

Distribution-guided Network Thresholding for Functional Connectivity Analysis in fMRI-based Brain Disorder Identification

Zhengdong Wang, Biao Jie*, Chunxiang Feng, Taochun Wang, Weixin Bian, Xintao Ding, Wen Zhou, Mingxia Liu*, *Senior Member, IEEE*

Abstract—Functional connectivity (FC) networks derived from resting-state functional magnetic resonance imaging (rs-fMRI) have been widely used in automated identification of brain disorders, such as Alzheimer’s disease (AD) and attention deficit hyperactivity disorder (ADHD). To generate compact representations of FC networks, various thresholding methods have been designed for FC network analysis. However, these studies usually use a pre-defined threshold or connection percentage to threshold whole FC networks, thus ignoring the diversity of temporal correlation (e.g., strong associations) between brain regions in subject groups. In this work, we propose a distribution-guided network thresholding learning (DNTL) method for FC network analysis in brain disorder identification with rs-fMRI. Specifically, for each connection of a pair of brain regions, we propose to determine its specific threshold based on the distribution of connection strength (i.e., temporal correlation) between subject groups (e.g., patients and normal controls). The proposed DNTL can adaptively yield an FC-specific threshold for each connection in an FC network, thus preserving diversity of temporal correlation among different brain regions. Experiment results on 365 subjects from two datasets (i.e., ADNI and ADHD-200) suggest that the DNT method outperforms state-of-the-art methods in brain disorder identification with rs-fMRI data.

Index Terms—Functional Connectivity, Thresholding, Brain Disorder, Identification, rs-fMRI

I. INTRODUCTION

HUMAN brain is a huge and complex system that depends on interactions among distributed brain regions [1]. This system can be characterized as a brain connectivity network, with nodes corresponding to brain regions and edges quantifying connections between brain regions. Thus, network analysis provides a meaningful way to investigate the organization of the brain, and also the association between brain functional deficit and structural disruption caused by brain disorders [2].

Manuscript received June 14, 2020; revised February 01, 2021, June 19, 2021 and August 05, 2021; accepted August 19, 2021. Date of current version August 20, 2021. This work was supported in part by the NSFC (Nos. 61976006, 61902003 and 61573023), NSF-AH (Nos. 1808085MF171 and 2108085MF206), and NGII (20190612).

Z. Wang, B. Jie, C. Feng, T. Wang, W. Bian, X. Ding, and W. Zhou are with the School of Computer and Information, Anhui Normal University, Wuhu 241003, China. M. Liu is with the Department of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599, USA.

*Corresponding authors: B. Jie (bjiao@nuaa.edu.cn) and M. Liu (mxliu1226@gmail.com).

Existing studies focus on studying the association between network properties and brain disorders, such as Alzheimer’s disease (AD) [3], mild cognitive impairment (MCI) [4], and attention deficit hyperactivity disorder (ADHD) [5].

As an advanced neuroimaging technique, functional magnetic resonance imaging (fMRI) provides us with an efficient way to study the functional interaction of distributed brain regions [6], and facilitate us to understand the pathology of brain diseases. By characterizing temporal correlation between brain regions via blood oxygen level-dependent (BOLD) signals in resting-state fMRI (rs-fMRI), functional connectivity (FC) networks have shown great promise in revealing brain functions [1], [7] and pathology of brain disorders [8]–[12], and have been successfully applied to brain disease analysis [13] and automated brain disorder identification [3], [14].

In the literature, graph object has shown its advantage in FC network analysis. For example, one can obtain many network measures (e.g., degree of node and clustering coefficient) to describe brain structural/functional connectivity properties, and to characterize the segregated and integrated nature of brain activity [15]. Also, the graph object provides a feasible solution to identify imaging biomarkers for brain disorder diagnosis [16], [17]. For example, graph object based on structural/functional connectivity networks has been successfully used to identify MCI individuals [18], [19] and ADHD patients [20]. Existing network measures are usually defined on *binary* graphs with connections/connectivity present (e.g., 1) or not (e.g., 0), while the connection strength (i.e., temporal correlation) between a specific pair of brain regions is generally measured continuously (e.g., via $[-1, 1]$). To facilitate the use of existing network measures, existing studies often employ various thresholding methods to construct binary networks for subsequent brain disease analysis [3], [18], [21].

In addition, to reduce the computational burden [22], the network thresholding strategy also provides two other advantages. *First*, with each connectivity in an FC network denoting the temporal correlation between brain regions, those uninformative/noisy connectivity could bring negative influence to identify patients from normal controls (NCs). Previous studies have shown that anatomical connections between brain regions are sparse, that is, a region may be sparsely connected to other regions [23]. Therefore, it is important to remove those uninformative/noisy connectivity via network threshold-

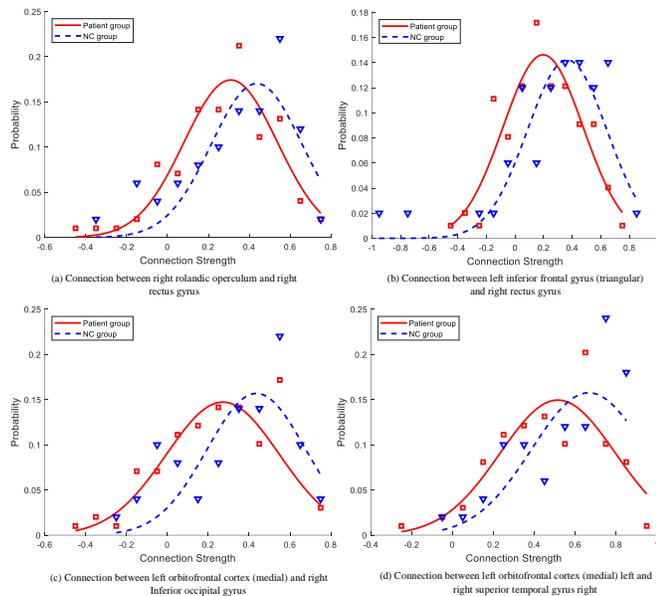


Fig. 1. Probability distributions of connection strength for four pairs ROIs in the MCI and NC subject groups from the real ADNI dataset. Here, for each pair of brain regions, we partition all possible values of connection strength into 20 intervals from -1 to 1 with a step size of 0.1 , and calculate the proportion of interval occurred in each subject group as the probability of connection strength. The red squares and blue triangles denote the probability of connection strength in MCI group and NC group, respectively.

ing [24]. Besides, it is convenient for users to extract network representations from thresholded FC networks [2], since most network measures are defined on binary networks [25], [26].

Existing thresholding methods can be roughly divided into two groups: 1) threshold-based approaches [3], [18], and 2) sparsity-based methods [22]. In the first category, FC networks are usually thresholded by using a pre-defined value, *i.e.*, an edge between brain regions exists if and only if the connection strength (*e.g.*, temporal correlation coefficient) is larger than a given threshold. In the second category, existing studies typically preserve a pre-defined percentage of connectivity with the strongest temporal correlation. Since these methods rely on a pre-defined threshold value or percentage of connections for network thresholding, they generally ignore the *diversity of temporal correlation* among different brain regions.

To visually show such diversity of temporal correlation in brain functional connectivity networks, we compute the probability distribution of connection strength among brain regions in the real ADNI database¹. The characteristics of these subjects are introduced in Section III-A. Figure 1 illustrates the obtained results for four pairs of brain regions for two groups (*i.e.*, MCI and NC). As shown in Fig. 1, the distributions of connection strength in the same subject group (*e.g.*, NC or MCI) are different for different pairs of brain regions. Also, the distributions of connection strength on the same pair of brain regions are different for both subject groups (*i.e.*, MCI and NC groups). These results demonstrate the diversity of temporal correlation among brain regions. In addition, Fig. 1 suggests that the distribution of

connection strength between the same pair of brain regions is very close to a normal/Gaussian distribution. Intuitively, network thresholding that considers the diversity information of temporal correlation could yield better representations of FC networks, and thus, help boost the performance for FC network based brain disorder identification.

In Fig. 2, we further illustrate two brain FC networks from an NC and MCI patient, respectively. Here, each node (denoted as a capital letter) corresponds to a brain region. As we can see from Fig. 2, the thresholded networks using both threshold-based and sparsity-based methods can not characterize the real difference of connections in two FC networks. When considering the diversity of temporal correlation among brain regions, it is reasonable to use FC-specific values to threshold the corresponding connections. From Fig. 2 (c), we can see that three FC-specific values are used for network thresholding by taking advantage of connection diversity among brain regions, while thresholded networks show the real difference between FC networks of the normal and the patient. However, it is a challenging task to determine all FC-specific thresholds for whole-brain FC networks.

In this work, we develop a Distribution-guided Network Thresholding (DNT) approach for rs-fMRI based functional connectivity analysis, and apply it to automated brain disorder identification. To be specific, we first split all training subjects into two groups based on their categories (*e.g.*, patient or NC). Then, for each pair of brain regions, we assume that the distribution of connection strength in each subject group follows a normal distribution, as shown in Fig. 1, and estimate this distribution using the corresponding training subjects. Finally, we construct an FC-specific threshold by employing both distributions across two subject groups. Different from existing methods, we define a threshold for each connectivity using its connection strength distribution information. Hence, our DNT can adaptively generate FC-specific thresholds for connections in FC networks, by preserving the diversity of temporal correlations between different brain regions. We evaluate the proposed DNT method on 365 subjects from the ADNI and ADHD-200 datasets with baseline rs-fMRI data. The experimental results demonstrate the efficacy of the proposed method in brain disorder identification.

The main contribution of this study is three-fold. *First*, we propose a distribution-guided network thresholding approach to analyze functional connectivity data. To the best of our knowledge, this is among the first attempts to develop a distribution-oriented network thresholding method by exploring diversity of temporal correlations in FC networks. *Second*, we develop a DNT based learning framework for automated brain disorder identification with rs-fMRI data. *Finally*, the proposed method is evaluated on 365 subjects from two rs-fMRI datasets, with results suggesting its effectiveness.

The rest of the paper is organized as follows. In Section II, we briefly review related studies. Then, we introduce subjects used in this study and present the proposed method in Section III. In Section IV, we introduce experimental setups and results. We further discuss the experimental results, the influence of parameters and limitations of the proposed method in Section V. Finally, we conclude this paper in Section VI.

¹<http://adni.loni.usc.edu/>

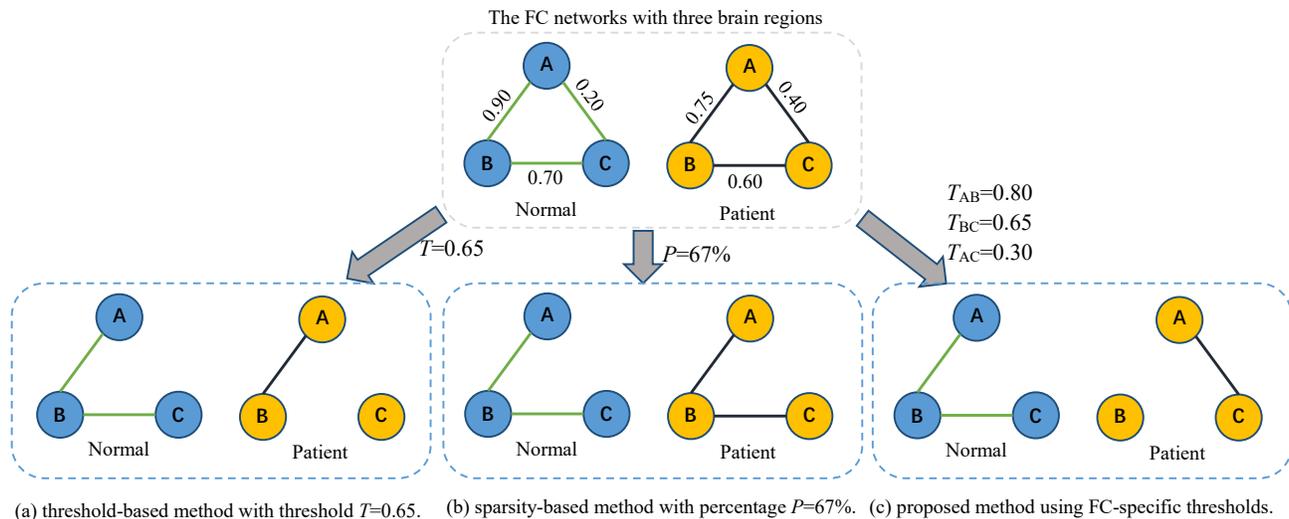


Fig. 2. Illustration of functional connectivity (FC) networks with three nodes/regions (top panel) of a normal (left) and an patient (right), as well as their corresponding thresholded networks (bottom panel) using (a) a threshold-based method (with the threshold of $T = 0.65$), (b) a sparsity-based method (with the percentage of $P = 67\%$), and (c) our DNT method using FC-specific thresholds. Note that there are different strengths for connections in FC networks of NC and patient groups, caused by brain disorders. Thus, thresholded FC networks (via threshold-based or sparsity-based methods) can not precisely characterize the real difference between FC networks of the NC and MCI subjects. Intuitively, it is reasonable to use FC-specific values to threshold each connection/connectivity in FC networks, as shown in (c).

II. RELATED WORK

A. Network Analysis for Brain Disorder Identification

Graph object-based methods and graph learning methods have been used for network analysis and shown a series of disrupted connectivity and network properties in AD [3], [4], ADHD [5] and schizophrenia [27]. For example, several studies have reported that there exist abnormal functional integration (e.g., characteristic path length), functional segregation (e.g., clustering coefficient), and “small-world characteristics” in AD and MCI patients [28], [29]. Tijms *et al.* [15] investigate and examine network properties that have been changed in AD patients. Altered network properties (e.g., increased local efficiencies and decreased tendency in global efficiencies) have also been studied in the brain of ADHD patients [5], [30].

Recent studies have applied functional connectivity networks to brain disorder identification using machine learning methods [20], [31]. In a typical FC network based classification pipeline, various network measures (e.g., node strength [32], clustering coefficients [19], The short path [18] and subnetworks [33]) are extracted from FC networks as features for training subsequent classifiers. For example, local clustering coefficients extracted from FC networks are used as features to identify MCI patients from NCs [17]. Network histogram features [34] and statistical&frequency features [35] are used for ADHD classification. Also, sixteen network measures have been extracted from FC networks to construct the network representation [18]. Recently, studies integrate multiple network properties for MCI classification [36]. These studies demonstrate the advantages of graph-based network analysis for neuroimaging analysis.

B. Network Thresholding for FC Analysis

To characterize topological properties and simplify representation learning of networks, studies usually use thresholding strategies (e.g., threshold-based and/or sparsity-based

methods) to transform original FC networks to binary networks, and then extract network measures as feature representations of subjects for FC network analysis [22].

In fact, the threshold or connection percentage is usually arbitrarily determined, so one has to study network properties by using a large range of thresholds to determine the optimal one, thus significantly increasing the computational burden. For example, Supekar *et al.* [3] explore the “small-world” properties of FC networks using thresholds in the range of [0.01, 0.99] with the step size of 0.01. Zanin *et al.* [18] optimize FC network representation according to the classification performance over a wide range of thresholds. Considering that network properties with different thresholds could be complementary to each other for classification, studies develop the multi-threshold strategy to extract multi-level network properties for brain disorder predication [21]. Compared with single-threshold based methods, this method can obtain better classification. However, these studies still ignore diversity information of temporal correlation in FC networks, as shown in Fig. 2.

III. METHOD

In Fig. 3, we present our DNTL framework for brain disorder identification based on rs-fMRI data with three steps: 1) fMRI pre-processing and functional connectivity network construction, 2) distribution-guided network thresholding (DNT) method, and 3) feature learning and model construction.

A. Materials

1) *Subjects*: Two datasets are used in this study: ADNI and ADHD-200. Specifically, ADNI contains 50 NC subjects, 56 early MCI (eMCI) and 43 late MCI (lMCI) subjects. A brief description for data acquisition in ADNI is as follows: 2.29-3.31 mm image resolution for in-plane, and 3.31 mm slice thickness, TE=30 ms, TR=2.2-3.1 s. Besides, ADHD-200 contains 118 ADHD and 98 NC subjects, with data

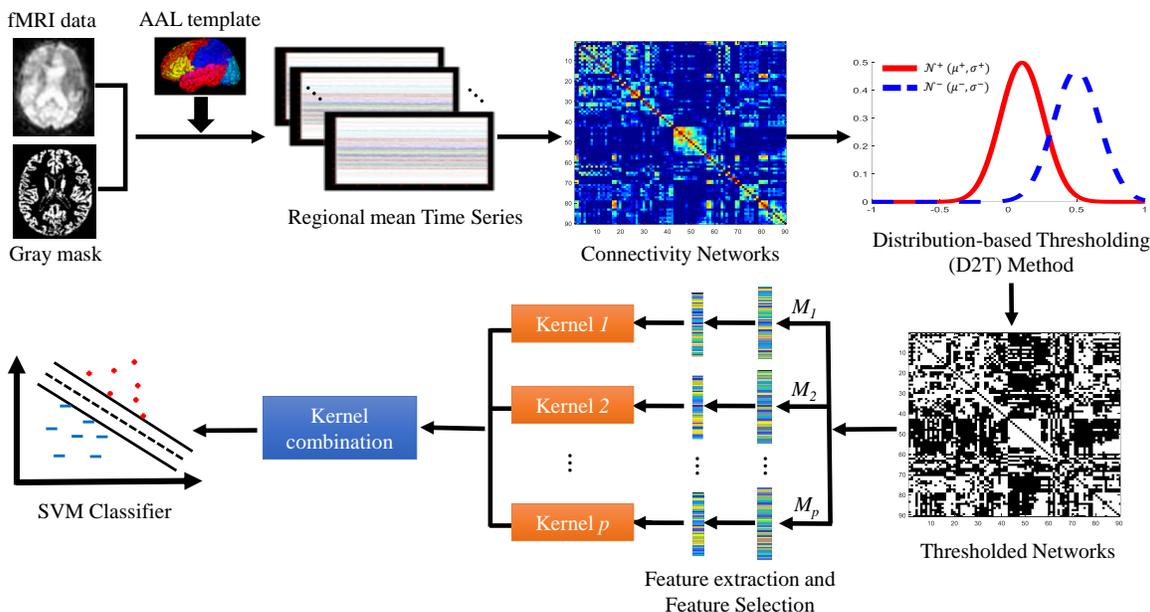


Fig. 3. Illustration of the proposed DNT-based learning (DNTL) framework for brain disease diagnosis, including three main steps: (1) image pre-processing and FC network construction, (2) Network thresholding, and (3) feature learning and model construction.

TABLE I

CHARACTERISTICS OF SUBJECTS IN THE ADNI AND ADHD-200 DATASETS. MMSE: MINI-MENTAL STATE EXAMINATION.

Dataset	group	# Subject	Age	Male/Female	MMSE
ADNI	IMCI	43	72.1 ± 8.2	26/17	27.2 ± 2.0
	eMCI	56	71.2 ± 6.8	21/35	28.1 ± 1.5
	NC	50	75.0 ± 6.9	21/29	28.9 ± 1.6
ADHD-200	ADHD	118	11.2 ± 2.7	25/93	-
	NC	98	12.2 ± 2.1	51/47	-

acquired as follows: 49×58 matrix size, 47 axial slices, 4 mm slice thickness, TE= 15 ms, TR=2 s, FOV=240 mm, flip angle = 90, the voxel size $3 \times 3 \times 4 \text{ mm}^3$. Table I reports the characteristics of studied subjects in these two datasets.

2) *Image Pre-processing*: For ADNI, following the work in [37], we pre-process resting-state functional MR images using the standard pipeline in FSL FEAT², including brain skull removal, slice time correction, motion correction, and spatial smoothing. Briefly, for each subject, we first discard the first 10 volumes, and perform slice timing for the remaining volumes, and align all images to the first volume for head motion correction. Here, we extract BOLD signals from the gray matter (GM) tissue. Accordingly, for each subject, we pre-process its T1-weighted MR image, including the brain skull removal, tissue segmentation into GM, white matter (WM) and cerebrospinal fluid (CSF), and co-register the first scan of the remaining fMRI of the same subject. Here, we mask fMRI data using the GM tissue of each subject. Finally, we partition the brain into 90 regions-of-interest (ROIs) according to the automated anatomical labeling (AAL) template [38]. For each ROI of each subject, we calculate its mean time series by averaging the GM-masked BOLD signals over all voxels. Then, we filter its mean time series within a frequency interval of $[0.025 \text{ Hz}, 0.100 \text{ Hz}]$ to increase the reliability of measurement [39] and to reduce low-frequency drift and high-frequency physiological noise [40].

For subjects from the ADHD-200 dataset, we follow the previous study [41], and directly use ROI-based time series from the Athena pre-processed data, with details shown online³. The data pre-processing steps include removing the first 4 image volumes, slice timing and head motion correction, exclude voxels at non-brain regions, co-registering the averaging echo-planar image into $4 \times 4 \times 4$ template space, eliminating the possible effect of WM, CSF, and head motion, temporal band-pass filtering (*i.e.*, $[0.009 \text{ Hz}, 0.08 \text{ Hz}]$), spatial smoothing with a 6 mm full width at half maximum Gaussian filter, and partitioning each brain into 90 regions based on the AAL template. Finally, we extract the ROI-based mean time series from GM-masked fMRI time series over all voxels in the particular ROI for representing each subject.

3) *Construction of Functional Connectivity Networks*: We use the pairwise Pearson correlation coefficient to measure temporal correlation of ROIs for constructing an FC network for each subject. Specifically, for each subject in ADNI and ADHD-200, we calculate the Pearson correlation coefficient between ROIs as the connection strength (*i.e.*, weight of an edge), with each ROI corresponding to a node in the network. Thus, we can build an FC network of each subject. Then, to improve the normality of correlation coefficients, we perform Fisher r-to-z transformation for the connections in FC networks.

B. Proposed Distribution-guided Network Thresholding

To capture the diversity of temporal correlations between ROI pairs, we propose a distribution-guided network thresholding (DNT) approach to determine FC-specific thresholds for connections in FC networks adaptively.

Given N training subjects and their class labels $Y = [y^1, y^2, \dots, y^N]$, $\mathcal{F} = [F^1, F^2, \dots, F^N]$ represents their FC networks (*i.e.*, adjacency matrices). Here, F^n is the FC

²<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FEAT>

³<http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline>

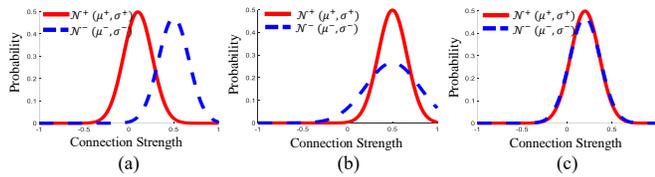


Fig. 4. Illustration of two normal distributions of connection strength for both the patient group and the NC group. Three possible cases are included: (a) different distributions with different means and their variances could be same or different, (b) different distributions with same mean but different variances, and (c) the same/similar distribution.

network of the n -th subject, while y^n denotes its class label (*i.e.*, patient/NC). All training subjects are partitioned into two groups (*i.e.*, patient and NC groups) based on their class labels, which are denoted as $\mathcal{F}^+ = [F^{1+}, F^{2+}, \dots, F^{N_1+}]$ and $\mathcal{F}^- = [F^{1-}, F^{2-}, \dots, F^{N_2-}]$, respectively. Here, N_1 and N_2 (with $N_1 + N_2 = N$) denote the number of subjects in the patient and NC groups, respectively.

For any pair of brain regions i and j , we assume that the distribution of connection strength in each subject group follows a normal distribution, *i.e.*, $\mathcal{N}^+(\mu_{ij}^+, \sigma_{ij}^+)$ for the patient group and $\mathcal{N}^-(\mu_{ij}^-, \sigma_{ij}^-)$ for NC groups. Then, we estimate these distributions based on training subjects in the corresponding group. For example, the distribution $\mathcal{N}^+(\mu_{ij}^+, \sigma_{ij}^+)$ can be estimated by

$$\mu_{ij}^+ = \frac{1}{N_1} \sum_{n=1}^{N_1} F_{ij}^{n+}, \quad (1)$$

$$\sigma_{ij}^+ = \frac{1}{N_1 - 1} \sum_{n=1}^{N_1} (F_{ij}^{n+} - \mu_{ij}^+)^2, \quad (2)$$

where F_{ij}^{n+} is the element of matrix F^{n+} and denotes connection strength between the i -th and j -th ROIs.

For both normal distributions (*i.e.*, \mathcal{N}^+ and \mathcal{N}^-), there are three possible cases, *i.e.*, 1) they are different distributions with different means, and their variances could be same or different, 2) they are different distributions with the same mean, but different variances, and 3) they are same/similar distribution. Figure 4 (a) shows the first case where the mean in two distributions are different, implying that the temporal correlation between regions i and j is changed in the patient group (when compared with that in the NC group). Thus, the optimal threshold of connection between this pair of brain regions should be the *intersection point* of two distributions, located in the range of $[\mu_{ij}^+, \mu_{ij}^-]$.

Figure 4 (b) illustrates the second case, where the means of two distributions are the same or similar, but their standard deviations (SDs) are different. Here, a large SD value means that the corresponding functional connectivity between regions i and j is unstable for subjects in the corresponding group. Figure 4 (c) shows the third case where two distributions are the same or very similar, suggesting that no significant difference exists in functional connectivity between patient and NC groups. For these two cases, it is challenging to find stable patterns of functional connectivity between patients and NCs. For simplicity, we can preserve or remove the corresponding connection according to specific applications, by setting the threshold as a constant c when performing thresholding. For

Algorithm 1: Distribution-guided Network Thresholding (DNT) method

Input: A set of FC networks \mathcal{F} with a response vector Y , and three parameters δ , θ and c
Output: A threshold matrix T

- 1 Partition all training subjects into two groups according to their class labels, denoted as \mathcal{F}^+ and \mathcal{F}^- .
- 2 **foreach** each pair of brain regions i and j **do**
- 3 Estimate the distributions of $\mathcal{N}^+(\mu_{ij}^+, \sigma_{ij}^+)$ and $\mathcal{N}^-(\mu_{ij}^-, \sigma_{ij}^-)$ by Eq. (1) and Eq. (2).
- 4 Compute $KL(\mathcal{N}^+, \mathcal{N}^-)$ using Eq. (4);
- 5 Calculate the threshold T_{ij} by Eq. (3).
- 6 **Return** T .

example, the connection can be removed with $c = 1$, or can be preserved with $c = -1$. In this work, we set $c = 1$ to remove the corresponding connection for network-based classification.

According to the previous discussions, for two distributions with different means (as shown in Fig. 4(a)), the optimal threshold should be the intersection points of those distributions. Therefore, the threshold T_{ij} for the paired brain regions i and j can be determined as follows

$$T_{ij} = \begin{cases} s_{ij}, & \text{if } KL(\mathcal{N}^+ || \mathcal{N}^-) > \delta \text{ and } |\mu_{ij}^+ - \mu_{ij}^-| > \theta; \\ c, & \text{otherwise;} \end{cases} \quad (3)$$

where δ and θ are two pre-defined positive values, and s_{ij} is the intersection point of two normal distributions, located in the range of $[\mu_{ij}^+, \mu_{ij}^-]$, which can be computed by solving univariate equation of $\mathcal{N}^+ - \mathcal{N}^- = 0$. The term $KL(\mathcal{N}^+ || \mathcal{N}^-)$ denotes the Kullback-Leibler divergence that calculates the similarity of two distributions, which can be defined as

$$KL(\mathcal{N}^+ || \mathcal{N}^-) = \log \left(\frac{\sigma_{ij}^+}{\sigma_{ij}^-} \right) + \frac{(\sigma_{ij}^-)^2 + (\mu_{ij}^+ - \mu_{ij}^-)^2}{2(\sigma_{ij}^+)^2} - \frac{1}{2}. \quad (4)$$

In Algorithm 1, we summarize the detailed process of DNT. For each connection between a pair of brain regions i and j , we construct an FC-specific threshold T_{ij} by using the distributions of connection strength in two subject groups. In this way, one can generate a threshold matrix T for the whole FC network, which can preserve the diversity information of temporal correlation among brain regions. Note that only training subjects are used for the threshold construction in the proposed DNT method.

C. Proposed DNT-based Classification Framework

1) **Construction of Thresholded Network:** To model the topological properties of FC networks, we threshold FC networks of all training subjects using the threshold matrix T constructed by the proposed DNT method. Specifically, the n -th FC network F^n is thresholded as follows

$$\tilde{F}_{ij}^n = \begin{cases} 0, & \text{if } F_{ij}^n < T_{ij} \text{ or } F_{ij}^n = 0; \\ 1, & \text{otherwise;} \end{cases} \quad (5)$$

where F_{ij}^n is an element of F^n (*i.e.*, corresponding to the connection strength between the brain regions i and j). In this way, connections with zero values will be removed from the network. Thus, we can obtain a thresholded network \tilde{F}^n . Note that only connections in FC networks are changed in the

thresholding process, thus the resulting thresholded networks still have the same nodes/ROIs as their original ones.

2) *Feature Extraction and Feature Selection*: To reduce the feature dimension, it is crucial to extract meaningful measures from constructed FC networks. Following previous work in [3], [42], in the study, we extract three kinds of network measures from thresholded networks, including the degree of a node, local clustering coefficient, and betweenness centrality, as features for subsequent classification.

Given the n -th thresholded network (matrix) $\tilde{F}^n \in R^{N \times N}$, the definitions of these network measures are as follows.

(1) Degree of node (DN): It is equal to the number of connections to the node i , and is defined as follows

$$d_i = \sum_{j=1}^N \tilde{F}_{ij}^n, \quad (6)$$

where \tilde{F}_{ij}^n is a element of \tilde{F}^n , corresponding to the connection value between the nodes i and j .

(2) Local clustering coefficient (CC): This metric reflects the prevalence of clustered connection around the node i , and is defined as follows

$$c_i = \frac{2}{d_i(d_i - 1)} \sum_{j,q=1}^N (\tilde{F}_{ij}^n \tilde{F}_{jq}^n \tilde{F}_{iq}^n)^{\frac{1}{3}}, \quad (7)$$

where d_i is the degree of node i defined in Eq. (6). Here, $c_i = 0$ for $d_i < 2$.

(3) Betweenness centrality (BC): It is defined as the fraction of all shortest paths in the network that pass through the node i , *i.e.*,

$$b_i = \frac{1}{(N-1)(N-2)} \sum_{\substack{j,q=1 \\ i \neq j, j \neq q, q \neq i}}^N \frac{t_{jq}(i)}{t_{jq}}, \quad (8)$$

where t_{jq} is the number of shortest paths between nodes j and q , and $t_{jq}(i)$ is the number of shortest paths between nodes j and q that pass through node i .

Based on Eqs. (6)-(8), we obtain three sets of network features for each subject. These features contain redundant and irrelevant features for classification. Therefore, we further perform a standard paired t -test method to screen out those uninformative features. Specifically, for each type of network features, we calculate the p -value of each feature using standard t -test on the training subjects, and discard those features with p -value larger than a pre-defined value (*e.g.*, 0.05). All surviving features are used for subsequent classification.

3) *Multi-Kernel Support Vector Machine*: Recent studies have shown that multi-kernel learning techniques can effectively integrate features from multiple modalities, and achieve better classification performance in comparison with single kernel learning methods [17], [43]. Following these studies, we use a multi-kernel learning technique to integrate three types of network features for classification.

Specifically, a basic kernel is first calculated for each type of network features (after feature selection via t -test) of training subjects. Then, the multi-kernel learning is performed via a linear combination of p ($p = 3$) kernels constructed on three

types of network measures, formulated as:

$$k(x^n, x^m) = \sum_{i=1}^p \alpha_i k_i(x_i^n, x_i^m), \quad (9)$$

where $k_i(x_i^n, x_i^m)$ is the basic kernel (*i.e.*, linear kernel in this study) for the m -th and n -th subjects based on the i -th type of features, and α_i is non-negative weight with $\sum_{i=1}^p \alpha_i = 1$.

Following [43], we calculate the optimal α_i using a grid search technique on the training subjects. Once obtaining the optimal parameter $\{\alpha_i\}_{i=1}^p$, we can transform multiple kernels into a single kernel, and directly adopt the traditional support vector machine (SVM) for classification.

IV. EXPERIMENT

A. Experimental Setup

Based on a 5-fold cross-validation (CV) strategy, we conduct four groups of binary classification tasks, *i.e.*, 1) MCI vs. NC, 2) eMCI vs. IMCI, 3) eMCI vs. NC, and 4) ADHD vs. NC classification. Four metrics are used for performance evaluation, including accuracy, sensitivity, specificity and area under the receiver operating characteristic (ROC) curve (AUC).

In the proposed DNTL method, we set the parameters $\delta = 0.05$, $\theta = 0.1$, and set parameter $c = 1$ when considering the aim of brain disease classification. In the step of feature selection, we select the features with p -value less than 0.05 for classification on training data. The optimal parameters α_i ($i = 1, 2, 3$) in multi-kernel learning is determined by an inner 5-fold CV on training subjects (via a grid search in the interval of $[0, 1]$ with a step size of 0.1). We use the linear SVM (with the parameter $C = 1$) for classification.

B. Methods for Comparison

We first compare the proposed DNTL with a baseline method without performing network thresholding (denoted as **Baseline**). Specifically, the Baseline method directly extracts network measures (*i.e.*, weighted clustering coefficients [25]) from original FC networks as feature representations, followed by a standard t -test method (with $p < 0.05$) for feature selection. Then, a linear SVM classifier with a default parameter (*i.e.*, $C = 1$) is used for classification in Baseline.

We also compare DTNL with two state-of-the-art methods for network thresholding, including a threshold-based method (called **THR**) and a sparsity-based method (called **SPA**). For the fairness of comparison, in the THR method, we use all possible thresholds within the range of $[0.01, 0.99]$ with the step size of 0.01. In the SPA method, we use all possible connection percentages within the range of $[1\%, 99\%]$ (step size: 1%). Then, we compute and report their best performances for both THR and SPA methods.

C. Results of Brain Disorder Identification

1) *Results with Different Network Measures*: Based on three network measures, Table II summarizes the classification performance of all four methods (*i.e.*, Baseline, THR, SPA, and DNTL), and Fig. 5 plots the ROC curves of these methods.

TABLE II

PERFORMANCE OF FOUR DIFFERENT METHODS IN FOUR CLASSIFICATION TASKS. ACC: ACCURACY; SEN: SENSITIVITY; SPE: SPECIFICITY.

Method	MCI vs. NC				IMCI vs. eMCI				eMCI vs. NC				ADHD vs. NC			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Baseline	67.6	100.0	4.0	0.51	64.1	60.7	68.0	0.61	60.6	51.2	67.9	0.56	63.0	56.1	68.6	0.56
THR	75.8	91.9	44.0	0.58	70.6	75.0	66.0	0.69	70.3	55.8	80.4	0.69	68.7	60.2	75.4	0.66
SPA	72.6	87.9	42.0	0.64	72.6	80.36	64.0	0.70	71.5	67.4	73.2	0.68	67.7	59.2	74.6	0.66
DNTL (Ours)	79.2	92.9	52.0	0.68	74.5	73.2	76.0	0.72	76.0	74.4	76.8	0.70	72.3	62.2	80.5	0.67

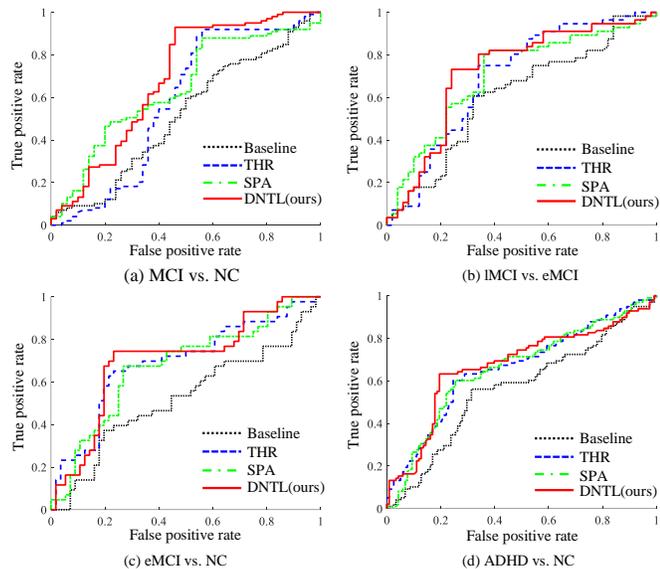


Fig. 5. ROC curves of four different methods in four classification tasks, *i.e.*, (a) MCI vs. NC classification, (b) eMCI vs. IMCI classification, (c) eMCI vs. NC classification, and (d) ADHD vs. NC classification.

In Fig. 6, we also present the results of four methods using every single type of network measure.

From Table II and Fig. 5, we can see that the DNTL outperforms the competing approaches in four classification tasks. For example, DNTL achieves the accuracy of 79.2%, 74.5%, 76.0% and 72.3% for MCI vs. NC, eMCI vs. IMCI, eMCI vs. NC and ADHD vs. NC classification, respectively, while the best accuracies yielded by the comparison methods are 75.8%, 73.7%, 71.5% and 68.7%, respectively. The underlying reason is that DNTL can take advantage of diversity information of temporal correlation between brain regions in different groups, thus determining optimal FC-specific thresholds to characterize FC networks more accurately.

From Fig. 6 and Table II, we can see that 1) compared with methods using any single network measure, methods using the combination of multiple network measures can achieve better classification performance. This indicates that the multiple network measure may convey different-yet-complementary information to further boost the classification performance. 2) Compared with the baseline method, methods with thresholding (*i.e.*, DNTL, SPA and THR) can obtain better results, suggesting that thresholding is essential for improving network-based classification performance. Also, Fig. 6 shows that our DNTL with a single network measure also outperforms the state-of-the-art two thresholding methods (*i.e.*, THR and SPA), suggesting the efficacy of our proposed method.

2) *Results with Different Thresholding Strategies*: In DNTL, we adopt the *intersection point* of two distributions to determine the optimal threshold. For comparison, we also test the

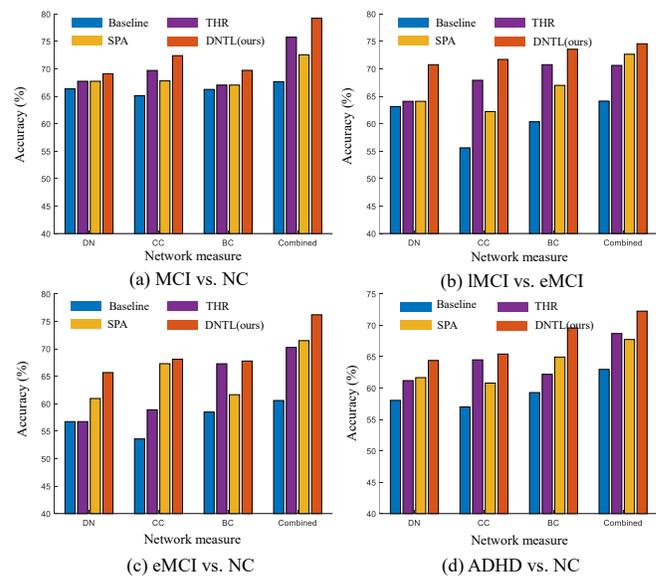


Fig. 6. Results of four methods with single/multiple network properties in four tasks: (a) MCI vs. NC classification, (b) eMCI vs. IMCI classification, (c) eMCI vs. NC classification, (d) ADHD vs. NC classification. Here, DN, CC, BC denote using the measures of Degree of Nodes, Clustering Coefficients, and Betweenness Centrality, respectively, while “Combined” denotes the method using all three types of measures.

results of DNTL using the average of means of two normal distributions as the optimal threshold (denoted as DNTL-mean). Fig. 7 plots the obtained classification accuracy. In both THR and SPA methods, one needs to pre-define the optimal thresholding parameters (*i.e.*, the threshold value for THR and connection percentage for SPA). We further report their results with respect to different thresholding parameters in Fig. 7. In Fig. 7 we report the results of Baseline method with sparse SVM [44] (instead of standard SVM) for classification (denoted as Baseline-SSVM).

As can be observed from Fig. 7 and Table II, the proposed DNTL consistently outperforms all competing methods (*i.e.*, DNTL-mean, SPA, THR and Baseline-SSVM) in four tasks. Besides, Fig. 7 suggests that the accuracies of THR and SPA are affected by different thresholds or connection percentages largely, indicating that the thresholding step is essential to capture network properties and subsequent network analysis. These results also suggest the advantage of DNTL in FC network based brain disorder identification.

D. Discriminative Brain Regions for Diagnosis

It is interesting to study the imaging biomarkers identified by the proposed DNTL method for brain disease analysis. We now investigate the discriminative features involved in the task of disease classification. Specifically, for the local

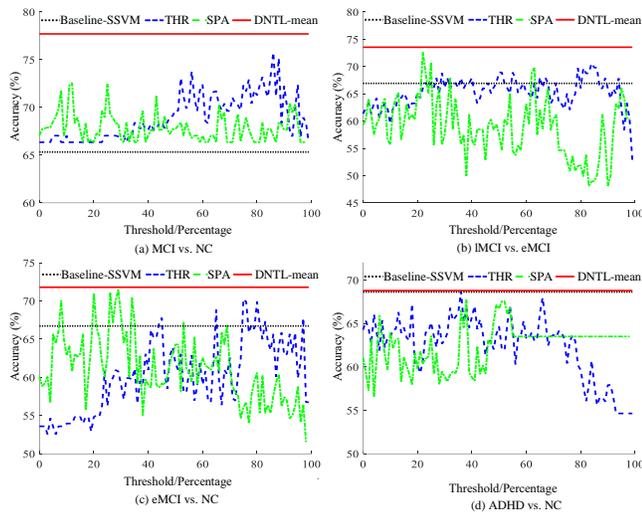


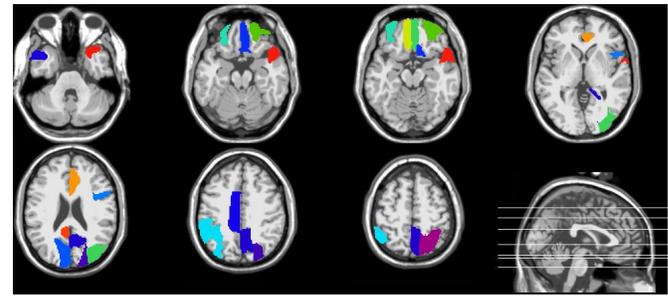
Fig. 7. Accuracy of four different methods in four classification tasks, where THR and SPA use different thresholding parameters.

clustering coefficient feature of each ROI, we calculate its number of occurrences (selected by feature selection of t -test) in all 5-fold CVs. Since the selected features (*i.e.*, ROIs) are different in each CV fold we treat those ROIs that always occur in all 5-fold CVs as discriminative brain regions. Tables S5 and S6 in *Supporting Information* list those brain regions for eMCI vs. NC classification and ADHD vs. NC classification, respectively. In Fig. 8, we further plot these regions in the template space. In addition, for node degree feature of each ROI, we use the same strategy to find the discriminative brain regions, as shown in Tables S7-S8 of *Supporting Information*. To present the overlap/different ROIs selected by two kinds of features, Figure S1 in *Supporting Information* plots the Venn diagram of selected ROIs using two kinds of features.

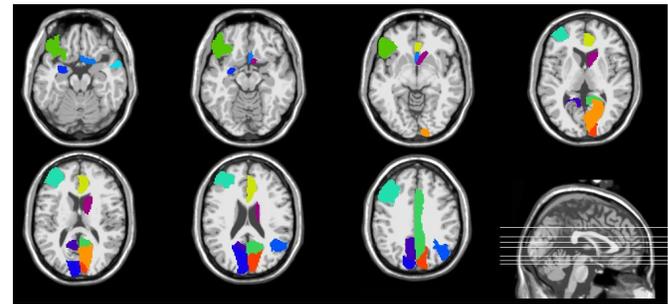
From Table S5 and Fig. 8 (a), one can observe that the discriminative brain regions include *frontal gyrus*, *opercular*, *cingulate*, *temporal pole*, and *parietal lobule*, which have been reported in previous studies [45], [46]. Table S6 and Fig. 8 (b) suggests that the discriminative regions, such as *frontal gyrus*, *cingulate*, *cuneus*, *precuneus* and *temporal pole*, have also been reported in previous ADHD studies [47], [48]. These results imply that those brain regions are highly associated with AD and ADHD. Furthermore, from Tables S5-S8 and Fig. S1, we can see that, in the same classification task, the discriminative brain regions are different for different network measure features, indicating that these features contain complementary information, and should be integrated to further improve the classification performance.

E. Functional Connectivity Analysis

We further study the functional connectivity difference between patient and NC groups. Specifically, we first compute the difference of connection strengths on all ROIs shown in Tables S5 and S6 of *Supporting Information* for two pairs of groups (*i.e.*, eMCI vs. NC, and ADHD vs. NC), respectively. Figure 9 shows the obtained results, where nodes denote ROIs and each connection indicates that the connectivity strength has been significantly changed (with p -value < 0.05) by the disorder when comparing the patient and NC groups.



(a) eMCI vs. NC



(b) ADHD vs. NC

Fig. 8. Discriminative ROIs identified by the proposed DNTL in the tasks of (a) eMCI vs. NC classification and (b) ADHD vs. NC classification.

Blue connections and red connections represent increased and decreased functional connectivity in the patient group, respectively, compared with the NC group. Then, to visually show the differences of FC networks between two groups, we calculate the average FC network for each group, and threshold the obtained average FC network using the thresholds by our proposed method in the first CV, with results reported in Figure S2 in *Supporting Information*.

From Fig. 9 (a) and Fig. S2 (a), we can see that the connection strengths in eMCI patients are significantly lower than those in NC group (corresponding to red edges). This indicates that FCs between these ROIs has decreased in eMCI patients, which are consistent with existing studies [29]. For most ROIs listed in Table S6, Fig. 9 (b) and Fig. S2 (b) suggests that their FCs have also decreased in ADHD group, as reported in previous studies [30]. These results further validate the reliability of our DNTL.

V. DISCUSSION

In this work, we propose a distribution-guided network thresholding (DNT) method for pre-processing FC networks. Different from previous studies, the proposed method can explore the diversity of temporal correlations among brain regions, and thus adaptively construct FC-specific thresholds for better characterizing FC networks. We further develop a DNT-based learning (DNTL) framework for brain disorder identification using rs-fMRI data, and evaluate the proposed method on two real datasets. The experimental results demonstrate that our method can significantly improve the performance of brain disorder classification. Note that the proposed method is a general approach for pre-processing brain networks, and can be used in both fields of machine-learning-based disease diagnosis/prognosis and group-match-based network analysis.

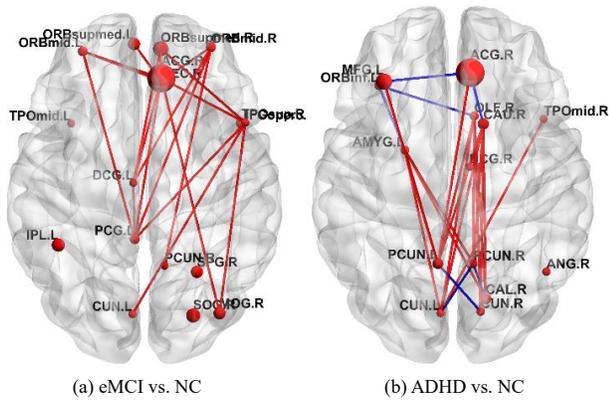


Fig. 9. Illustration of discriminative functional connectivity (*i.e.*, with connection strength significantly changed by brain disorders, p -value < 0.05) between ROIs for (a) eMCI vs. NC groups, and (b) ADHD vs. NC groups. Each node denotes a specific ROI. Each edge indicates that the functional connectivity with connection strength significantly changed (with p -value < 0.05) by the disorder when comparing the patient and NC groups. Blue and red edges represent the increased and decreased functional connectivity in patient groups, respectively.

A. Significance of Results

For the FC network analysis, thresholding is a fundamental step for exploring network topological properties. However, existing thresholding methods typically apply a pre-defined value or connection percentage for network thresholding, thus ignoring the diversity of temporal correlations between different pairs of ROIs. Besides, it is challenging to determine the optimal threshold or connection percentage, since there is no gold standard rule for selecting an optimal value. To this end, we design a distribution-guided network thresholding method to adaptively determine the FC-specific thresholds for connections in FC networks, and apply it to brain disorder diagnosis. Recent studies have been focused on brain diseases classification based on rs-fMRI data. For example, in [33] the accuracy of 82.6% and 74.8% was reported for MCI vs. NC and eMCI vs. IMCI classification on ADNI dataset. The accuracy of 65% and 68% on ADHD-200 was reported in [34] and [35], respectively. In contrast, our proposed method, respectively, achieves the accuracy of 79.2%, 76.0% and 72.3% for MCI vs. NC, eMCI vs. NC and ADHD vs. NC classification, which are comparable to the best results reported in those recent studies.

Based on the proposed method, we identify some disease-related brain regions, which have been reported in previous studies. For example, in eMCI vs. NC classification, the important brain regions include frontal gyrus [45], cingulate [49], [50], cuneus [51], angular gyrus [3], precuneus [46], temporal pole [52]. Furthermore, we analyze the connectivity between identified brain regions, and find the decreased functional connectivity in the patient groups (*i.e.*, eMCI and ADHD groups) compared with the NC groups, which is consistent with existing studies [29], [30]. These results indicate that functional connectivity among brain regions may be affected by brain disorders, thus leading to changed functional integration and segregation of the brain.

TABLE III

CLASSIFICATION ACCURACY OF FOUR METHODS WITHOUT FEATURE SELECTION IN FOUR CLASSIFICATION TASKS.

Method	MCI vs. NC (%)	IMCI vs. eMCI (%)	eMCI vs. NC (%)	ADHD vs. NC (%)
Baseline	66.4	63.1	53.6	60.6
THR	66.4	63.2	56.4	62.3
SPA	66.4	65.1	59.9	63.1
DNTL (Ours)	71.1	66.0	65.7	66.1

B. Influence of Multi-Kernel Weights

To evaluate the effect of weights in multi-kernel learning (*i.e.*, α_1 , α_2 , and α_3 in Eq. (9)) on the classification performance, we record the accuracy of the proposed DNTL using all combination of possible values within the range of $[0, 1]$ (step size: 0.1) and the constraint of $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Figure 10 reports the results of DNTL in four classification tasks with respect to different multi-kernel weights of three types of network measures. In each sub-figure in Fig. 10, results in the top right, bottom left and top left denote the accuracies of DNTL using only one type of network measures, *i.e.*, degree of node (with $\alpha_1 = 1$), local clustering coefficient (with $\alpha_2 = 1$), and betweenness centrality (with $\alpha_3 = 1$). Figure 10 suggests that we can get the best results when using the combination of three types of network measures instead of just a single metric. Besides, using many different weight combinations, our DNTL can yield stable results in four classification tasks.

C. Effect of Feature Selection

In this study, we employ the t -test feature selection method to pre-process features. To evaluate the influence of feature selection, we perform two additional experiments: 1) without feature selection step. Specifically, we record the classification performance of our DNTL without any feature selection. That is, we directly use multi-kernel SVM on three types of network measures extracted from thresholded FC networks for classification. Table III reports the results of our DNTL and three competing methods in four classification tasks, where no feature selection is used in these four methods. 2) With a different feature selection method. That is, we perform the proposed DNTL method using LASSO-based feature selection [53], instead of t -test method. Table S1 in the Supporting Information reports the obtained results. For better comparison, in Table S1, we report the results of three competing methods with LASSO-based feature selection.

As shown in Table III and Table S1, the proposed DNTL outperforms three competing methods. Furthermore, from Tables II, III and Table S1, we can see that, compared with methods without performing feature selection, methods with feature selection can obtain better classification performance, indicating the importance of feature selection. In addition, From Table II and Table S1, we can see that the proposed method with different feature selection methods can achieve a similar performance, demonstrating the robustness of the proposed method for different feature selection methods.

D. Effect of Parameters

In the proposed DNTL method, there include three parameters, *i.e.*, δ , θ and c . The parameter δ controls the

similarity of two normal distributions, the parameter θ controls the distance between means of two distributions, and the parameter c is used for preserving (with $c = -1$) or removing (with $c = 1$) the corresponding connections with similar distributions between groups. Considering network-based classification tasks, we set $c = 1$ in our experiment. To investigate the effects of parameters δ , θ on the performance of the proposed method, we vary the values of δ within the range of $\{0.01, 0.05, 0.08, 0.10\}$ and θ within the range of $\{0.05, 0.10, 0.12, 0.15\}$, and report the corresponding results of DNTL in Fig. 11. This figure suggests that given a fixed θ , the accuracy results slightly change with different δ , indicating that the proposed method is very robust for the parameter δ . Besides, given a fixed δ , the accuracy is largely affected by the parameter θ , suggesting that the value of θ is important for the proposed method. This is reasonable since the parameter θ controls the distance of two distribution means, thus affecting the determination of the optimal thresholds.

E. Repeatability using Different Brain Atlases

To evaluate performance of the proposed method using different brain atlases, we perform the same experiments using the functional atlases proposed in [54], [55], which partitions the brain into 200 and 160 ROIs, respectively. We adopt the same preprocessing steps with AAL atlas, and calculate the regional mean time series of these ROIs to compute the functional connectivity. Table S2 in the *Supporting Information* reports the classification accuracy of all methods (including the proposed method and the competing methods) for MCI vs NC classification. From Table S2, we can see that, compared with the competing methods, the proposed method can obtain better classification accuracy, suggesting robustness of the proposed method for different number of ROIs and varied spatial scales.

F. Effect of Distribution Assumption

In the proposed method, we assume that the distribution of strength of connections in each subject group is a normal distribution. To evaluate the effect of this distribution assumption, we test performance of the proposed method using Weibull distribution instead of normal distribution. The experimental results show that the proposed method can achieve the accuracy of 78.80%, 76.4%, 75.9% and 71.8% for MCI vs. NC, eMCI vs. IMCI, eMCI vs. NC and ADHD vs. NC classification, respectively. These results are still better the results of competing methods (as shown in Table II), further demonstrating efficacy of the proposed method.

G. Influence of Classifier

Following the previous work [16], we train a linear SVM for classification. To evaluate the effect of different classifiers on performance of the proposed method, we perform two additional experiments: 1) using a nonlinear SVM, and 2) using different classifiers. In the first experiment, we perform the proposed DNTL method using the SVM with RBF kernel (with $\sigma = 1$), instead of the linear SVM, for classification. In the second experiment, we concatenate all three network

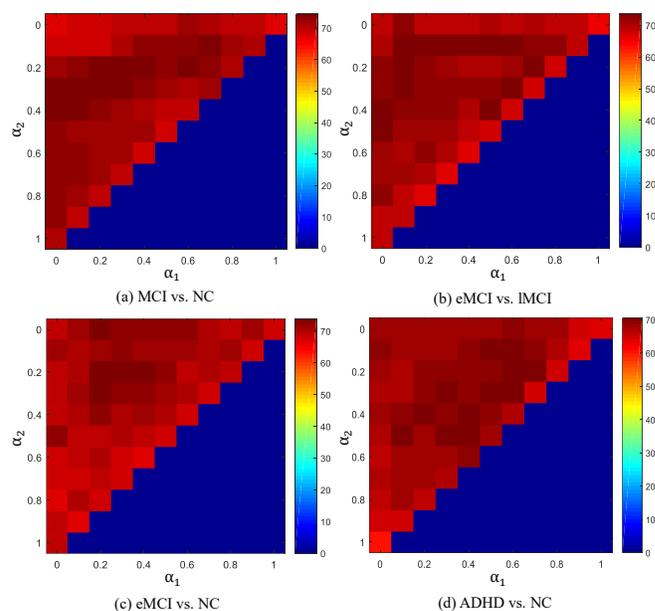


Fig. 10. Classification accuracy of the proposed DNTL using different multi-kernel weights (with $\alpha_1 + \alpha_2 + \alpha_3 = 1$) for three types of network measures in four tasks, including (a) MCI vs. NC, (b) IMCI vs. eMCI, (c) eMCI vs. NC, and (d) ADHD vs. NC classification.

measures from thresholded networks into a feature vector, and perform feature selection of t -test with p -value < 0.05 , followed by a decision tree using standard CART algorithm for classification. Tables S3-S4 in the *Supporting Information* report the classification accuracy results of four methods.

From Table II, Tables S3-S4, we have the following observations. 1) The proposed method outperforms the competing methods in four tasks, suggesting the efficacy of the proposed method. 2) Compared with decision tree method, the SVM-based methods can yield better classification performance, which shows the efficacy of kernel-based methods. 3) Compared with the linear SVM, the method using SVM with RBF kernel can achieve better results, indicating that the feature distribution of subjects could be linearly inseparable.

H. Limitations and Future Work

There are several limitations in the current study. *First*, we simple set $T_{ij} = 1$ (i.e., $c = 1$) when two group distributions are the same/similar in the DNTL algorithm. Such network thresholding strategy may lead to the loss of network information. It is interesting to determine the optimal value of c via employing traditional thresholding method for further improving classification performance. *Second*, we use the standard t -test method to select features from each type of network measures, thus ignoring the underlying association among different types of network measures. In the future, we plan to jointly select features from multi-modality/multi-task data for brain disease classification [43]. *Third*, we focus on extracting human-engineered network measures in this work, while several deep learning approaches have been recently applied to brain connectivity analysis [56]. In the future, we will employ deep learning techniques to learn the topological properties of FC networks for brain disorder identification. *Finally*, we extract three network measures as features for

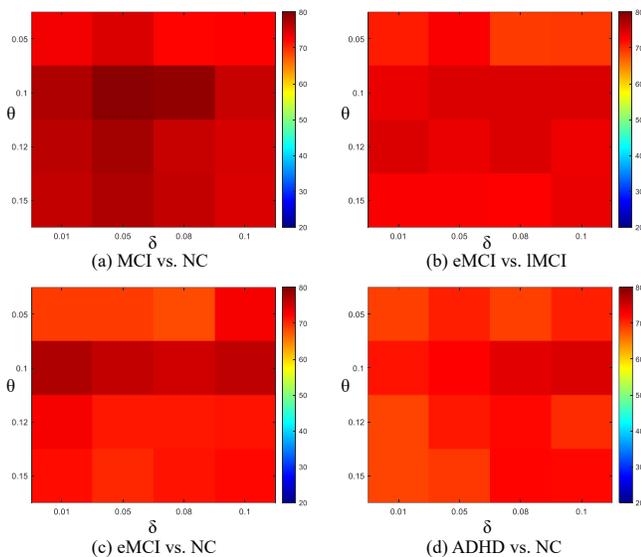


Fig. 11. The accuracy achieved by the proposed DNTL method with different values of δ and θ in four classification tasks.

network-based classification, and evaluate the performance of the proposed method on 365 subjects from two rs-fMRI datasets. In the further work, we will evaluate the proposed method using more network measures (*e.g.*, path length, shortest path, etc) on dataset with larger sample sizes.

VI. CONCLUSION

In this work, we propose a distribution-guided network thresholding (DNT) method for FC network analysis, and develop a DNT-based learning (DNTL) framework for brain disorder identification based on rs-fMRI data. Our DNT can adaptively determine the FC-specific threshold for each connection in FC networks, thus preserving the diversity of temporal correlation between different pairs of ROIs as well as different subject groups. Experiment results on 365 subjects from two public rs-fMRI datasets suggest that our method can improve the performance of brain disorder identification.

VII. ACKNOWLEDGMENT

Part of the data used in this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The investigators within the ADNI contributed to the design and implementation of ADNI and provided data but did not participate in analysis or writing of this article. A complete listing of ADNI investigators can be found online (https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

REFERENCES

- [1] O. Sporns, "Contributions and challenges for network models in cognitive neuroscience," *Nature Neuroscience*, vol. 17, no. 5, pp. 652–660, 2014.
- [2] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009.
- [3] K. Supekar, V. Menon, D. Rubin, M. Musen, and M. D. Greicius, "Network analysis of intrinsic functional brain connectivity in Alzheimer's disease," *PLoS Computational Biology*, vol. 4, no. 6, pp. e1000100:1–11, 2008.

- [4] F. Bai, Z. Zhang, D. R. Watson, H. Yu, Y. Shi, Y. Yuan, Y. Zang, C. Zhu, and Y. Qian, "Abnormal functional connectivity of hippocampus during episodic memory retrieval processing network in amnesic mild cognitive impairment," *Biological Psychiatry*, vol. 65, no. 11, pp. 951–958, 2009.
- [5] K. Konrad and S. B. Eickhoff, "Is the ADHD brain wired differently? A review on structural and functional connectivity in attention deficit hyperactivity disorder," *Human Brain Mapping*, vol. 31, no. 6, pp. 904–916, 2010.
- [6] S. O. Dumoulin, A. Fracasso, W. V. D. Zwaag, J. C. W. Siero, and N. Petridou, "Ultra-high field MRI: Advancing systems neuroscience towards mesoscopic human brain function," *NeuroImage*, vol. 168, pp. 345–357, 2018.
- [7] R. Flanagan, L. Lacasa, E. K. Towilson, S. H. Lee, and M. A. Porter, "Effect of antipsychotics on community structure in functional brain networks," *Journal of Complex Networks*, vol. 7, no. 6, pp. 932–960, 2019.
- [8] C. Stam, B. Jones, G. Nolte, M. Breakspear, and P. Scheltens, "Small-world networks and functional connectivity in Alzheimer's disease," *Cerebral Cortex*, vol. 17, no. 1, pp. 92–99, 2007.
- [9] M. R. Arbabshirani, K. A. Kiehl, G. D. Pearlson, and V. D. Calhoun, "Classification of schizophrenia patients based on resting-state functional network connectivity," *Frontiers in Neuroscience*, vol. 7, no. 133, pp. 1–16, 2013.
- [10] M. Wang, D. Zhang, J. Huang, P.-T. Yap, D. Shen, and M. Liu, "Identifying autism spectrum disorder with multi-site fMRI via low-rank domain adaptation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 3, pp. 644–655, 2019.
- [11] M. Wang, C. Lian, D. Yao, D. Zhang, M. Liu, and D. Shen, "Spatial-temporal dependency modeling and network hub detection for functional MRI analysis via convolutional-recurrent network," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 8, pp. 2241–2252, 2019.
- [12] D. Yao, J. Sui, M. Wang, E. Yang, Y. Jiaerken, N. Luo, P.-T. Yap, M. Liu, and D. Shen, "A mutual multi-scale triplet graph convolutional network for classification of brain disorders using functional or structural connectivity," *IEEE Transactions on Medical Imaging*, vol. 40, no. 4, pp. 1279–1289, 2021.
- [13] M. R. Brier, J. B. Thomas, A. M. Fagan, J. Hassenstab, D. M. Holtzman, T. L. Benzinger, J. C. Morris, and B. M. Ances, "Functional connectivity and graph theory in preclinical Alzheimer's disease," *Neurobiology of Aging*, vol. 35, no. 4, pp. 757–768, 2014.
- [14] B. Jie, M. Liu, and D. Shen, "Integration of temporal and spatial properties of dynamic connectivity networks for automatic diagnosis of brain disease," *Medical Image Analysis*, vol. 47, pp. 81–94, 2018.
- [15] B. M. Tijms, A. M. Wink, W. de Haan, W. M. van der Flier, C. J. Stam, P. Scheltens, and F. Barkhof, "Alzheimer's disease: Connecting findings from graph theoretical studies of brain networks," *Neurobiology of Aging*, vol. 34, no. 8, pp. 2023–2036, 2013.
- [16] B. Jie, C. Y. Wee, D. Shen, and D. Zhang, "Hyper-connectivity of functional networks for brain disease diagnosis," *Medical Image Analysis*, vol. 32, pp. 84–100, 2016.
- [17] C.-Y. Wee, P.-T. Yap, D. Zhang, K. Denny, J. N. Browndyke, G. G. Potter, K. A. Welsh-Bohmer, L. Wang, and D. Shen, "Identification of MCI individuals using structural and functional connectivity networks," *NeuroImage*, vol. 59, no. 3, pp. 2045–2056, 2012.
- [18] M. Zanin, P. Sousa, D. Papo, R. Bajo, J. Garcia-Prieto, F. del Pozo, E. Menasalvas, and S. Boccaletti, "Optimizing functional network representation of multivariate time series," *Scientific Reports*, vol. 2, no. 630, pp. 1–7, 2012.
- [19] C.-Y. Wee, P.-T. Yap, W. Li, K. Denny, J. N. Browndyke, G. G. Potter, K. A. Welsh-Bohmer, L. Wang, and D. Shen, "Enriched white matter connectivity networks for accurate identification of MCI patients," *NeuroImage*, vol. 54, no. 3, pp. 1812–1822, 2011.
- [20] D. Dai, J. Wang, J. Hua, and H. He, "Classification of ADHD children through multimodal magnetic resonance imaging," *Frontiers in Systems Neuroscience*, vol. 6, p. 63, 2012.
- [21] B. Jie, D. Zhang, C. Y. Wee, and D. Shen, "Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification," *Human Brain Mapping*, vol. 35, no. 7, pp. 2876–97, 2014.
- [22] K. A. Garrison, D. Scheinost, E. S. Finn, X. Shen, and R. T. Constable, "The (in)stability of functional brain network measures across thresholds," *NeuroImage*, vol. 118, no. 15, pp. 651–661, 2015.
- [23] O. Sporns, "The human connectome: A complex network," *Annals of the New York Academy of Sciences*, vol. 1224, no. 1, pp. 109–125, 2011.
- [24] S. L. Simpson, D. B. Bowman, and P. J. Laurienti, "Analyzing complex functional brain networks: Fusing statistics and network science to understand the brain," *Statistics Surveys*, vol. 7, pp. 1–36, 2013.

- [25] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, 2010.
- [26] M. A. Serrano, M. Boguna, and A. Vespignani, "Extracting the multiscale backbone of complex weighted networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 16, pp. 6483–8, 2009.
- [27] E. Damaraju, E. Allen, A. Belger, J. Ford, S. McEwen, D. Mathalon, B. Mueller, G. Pearlson, S. Potkin, A. Preda *et al.*, "Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia," *NeuroImage: Clinical*, vol. 5, pp. 298–308, 2014.
- [28] E. J. Sanz-Arigita, M. M. Schoonheim, J. S. Damoiseaux, S. A. Rombouts, E. Maris, F. Barkhof, P. Scheltens, and C. J. Stam, "Loss of 'small-world' networks in Alzheimer's disease: Graph analysis of fMRI resting-state functional connectivity," *PloS One*, vol. 5, no. 11, pp. e13 788:1–14, 2010.
- [29] J. Wang, X. Zuo, Z. Dai, M. Xia, Z. Zhao, X. Zhao, J. Jia, Y. Han, and Y. He, "Disrupted functional brain connectome in individuals at risk for Alzheimer's disease," *Biological Psychiatry*, vol. 73, no. 5, pp. 472–481, 2013.
- [30] L. Wang, C. Zhu, Y. He, Y. Zang, Q. Cao, H. Zhang, Q. Zhong, and Y. Wang, "Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder," *Human Brain Mapping*, vol. 30, no. 2, pp. 638–649, 2009.
- [31] C.-Y. Wee, P.-T. Yap, and D. Shen, "Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological change patterns," *Human Brain Mapping*, vol. 34, no. 12, pp. 3411–3425, 2013.
- [32] G. Chen, B. D. Ward, C. Xie, W. Li, Z. Wu, J. L. Jones, M. Franczak, P. Antuono, and S.-J. Li, "Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging," *Radiology*, vol. 259, no. 1, pp. 213–221, 2011.
- [33] B. Jie, M. Liu, D. Zhang, and D. Shen, "Sub-network kernels for measuring similarity of brain connectivity networks in disease diagnosis," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2340–2353, 2018.
- [34] B. Solmaz, S. Dey, A. R. Rao, and M. Shah, "ADHD classification using bag of words approach on network features," in *Medical Imaging 2012: Image Processing*, vol. 8314. International Society for Optics and Photonics, 2012, pp. 83 144T:1–7.
- [35] S. Itani, F. Lecron, and P. Fortemps, "A multi-level classification framework for multi-site medical data: Application to the adhd-200 collection," *Expert Systems with Applications*, vol. 91, pp. 36–45, 2018.
- [36] B. Jie, D. Zhang, W. Gao, Q. Wang, C.-Y. Wee, and D. Shen, "Integration of network topological and connectivity properties for neuroimaging classification," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 576–589, 2014.
- [37] D. Zhang, J. Huang, B. Jie, J. Du, L. Tu, and M. Liu, "Ordinal pattern: A new descriptor for brain connectivity networks," *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1711–1722, July 2018.
- [38] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, 2002.
- [39] X. N. Zuo, A. D. Martino, C. Kelly, Z. E. Shehzad, D. G. Gee, D. F. Klein, F. X. Castellanos, B. B. Biswal, and M. P. Milham, "The oscillating brain: Complex and reliable," *NeuroImage*, vol. 49, no. 2, pp. 1432–1445, 2010.
- [40] D. Cordes, V. M. Haughton, K. Arfanakis, J. D. Carew, P. A. Turski, C. H. Moritz, M. A. Quigley, and M. E. Meyerand, "Frequencies contributing to functional connectivity in the cerebral cortex in 'resting-state' data," *Ajnr American Journal of Neuroradiology*, vol. 22, no. 7, pp. 1326–1333, 2001.
- [41] P. Bellec, C. Chu, F. Chouinard-Decorte, Y. Benhajali, D. S. Margulies, and R. C. Craddock, "The neuro bureau adhd-200 preprocessed repository," *NeuroImage*, vol. 144, pp. 275–286, 2017.
- [42] M. Barthélemy, "Betweenness centrality in large complex networks," *European Physical Journal B*, vol. 38, no. 2, pp. 163–168, 2004.
- [43] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.
- [44] J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani, "1-norm support vector machines," *Advances in Neural Information Processing Systems*, vol. 16, no. 1, pp. 1–8, 2004.
- [45] Y. Feng, L. Bai, Y. Ren, S. Chen, and J. Tian, "fMRI connectivity analysis of acupuncture effects on the whole brain network in mild cognitive impairment patients," *Magnetic Resonance Imaging*, vol. 30, no. 5, pp. 672–682, 2012.
- [46] F. Bai, W. Liao, D. R. Watson, Y. Shi, Y. Wang, C. Yue, Y. Teng, D. Wu, Y. Yuan, J. Jia *et al.*, "Abnormal whole-brain functional connection in amnesic mild cognitive impairment patients," *Behavioural Brain Research*, vol. 216, no. 2, pp. 666–672, 2011.
- [47] G. Bellelli, P. Mazzola, M. Corsi, A. Zambon, G. Corrao, G. Castoldi, G. Zatti, and G. Annoni, "The combined effect of adl impairment and delay in time from fracture to surgery on 12-month mortality: An observational study in orthogeriatric patients," *Journal of the American Medical Directors Association*, vol. 13, no. 7, pp. 664.e9 – 664.e14, 2012.
- [48] B. E. Depue, G. C. Burgess, E. G. Willcutt, L. C. Bidwell, L. Ruzic, and M. T. Banich, "Symptom-correlated brain regions in young adults with combined-type ADHD: Their organization, variability, and relation to behavioral performance," *Psychiatry Research: Neuroimaging*, vol. 182, no. 2, pp. 96 – 102, 2010.
- [49] Y. Zhou, J. H. Dougherty, K. F. Hubner, B. Bai, R. L. Cannon, and R. K. Hutson, "Abnormal connectivity in the posterior cingulate and hippocampus in early Alzheimer's disease and mild cognitive impairment," *Alzheimer's & Dementia*, vol. 4, no. 4, pp. 265–270, 2008.
- [50] S. D. Han, K. Arfanakis, D. A. Fleischman, S. E. Leurgans, E. R. Tuminello, E. C. Edmonds, and D. A. Bennett, "Functional connectivity variations in mild cognitive impairment: Associations with cognitive function," *Journal of the International Neuropsychological Society*, vol. 18, no. 1, pp. 39–48, 2012.
- [51] T. Zhang and C. Davatzikos, "Optimally-discriminative voxel-based morphometry significantly increases the ability to detect group differences in schizophrenia, mild cognitive impairment, and Alzheimer's disease," *NeuroImage*, vol. 79, pp. 94–110, 2013.
- [52] F. Nobili, D. Salmasso, S. Morbelli, N. Girtler, A. Piccardo, A. Brugnolo, B. Dessi, S. A. Larsson, G. Rodriguez, and M. Pagani, "Principal component analysis of FDG PET in amnesic MCI," *European Journal of Nuclear Medicine & Molecular Imaging*, vol. 35, no. 12, p. 2191, 2008.
- [53] H. Lee, D. S. Lee, H. Kang, B. Kim, and M. K. Chung, "Sparse brain network recovery under compressed sensing," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1154–1165, 2011.
- [54] R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, and H. S. Mayberg, "A whole brain fMRI atlas generated via spatially constrained spectral clustering," *Human Brain Mapping*, vol. 33, no. 8, pp. 1914–1928, 2012.
- [55] N. U. Dosenbach, B. Nardos, A. L. Cohen, D. A. Fair, J. D. Power, J. A. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, C. N. Lessov-Schlaggar *et al.*, "Prediction of individual brain maturity using fmri," *Science*, vol. 329, no. 5997, pp. 1358–1361, 2010.
- [56] J. Kawahara, C. J. Brown, S. P. Miller, B. G. Booth, V. Chau, R. E. Grunau, J. G. Zwicker, and G. Hamarneh, "BrainNetCNN: Convolutional neural networks for brain networks; Towards predicting neurodevelopment," *NeuroImage*, vol. 146, pp. 1038–1049, 2016.