Identifying Autism Spectrum Disorder from Resting-State fMRI Using Deep Belief Network

Zhi-An Huang, Zexuan Zhu, *Senior Member, IEEE*, Chuen Heung Yau, and Kay Chen Tan, *Fellow, IEEE*

Abstract—With the increasing prevalence of autism spectrum disorder (ASD), it is important to identify ASD patients for effective treatment and intervention, especially in early childhood. Neuroimaging techniques have been used to characterize complex biomarkers based on the functional connectivity anomalies in ASD. However, the diagnosis of ASD still adopts the symptom-based criteria via clinical observation. The existing computational models tend to achieve unreliable diagnostic classification on large-scale aggregated datasets. In this work, we propose a novel graph-based classification model using deep belief network (DBN) and the Autism Brain Imaging Data Exchange (ABIDE) database, which is a world-wide multi-site functional and structural brain imaging data aggregation. The remarkable connectivity features are selected through a graph extension of K-nearest neighbors, and then refined by a restricted path-based depth first search algorithm. Thanks to the feature reduction, lower computational complexity could contribute to the shortening of training time. Automatic hyperparameter tuning technique is introduced to optimize the hyperparameters of DBN by exploring the potential parameter space. The simulation experiments demonstrate the superior performance of our model, which is 6.4% higher than the best result reported on ABIDE database. We also propose to use data augmentation and over-sampling technique to further identify possible subtypes within ASD. The interpretability of our model enables the identification of the most remarkable autistic neural correlation patterns from data-driven outcomes.

Index Terms—Autism spectrum disorder (ASD), functional magnetic resonance imaging (fMRI), computational diagnostic model (CDM), deep belief network (DBN), functional connectivity (FC)

I. INTRODUCTION

AUTISM spectrum disorder (ASD) is a range of common childhood neurodevelopmental disorders including autism, Asperger's syndrome, and other related conditions. ASD individuals typically present with variable deficits of clinical syndromes in restrictive interests, social communication, and repetitive behaviors. The rising incidence of ASD

Z.-A. Huang and K.C. Tan are with the Department of Computer Science, City University of Hong Kong, Kowloon Tong, Hong Kong; and the City University of Hong Kong Shenzhen Research Institute (e-mail: zahuang2-c@my.cityu.edu.hk; kaytan@cityu.edu.hk). Z. Zhu is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China; Shenzhen Pengcheng Laboratory, Shenzhen 518055, China; and also with the SZU Branch, Shenzhen Institute of Artificial Intelligence and Robotics for Society, Shenzhen University, Shenzhen 518060, China (e-mail: zhuzx@szu.edu.cn). C.H. Yau is with the School of Chinese Medicine, The University of Hong Kong, Pok Fu Lam, Hong Kong (e-mail: annyau333@gmail.com).

has attracted public concern. The serious impairments of ASD can last throughout one's life and represent a major burden of health and finance globally. The lifetime costs of treating an ASD patient in the United States has exceeded one million dollars [1]. Early treatments and intervention can alleviate the symptoms and improve the life quality of the patients. Although the neurogenetics and neurobiology of ASD have been widely investigated, the pathology of ASD still remains uncertain. Due to the etiological heterogeneity of ASD, no neuropathological or unified structural traits have been conclusively characterized [2]. The symptom-based diagnostic criteria are still fundamental for the clinical diagnosis of ASD. ASD patients are typically diagnosed by using the criteria checklist of the Diagnostic and Statistical Manual (DSM). Despite the ever-changing diagnostic metrics and categories over the past decades, the diagnosis of ASD, at its core, still suffers from significant limitations in observational strategies. The recent studies have revealed that disparities between individuals with autism and those without autism have reduced over time, which could be attributed to the continuous changes in the definition of autism [3]. In other words, the traditional clinical diagnosis could more likely lead to misdiagnosis due to the subjectivity and complexities involved in the diagnosis process. Additionally, the global shortfall in well-qualified professionals of ASD could delay the early diagnosis and treatment by aggravating related symptoms.

The non-invasive neuroimaging techniques provide data for capturing neural patterns of brain structure and function. As a kind of neuroimaging techniques, functional magnetic resonance imaging (fMRI) [4] is extensively employed to investigate functional variations and structural alterations in the autistic brain activity. The rationale of fMRI is to measure the blood-oxygen-level dependent (BOLD) contrast related to the energy use by brain cells. Resting state fMRI (rs-fMRI), serving as one of the fMRI paradigm, provides the data of subjects' baseline BOLD variance. Numerous studies have proven the feasibility of using rs-fMRI to uncover pairwise interactions between regions of interest (ROIs) in psychiatric illnesses such as Alzheimer disease [5], attention deficit hyperactivity disorder (ADHD) [6], autism [7], and others [8]. Rs-fMRI signals have shown great potential in identifying neuropathology diagnostic biomarkers [9]. Leveraging the data integrated from multiple datasets could be a silver-bullet solution for clinical applications through sharing and consolidating independent data samples across different studies. This can greatly contribute to the diagnosis, determination of risk prognosis, and monitoring of treatment response of ASD [10]. However, controversies still surround preprocessing pipelines

including bandpass filtering and global signal regression. Without a unified set of standard, data aggregation solution could be debatable with regards to the reproducibility and generalizability of data sources. Uncontrolled variations can inevitably exist in inter-site data ranging from MRI acquisition protocols (e.g. MRI vendors, flip angle, scan time and basic pulse sequence parameters), to recruitment strategies (e.g. age tendency and IQ-range), and to participant instructions (e.g. eye open/closed conditions) [11]. This impedes the effort to leverage aggregated MRI data in research. The advances in computing power and methodology [12] have enabled the development of useful computational model to solve the aforementioned issues.

In this paper, we propose a novel graph-based classification model using deep belief network (DBN) [13]. To train our model, we leverage the data of the Autism Brain Imaging Data Exchange (ABIDE) database, which is a world-wide multi-site functional and structural brain imaging data aggregation [14]. First, the ASD remarkable functional connectivities (FCs) are selected by the graph-based feature selection (GBFS) method based on the both external and internal measures. Then, a restricted path-based depth first search (RP-DFS) algorithm is implemented to further explore the topological information implied in graph. Finally, a three-layer DBN with automatic hyperparameter tuning is proposed for the identification of ASD patients. The result of mean accuracy of 0.764 ± 0.022 based on 10-fold cross validation (CV) shows that our model outperforms the state-of-the-art methods. The receiver operating characteristic (ROC) curve and the area under ROC curve (AUC) are used to evaluate the classification performance for the best-case and worst-case scenarios. Leave-one-site-out test well demonstrates the applicability and generalization of the proposed model to a new different site. The proposed model is also demonstrated to have moderate improvement in predictive capability for identifying ASD subtypes. No significant differences are observed between accuracy and the confounding factors, e.g. age, gender, and full-scale intelligence quotient (FIQ) scores. In addition, the interpretability of our model enables us a statistical analysis for uncovering the correlation patterns in autistic brains. Subsequently, some rules that we found are presented in this work. The main contributions of this paper can be summarized as follows:

1) The GBFS method is proposed to effectively select remarkable connections in ASD brain based on both external and internal measures. Moreover, we further design a DBN classifier with automatic hyperparameter tuning, which is more accurate and efficient as compared with other state-of-the-art algorithms.

2) We first try to identify possible subtypes within ASD on imbalanced rs-fMRI datasets. To tackle this issue, data augmentation and oversampling technique are introduced based on the proposed model. Experiment results suggest our work obtain moderate improvement in predictive capability for identifying ASD subtypes.

3) Based on the real-world data of ABIDE, we conduct a series of comprehensive experiments to validate the superior performance of our model against the state-of-the-art methods. The interpretability of our model enables us to prioritize the seminal correlation patterns in autistic brains, some of which have been manually validated by published literatures.

The rest of this paper is organized as follows. Section II reviews the related work. Section III describes the proposed framework including GBFS, RP-DFS and DBN model with automatic hyperparameter tuning. The experimental results and analysis are discussed in Section IV. Finally, Section V concludes the paper.

II. RELATED WORK

More recently, there has been an increasing interest to investigate brain FC as neurological biomarkers in numerous studies for the classification of mental sates (e.g. the presentation of emotions [15], learning [16], and semantic categories [17]), and mental disorders (e.g. ASD, schizophrenia, and major depressive disorder (MDD) [18]). Based on the discovery of neurological biomarkers, extensive efforts have been made to developing the computer-aided diagnostics models [19], [20], [21]. In the following, we briefly review some extensively used feature selection strategies and classification models based on brain disorder studies.

A. Feature Selection Strategies

To select discriminative and effective features on fMRI data is particularly important to well-performing classifiers for identification of mental disorder [22]. Especially for FCs, usually tens of thousands of features are defined whereas only a small percentage of them carry valuable information towards the goal. Nielsen *et al.* [23] extracted 7265 features from 26.4 million connections by computing the pairwise correlation coefficient between each generated ROI. Feature selection strategies can be mainly categorized into three classes: filter methods, wrapper methods, and embedded methods [24].

Filter methods are widely applied for feature selection thanks to their effectiveness in computational expense and their robustness to overfitting. The basic principle is to assign proxy measures (e.g. statistical test [24], Fisher score [25], and correlation coefficient [26]) to features using the general characteristics of datasets, so as to harvest optimal feature subsets with top scores. The group-level statistical test is the most popular type of filter methods such as *t*-test, ranksumtest, and Welch's *t*-test. The primary issue with this strategy is that remarkable features are selected by their *p*-values, which sometimes disable to reflect those with the largest discrimination power. That is, features with small *p*-values (i.e., high confidence) might result in a poor classification performance [27]. Correlation-based feature selection is another type of filter methods to rank features based on the assumption that optimal feature subsets contain highly correlated information with regard to the classification so as to distinguish between instances [26]. Abraham *et al.* use a regularized covariance estimator to estimate the connectivity coefficients between ROI time series based on three different measure methods [11]. This strategy can achieve satisfactory performance thanks to the low-order polynomial run time, applicability to binary or

continuous data, as well as noise-tolerance and robustness to FC feature interactions.

Wrapper methods select decreasing numbers of features with different possible combinations depending on the performance of classifiers. Recursive feature elimination (RFE) [28] and genetic algorithm (GA) [29] are two extensively used wrapper methods to produce seminal feature combinations. RFE optimizes classifiers by recursively considering smaller and smaller possible feature combinations. To explore an optimal subset of features, the work in [30] determines the number of feature variables for classification by the highest accuracy as recursively increasing the thresholds at seven different effect sizes. GA initializes binary encoded populations (representing a feature included or not) and then generates high-quality feature combinations by bio-inspired operators i.e., mutation, crossover and selection. The third class, embedded methods are quite similar to wrapper methods since they also integrate feature selection into the learning of specific classifiers for decision processes. The major difference to wrapper methods is that an intrinsic model building metric is leveraged during learning. The least absolute shrinkage and selection operator (LASSO) method is the most popular embedded method to add a penalty against complexity for reducing the degree of variance or overfitting of a linear model [31]. In recent neuroimaging studies [32], [33], multi-task feature selection as an emerging approach aims to integrate data sharing highly consistent feature patterns between modality by ideally ignoring the data-dependent noise. Based on sparse learning, Wang *et al.* [34] used the group $\ell_{2,1}$-norm regularizer to impose the sparsity between all features and non-sparsity between tasks for obtaining important features. In theory, wrapper methods and embedded methods tend to yield better feature subsets specific for a particular classifier than filter methods do. However, the quantity of possible feature combinations exponentially grows as the number of features increases. Wrapper methods and embedded methods are more computationally expensive than filter methods due to the repeated learning steps and cross-validation. Furthermore, features selected by them could not well characterize the reproducible autistic connectivity patterns due to their dependent relationships with specific classifiers. Since the purpose of this study is to explore reproducible biomarkers with discriminative power for effective classification, a heuristic graph-based filter feature selection method is proposed based on both external and internal measures on rs-fMRI data.

B. Classification models

In previous literature [20], [23], traditional machine learning (ML) classifiers with hand-engineered features are commonly utilized to uncover the autistic function patterns. The ongoing development of ML toolkits (e.g. scikit-learn [35]) provide easy-to-implement ML methods and allow researchers to compare the feasibility and effectiveness of diverse ML methods, such as random forest (RF) and SVM [36]. Plitt *et al.* explored 9 ML classification algorithms based on the statistical significance of accuracy, positive predictive value (PPV), and negative predictive value (NPV) [19]. However,

most ML classifiers rely on the "shallow" or linear models. They are incapable of capturing the topological information within brain networks as well as the relationships between connectivity features and clinical traits, especially on large-scale datasets [37].

In recent years, deep learning (DL) techniques [38], [39] have emerged as a promising approach with an outstanding performance comparable to or even in some cases superior to human experts. A growing number of studies [40], [39], [30] attempt to employ DL algorithms for the classification of brain disorders and address the issues of generalizability and subjectivity brought by ML classifiers. Heinsfeld *et al.* [38] concluded that DL algorithms should minimize the human intervention to automatically extract relevant features using unsupervised learning methods. They use two stacked denoising autoencoders to transfer 19,900 features into deep neural network (DNN) [41] achieving the average accuracy of 70%, which is the best result reported to date using the whole ABIDE database. The well-performing DL classifiers benefit from their hierarchical structure with different levels of complexity, and non-linear transformations with different levels of abstraction provided by hidden layers [24], [42]. This allows us to characterize the reproducible autistic connectivity patterns or "fingerprints" served as biomarkers, to effectively discriminate ASD patients from the control. Nevertheless, determining appropriate models' configuration setting is quite challenging by the many hyperparameter choices one must make, e.g., learning rates and batch sizes. Automatic hyperparameter tuning technique is of crucial importance for utilizing DL algorithms in practice. In this work, we propose a three-layer DBN model with automatic hyperparameter tuning technique for classification and further improve the accuracy through tackling this issue.

III. MODEL DESIGN

The flowchart of the proposed model is shown in Fig. 1. First of all, by the preprocessing pipeline, the mean time series of ROIs are extracted from the raw rs-fMRI data and then converted into two-dimensional feature matrices T, where T_{ik} represents the ith ROI's mean time series of kth timestamp. GBFS is used to select remarkable autistic functional connections for graph construction based on both external and internal measures. Then we exploit topological information to enrich the graph representation by using RP-DFS method. Finally, a three-layer DBN model is built and its configuration setting is optimized by automatic hyperparameter tuning.

A. Graph-Based Feature Selection

In this section, the graph-based feature selection (GBFS) is proposed to select remarkable autistic functional connections based on the preprocessed datasets. Incorporating appropriate FC features is necessary for the classification of brain disorders. If we calculate all FCs between voxels, the number of features adds up to tens of thousands. The irrelevant features with redundant information would be detrimental to the classification performance. In this work, the Pearson
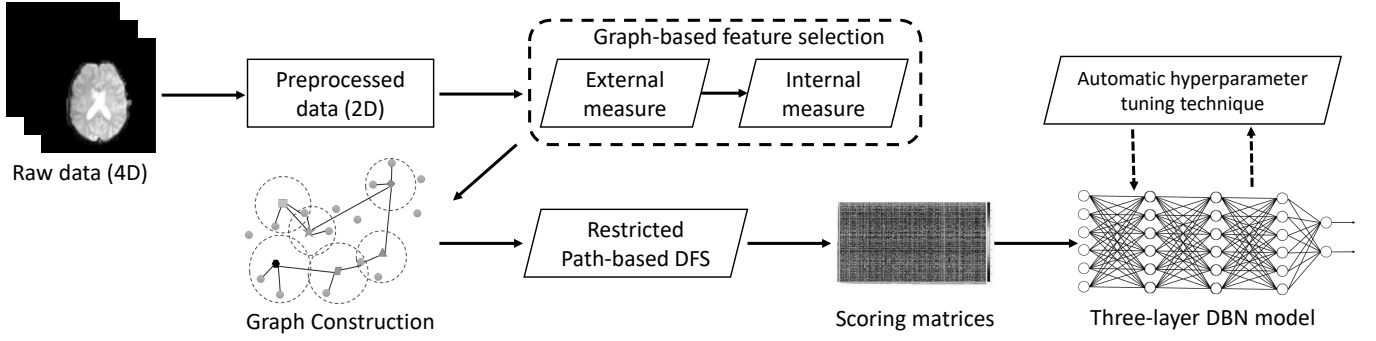
Fig. 1. The flowchart of the proposed model.

correlation coefficient (PCC) is used to measure the levels of co-activation between each ROI' time series as follows:

$$FC(i,j) = \frac{\sum_{k=1}^{N}(T_{ik} - \overline{T_i})(T_{jk} - \overline{T_j})}{\sqrt{\sum_{k=1}^{N}(T_{ik} - \overline{T_i})^2 \sum_{k=1}^{N}(T_{jk} - \overline{T_j})^2}} \quad (1)$$

where N represents the total number of timestamp and varies from different institutes. $FC(i,j)$ means the FC correlation between any two ROIs ranging from 1 (highly correlated) to -1 (highly anti-correlated). In this way, the generated connectivity vector for each subject contains 19,900 features/connections (200 ROIs, $\frac{200 \times (200-1)}{2} = 19,900$). The remarkable autistic connections of ROIs are selected based on both external and internal measures.

First, the external measure is a data-driven approach for globally identifying the unique neural patterns associated with ASD. As mental disorders can be considered as disturbances or disruptions of the normal operation of brain activity, the remarkable connections need to satisfy two following criteria: 1) Remarkable connections must be distinctive enough in terms of autistic average level. All ASD datasets are merged to calculate the mean value for each connection, denoted as $mean(FC_{ASD})$, and the global mean value, denoted as $mean_{ASD}$, as well as the global standard deviation (STD) as STD_{ASD} from all connections. Then those connections with $mean(FC_{ASD}(i,j))$ higher or lower than $mean_{ASD}$ for α-times STD_{ASD} are considered to satisfy this condition. A default filter factor α is used to control the number of selected connections. 2) Remarkable connections must be simultaneously discriminative between ASD group and typical control (TC) group. In other words, combined with the first criteria, the selected connections should be remarkable in ASD group while unremarkable in TC group. Accordingly, these two criteria can be mathematically subject to

$$\begin{cases} \|mean(FC_{ASD}(i,j)) - mean_{ASD}\| > \alpha * STD_{ASD} \\ \|mean(FC_{TC}(i,j)) - mean_{TC}\| \leq \alpha * STD_{TC}. \end{cases}$$
$$(2)$$

α is empirically set to 1 in our work via 10-fold CV. Each connection is examined to determine whether it should be selected as a remarkable connection based on these two criteria. As such, only 81 qualified connections are selected to perform the internal measure in the next step. It is worth to note that when we repeat the same procedures for TC group based on the opposite version of Eq. (2), the remarkable
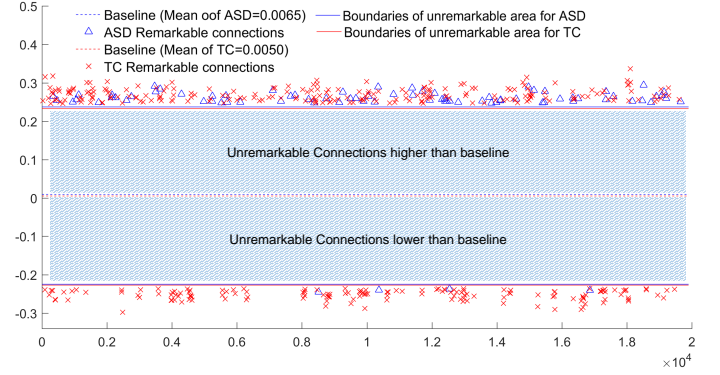
Fig. 2. 81 and 380 remarkable connections are highlighted in ASD group and TC group, respectively. ($STD_{ASD} = 0.2407$ and $STD_{TC} = 0.2437$)

connections of TC group that are remarkable in TC group while unremarkable in ASD group achieves up to 380 (as shown in Fig. 2). The brain networks for both ASD and TC groups are visualized in Fig. 3 using BrainNet Viewer [43]. Most notably, 115 and 37 remarkable connections in TC group are associated with the frontal and posterior cortical areas, respectively. Whereas in ASD group, only 18 and 4 remarkable connections are associated with the frontal and posterior cortical areas, respectively. This finding is consistent with the theory of frontal-posterior underconnectivity in autism [44], which attributes the disorder to lower synchronization caused by lower communication bandwidths between frontal and posterior areas in the autistic brain. Therefore, the external measure can explore the reproducible neurological biomarkers with discriminative power in deciphering remarkable autistic neural correlation patterns from data-driven outcomes.

Second, a graph extension of K-nearest neighbors as an internal measure is applied for graph construction on communities of potential ROIs between those remarkable connections selected by the external measure. As shown in Fig.4, this approach takes advantage of spatial distribution information by detecting the normalized neighborhood of K-nearest ROIs as the "receptive fields" for connected communities. For a target remarkable connection, their neighbor ROIs do not belong to this remarkable connection. Take the diamond-shape node (denoted as $Node_d$) and the triangle-shape node (denoted as $Node_t$) in Fig.4 as examples, a graph is initialized with edges
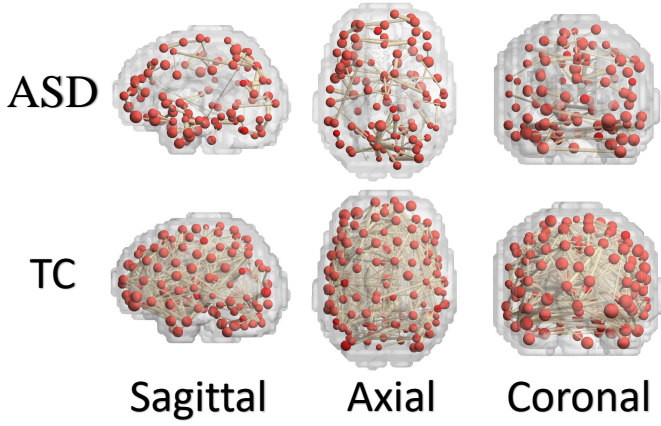
Fig. 3. The visualization of brain networks for ASD and TC groups from views of sagittal, axial and coronal planes.
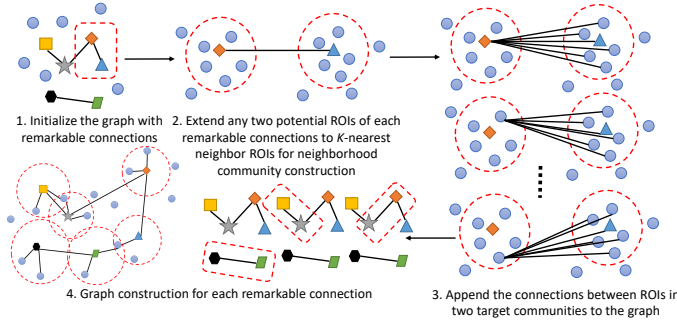

Fig. 4. Illustration of the internal measure pipeline by a graph extension of K-nearest neighbors

connecting $Node_t$ and $Node_d$. We assume that the autistic neuronal activity could function as the collaborative activation pattern in potential associated areas, rather than the single region-to-region connectivity interaction. The subgraph extension is conducted by extending both $Node_t$ and $Node_d$ to their K-nearest neighbor ROIs forming communities (subgraphs) based on Euclidean distance. The FCs of ROIs between the subgraphs of $Node_t$ and $Node_d$ are appended to the graph with their weights that are estimated by PCC. After using GBFS for each remarkable connection, the number of selected FC features should be $81 \times (K+1)^2 = 3,969$, where K is empirically set to 6 in this work. After filtering out 681 repeated connections, we only consider 16.52% ($\frac{3,288}{19,900} \times 100\%$) of the original features associated with the autistic brain activity. The pseudo-code is shown in Algorithm 1.

B. Restricted Path-Based Depth First Search Algorithm

Network topology in graph is helpful to unravel the implied information, which effectively enriches the connectivity matrices. We extend our previous work [45] and redesign a restricted path-based depth first search algorithm named RP-DFS for refining FC feature matrices to traverse all potential paths between both target ROIs of each remarkable connection. A path (denoted as P) is defined as a set of connections

---

Algorithm 1 The pseudo-code of GBFS

Input: The # of nearest neighbors K, a feature matrix T.
Output: Graph $G(V, E)$.
1: Initialize remarkable connection set $\mathcal{S} = \{\}$
2: Split all datasets into ASD group and TC group
3: Calculate $mean(FC_{ASD})$, $mean_{ASD}$, and STD_{ASD} in ASD group as well as $mean(FC_{TC})$, $mean_{TC}$, and STD_{TC} in TC group
4: **for** $i = ROI_1, ..., ROI_{200}$ **do**
5:    **for** $j = i + 1, ..., ROI_{200}$ **do** //19,900 in total
6:       **if** $FC(i, j)$ satisfies formula (2) **then**
7:          $\mathcal{S} \leftarrow \mathcal{S} \cup edge(i, j)$ //add remarkable links
8:       **end**
9:    **end**
10: **end** //End of the external measure
11: Initialize Graph G: $G(V, E) = \varnothing$
12: **for** each $edge(i, j) \in \mathcal{S}$ **do**
13:    Pick K-nearest neighbor ROIs for i, j as U_i, U_j
14:    **for** each ROI $p \in U_i$ **do**
15:       **for** each ROI $q \in U_j$ **do**
16:          **if** $edge(p, q) \notin G$ **then**
17:             $edge(p, q).weight \leftarrow FC(p, q)$
18:             $G.V \leftarrow G.V \cup \{p, q\}$ //add nodes
19:             $G.E \leftarrow G.E \cup edge(p, q)$ //add edges
20:          **end**
21:       **end**
22:    **end**
23: **end** //End of the internal measure

---

between two target ROIs within restricted lengths (steps). Given two target ROIs as R_i and R_j, this algorithm is mainly based on two assumptions: 1) If R_i and R_j are connected with another ROI R_k, denoted as $R_i \leftrightarrow R_k$ and $R_j \leftrightarrow R_k$, but disconnected with each other, R_i and R_j are considered to have a semi-remarkable connection in graph with a weight that should be relatively reduced as the path elongates. 2) The more restricted paths are found to connect two target ROIs, the more likely they have a strong correlation. The accumulative contributions are integrated from all potential paths between R_i and R_j as a final score. The score can be formulated as

$$score(R_i, R_j) = \sum_{t=1}^{N_{ij}} (\prod W(P_t))^{F_{decay}(P_t)} \quad (3)$$

$$F_{decay}(P) = \beta \times (len(P) - 1) \quad (4)$$

where N_{ij} is the maximum restricted length of P between R_i and R_j. β is the decay factor ranging from 1.5 to 3 according to the previous research [46]. $W(P_t)$ represents the weight of tth connection in path P. $F_{decay}(P)$ is the exponential decay function assigning less confidence to longer paths. This algorithm is easy to implement as a recursive computation. The principle of acyclicity ensures that no ROIs are repeatedly visited in each restricted path. Considering the computational complexity, the maximum length of paths is set to 2 (i.e. $N_{ij} = 2$) in this work. Besides those $3,288$ remarkable connections like $R_i \leftrightarrow R_k$, there are $7,712$ semi-

remarkable connections like $R_i \leftrightarrow R_k \leftrightarrow R_j$ explored by this algorithm. As such, total 125,639 potential paths are involved in the remarkable and semi-remarkable connections in the graph for each dataset. The final scores of the involved connections $(3,288 + 7,712 = 11,000)$ are converted into a two-dimensional matrix of size 110×100 for classification. The algorithm is shown in Algorithm 2.

Algorithm 2 The pseudo-code of RP-DFS

Input: Graph $G(V, E)$, *Max_len*
Output: A scoring matrix *T'*.
1: Initialize a scoring matrix *T'* and a path set *PS*
2: **for** *len*=1,...,*Max_len* **do**
3:    **for** each $Node_i \in G.V$ **do**
4:      **for** each $Node_j \in G.V$ **do**
5:        $PS_{ij} \leftarrow$ getAllPaths($Node_i$,$Node_j$,$G(V, E)$,*len*)
6:        **for** each path $\in PS_{ij}$ **do**
7:          $T' \leftarrow T' + Score(Node_i,Node_j)$ via Eq.3 & 4
8:        **end**
9:      **end**
10:    **end**
11: **end**
12: //define function
13: **getAllPaths**(*Origin_Node*, *Target_Node*, $G(V, E)$, *len*):
14:    **return allPaths**(*Target_Node*, *list(Origon_Node)*, *set(Origin_Node)*, $G(V, E)$, *list()*, *len*)
15: //define function
16: **allPaths**(*Target_Node*, *currentPath*, *used_Nodes*, $G(V, E)$, *answerPath*, *len*):
17:    *Last_Node* ← the last node of *currentPath*
18:    **if** *Last_Node*=*Target_Node* **then**
19:      *answerPath* ← *answerPath* ∪ *currentPath*
20:    **else**
21:      **for** each *Neighbor_Node* $\in G.V(Last\_Node)$ **do**
22:        **if** *Neighbor_Node* ∉ *used_Nodes* **and** *currentPath.len()*<*len* **then**
23:          *currentPath* ← *currentPath*∪*Neighbor_Node*
24:          *used_Nodes* ← *used_Nodes*∪*Neighbor_Node*
25:          **allPaths**(*Target_Node*, *currentPath*, *used_Nodes*, $G(V, E)$, *answerPath*, *len*) //recursion
26:          *used_Nodes* ← *used_Nodes**Neighbor_Node*
27:          *currentPath.pop()*
28:        **end**
29:      **end**
30:    **return** *answerPath*

C. DBN model with Automatic Hyperparameter Tuning

Based on the refined feature matrices, DL models are capable of automatically learning the optimal representation for the identification of ASD. In this section, we use a three-layer DBN model with automatic hyperparameter tuning technique for classification (see Fig. 5). As one of the most effective DL models [13], DBN is composed of three restricted Boltzmann machine (RBM) [47] layers of hidden units with communications between concatenated layers but not between units for each layer. Basically, the training of DBN performs in two

stages: unsupervised pretraining and supervised fine-tuning [48]. First in the unsupervised pretraining, the layers of DBN iteratively learn to probabilistically reconstruct the visible units without labels. The activation functions of visible layers and of hidden layers in RBMs are set to the affine function and sigmoid functions, respectively. Then the supervised learning of DBNs fine-tunes the resulting network with labels using the root mean square propagation (RMSProp) [49] and momentum technique as backpropagation methods, where the activation function is the rectified linear unit (ReLU) [50]. The outputs are derived from the softmax function yielding the probability of being one class (i.e. ASD or TC). Hence, there are only two units in the output layer based on one-hot encoding. To calculate the difference between estimated and true values, both unsupervised and supervised learning employ the same loss function, i.e., the cross entropy. Dropout regularization is used for reducing overfitting.

The well-known issue of using DL in classification is the strenuous and time-consuming hyperparameter tuning, which varies from different problems. Feature reduction leads to the hierarchical structure with lower levels of computational complexity that shortens the training time based on the GPU acceleration. Automatic hyperparameter tuning technique is first introduced for computational diagnosis of mental disorders using Bayesian optimization (BO) with Gaussian processes (GPs) [51]. Thanks to its useful surrogate model and good practices, we take advantages of speed to increase the number of iterations for further improving the classification accuracy. The BO-based optimizer is used to guide the potential search direction towards seminal hyperparameter sets in problem space. Three continuous and four discrete hyperparameter values shown in Fig. 5 are optimized. Based on [52], we consider mathematically the problem of seeking a global maximizer of an unknown objective function f as:

$$x^* = \arg\max_{x \in \mathcal{X} \subseteq \mathbb{R}^d} f(x) \tag{5}$$

where $\mathcal{X}$ is the certain design space of interest in global optimization. Since f is unknown, the sequential design strategy of BO treats f as a random function at first and places a prior over it. After gathering the tested data from function evaluations, the posterior distribution over f is approximated by the updated prior. In turn, the observation of posterior distribution is used for the construction of an acquisition function that determines the next query point (combination).

IV. EXPERIMENTS

For evaluation of the efficiency and effectiveness of our model, all fMRI datasets from ABIDE are used to calculate the evaluation matrices (e.g., accuracy, sensitivity, specificity, ROC curve and AUC) via 10-fold CV. Seven state-of-the-art models are compared with the proposed model for the performance comparison. Then differential evolution (DE) [53], [54] and particle swarm optimization (PSO) [55], [56] are applied as the global optimizers to compare with BO for the performance evaluation of hyperparameter tuning. We conduct a leave-one-site-out test to validate the applicability and generalization of the proposed model to the data from a new

- Activation function of visible layers in step 1: Affine
- Activation function of hidden layers in step 1: Sigmoid
- Hidden activation function in steps 2 and 3: ReLU
- Output activation function in steps 2 and 3: Softmax
- Loss function: Cross entropy
- Optimization algorithm: RMSProp and momentum
- Epochs: 50

❑ **Parameter ranges:**   Next query combination, e.g.:
1. Batch size: [8, 32] ------------------------→ (14)
2. Learning rate: [10^{-3}, 10^{-4}] ----------------→ (0.0005)
3. Momentum rate: [0.1, 0.99] ---------------→(0.87)
4. Dropout rate: [0.1, 0.5] --------------------→ (0.3)
5. Unit # of hidden layer 1: [800, 2000] --------→ (1730)
6. Unit # of hidden layer 2: [300, 700]----------→ (600)
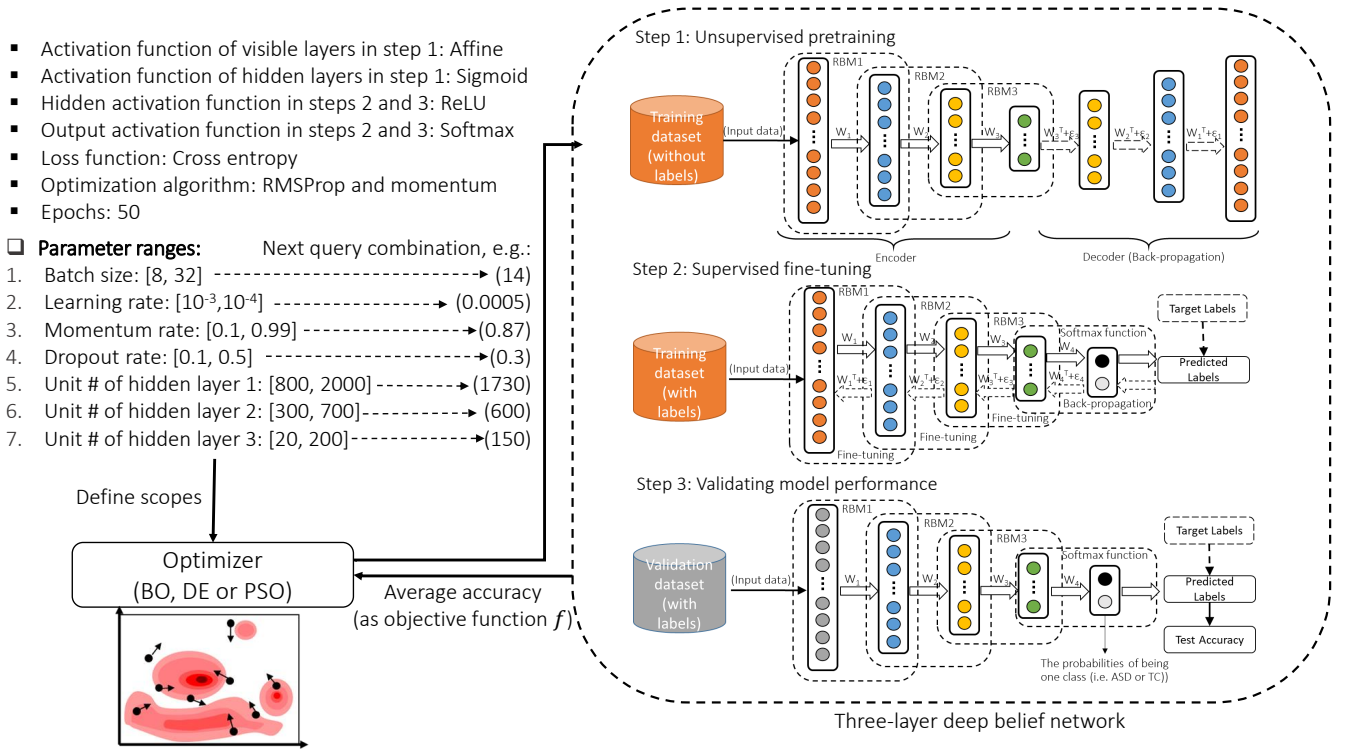7. Unit # of hidden layer 3: [20, 200]-----------→(150)

Fig. 5. A three-layer DBN classifier with automatic hyperparameter tuning of optimizer based on BO, DE or PSO

different site. The proposed model is leveraged to first identify three possible subtypes within ASD using data augmentation and over-sampling technique. We further conduct an analysis of the effects of confounding factors (e.g., age and gender) on classification accuracy. Finally, the predicted autistic neural patterns are manually validated based on the results reported in previous literatures.

A. ABIDE Dataset and Data Processing

The Autism Brain Imaging Data Exchange I (ABIDE) [14] database is a multi-site open-access consortium collected from 17 different international brain imaging laboratories. In this study, we include all valid rs-fMRI datasets from 505 ASD patients and 530 TCs along with their key phenotypical information including age, subject gender, handedness and full scale IQ. There is no consensus on the best way to preprocess rs-fMRI data. To better strengthen the data-sharing effort, the Preprocessed Connectomes Project (http://preprocessed-connectomes-project.org/abide/) publicly release the preprocessed version of ABIDE by five different teams using their preferred preprocessing strategies [52]. We select the data preprocessed through the Configurable Pipeline for the Analysis of Connectomes (C-PAC). This pipeline includes the slice time correction, motion realignment, global mean intensity normalization, nuisance variable regression, band-pass filtering (0.01-0.1Hz) and functional image transformation. The nuisance variable regression is modeled with 24 motion parameters, 5 principal components of CompCor [57] and low-frequency drifts of linear and quadratic trends, to clean confounding variations from fMRI signal. The statistical

TABLE I
PERFORMANCE COMPARISON BETWEEN OUR MODEL AND OTHER PREVIOUS STUDIES BASED ON THE ABIDE DATABASE.

Model	Classifier	Validation	sample #	Accuracy(STD)
Ours	DBN	10-fold CV	1035	**0.764 (0.022)**
Heinsfeld et al. [38]	DNN	10-fold CV	1035	0.700 (N.A.)
Dvornek et al. [39]	LSTM	10-fold CV	1035	0.685 (0.055)
Plitt et al. [19]	L-SVMs	10-fold CV	178	0.697 (0.027)
Chen et al. [20]	RFE-SVM	Train/Val	252	0.660 (N.A.)
Abraham et al. [27]	ℓ_2-SVC	10-fold CV	871	0.669 (0.027)
Nielsen et al. [23]	LOO linear	LOOCV	964	0.600 (N.A.)
Ghiassian et al. [59]	RBF-SVM	Train/Val	1035	0.592 (N.A.)

derivatives are normalized to MNI152 template space (3 mm isotropic) and spatially smoothed with a 6-mm Gaussian kernel of full width at half maximum (FWHM). The mean time series for each subject are extracted by seed voxel signals in each non-overlapping ROIs. There are seven ROI atlases providing different solutions to extract the mean time series of functional data. Among these seven sets of ROIs, we only use the Craddock 200 (CC200) functional parcellation atlas for extraction in this work [58]. This data-driven parcellation atlas leverages the two-stage spatially-functional procedure to partition individual-level connectivity graphs into 200 regions via normalized cut spectral clustering.

B. Ten-Fold Cross Validation and Leave-One-Site-Out Test

We implement 10-fold CV to investigate the performance of the proposed model. For each fold, the whole ABIDE

database is split into training/validation/testing sets with a proportion of 8:1:1. The average elapsed time of training with the whole ABIDE database is about 96 seconds using one Intel(R) Core(TM) CPU i7-8700K @ 3.70GHz and one NVIDIA GeForce RTX 2080 Ti GPU. Since the training time is shortened by the feature reduction, sufficient numbers of iterations are allowed to find a better network parameter configuration for higher classification accuracy. For sufficient exploration in each fold CV, here we limit the number of optimization iterations to 300 function evaluations (10 generations for DE and PSO with population size of 30), where one function evaluation invokes one full training and evaluation of DBN model. As we only focus on the ABIDE database in this work, the proposed DBN model is only compared with the other previous studies, whose results are also based on the ABIDE database. As shown in Table I, the four previous studies that do not use the whole ABIDE database tend to perform the ASD identification based on their specific sophisticated sampling criteria, which limits generalizability of such methods. Arbabshirani *et al.* [27] demonstrated that the main bottleneck of this field is yet the limited sample size and the high classification accuracy could degrade significantly when the sample data include over 100 participants. Therefore, the studies not using the whole ABIDE database could be more likely to suffer from overfitting based on the sophisticated sampling criteria. The main goal of our model is to achieve reliable diagnostic classification on large-scale aggregated datasets by extracting the reproducible autistic connectivity patterns. It is more desirable to put more focus on studies based on the whole ABIDE database. As we can see that, two DL-based models [38], [39] outperform all compared ML-based models [19], [20], [11], [23], [59] by using DNN and long short-term memory (LSTM) network, respectively. Besides 10-fold CV, leave-one-out cross validation (LOOCV) and single training/validation set (Train/Val) are also utilized for validation. Our DBN model achieves a mean classification accuracy of 0.764 ± 0.022 (sensitivity: 0.778, specificity: 0.750) in terms of 10-fold CV, which has been the highest accuracy achieved so far. In this work, the sensitivity tells us what percentage of ASD subjects are correctly identified while the specificity indicates the percentage of people without ASD are correctly identified.

Since hyperparameter tuning in this case can be considered as a single objective optimization problem, DE and PSO are population-based metaheuristic algorithms, which can achieve outstanding optimization performance when sufficient numbers of iterations are allowed. So DE and PSO are used to compare with BO for performance evaluation. Fig. 6 shows the performance of our model with three different optimizers as the number of generations increases. Random search is also used for the performance comparison as a baseline. For the proper division of ABIDE data into training/validation/testing sets, we perform two ways of stratified random sampling by maintaining the balance of label distribution (shown in Fig. 6a) and each site frequency (shown in Fig. 6b), respectively. The effect of the stratified random sampling based on each site frequency is slightly worse than that of the stratified random sampling based on label distribution. As shown in Fig. 6b, the
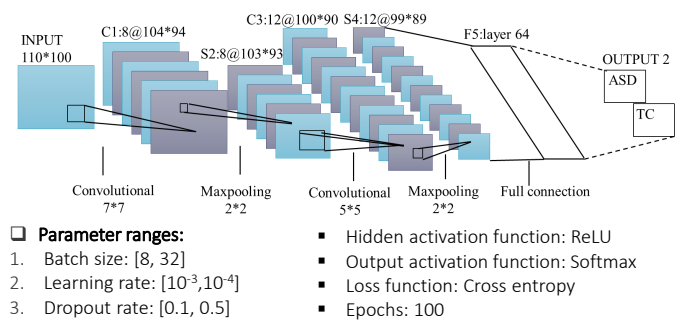

Fig. 7. The architecture and setting of CNN are designed by manual tunning. Batch size, learning rate and dropout rate are optimized by the BO-based optimizer.

BO-based optimizer achieves the highest accuracy of 0.762 while the other three optimizers show different levels of performance degradation, which could be attributed to the sample number imbalance between each site. Within two generations, all compared optimizers tend to show slower convergence. Consequently, we adopt the label distribution-based stratified random sampling for data splitting in the following simulation experiments. In Fig. 6a, within two generations, the BO-based optimizer shows faster convergence while the other two optimizers perform like a random search. Then the BO-based optimizer gets stuck on local optima while the other two optimizers manage to find a near-optimal global solution. Finally, the DE-based optimizer obtains the highest accuracy of 0.764, which is comparable to the PSO-based optimizer's accuracy 0.763. And the accuracy of BO-based optimizer is 0.756. Considering the expensive computational cost of DBN training, BO tends to obtain better trade-off of accuracy and efficiency.

The evaluation of most previous studies is based on one time CV schema. However the effect of random sample divisions in CV should be considered. To reduce such bias, ten times of 10-fold CV are conducted on our model with BO-based optimizer (see Table II). Only one generation (i.e. 30 iterations) is allowed for each fold CV in order to evaluate its feasibility and practicability within finite time. RF, SVM and CNN (convolutional neural network [60], [61]) models are used for performance comparison under the same feature selection framework. The architecture and setting of CNN are shown in Fig. 7. For fair test, the parameters of RF and SVM are optimized from all potential parameter combinations via GridSearchCV, a grid-search module from scikit-learn [35]. As the best model reported to date using the whole ABIDE database, the Heinsfeld's framework with different classifiers is also used for comparison. The results derived from their paper [38] are based on one time 10-fold CV. For a comprehensive assessment of our model, the results of three scenarios are analyzed, i.e., the mean-case, best-case and worst-case scenarios. As we can see in Table II, random sample divisions do cause a certain impact on accuracy ranging from 3.3% to 1.3% in our work with different classifiers. Under the proposed feature selection framework, RF and SVM achieve

(a) Stratified random sampling based on label distribution.

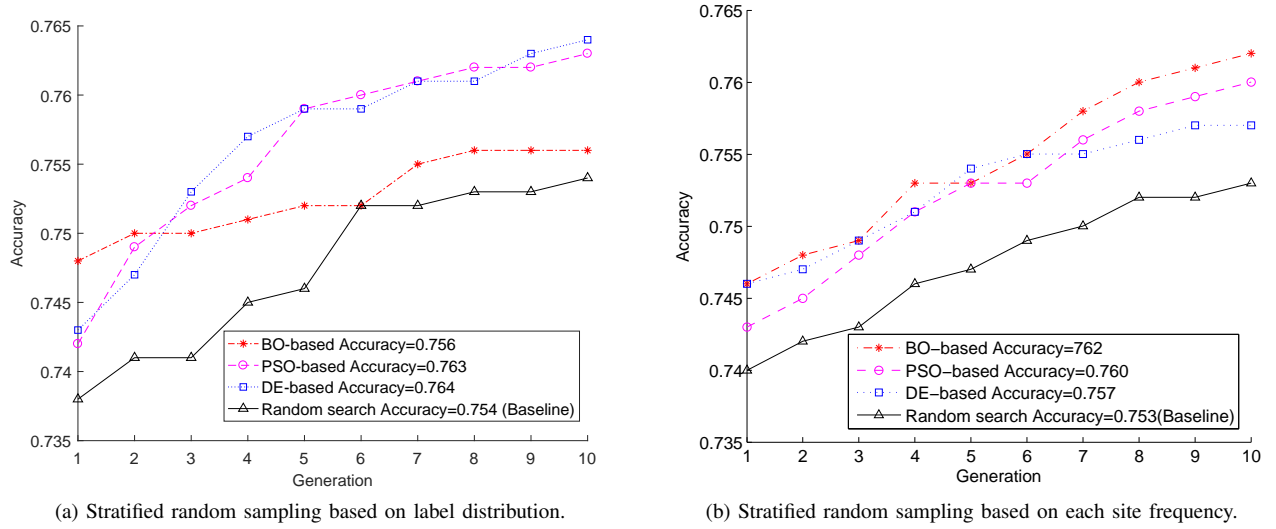(b) Stratified random sampling based on each site frequency.

Fig. 6. The optimal convergence achieved by our model with three different optimizers as the generation number of hyperparameter tuning increases.

TABLE II
PERFORMANCE COMPARISON BETWEEN OUR WORK AND HEINSFELD'S WORK BASED ON DIFFERENT CLASSIFIERS.

Ten times 10-fold CV in our work under the same feature selection framework						One time 10-fold CV in Heinsfeld's work [38]			
Ours	Mean			Best	Worst				
Classifier	Sensitivity	Specificity	Accuracy (STD)	Accuracy (STD)	Accuracy (STD)	Classifier	Sensitivity	Specificity	Accuracy (STD)
RF	0.734	0.580	0.656 (0.046)	0.665 (0.031)	0.650 (0.024)	RF	0.69	0.58	0.63 (N.A.)
SVM	0.691	0.646	0.670 (0.047)	0.682 (0.041)	0.649 (0.057)	SVM	0.68	0.62	0.65 (N.A.)
DBN	0.762	0.731	0.745 (0.031)	0.751 (0.029)	0.738 (0.027)	DNN	0.74	0.63	0.70 (N.A.)
CNN	0.745	0.726	0.733 (0.046)	0.736 (0.041)	0.718 (0.052)				

the mean accuracies of 0.656 and 0.670, respectively, which are nearly 2% higher than those in Heinsfeld's work. Note that the Heinsfeld's work employed the entire connectivity features whereas the proposed model uses only 16.52% of them. This result well demonstrates the effectiveness of our model. As expected, the DL-based classifiers outperform the traditional ML-based classifiers. DBN achieves the best performance with the mean accuracy of 0.745±0.031. The translation invariance property of CNN may be more suitable to resolve image classification problems, instead of such graph-structured feature data. This result demonstrates the successful application of DBN to complex brain imaging data, showing high potential of neural networks for learning and detecting the underlying relationships between the extracted features acquired from neuroimaging data and the psychological representations.

The contributions of the three proposed methods are also evaluated in Table III, i.e. GBFS, RP-DFS and the automatic hyperparameter tuning of BO. Given the reliable performance of the GBFS method, three representative methods are used for performance comparison, i.e., the RF wrapper method [62], the t-test filter method [63], and the chi-squared statistic filter method [64]. As shown in Table. IV, the proposed GBFS method achieves the reliable overall performance with the highest mean accuracy and specificity. The other three compared methods can obtain high sensitivities. Specifically, the RF wrapper method obtains the highest sensitivity of

TABLE III
PERFORMANCE COMPARISON FOR USING *GBFS*, *RP-DFS* AND *BO* IN DBN MODEL VIA 10 TIMES 10-FOLD CV.

	Mean accuracy	Sensitivity	Specificity
DBN with *GBFS*	0.716 (0.035)	0.750	0.676
DBN with *GBFS,RP-DFS*	0.725 (0.037)	0.740	0.710
DBN with *GBFS,RP-DFS,BO*	0.745 (0.031)	0.762	0.731

TABLE IV
PERFORMANCE COMPARISON BETWEEN GBFS AND THE REPRESENTATIVE FEATURE SELECTION METHODS VIA 10 TIMES 10-FOLD CV.

Method	Class	Mean accuracy	Sensitivity	Specificity
GBFS	Filter	**0.745 (0.031)**	0.762	**0.731**
RF	Wrapper	0.734 (0.042)	**0.806**	0.641
T-test	Filter	0.693 (0.058)	0.764	0.598
Chi-squared	Filter	0.687 (0.069)	0.768	0.606

0.806. However, they all fail to manage a trade-off between sensitivity and specificity, which likely fluctuates their mean accuracies with relatively high STDs. The comparison results demonstrate the superiority of the proposed method for FC feature selection.

Since the primary goal of this work is the intelligent auxiliary diagnosis for individuals with ASD, the sensitivity

TABLE V
PERFORMANCE COMPARISON FOR DEEPER LEARNING NETWORK IN DBN
VIA 10 TIMES 10-FOLD CV.

Layers	3	4	5
Average accuracy	0.745 (0.031)	0.737 (0.025)	0.727 (0.026)

TABLE VI
LEAVE-ONE-SITE-OUT TEST.

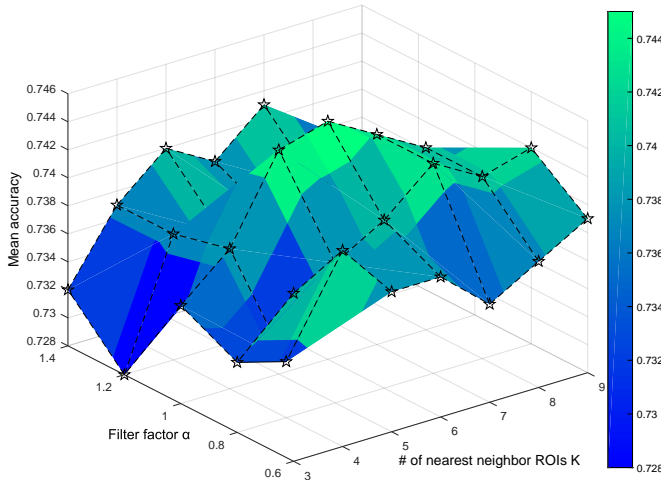| Site	Vendor	FD	# of ASD	# of TC	Accuracy
KKI	Philips	0.17	22 (8-13)	33 (8-13)	0.792
LEUVEN	Philips	0.09	29 (12-32)	35 (12-32)	0.810
SBL	Philips	0.16	15 (22-64)	15 (20-41)	0.667
TRINITY	Philips	0.11	24 (12-26)	25 (12-26)	0.766
NYU	Siemens	0.07	79 (7-39)	105 (7-32)	0.743
OHSU	Siemens	0.10	13 (8-15)	15 (8-12)	0.731
OLIN	Siemens	0.18	20 (11-24)	15 (10-23)	0.765
PITT	Siemens	0.15	30 (9-35)	27 (9-33)	0.786
UCLA	Siemens	0.19	62 (8-18)	47 (9-18)	0.786
MAX_MUN	Siemens	0.13	24 (7-58)	33 (7-48)	0.673
CMU	Siemens	0.29	14 (19-39)	13 (20-40)	0.889
USM	Siemens	0.14	58 (11-50)	43 (9-39)	0.859
CALTECH	Siemens	0.07	19 (17-45)	19 (17-56)	0.811
YALE	Siemens	0.11	28 (7-18)	28 (8-18)	0.839
SDSU	GE	0.09	14 (12-17)	22 (9-17)	0.806
STANFORD	GE	0.11	20 (8-13)	20 (8-12)	0.795
UM	GE	0.14	68 (8-19)	77 (8-29)	0.771
Mean	N.A.	0.14	31.7	33.6	0.782

Fig. 9. Parameter tuning of α and K via 10-fold CV. The lighter of the color you observe, the higher mean accuracy of the area it represents.

is more preferable to the specificity. It means that the missed diagnosis could bring more severe consequences than the mis-diagnosis does. ROC and AUC are widely used to evaluate the performance of binary classification problems by calculating the true positive rate (sensitivity) and false positive rate (1-specificity) as changing the thresholds. The ROC curve and AUC value are used to observe the performance of our model for the best-case and worst-case scenarios via 10-fold CV (see Fig 8). As a result, DBN achieves the highest AUC values of 0.7535 and 0.7425 in the best-case and worst-case scenarios, respectively. DBN and CNN can obtain a good true positive rate while maintaining a relatively low false positive rate. Table II also demonstrates that the sensitivities of DBN and CNN are higher than their specificities. Moreover, to investigate the potential of deeper learning network in DBN, we conduct a contrast experiment by gradually increasing hidden layers of the architecture from three to five. Based on the result of Table. V, the three-layer architecture is thereby adopted in DBN model. The filter factor α and the number of nearest neighbor ROIs K are two key parameters for the proposed GBFS method. The selection of α and K is determined based on the experiment result of 10-fold CV, as shown in Fig. 9. As we can see, the best performance is achieved with the combination of $\alpha = 1$ and $K = 6$. Therefore, the values of α and K are empirically set to 1 and 6, respectively.

Since ABIDE is the repository aggregated across 17 international sites without prior coordination, the proposed model should be reliable enough against the effects brought by different experimental settings and participants. Leave-one-site-out test is used to evaluate the applicability and generalization of the proposed model to a new different site. As for leave-one-site-out test, the datasets from one site are used to test the classification accuracy while the remaining datasets from other sites are split into training/validation sets in a proportion of 8:2. MRI vendors, mean framework displacement (FD) and phenotypic information are selected as three representative factors. Age ranges are correspondingly marked in brackets. Mean FD is a measure of subject head motion to compare the change between the current and previous volumes. The lower value of the mean FD indicates the less subject head motion. As we can see in Table VI, the proposed model achieves the mean accuracy of 0.782. Four sites show significantly lower accuracies than the mean: SBL, NYU, OHSU and MAX_MUN. Note that the accuracies of SBL, OHSU and MAX_MUN are also significantly lower the mean in Heinsfeld's work [38]. This means that these sites might hold site-specific variability and heterogeneity being absent in other sites. The result suggest that there is no significant effect of MRI vendor, mean FD and phenotypic information on the classification performance. It could be concluded that the proposed model holds a high reliability against the uncertainties from new sites.

C. Identification of Possible Subtypes within ASD

As a range of mental disorders, ASD can be further diagnosed as autism, Asperger syndrome, pervasive developmental disorder not otherwise specified (PDD-NOS) and etc, based on domains of impairment in ASD (see Table VII) [2]. The subtle changes of their clinical presentations are challenging to be detected and distinguished via symptom-based diagnostic criteria, and that obstructs more precise therapeutic decision-making for individuals with ASD. To the best of our knowl-

(a) The best-case scenario for the four classifiers in our work

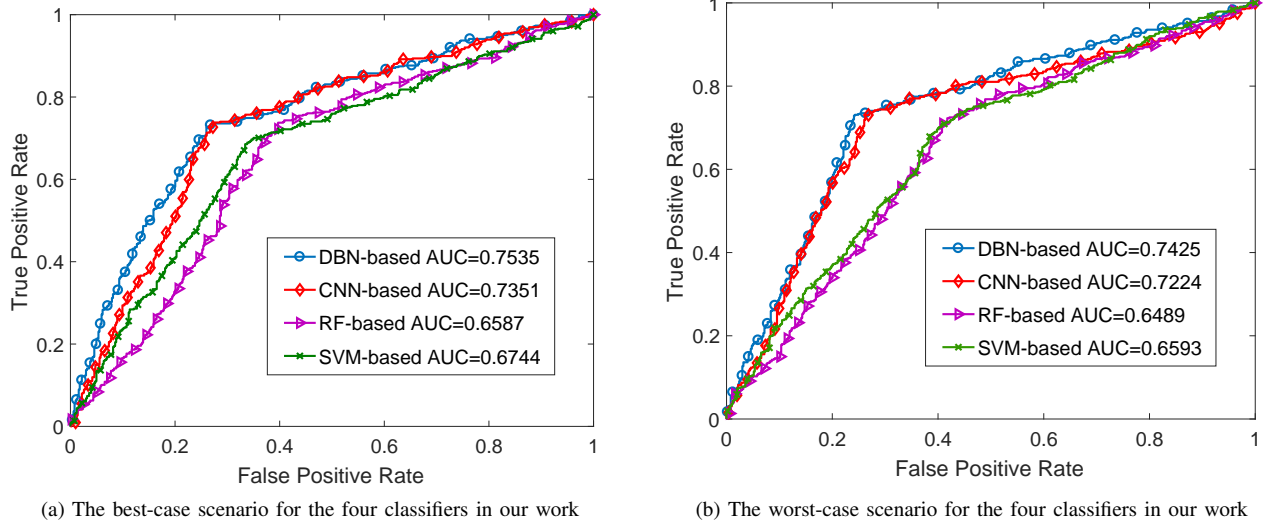(b) The worst-case scenario for the four classifiers in our work

Fig. 8. The classification performance of best-case and worst-case scenarios for the four classifiers evaluated by ROC curve and AUC in 10-fold CV.

edge, no computational models have been proposed to identify possible subtypes within ASD using fMRI data. In this study, we try to leverage the proposed model to address this issue as a multi-class classification problem. Table VII shows that ABIDE database is imbalanced for consisting of 323 autism samples, 87 Asperger samples and 35 PDD-NOS samples, with a ratio of 9.2:2.5:1.

To balance the dataset before training, data augmentation and oversampling are used (see Fig. 10). First, we randomly crop the raw input time course to the same sequence length for each sample of Asperger and PDD-NOS. Considering the length of the shortest time-series, we make such a fourfold augmentation to yield more sample data for training and testing. However, PDD-NOS samples still make up a small percentage of 17% of the augmented datasets, where the amount of Asperger samples is comparable to autism samples'. It means that more PDD-NOS samples are required for training the model. The synthetic minority over-sampling technique (SMOTE) [65] is introduced to synthesize new minority instances (i.e. PDD-NOS samples) between a pair of one minority instance and one of its K nearest neighbors. To be fair, SMOTE is not used for the test set, preventing a high rate of false-positive results. Finally, the proposed model is implemented on the augmented datasets likewise. To better evaluate such imbalanced data, F1-measure is used as a measure metrics for each ASD subtype as follows.

$$
\begin{cases}
F1 - measure = \dfrac{2 \times precision \times recall}{(precision + recall)} \\[2ex]
precision = \dfrac{True\ positive}{True\ positive + False\ positive} \\[2ex]
recall = \dfrac{True\ positive}{True\ positive + False\ negative}
\end{cases}
\tag{6}
$$

Macro-average accuracy is also used to globally evaluate the performance by weighting toward the minority class, i.e.,

the PDD-NOS samples. As shown in Table VIII, all compared classifiers obtain the reliable performance in samples of autism and Asperger by the same feature selection of GBFS. DBN classifier achieves the highest macro-average accuracy of 0.838. All classifiers in different extent show biased predictive capability that have a higher F1-measure over the majority classes but poorer F1-measure over the minority class. However, DBN classifier shows moderate improvement in prediction performance on PDD-NOS samples. By taking out the minority class i.e. PDD-NOS, we also use the proposed model to identify autism and Asperger syndrome. In 10 times 10-fold CV, the proposed model achieves the satisfactory mean accuracy of 0.896±0.030 with high sensitivity of 0.885 and high specificity of 0.907. It demonstrates that our model can make a reliable identification for autism and Asperger syndrome.

It also needs to note that there are some downsides to perform data augmentation and oversampling on datasets. Since we make a fourfold augmentation to yield more sample data for training and testing, the extracted FC features are derived from shorter mean time series of rs-fMRI data, i.e., a quarter of the original ones. The quality of extracted feature could be affected to a certain extent. Furthermore, the oversampling technique can create the synthetic examples without considering the majority class (i.e., autism, Asperger syndrome in this study). It could lead to overfitting to the training data. To be sure, every approach has its own downsides. Nevertheless, using data augmentation and oversampling in our model is quite effective to increase training samples and equilibrate class weights by mitigating the issue of class imbalance. This work is expected to provide valuable insights into the identification of possible subtypes within ASD using fMRI data.

D. Effects of Confounding Factors on Classification Accuracy

Some neuroscientific work may specify recruitment strategies and participant instructions for particular research purposes. For example, in general, children with ASD should
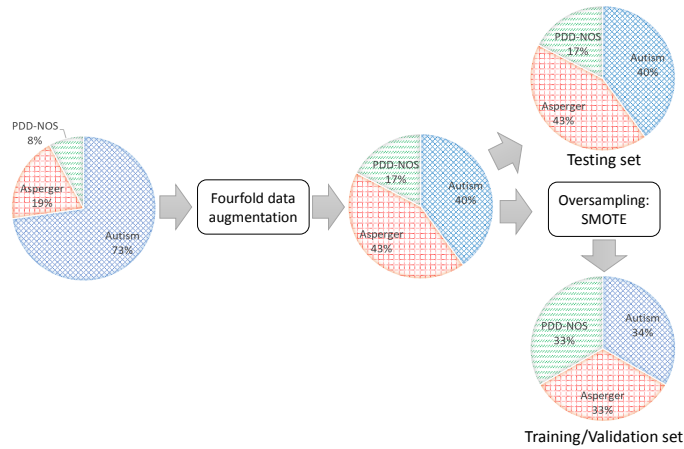
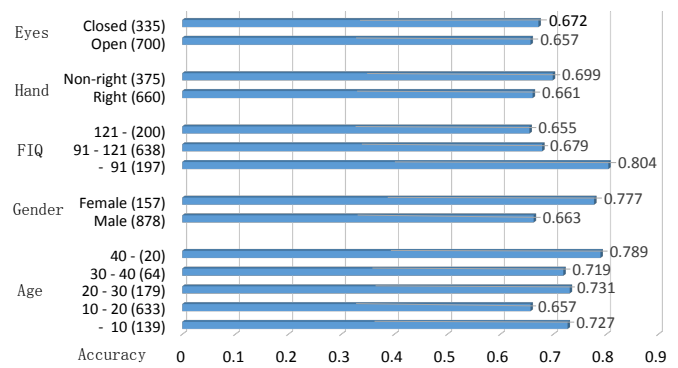Fig. 10. Data augmentation and SMOTE are used for balancing the datasets.

Fig. 11. The effects of eye status, handedness, FIQ, gender and age are evaluated on prediction accuracy as confounding factors. The sample numbers are given in brackets.

TABLE VII
DOMAINS OF IMPAIRMENT IN ASD AND SAMPLE NUMBERS OF ASD SUBTYPES IN ABIDE.

	Autism	Asperger	PDD-NOS
social communication	required	required	required
language	required	normal	variable
repetitive behaviors	required	required	variable
Sample # in ABIDE	323	87	35

TABLE VIII
COMPARISON RESULTS FOR IDENTIFICATION OF ASD SUBTYPES VIA 10 TIMES 10-FOLD CV.

	Autism	Asperger	PDD-NOS	Total
Classifier	F1-measure			Macro-average accuracy
DBN	0.940	0.902	0.603	0.838
CNN	0.922	0.915	0.547	0.828
RF	0.880	0.806	0.362	0.761
SVM	0.933	0.887	0.471	0.802

be diagnosed before 6 years old otherwise it would be too late for effective treatments. When testing the sample sets with a certain condition (e.g. age<6), a reliable model should keep robustness and show no significant difference in accuracy based on the remaining samples as training data. Therefore, it is necessary to evaluate the effects of confounding factors on classification accuracy. As shown in Fig. 11, basically, there are no significant differences observed between accuracy and the confounding factors including eye status, handedness, FIQ scores, gender, and age. The imbalanced datasets have a negative effect on prediction accuracy. It should be noted that the accuracy of the low FIQ group (< 91) is 14.9% higher than that of the high FIQ group (> 121) based on the balanced sample data. This result suggests that intellectual disability could be a good diagnostic feature for ASD subjects.

E. Model Interpretation: Neural Patterns in The Autistic Brain

Based on the connectivity matrices refined by RP-DFS, the correlations between two ROIs are represented by the accumulated scores from all potential paths. Given the assumptions of such correlation patterns, those connections with high accumulated scores could be highly connected (positively correlated). Otherwise, the connections with low accumulated scores tend to be underconnected (negatively correlated). To derive the reproducible autistic neural patterns, the top-10 ranked remarkable connections are selected based on their mean scores from the ASD group (see Fig. 12). Then their ROIs are manually validated according to the results reported in the previous literatures. Table IX demonstrates the manual validation of those involved ROIs that have been confirmed to have associations with ASD via Digital Object Identifiers (DOIs). ROI numbers of CC200 atlas are given in brackets. The result shows that 65% (13/20) and 60% (12/20) of the involved ROIs are confirmed to have highly connected function and underconnected function in the autistic brain, respectively. Furthermore, these experimental data are consistent with the findings from previous literatures. For examples, in adults with ASD, a specific focal increase in cortical thickness is found at the left fusiform gyrus (122) [66]. It could result in the associated impairments in face processing for ASD subjects. In this work, the left fusiform gyrus (122) is predicted to have a strong positive correlation (1st in the prediction list) with the right precentral gyrus (115) in the autistic brain. Moreover, the area of reduced gray matter volumes in autistic children is found in the left superior frontal gyrus (173) by the analysis of voxel-based morphometry [67]. In this work, the left superior frontal gyrus (173) is predicted to have a strong negative correlation (4th in the prediction list) with the left precuneous cortex (19) in the autistic brain. These predicted autistic neural patterns are anticipated to serve as reproducible biomarkers offering insights into the pathophysiological mechanism of ASD.

V. CONCLUSION

We presented a novel graph-based computational model for identification of ASD using rs-fMRI data. A GBFS method

TABLE IX
MANUAL VALIDATION FOR THOSE INVOLVED ROIS IN PREDICTED AUTISTIC CONNECTIONS BASED ON PREVIOUS LITERATURES.

ROI	DOI	ROI	DOI	Score
Top-10 positively correlated predicted connections in the autistic brain				
Left Fusiform Gyrus (122)	10.1001/archgenpsychiatry.2010.31	Right Precentral Gyrus (115)	10.1097/01.wnr.0000233087.15710.87	1.8842 (1)
Left Superior Temporal Gyrus (146)	10.1016/j.pscychresns.2005.12.009	Right Thalamus (18)	10.1016/j.brainres.2009.12.081	1.6430 (2)
Right Middle Temporal Gyrus (69)	10.1017/S1355617708081216	Right Middle Temporal Gyrus (107)	10.1017/S1355617708081216	1.4668 (3)
Left Middle Cingulate Cortex (76)	unconfirmed	Right Middle Temporal Gyrus (85)	10.1017/S1355617708081216	1.4565 (4)
Right Caudate Nucleus (135)	10.1186/s11689-015-9107-8	Right Precentral Gyrus (115)	10.1007/s10803-011-1221-1	1.3202 (5)
Right Middle Temporal Gyrus (14)	10.1017/S1355617708081216	Left Middle Occipital Gyrus (131)	unconfirmed	1.3064 (6)
Left Middle Frontal Gyrus (125)	unconfirmed	Left Middle Frontal Gyrus (61)	unconfirmed	1.3018 (7)
Left Postcentral Gyrus (88)	10.1016/j.biopsych.2015.10.020	Left Cerebellum (36)	unconfirmed	1.2491 (8)
Right Inferior Frontal Gyrus (53)	10.1016/j.neuroimage.2004.12.022	Left Cerebellum (52)	unconfirmed	1.2473 (9)
Left Middle Cingulate Cortex (76)	unconfirmed	Left Temporal Pole (78)	10.1016/j.biopsych.2012.12.013	1.2431 (10)
Top-10 negatively correlated predicted connections in the autistic brain				
Left Precentral Gyrus (90)	10.1136/jnnp.2010.239111	Left Middle Occipital Gyrus (131)	unconfirmed	-0.2706 (1)
Right Frontal Orbital Cortex (71)	unconfirmed	Right Parahippocampal Gyrus (155)	10.1016/j.neuroimage.2009.04.069	-0.2696 (2)
Right Frontal Orbital Cortex (71)	unconfirmed	Right Caudate Nucleus (135)	10.1186/s11689-015-9107-8	-0.2480 (3)
Left Superior Frontal Gyrus (173)	10.1016/j.brainresbull.2010.12.002	Left Precuneus Cortex (19)	unconfirmed	-0.2479 (4)
Right Caudate Nucleus (135)	10.1186/s11689-015-9107-8	Left Precentral Gyrus (90)	10.1136/jnnp.2010.239111	-0.2447 (5)
Left Putamen (4)	10.1007/s12264-017-0118-1	Right Parahippocampal Gyrus (155)	10.1016/j.neuroimage.2009.04.069	-0.2436 (6)
Left Middle Temporal Gyrus (2)	10.1016/j.expneurol.2003.09.010	Right Subcallosal Cortex	unconfirmed	-0.2423 (7)
Right Frontal Pole (113)	10.1016/j.biopsych.2013.06.018	Left Frontal Pole (183)	10.1016/j.neuroimage.2004.02.029	-0.2421 (8)
Left Precuneus Cortex (19)	unconfirmed	Left Frontal Orbital Cortex (57)	unconfirmed	-0.2417 (9)
Left Middle Occipital Gyrus (131)	unconfirmed	Left Putamen (4)	10.1007/s12264-017-0118-1	-0.2385 (10)

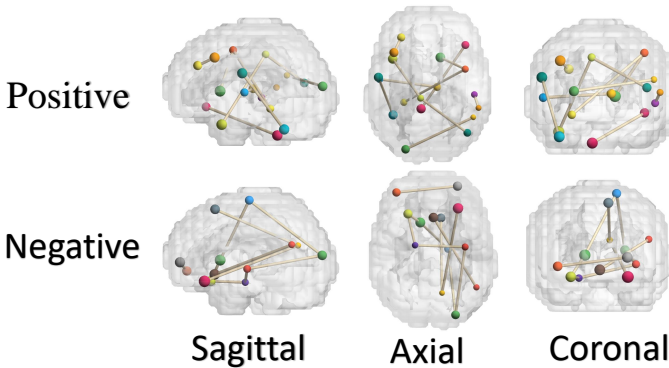Positive

Negative

Sagittal　Axial　Coronal

Fig. 12. The most positively/negatively correlated predicted connections with highest/lowest scores.

was proposed to highlight the remarkable connections via both external and internal measure. To take advantage of topological information implied in graph, RP-DFS was proposed to further refine the remarkable connectivity matrices. Finally, a three-layer DBN model with automatic hyperparameter tuning technique was applied for classification. Compared with other state-of-the-art methods, our model achieved the highest mean accuracy of 0.764±0.022. The reliable performance was fully demonstrated by the comprehensive experiments. This work also provides an insight into the identification of possible subtypes within ASD. By a statistical analysis, the interpretability of our model enabled to uncover the correlation patterns in the autistic brain. The current work is anticipated to serve as a can-

didate tool for identifying ASD subjects using rs-fMRI data. We expect that the proposed model could offer insights into computer-aided diagnosis of complex psychological disorders using neural network models.

For future work, it would be interesting to develop a multi-task learning classification framework to capture the site-shared and site-specific feature patterns from multi-site imaging data. The merits of multimodal data fusion [68] can be utilized to improve the spatio-temporal resolution of characterization of brain state by combining rs-fMRI and structural MRI (sMRI) data.

REFERENCES

[1] S. Greenspan, "Autism spectrum disorder," 2015.
[2] D. H. Geschwind, "Advances in autism," *Annual review of medicine*, vol. 60, pp. 367–380, 2009.
[3] E.-M. Rødgaard, K. Jensen, J.-N. Vergnes, I. Soulières, and L. Mottron, "Temporal changes in effect sizes of studies comparing individuals with and without autism: A meta-analysis," *JAMA psychiatry*, 2019.
[4] S. A. Huettel, A. W. Song, G. McCarthy *et al.*, *Functional magnetic resonance imaging*. Sinauer Associates Sunderland, MA, 2004, vol. 1.
[5] 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.
[6] Z. Yu-Feng, H. Yong, Z. Chao-Zhe, C. Qing-Jiu, S. Man-Qiu, L. Meng, T. Li-Xia, J. Tian-Zi, and W. Yu-Feng, "Altered baseline brain activity in children with adhd revealed by resting-state functional mri," *Brain and Development*, vol. 29, no. 2, pp. 83–91, 2007.
[7] V. L. Cherkassky, R. K. Kana, T. A. Keller, and M. A. Just, "Functional connectivity in a baseline resting-state network in autism," *Neuroreport*, vol. 17, no. 16, pp. 1687–1690, 2006.

[8] J. S. Anderson, J. A. Nielsen, A. L. Froehlich, M. B. DuBray, T. J. Druzgal, A. N. Cariello, J. R. Cooperrider, B. A. Zielinski, C. Ravichandran, P. T. Fletcher *et al.*, "Functional connectivity magnetic resonance imaging classification of autism," *Brain*, vol. 134, no. 12, pp. 3742–3754, 2011.

[9] M. Greicius, "Resting-state functional connectivity in neuropsychiatric disorders," *Current opinion in neurology*, vol. 21, no. 4, pp. 424–430, 2008.

[10] F. X. Castellanos, A. Di Martino, R. C. Craddock, A. D. Mehta, and M. P. Milham, "Clinical applications of the functional connectome," *Neuroimage*, vol. 80, pp. 527–540, 2013.

[11] A. Abraham, M. P. Milham, A. Di Martino, R. C. Craddock, D. Samaras, B. Thirion, and G. Varoquaux, "Deriving reproducible biomarkers from multi-site resting-state data: an autism-based example," *NeuroImage*, vol. 147, pp. 736–745, 2017.

[12] D. Nie, L. Wang, Y. Gao, J. Lian, and D. Shen, "Strainet: Spatially varying stochastic residual adversarial networks for mri pelvic organ segmentation," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 5, pp. 1552–1564, 2018.

[13] G. E. Hinton, "Deep belief networks," *Scholarpedia*, vol. 4, no. 5, p. 5947, 2009.

[14] A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto *et al.*, "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism," *Molecular psychiatry*, vol. 19, no. 6, p. 659, 2014.

[15] K. S. Kassam, A. R. Markey, V. L. Cherkassky, G. Loewenstein, and M. A. Just, "Identifying emotions on the basis of neural activation," *PloS one*, vol. 8, no. 6, p. e66032, 2013.

[16] A. J. Bauer and M. A. Just, "Monitoring the growth of the neural representations of new animal concepts," *Human brain mapping*, vol. 36, no. 8, pp. 3213–3226, 2015.

[17] A. J. O'toole, F. Jiang, H. Abdi, and J. V. Haxby, "Partially distributed representations of objects and faces in ventral temporal cortex," *Journal of cognitive neuroscience*, vol. 17, no. 4, pp. 580–590, 2005.

[18] N. Yahata, K. Kasai, and M. Kawato, "Computational neuroscience approach to biomarkers and treatments for mental disorders," *Psychiatry and clinical neurosciences*, vol. 71, no. 4, pp. 215–237, 2017.

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

[20] C. P. Chen, C. L. Keown, A. Jahedi, A. Nair, M. E. Pflieger, B. A. Bailey, and R.-A. Müller, "Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism," *NeuroImage: Clinical*, vol. 8, pp. 238–245, 2015.

[21] M. Mahmud, M. S. Kaiser, A. Hussain, and S. Vassanelli, "Applications of deep learning and reinforcement learning to biological data," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 6, pp. 2063–2079, 2018.

[22] M. Kachuee, S. Darabi, B. Moatamed, and M. Sarrafzadeh, "Dynamic feature acquisition using denoising autoencoders," *IEEE transactions on neural networks and learning systems*, 2018.

[23] J. A. Nielsen, B. A. Zielinski, P. T. Fletcher, A. L. Alexander, N. Lange, E. D. Bigler, J. E. Lainhart, and J. S. Anderson, "Multisite functional connectivity mri classification of autism: Abide results," *Frontiers in human neuroscience*, vol. 7, p. 599, 2013.

[24] Y. Du, Z. Fu, and V. D. Calhoun, "Classification and prediction of brain disorders using functional connectivity: Promising but challenging," *Frontiers in neuroscience*, vol. 12, 2018.

[25] A. Khazaee, A. Ebrahimzadeh, and A. Babajani-Feremi, "Identifying patients with alzheimers disease using resting-state fmri and graph theory," *Clinical Neurophysiology*, vol. 126, no. 11, pp. 2132–2141, 2015.

[26] E. Challis, P. Hurley, L. Serra, M. Bozzali, S. Oliver, and M. Cercignani, "Gaussian process classification of alzheimer's disease and mild cognitive impairment from resting-state fmri," *NeuroImage*, vol. 112, pp. 232–243, 2015.

[27] M. R. Arbabshirani, S. Plis, J. Sui, and V. D. Calhoun, "Single subject prediction of brain disorders in neuroimaging: promises and pitfalls," *NeuroImage*, vol. 145, pp. 137–165, 2017.

[28] Y. Du, G. D. Pearlson, J. Liu, J. Sui, Q. Yu, H. He, E. Castro, and V. D. Calhoun, "A group ica based framework for evaluating resting fmri markers when disease categories are unclear: application to schizophrenia, bipolar, and schizoaffective disorders," *Neuroimage*, vol. 122, pp. 272–280, 2015.

[29] J. Yang and V. Honavar, "Feature subset selection using a genetic algorithm," in *Feature extraction, construction and selection*. Springer, 1998, pp. 117–136.

[30] T. Iidaka, "Resting state functional magnetic resonance imaging and neural network classified autism and control," *Cortex*, vol. 63, pp. 55–67, 2015.

[31] V. Fonti and E. Belitser, "Feature selection using lasso," *VU Amsterdam Research Paper in Business Analytics*, 2017.

[32] J. Wang, Q. Wang, J. Peng, D. Nie, F. Zhao, M. Kim, H. Zhang, C.-Y. Wee, S. Wang, and D. Shen, "Multi-task diagnosis for autism spectrum disorders using multi-modality features: A multi-center study," *Human brain mapping*, vol. 38, no. 6, pp. 3081–3097, 2017.

[33] D. Zhang, D. Shen, A. D. N. Initiative *et al.*, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease," *NeuroImage*, vol. 59, no. 2, pp. 895–907, 2012.

[34] H. Wang, F. Nie, H. Huang, S. L. Risacher, A. J. Saykin, L. Shen, and A. D. N. Initiative, "Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning," *Bioinformatics*, vol. 28, no. 12, pp. i127–i136, 2012.

[35] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[36] X. Pan and Y. Xu, "A novel and safe two-stage screening method for support vector machine," *IEEE transactions on neural networks and learning systems*, 2018.

[37] T. Zhai, F. Koriche, H. Wang, and Y. Gao, "Tracking sparse linear classifiers," *IEEE transactions on neural networks and learning systems*, 2018.

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

[39] N. C. Dvornek, P. Ventola, K. A. Pelphrey, and J. S. Duncan, "Identifying autism from resting-state fmri using long short-term memory networks," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2017, pp. 362–370.

[40] M. Khosla, K. Jamison, A. Kuceyeski, and M. R. Sabuncu, "3d convolutional neural networks for classification of functional connectomes," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 137–145.

[41] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, 2017.

[42] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, "Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2306–2318, 2016.

[43] M. Xia, J. Wang, and Y. He, "Brainnet viewer: a network visualization tool for human brain connectomics," *PloS one*, vol. 8, no. 7, p. e68910, 2013.

[44] M. A. Just, T. A. Keller, V. L. Malave, R. K. Kana, and S. Varma, "Autism as a neural systems disorder: a theory of frontal-posterior underconnectivity," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 4, pp. 1292–1313, 2012.

[45] Z.-A. Huang, X. Chen, Z. Zhu, H. Liu, G.-Y. Yan, Z.-H. You, and Z. Wen, "Pbhmda: path-based human microbe-disease association prediction," *Frontiers in microbiology*, vol. 8, p. 233, 2017.

[46] W. Ba-Alawi, O. Soufan, M. Essack, P. Kalnis, and V. B. Bajic, "Daspfind: new efficient method to predict drug–target interactions," *Journal of cheminformatics*, vol. 8, no. 1, p. 15, 2016.

[47] C. P. Chen and S. Feng, "Generative and discriminative fuzzy restricted boltzmann machine learning for text and image classification," *IEEE transactions on cybernetics*, 2018.

[48] C. Zhang, K. C. Tan, H. Li, and G. S. Hong, "A cost-sensitive deep belief network for imbalanced classification," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 1, pp. 109–122, 2018.

[49] F. Zou, L. Shen, Z. Jie, W. Zhang, and W. Liu, "A sufficient condition for convergences of adam and rmsprop," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 127–11 135.

[50] Y. Li and Y. Yuan, "Convergence analysis of two-layer neural networks with relu activation," in *Advances in Neural Information Processing Systems*, 2017, pp. 597–607.

[51] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, 2012, pp. 2951–2959.

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

[53] L. Tang, X. Wang, and Z. Dong, "Adaptive multiobjective differential evolution with reference axis vicinity mechanism," *IEEE transactions on cybernetics*, vol. 49, no. 9, pp. 3571–3585, 2018.

[54] S. X. Zhang, S. Y. Zheng, and L. M. Zheng, "An efficient multiple variants coordination framework for differential evolution," *IEEE transactions on cybernetics*, vol. 47, no. 9, pp. 2780–2793, 2017.

[55] L. Li, L. Chang, T. Gu, W. Sheng, and W. Wang, "On the norm of dominant difference for many-objective particle swarm optimization." *IEEE transactions on cybernetics*, 2019.

[56] Y.-F. Zhang and H.-D. Chiang, "A novel consensus-based particle swarm optimization-assisted trust-tech methodology for large-scale global optimization," *IEEE transactions on cybernetics*, vol. 47, no. 9, pp. 2717–2729, 2016.

[57] Y. Behzadi, K. Restom, J. Liau, and T. T. Liu, "A component based noise correction method (compcor) for bold and perfusion based fmri," *Neuroimage*, vol. 37, no. 1, pp. 90–101, 2007.

[58] R. C. Craddock, G. A. James, P. E. Holtzheimer III, 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.

[59] S. Ghiassian, R. Greiner, P. Jin, and M. R. Brown, "Using functional or structural magnetic resonance images and personal characteristic data to identify adhd and autism," *PloS one*, vol. 11, no. 12, p. e0166934, 2016.

[60] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, 2018.

[61] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in mri images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[62] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, no. 1, pp. 3–42, 2006.

[63] R. Chaves, J. Ramírez, J. Górriz, M. López, D. Salas-Gonzalez, I. Alvarez, and F. Segovia, "Svm-based computer-aided diagnosis of the alzheimer's disease using t-test nmse feature selection with feature correlation weighting," *Neuroscience letters*, vol. 461, no. 3, pp. 293–297, 2009.

[64] H. G. Schulze, R. B. Foist, A. Ivanov, and R. F. Turner, "Chi-squared-based filters for high-fidelity signal-to-noise ratio enhancement of spectra," *Applied spectroscopy*, vol. 62, no. 8, pp. 847–853, 2008.

[65] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[66] I. Dziobek, M. Bahnemann, A. Convit, and H. R. Heekeren, "The role of the fusiform-amygdala system in the pathophysiology of autism," *Archives of general psychiatry*, vol. 67, no. 4, pp. 397–405, 2010.

[67] P. Mengotti, S. DAgostini, R. Terlevic, C. De Colle, E. Biasizzo, D. Londero, A. Ferro, G. Rambaldelli, M. Balestrieri, S. Zanini *et al.*, "Altered white matter integrity and development in children with autism: a combined voxel-based morphometry and diffusion imaging study," *Brain research bulletin*, vol. 84, no. 2, pp. 189–195, 2011.

[68] O. Dekhil, M. Ali, Y. El-Nakieb, A. Shalaby, A. Soliman, A. Switala, A. Mahmoud, M. Ghazal, H. Hajjdiab, M. F. Casanova *et al.*, "A personalized autism diagnosis cad system using a fusion of structural mri and resting-state functional mri data," *Frontiers in psychiatry*, vol. 10, 2019.

Zhi-An Huang received the B.Eng. degree in software engineering from Shenzhen University, Shenzhen, China, in 2018. He is currently working toward the Ph.D. degree with Department of Computer Science, City University of Hong Kong, Hong Kong.

His research interests include artificial intelligence, machine learning, bioinformatics, and medical imaging analysis.

Zexuan Zhu (M'12-SM'20) received the B.S. degree in computer science and technology from Fudan University, China, in 2003 and the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore, in 2008.

He is currently a Professor with the College of Computer Science and Software Engineering, Shenzhen University, China. His research interests include computational intelligence, machine learning, and bioinformatics. Dr. Zhu is an Associate Editor of IEEE Transactions on Evolutionary Computation and IEEE Transactions on Emerging Topics in Computational Intelligence. He is also the Chair of the IEEE CIS Emergent Technologies Task Force on Memetic Computing.

Chuen Heung Yau graduated from Hunan University of Traditional Chinese Medicine in July 1985 with a bachelor's degree in medicine; University of Hong Kong in February 2002 with a master's degree in Chinese medicine (acupuncture). PHD candidate of Hunan University of Traditional Chinese Medicine. Worked as a doctor in Hunan Provincial Traditional Chinese Medicine Hospital from August 1985 to July 1992; as a lecturer and deputy chief Chinese medicine practitioner in the School of Chinese Medicine of Hong Kong Baptist University from April 1999 to July 2019; from August 2019 to present works in the school of Chinese medicine university of Hong Kong as a Lecturer. Published 17 papers, 1 book, 2 chapters of the books. The research interest is mainly in the psychoanalysis of children with autism and the observation of the effect of scalp acupuncture on it.

Kay Chen Tan (Fellow, IEEE) received the B.Eng. degree (First Class Hons.) in electronics and electrical engineering and the Ph.D. degree in evolutionary computation and control systems from the University of Glasgow, Glasgow, U.K., in 1994 and 1997, respectively.

He is a Full Professor with the Department of Computer Science, City University of Hong Kong, Hong Kong. He has published over 200 refereed articles and six books.

Prof. Tan is the Editor-in-Chief of the IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, was the Editor-in-Chief of the IEEE Computational Intelligence Magazine from 2010 to 2013, and currently serves as the Editorial Board Member of over ten journals. He was an Elected Member of IEEE CIS AdCom from 2017 to 2019.