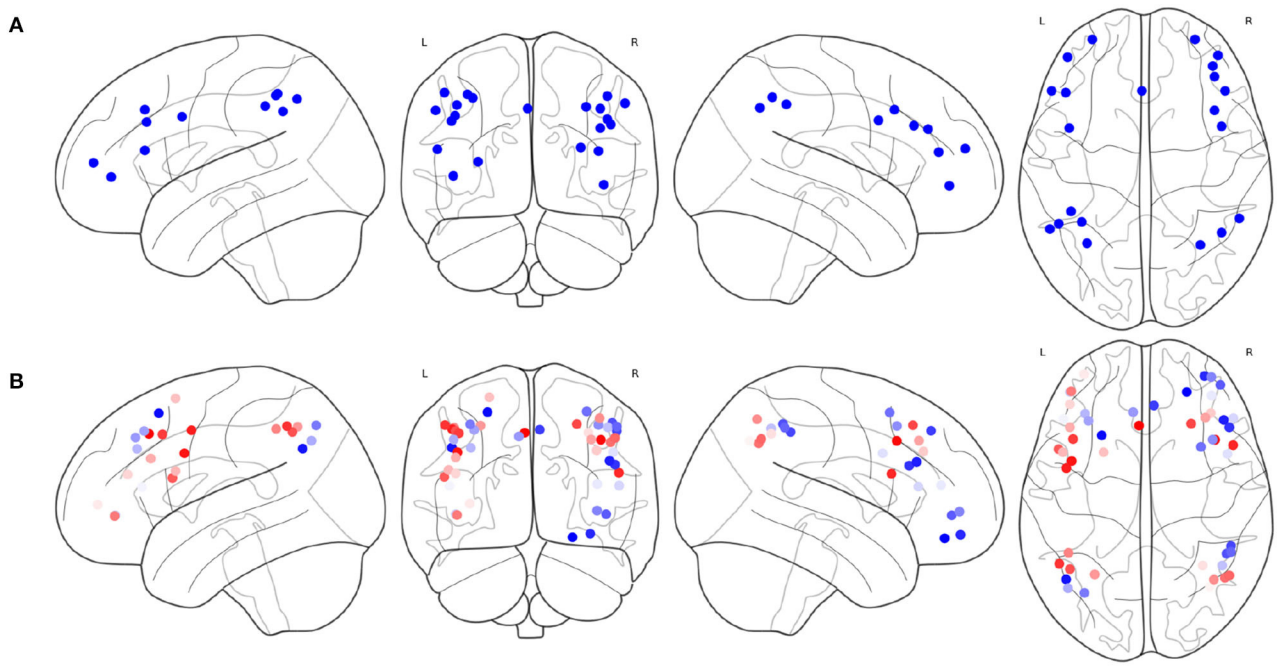

FIGURE 6 | Results of the main FPN and supplementary FPNs, **(A)** the ROIs in *MainROIs*, i.e., the FPN in Dosenbach-160, **(B)** the ROIs in all the supplementary FPNs, i.e., the FPNs in AAL, Power-264, and Willard-499. ROI with warmer colors has a higher priority in *SuppROIs*. Conversely, ROI with cooler color has a lower priority in *SuppROIs*.

permission from the Ethics Committee of Xuanwu Hospital of Capital Medical University, the cognitive task was implemented and the fMRI data were collected by the WIC team in the hospital.

The preprocessing of the fMRI data was conducted with the software of Statistical Parameter Mapping (Friston et al., 2007) in four steps, namely slice time correction, head motion correction, spatial normalization, and smoothing. The Python software package and other relative software packages were used for the graph analysis of the fMRI data and the implementation of the cognitive networks fusion algorithm. The NiBabel was used for the basic manipulation of neuroimaging files like fMRI data and the SciPy (Virtanen et al., 2020) was used for Pearson coefficient calculation, FDR correction, and statistical test. Topological indices in the graph analysis were calculated by NetworkX (Hagberg et al., 2008), and the results were visualized *via* NiLearn.

3.2. Results

As shown in **Figure 5** all the instances of *FPN*, FPN_d has the optimal performance of the statistical test (*P-Value* = 0.00032) between the degree centralities of the functional connectivity matrices under the two mental arithmetic cognitive states, followed by *CombinedFPN* (*P-Value* = 0.000328), FPN_a (*P-Value* = 0.000486), FPN_p (*P-Value* = 0.000669), and FPN_w (*P-Value* = 0.001053), respectively. Consequently, FPN_d is set as *MCN*, and the remaining instances of *FPN* are used as *SCN* in the following fusion of cognitive networks.

The result of **Algorithm 1** is shown in **Figure 6**. FPN_d is selected as the main FPN and all the ROIs in FPN_d are in the returned non-priority list *MainROIs* as in **Figure 6A**. FPN_a, FPN_p, and FPN_w are all selected as the supplementary FPNs, all the ROIs are in the returned priority list *SuppROIs*. As shown in **Figure 6B**, the ROI with warmer color has a higher priority and will be in closer propinquity to the head of *SuppROIs*, while the ROI with cooler color has a lower priority and will be in closer propinquity to the tail of *SuppROIs*. The position of the ROI in *SuppROIs* determines the time when it will be added into the candidate FPN in the iterative process of **Algorithm 2**.

The iterative process of **Algorithm 2** is displayed in **Figure 7**. FPN_d is used as the initial candidate fused FPN, and the ROIs in *SuppROIs* are added into the candidate fused FPN one by one. The performance of the candidate fused FPN in discriminating the graph properties under the two mental arithmetic conditions is statistically tested during each iteration. The results show that the candidate fused FPN, formed by the ROIs of FPN_d and the first 30 ROIs in *SuppROIs*, has the optimal performance. The triangular part of the inferior frontal gyrus from the right hemisphere of AAL is eventually added into the resulting fused FPN with 51 ROIs. As shown in **Table 2**, the numbers of ROIs belonging to the frontal and parietal lobes in the fused FPN are 33 and 18, respectively. More specifically, among the 33 ROIs in the frontal lobe, the numbers from the FPNs in brain atlas of AAL, Dosenbach-160, Power-264, and Willard-499 are 5, 13, 11, and and 4 respectively while the corresponding numbers are 2, 8, 5, and 3 in the parietal lobe respectively.

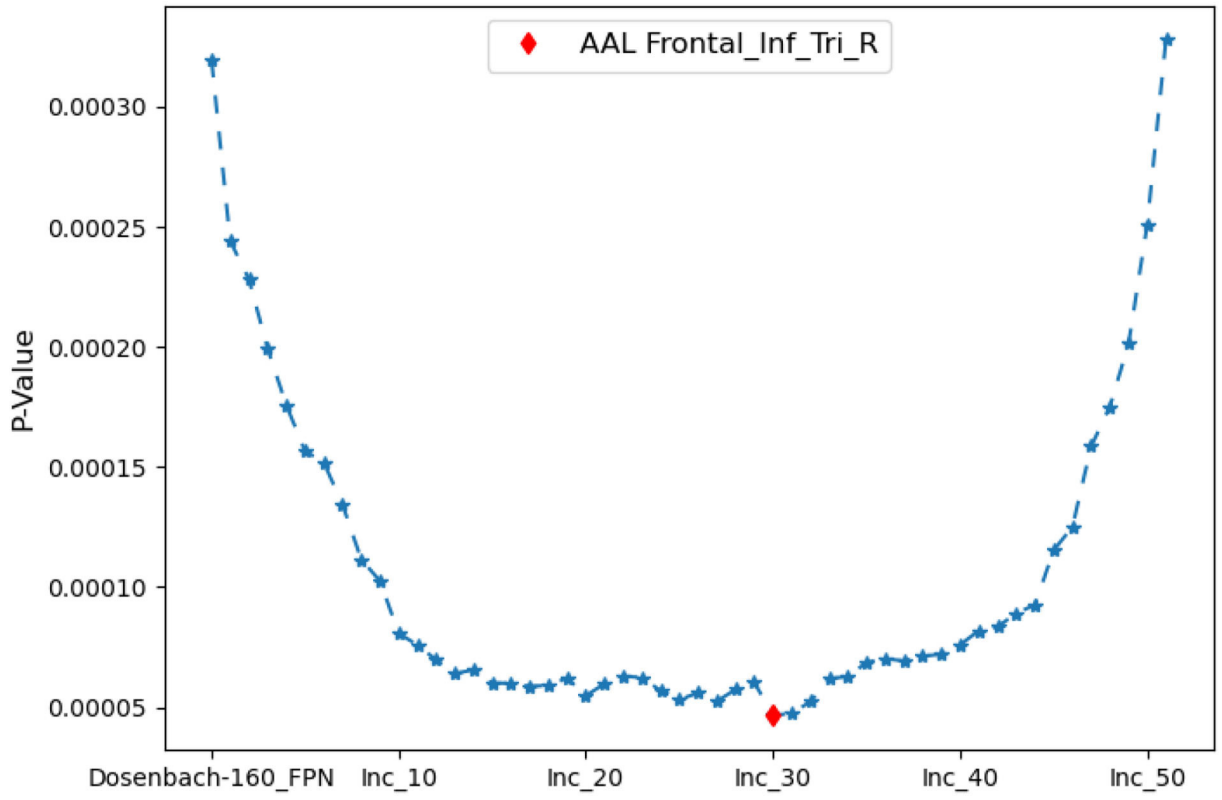

FIGURE 7 | The iterative statistical test of the graph properties of the candidate fused FPNs in discriminating the cognitive states. When Dosenbach-160 is set as the initial candidate fused FPN, and the top thirty of all ROIs in AAL, Power-264, and Willard-499 are added to Dosenbach-160, the candidate fused FPN can produce the most significant statistical performance. The last ROI to be added to the fused FPN is the triangular part of the inferior frontal gyrus from the right hemisphere of AAL.

The spacial differences can be intuitively identified between the distributions of the ROIs of the generated FPNs with a close eye on **Figure 8**. **Figures 8A–D** display the FPNs with the top 51 ROIs chosen by ExtraTrees, AdaBoost, RandomForest, and XGB, respectively. The fused FPN calculated by **Algorithm 2** is shown in **Figure 8E**. The numbers of ROIs located in the frontal lobe and parietal lobe are 31/20 (ExtraTrees), 36/15 (AdaBoost), 35/16 (RandomForest), 35/16 (XGB), and 33/18 (**Algorithm 2**). All these algorithms are conducted on the basis of *CombinedFPN*. Since *CombinedFPN* yields 47 ROIs and 25 ROIs in the frontal lobe and parietal lobe, respectively, it could be judged that the distribution ratio of ROIs in the two lobes generated from **Algorithm 2** is close to that of *CombinedFPN*.

On the other hand, compared with the adopted machine learning algorithms, **Algorithm 2** has the optimal performance, as shown in **Figure 9** and **Table 3**, of the statistical test ($P\text{-}Value = 4.7e\text{-}05$) between the degree centralities of the functional connectivity matrices under the two mental arithmetic cognitive states, followed by that of ExtraTrees ($P\text{-}Value = 0.000291$), XGB ($P\text{-}Value = 0.000353$), RandomForest ($P\text{-}Value = 0.000372$), and AdaBoost ($P\text{-}Value = 0.000453$). Such a performance is even better than that of *CombinedFPN* ($P\text{-}Value = 0.000328$). In a word, it can be safely concluded that **Algorithm 2** can choose the FPN better representing the cognitive states of mental arithmetic.

4. DISCUSSION

In the present study, the graph properties of the multi-source FPNs, combined FPN, fused FPN, and the FPNs generated by the adopted machine learning methods can be effectively used to discriminate the graph properties under different cognitive states in mental arithmetic task. Such a result lends support to the previous studies about the major dependence of adults' arithmetic ability on the FPN. To date, many research methods have been adopted to explore the possible role of the FPN in mental arithmetic, such as the meta-analysis of the brain regions involved in numbers and mental arithmetic (Arsalidou and Taylor, 2011), the pathway analysis on the brain mental arithmetic (Dehaenea and Cohen, 1997), and the structural connection analysis on the code model involved in mental arithmetic (Klein et al., 2013, 2016). All these studies proved that the mental arithmetic processing can activate the adult brain's FPN, which consists of the superior parietal lobule (SPL) and inferior parietal lobule (IPL) in the parietal regions, and inferior frontal gyrus (IFG), middle frontal gyrus (MFG), and left superior frontal in frontal regions.

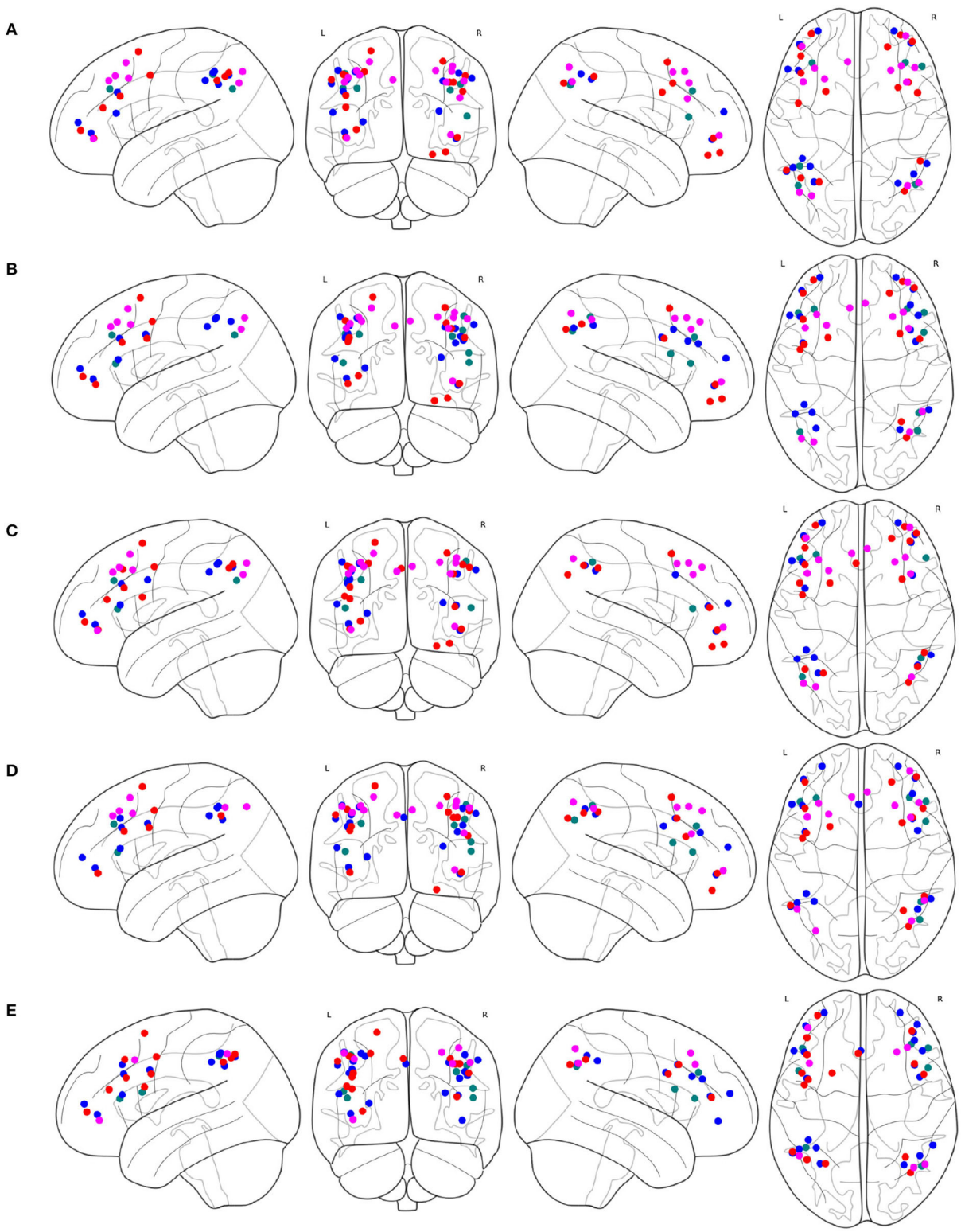

FIGURE 8 | FPNs with 51 ROIs calculated by different methods. ROIs with colors teal, blue, red, magenta come from AAL, Dosenbach-160, Power-264, and Willard-499, respectively, **(A)** 6, 14, 17, and 14 ROIs are selected by ExtraTrees from the four brain atlases, respectively, **(B)** 8, 17, 13, and 13 ROIs are selected by AdaBoost from the four brain atlases, respectively, **(C)** 5, 15, 18, and 13 ROIs are selected by RandomForest from the four brain atlases, respectively, **(D)** 7, 17, 14, and 13 ROIs are selected by XGB from the four brain atlases, respectively, **(E)** 7, 21, 16, and 7 ROIs are selected by the cognitive networks fusion algorithm from the 4 brain atlases, respectively.

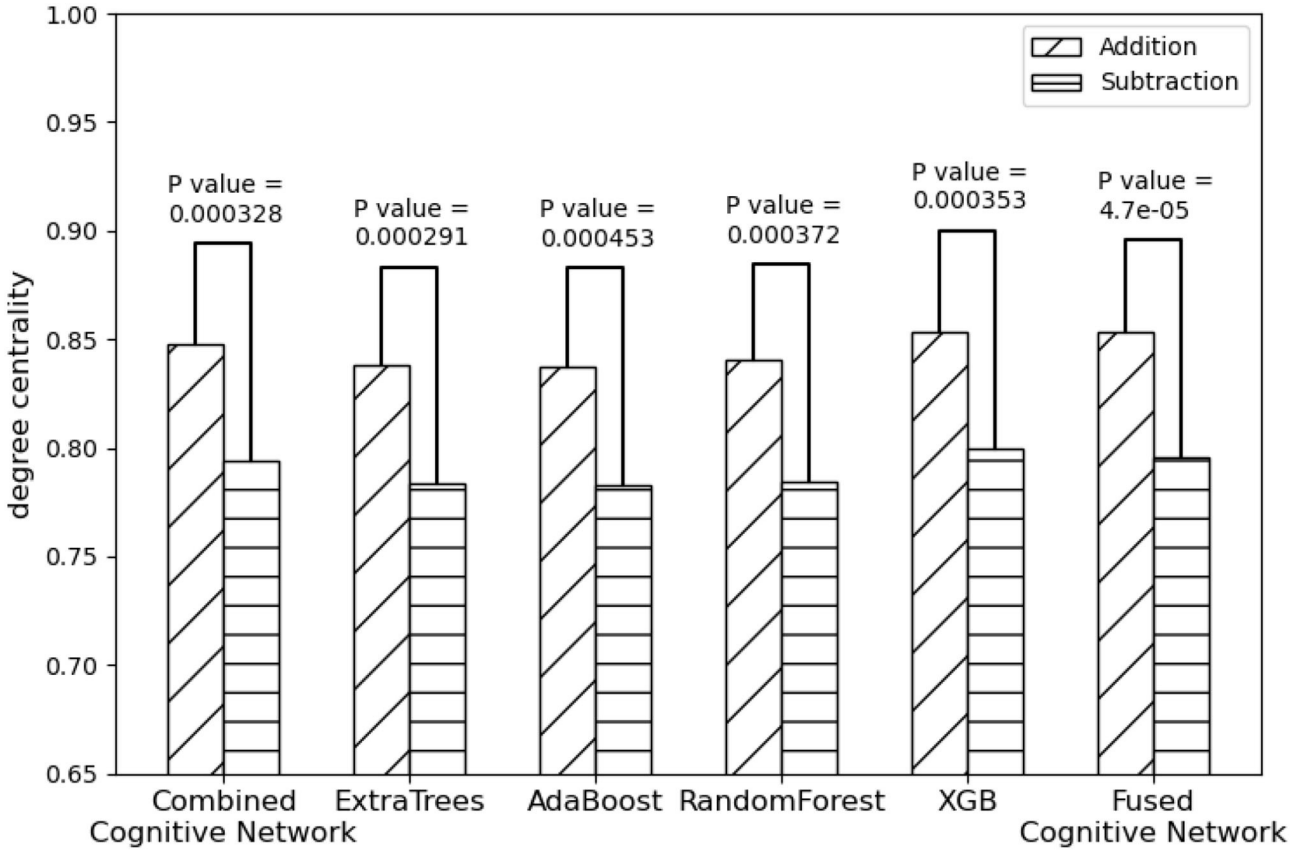

FIGURE 9 | Statistical comparison results of the FPNs' topological properties between the addition cognitive state and the subtraction cognitive state. Among them, the fused FPN by the cognitive networks fusion algorithm has yielded the sharp differences with the biggest statistical significance between the two cognitive states of mental arithmetic, followed by the FPNs generated by ExtraTrees, CCN (Combined Cognitive Network), XGB, RandomForest, and AdaBoost.

The FPN plays an important part in adults' mental arithmetic processing, and a similar network was also detected in children's mental arithmetic processing with experiment and retrospective analysis (Peters and De Smedt, 2018). It is generally considered that the FPN is in charge of perceiving the top-down activity regulation of the cortex for attention preparation and memory orientation. CON also plays a vital role in the cognitive control, whose downstream effect may be attributed to the output gating of memory. Thus, both the FPN and CON were indispensable in controlling working memory (Wallis et al., 2015). By adopting a functional connectivity analysis, the working memory in the mental arithmetic tasks was also explored and the collaborative work between the frontal lobes and parietal lobes in working memory tasks was detected as well (Hagiwara et al., 2016). However, the DMN was found to be passivated in the mental arithmetic processing, which might be caused by the inhibitory effect of functional network activation during the cognitive tasks (Dimitriadis et al., 2010).

With regards to the big variations in the topological structure of the FPN, it is likely to result from the significant differences in the graph properties between the cognitive states of addition and subtraction in performing the mental arithmetic processing. *Via* the graph analysis and statistical analysis, significant differences can be identified in the adopted metric of the FPN's degree centrality from different brain atlases under the two mental arithmetic cognitive states. Such a result is consistent with the findings of previous studies. For example, the subjects achieved obvious improvements in their mathematics skills and the FPN activities after attending the adaptive number-sense training. It was found that the activation of the subjects' bilateral parietal lobe significantly increased, while the activation of their frontal striatum and middle temporal lobe decreased considerably (Kesler et al., 2011). Moreover, more activation of the FPN is thought to be generated in the numerical inductive reasoning, such as the mental arithmetic process, because more exchanges might be transacted between the intermediate representations and long-term declarative knowledge in the process of numerical rule recognition (Liang et al., 2016). On top of that, the white matter dispersion property of the FPN was also detected to be effective in the prediction of children's mental arithmetic ability (Tsang et al., 2009).

5. CONCLUSION

Focusing on the multi-source cognitive networks, this study takes the single-source cognitive network with the

TABLE 3 | Performance of each method in discriminating cognitive states.

Method	Mean of degree centrality		*P*-Value
	Addition	Subtraction	
Combined Cognitive Network	0.8474	0.7941	0.000328
ExtraTress	0.8384	0.7833	0.000291
AdaBoost	0.8371	0.7830	0.000453
RandomForest	0.8401	0.7845	0.000372
XGB	0.8536	0.7999	0.000353
Fused Cogntive Network	0.8536	0.7957	0.000047

optimal performance as the main cognitive network through synthesizing the multi-source cognitive networks. The ROIs in the supplementary cognitive networks are sorted and integrated into the main cognitive network iteratively, so as to search for the fused cognitive network with the optimal performance in discriminating the graph properties under different cognitive states. The potential advantages of the present research method can be summarized as follows:

1. The distribution of the obtained ROIs in the fused FPN is spatially closer to that of the combined cognitive network, and the ROIs selected are better balanced between the frontal lobe and parietal lobe.
2. The fused cognitive network is constructed under the framework of the main cognitive network by integrating the ROIs with top priority in the supplementary cognitive networks. In the analysis of the fMRI data, the fused cognitive network relies on the main cognitive network for the major interpretation, together with the ROIs of supplementary cognitive networks for the complementary explanation.
3. Compared with other typical machine learning algorithms, the proposed method can yield better performance and the results bear more self-consistency to those obtained in cognitive neuroscience.

On the whole, it has proved that the proposed method can produce a satisfactory evaluation performance and provide a more reasonable interpretation for the related cognitive neuroscience research. However, the potential impact of the fused cognitive network on the cognitive computing model waits for further explorations. Additionally, the generality of such a proposed method also waits for further validations with diverse brain atlases and various fMRI datasets.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Xuanwu Hospital of Capital Medical University Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XZ: conceptualization of this study, methodology design and implementation, analysis and interpretation of data, and writing–original draft. YY: cognitive task design and implementation, acquisition of fMRI data, and interpretation of data. HK, JC, and JH: methodology design and implementation. PL: conceptualization of this study and interpretation of data. NZ: conceptualization of this study, methodology design, interpretation of data, and final approval of the version. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (61420106005), the National Key Research and Development Program of China (No. 2020YFB2104402), the JSPS Grants-inAid for Scientific Research of Japan (19K12123), and the Major Projects of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province, and the Beijing Natural Science Foundation (No. 4222022).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.866734/full#supplementary-material

REFERENCES

Arsalidou, M., and Taylor, M. J. (2011). Is 2+2=4? meta-analyses of brain areas needed for numbers and calculations. *Neuroimage* 54, 2382–2393. doi: 10.1016/j.neuroimage.2010.10.009

Colom, R., Karama, S., Jung, R. E., and Haier, R. J. (2010). Human intelligence and brain networks. *Dialogues Clin. Neurosci.* 12, 489–501. doi: 10.31887/DCNS.2010.12.4/rcolom

Dehaenea, S., and Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex* 33, 219–250. doi: 10.1016/S0010-9452(08)70002-9

Dimitriadis, S. I., Laskaris, N. A., Tsirka, V., Vourkas, M., and Micheloyannis, S. (2010). What does delta band tell us about cognitive processes: a mental calculation study. *Neurosci. Lett.* 483, 11–15. doi: 10.1016/j.neulet.2010.07.034

Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., and Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends Cogn. Sci.* 12, 99–105. doi: 10.1016/j.tics.2008.01.001

Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., et al. (2010). Prediction of individual brain maturity using fMRI. *Science* 329, 1358–1361. doi: 10.1126/science.1194144

Fornito, A. (2016). Graph theoretic analysis of human brain networks. *fMRI Techn. Protocols* 119, 283–314. doi: 10.1007/978-1-4939-5611-1_10

Friston, K., Ashburner, J., Kiebel, S., Nichols, S., and Penny, W. (2007). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. London: Academic Press.

Gao, W., and Lin, W. (2012). Frontal parietal control network regulates the anti-correlated default and dorsal attention networks. *Hum. Brain Mapp*. 33, 192–202. doi: 10.1002/hbm.21204

Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). "Exploring network structure, dynamics, and function using networkx," in *Proceedings of the 7th Python in Science Conference* (Pasadena, CA), 11–15. Available online at: https://conference.scipy.org/proceedings/SciPy2008/index.html

Hagiwara, R., Hiwa, S., and Hiroyasu, T. (2016). "Functional connectivity analysis of working memory during a mental arithmetic task," in *Proceedings of the 2016 Annual International Neuroinformatics Coordinating Facility Congress* (Reading, UK), 32.

Kesler, S. R., Sheau, K., Koovakkattu, D., and Reiss, A. L. (2011). Changes in frontal-parietal activation and math skills performance following adaptive number sense training: preliminary results from a pilot study. *Neuropsychol. Rehabil.* 21, 433–454. doi: 10.1080/09602011.2011.578446

Klein, E., Moeller, K., Glauche, V., Weiller, C., and Willmes, K. (2013). Processing pathways in mental arithmetic–evidence from probabilistic fiber tracking. *PLoS ONE* 8, e55455. doi: 10.1371/journal.pone.0055455

Klein, E., Suchan, J., Moeller, K., Karnath, H. -O., Knops, A., Wood, G., et al. (2016). Considering structural connectivity in the triple code model of numerical cognition: differential connectivity for magnitude processing and arithmetic facts. *Brain Struct. Funct*. 221, 979–995. doi: 10.1007/s00429-014-0951-1

Lang, S., Gaxiola-Valdez, I., Opoku-Darko, M., Partlo, L. A., Goodyear, B. G., Kelly, J. J. P., et al. (2017). Functional connectivity in frontoparietal network: Indicator of preoperative cognitive function and cognitive outcome following surgery in patients with glioma. *World Neurosurg.* 105, 913.e2–922.e2. doi: 10.1016/j.wneu.2017.05.149

Li, W., Yang, P., Ngetich, R. K., Zhang, J., Jin, Z., and Li, L. (2021). Differential involvement of frontoparietal network and insula cortex in emotion regulation. *Neuropsychologia* 161, 107991. doi: 10.1016/j.neuropsychologia.2021.107991

Liang, P., Jia, X., Taatgen, N. A., Borst, J. P., and Kuncheng, L. (2016). Activity in the fronto-parietal network indicates numerical inductive reasoning beyond calculation: an fMRI study combined with a cognitive model. *Sci. Rep*. 6, 25976. doi: 10.1038/srep25976

Marek, S., and Dosenbach, N. U. (2018). The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. *Dialogues Clin. Neurosci*. 20, 133–141. doi: 10.31887/DCNS.2018.20.2/smarek

Mizuhara, H., and Yamaguchi, Y. (2007). Human cortical circuits for central executive function emerge by theta phase synchronization. *Neuroimage* 36, 232–244. doi: 10.1016/j.neuroimage.2007.02.026

Naghavi, H. R., and Nyberg, L. (2005). Common fronto-parietal activity in attention, memory, and consciousness: shared demands on integration? *Conscious Cogn*. 14, 390–425. doi: 10.1016/j.concog.2004.10.003

Oliver, I., Hlinka, J., Kopal, J., and Davidsen, J. (2019). Quantifying the variability in resting-state networks. *Entropy* 21, 882. doi: 10.3390/e21090882

Peters, L., and De Smedt, B. (2018). Arithmetic in the developing brain: a review of brain imaging studies. *Dev. Cogn. Neurosci*. 30, 265–279. doi: 10.1016/j.dcn.2017.05.002

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., et al. (2011). Functional network organization of the human brain. *Neuron*. 72, 665–678. doi: 10.1016/j.neuron.2011.09.006

Richiardi, J., Altmann, A., Milazzo, A.-C., Chang, C., Chakravarty, M. M., Banaschewski, T., et al. (2015). Correlated gene expression supports synchronous activity in brain networks. *Science* 348, 1241–1244. doi: 10.1126/science.1255905

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *J. Neurosci*. 27, 2349–2356. doi: 10.1523/JNEUROSCI.5587-06.2007

Sheffield, J. M., Repovs, G., Harms, M. P., Carter, C. S., Gold, J. M., MacDonald, A. W., et al. (2015). Fronto-parietal and cingulo-opercular network integrity and cognition in health and schizophrenia. *Neuropsychologia* 73, 82–93. doi: 10.1016/j.neuropsychologia.2015.05.006

Tsang, J. M., Dougherty, R. F., Deutsch, G. K., Wandell, B. A., and Ben-Shachar, M. (2009). Frontoparietal white matter diffusion properties predict mental arithmetic skills in children. *Proc. Natl. Acad. Sci. U.S.A*. 106, 22546–22551. doi: 10.1073/pnas.0906094106

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., and Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *J. Neurophysiol.* 100, 3328–3342. doi: 10.1152/jn.90355.2008

Virtanen, P., Gommers, R., Oliphant, T. E., Reddy, T., Cournapeau, D., Burovski, E., et al. (2020). Scipy 1.0: fundamental algorithms for scientific computing in python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2

Wallis, G., Stokes, M., Cousijn, H., et al. (2015). Frontoparietal and cingulo-opercular networks play dissociable roles in control of working memory. *J Cogn Neurosci*. 27, 2019–2034. doi: 10.1162/jocn_a_00838

Yang, Y., Zhong, N., Friston, K., Imamura, K., Lu, S., Li, M., et al. (2017). The functional architectures of addition and subtraction: network discovery using fMRI and DCM. *Hum. Brain Mapp*. 38, 3210–3225. doi: 10.1002/hbm.23585

Zanto, T. P., and Gazzaley, A. (2013). Fronto-parietal network: flexible hub of cognitive control. *Trends Cogn. Sci*. 17, 602–603. doi: 10.1016/j.tics.2013.10.001

Zhong, N., Bradshaw, J. M., Liu, J., and Taylor, J. G. (2011). Brain Informatics. *IEEE Intell. Syst*. 26, 16–21. doi: 10.1109/MIS.2011.83

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership