Multi-site clustering and nested feature extraction for identifying autism spectrum disorder with resting-state fMRI

Nan Wang\textsuperscript{a}, Dongren Yao\textsuperscript{b}, Lizhuang Ma\textsuperscript{a,c,r}, Mingxia Liu\textsuperscript{b,a}

\textsuperscript{a} East China Normal University, Shanghai 200062, China
\textsuperscript{b} Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
\textsuperscript{c} Shanghai Jiao Tong University, Shanghai 200240, China

\textbf{ABSTRACT}

Brain functional connectivity (FC) derived from resting-state functional magnetic resonance imaging (rs-fMRI) has been widely employed to study neuropsychiatric disorders such as autism spectrum disorder (ASD). Existing studies usually suffer from (1) significant data heterogeneity caused by different scanners or studied populations in multiple sites, (2) curse of dimensionality caused by millions of voxels in each fMRI scan and a very limited number (tens or hundreds) of training samples, and (3) poor interpretability, which hinders the identification of reproducible disease biomarkers. To this end, we propose a Multi-site Clustering and Nested Feature Extraction (MC-NFE) method for fMRI-based ASD detection. Specifically, we first divide multi-site training data into ASD and healthy control (HC) groups. To model inter-site heterogeneity within each category, we use a similarity-driven multiview linear reconstruction model to learn latent representations and perform subject clustering within each group. We then design a nested singular value decomposition (SVD) method to mitigate inter-site heterogeneity and extract FC features by learning both local cluster-shared features across sites within each category and global category-shared features across ASD and HC groups, followed by a linear support vector machine (SVM) for ASD detection. Experimental results on 609 subjects with rs-fMRI from the ABIDE database with 21 imaging sites suggest that the proposed MC-NFE outperforms several state-of-the-art methods in ASD detection. The most discriminative FCs identified by the MC-NFE are mainly located in default mode network, salience network, and cerebellum region, which could be used as potential biomarkers for fMRI-based ASD analysis.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Autism spectrum disorder (ASD) is a developmental disability that can cause significant social, communication and behavioral challenges. As one of the largest worldwide early-onset neurodevelopmental disorders (Amaral \textit{et al.}, 2008; Monk \textit{et al.}, 2009; Ecker \textit{et al.}, 2010; Wee \textit{et al.}, 2016), ASD is generally characterized by social interaction difficulties, communication barriers and limited repetitive interests (Chen \textit{et al.}, 2015; Jung \textit{et al.}, 2015; Alaerts \textit{et al.}, 2015). Recent studies suggest that ASD develops from a combination of genetic and nongenetic, or environmental influences (Nielsen \textit{et al.}, 2013; Wang \textit{et al.}, 2019b). The current diagnosis of ASD is primarily based on questionnaires and clinical observations that are prone to be subjective and susceptible. It is highly desired to discover objective imaging biomarkers for ASD detection, promoting early intervention and effective treatment of the disease.

Resting-state functional magnetic resonance imaging (rs-fMRI) can measure hemodynamic changes caused by neural activity for the whole brain at a series of time points (Friston \textit{et al.}, 1994). It provides a non-invasive solution to objectively quantify physical disorders that cause obvious mental illness, and has been widely employed in the research of brain dysfunction diseases (Abraham \textit{et al.}, 2017; Jie \textit{et al.}, 2018; Xiang \textit{et al.}, 2020; Wang \textit{et al.}, 2021). Brain functional connectivity (FC) derived from rs-fMRI data can depict large-scale abnormality or dysfunction in brain FC networks.

With brain FC data as input, many machine learning methods have been proposed to analyze brain FC networks and produced promising results in discriminating various mental disorders from age-matched healthy controls (HCs) (Philip \textit{et al.}, 2012; Feckzo \textit{et al.}, 2018; Kleinhans \textit{et al.}, 2011; Yao \textit{et al.}, 2019; Wang \textit{et al.}, 2019b). In particular, some machine learning technologies such as deep neural network (DNN) (Li \textit{et al.}, 2018b; Guo \textit{et al.}, 2017; Jiang and Zhao, 2017), long short term memory (LSTM) (Li \textit{et al.}, 2020;
Tao and Shyu, 2019; El-Gazzar et al., 2019), and convolutional neural network (CNN) (Ktiena et al., 2017; Khosla et al., 2019; Anirudh and Thiagarajan, 2019) have been used for brain FC analysis and ASD detection. For instance, Heinsfeld (Heinsfeld et al., 2018) used a DNN to extract low-dimensional FC representations for ASD detection with 1,035 subjects from autism brain imaging data exchange (ABIDE) (Di Martino et al., 2014).

Dvornek et al. (Dvornek et al., 2017) used a recurrent neural network with a long short-term memory (LSTM) architecture for ASD classification based on rs-fMRI time series. Parisot et al. (Parisot et al., 2017) applied a 3D CNN model for ASD analysis based on stochastic parcellation and seven atlases, where an ensemble learning strategy corresponding to different brain parcellations was used for disease prediction.

Even though these methods have achieved promising results, existing studies usually suffer from three challenges: (1) significant heterogeneity of multi-site data caused for instance by the use of different scanners, protocols or populations in different imaging sites; (2) curse of dimensionality caused by millions of voxels in each fMRI scan and a very limited number (tens or hundreds) of training samples; and (3) poor interpretability for brain FC dysfunction found in experiments related to ASD, especially when using multi-site fMRI data. These will inevitably hinder the identification of reproducible disease biomarkers from rs-fMRI data.

To this end, we propose a novel multi-site clustering and nested feature extraction (MC-NFE) method for ASD detection (see Fig. 1). As shown in Fig. 1, we first divide multi-site training samples (with each represented by an FC network) into ASD and HC groups. A similarity-driven multiview linear reconstruction (SIMLR) model is used to cluster heterogeneous brain FC networks into non-overlapping subspaces within each group. We then design the nested singular value decomposition (SVD) method to learn compact FC features, followed by a linear support vector machine (SVM) for ASD detection. Experimental results on 609 subjects from ABIDE with 21 sites and rs-fMRI data demonstrate the effectiveness of the proposed MC-NFE in ASD detection. The identified discriminative FCs are mainly located in the default mode network, salience network, and cerebellum region, which could be used as potential biomarkers for fMRI-based ASD analysis.

The contributions of this work can be summarized as follows:

(1) We propose a general framework to model inter-site heterogeneity for functional connectivity (FC) based brain disease identification. Specifically, we first partition the population into a patient group and a healthy control group, followed by subpopulation clustering within each group/category and informative FC feature extraction. This helps model the cross-site data heterogeneity within each category. It is straightforward to apply this framework to other multi-site applications.

(2) We design a nested SVD strategy to mitigate inter-site heterogeneity and extract FC features, by learning both local cluster-shared features across sites for each category and global category-shared features within each specific group (i.e., ASD or HC). This helps alleviate the problem that inter-category differences may be more significant than inter-category differences.

(3) With the proposed method, one can identify and visually show disease-associated functional connectivities. Thus, our approach has great interpretability in detecting ASD-related brain FC abnormality. This would be very helpful for fMRI-based brain disease analysis in clinical practice.

The remainder of this paper is organized as follows. Section 2 reviews the most relevant studies. In Section 3, we introduce materials used in this work and the proposed method. In Section 4, we compare the proposed method with several competing methods for ASD detection. Section 5 analyzes the influence of several major components of MC-NFE and presents several limitations of the current work and future work. Finally, this paper is concluded in Section 6.
2. Related work

2.1. Machine learning methods for ASD detection

Brain functional connectivity (FC) derived from rs-fMRI data helps depict abnormality or dysfunction in brain connectivity networks. Based on brain FC networks, many machine learning techniques have been developed for ASD diagnosis. For instance, Wee et al. (Wee et al., 2012) proposed to extract weighted local clustering coefficients from brain FC networks and use a support vector machine (SVM) for classification. Jie et al. (Jie et al., 2014) used Weisfeiler-Lehmahn graph kernel (Shervashidze et al., 2011) to model pairwise relationship of brain FC networks, followed by an SVM for classification. Cao et al. (Cao et al., 2016) used a discriminative subgraph mining method for brain FC network-based brain disorder diagnosis, and employed a feature evaluation criterion to estimate importance of subgraph patterns/features.

Unlike previous studies that use predefined FC features, many deep learning methods have been developed to extract FC features in a data-driven manner. Khosla et al. (Khosla et al., 2018) employ a volumetric CNN framework that takes advantage of 3D spatial structure of rs-fMRI data for disease classification. Several author-encoder (AE) based methods (Wang et al., 2016; Hazzelt et al., 2017) have been used for ASD analysis by learning new FC features before inputting FC network data into classification models. Choi (Choi, 2017) used variational autoencoders (VAE) to identify multivariate and nonlinear functional connectivity patterns of ASD, and performed classification on the ABIDE dataset. Bi et al. (Bi et al., 2018) developed a random neural network cluster model that consists of multiple neural networks to classify ASD patients and healthy controls.

Recurrent neural network (RNN) is also used to combine temporal information and spatial characteristics of resting-state fMRI data, not just the spatial patterns in the connected group (Chen and Hu, 2018). Wang et al. (Wang et al., 2019a) developed a spatial-temporal convolutional-recurrent network for disease prediction with rs-fMRI time series in an end-to-end manner. Although many attempts have been made to explain and intuitively understand brain FC network alteration, the existing models usually have poor interpretability, and thus cannot well explain the underlying functional connectivity dysfunctions associated with ASD.

2.2. Feature extraction and selection of brain FC networks

Brain functional connectivity networks may be extremely high-dimensional (due to 4D nature of rs-fMRI data) and have significant heterogeneity (caused by different studied subjects or scanning parameters in multiple imaging sources). This introduces many challenges in traditional classification models. Effective feature extraction and selection methods have become indispensable components. Morris et al. (Morris and Rekik, 2017) used a set of cortical attributes to simulate morphological brain functional connectivity. They also utilized a sparse graph embedding method to map high-dimensional connectomic features into a low-dimensional space, by preserving the locality of original data. Dryburgh et al. (Dryburgh et al., 2019) used a connectivity based predictive model to predict functional intelligence scores based on fMRI data. However, these methods are usually not scalable due to their high computational costs. Georges et al. (Georges and Rekik, 2018) proposed a multi-graph architecture to model the relationship between a set of feature selection methods, where each graph quantifies the feature reproducibility between graph nodes at a fixed number of top-ranked features. This work is an interesting attempt to generate reproducible brain FC network features that accurately differentiate between two brain states. However, the existing studies cannot identify connectomic features that are reproducible and stable across heterogeneous datasets.

On the other hand, several studies have tried to interpret identified functional connectivities that are informative for ASD diagnosis. Bach et al. (Bach et al., 2015) proposed to assign decision of a neural network to its input through hierarchical correlation propagation. Mahendran et al. (Mahendran and Vedaldi, 2015) used an approximated inversion of a network to explain its decision. Sundararajan et al. (Sundararajan et al., 2017) proposed an integrated gradient to explain the decision of a model. However, these models have only very limited interpretability, since they cannot directly identify informative functional connectivity associated with ASD. To enhance models’ interpretability, we propose to learn group-specific FC features for ASD patients and healthy controls, where learned FC features can be mapped to original FC space to facilitate detection of functional connectivity differences for ASD identification.

3. Materials and methodology

3.1. Materials and image preprocessing

A total of 609 subjects with rs-fMRI data recruited from 21 sites in ABIDE-I and ABIDE-II (Di Martino et al., 2014) are used in this work, including 280 ASD patients and 329 HC subjects. The demographic information of the studied subjects is given in Table 1. The age was 7 to 27 years old, and the gender distribution was unbalanced (with 499 males and 110 females). It suggests that there exists significant inter-site heterogeneity caused by populations in this database. We use two-tailed two sample t-test algorithm to evaluate the statistical significance of the difference between the ASD and HC groups in terms of gender and age, respectively, and use the asterisk ‘∗’ to mark non-statistically significant differences between the two groups in Table 1.

The rs-fMRI data are pre-processed using the Data Processing Assistant for Resting-State fMRI (DPARSF). The pipeline is listed as follows: (1) discarding the first 10 volumes; (2) correcting for slice-dependent delays; (3) correcting head motion; (4) normalization with an EPI template in the MNI space, and resampling to the 3 × 3 × 3 mm³ resolution; (5) spatial smoothing with a 4 mm full width half maximum Gaussian kernel; (6) linear detrending and temporal band-pass filtering (0.01 – 0.10 Hz) for BOLD signals; and (7) regressing out nuisance signals of head motion parameters, white matter, cerebrospinal fluid (CSF) and global signals. The registered fMRI volumes were partitioned into 64 regions-of-interest (ROIs) using the Bootstrap Analysis of Stable Clusters (BASC-064) template. We construct a 64 × 64 FC network for each subject, where each node is an ROI and the edge weight is the Pearson’s correlation between the time series of BOLD signals of paired ROIs. We use the upper triangle of the FC matrix to represent a subject, yielding a 2,016-dimensional feature vector.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Demographic information of studied subjects from 21 sites in ABIDE. (Std: standard deviation). ASD: autism spectrum disorder; HC: healthy control. The term ‘∗’ denotes that there is no significant difference (p &gt; 0.05) between ASD and HC groups in terms of gender/age via two-tailed two sample t-test.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>ASD</td>
</tr>
<tr>
<td>Subject #</td>
<td>280</td>
</tr>
<tr>
<td>Gender* (Female/Male)</td>
<td>38/242</td>
</tr>
<tr>
<td>Age* (Mean±Std)</td>
<td>17.09±8.14</td>
</tr>
</tbody>
</table>
3.2. Proposed method

Fig. 1 provides an overview of our multi-site clustering and nested feature extraction (MC-NFE) model that can identify discriminatory functional connectivities for ASD identification, including: (1) subject grouping, (2) multi-site clustering of FC networks, (3) nested feature extraction, and (4) biomarker identification and disease diagnosis. More details are given below.

3.2.1. Subject grouping

The distribution of functional connectivity network data may have significant heterogeneity among different imaging sites due to different scanners or populations. Such inter-site differences may be more significant than inter-category differences, while it has been proven that mitigating the inter-site heterogeneity helps capture more consistent disease-related FC patterns (Nielsen et al., 2013; Wang et al., 2019b).

To facilitate the modeling of inter-site heterogeneity, we propose to divide multi-site training samples (with each represented by an FC network) into an ASD group (denoted as A) and an HC group (denoted as H) based on their category labels. Note that we only use class labels of training samples for model construction, and no test data are available in the training stage. To mitigate inter-site heterogeneity and also identify group-level differences, we will first extract local cluster-shared FC features for each subpopulation/cluster (via SIMLR and the 1st-level SVD) and then extract global category-shared FC features for each group/category (via the 2nd-level SVD), followed by ASD identification based on the learned new FC network features (see Fig. 1).

3.2.2. Multi-Site clustering of FC networks

To model inter-site differences, we propose a multi-site clustering method to partition subjects into several subpopulations. Within the ASD/HC group, we use multiple Gaussian kernels to measure inter-subject distance in the original FC network feature space, considering that multiple kernels are superior in providing multi-scale representations of biomedical data (Gonen and Alpaydin, 2011). Denote $x^i$ and $x^j$ as 2,016-dimensional feature vectors of the $i$-th and the $j$-th FC networks (with each one corresponding to a subject), respectively, and these two networks are from the same group (i.e., ASD or HC). The $g$-th ($g=1, \ldots, G$) Gaussian kernel is defined as follows:

$$K_g(x^i, x^j) = \frac{1}{\epsilon_g^{(g)}} \exp\left(-\frac{|x^i - x^j|^2}{2\epsilon_g^{(g)}}\right),$$

(1)

where $\epsilon_g^{(g)} = \sigma_1^{(g)}(\mu_1^{(g)} + \mu_1^{(g)})/2$, $\sigma$ is a tuning parameter with $\mu_1^{(g)} = \frac{1}{\sum_{k=1}^{k_k} k_k} KNN(x^i)$, and $KNN(x^i)$ is the top $k$ neighbors of the $i$-th subject. Each kernel is equipped by two parameters, i.e., $\sigma$ and $k$. By empirically setting $k=10$ and $\sigma = \{1.0, 1.25, \ldots, 2\}$, we can obtain a total of $G=5$ kernels. With Eq. 1, we can evaluate the kernel-based distance between two FC networks (corresponding to two subjects) at a specific spatial scale in the original FC feature space.

To model inherent distribution information of multi-site data within each group, we employ a similarity-driven multi-view linear regression (SIMLR) model (Wang et al., 2018) to learn the intrinsic similarity matrix $S \in \mathbb{R}^{N \times N}$ with $N$ subjects. This model assumes that if $N_i$ separable populations exist among $N$ subjects, $S$ should have an approximate block-diagonal low-rank structure with $N_i$ blocks (with similar subjects belonging to the same subpopulation). It models the FC network manifold by learning the weights $\{w_g\}_{g=1}^G$ associated with $G$ kernels to capture various statistical characteristics of input data, formulated as follows:

$$\min_{\mathbf{S}, \mathbf{w}} \sum_{i=1}^N \sum_{j=1}^N \sum_{g=1}^G -w_g K_g(x^i, x^j) S_{ij} + \|S\|_F^2 + \|\mathbf{t}(\mathbf{L}^T (\mathbf{I}_N - S) \mathbf{L})\|_1$$

s.t. $\sum_{g=1}^G w_g = 1, w_g \geq 0, \mathbf{L}^T \mathbf{L} = \mathbf{I}_N, \sum_{j=1}^N S_{ij} = 1, S_{ij} \geq 0,$

(2)

where $\mathbf{t}(\cdot)$ is the trace of a matrix, $\mathbf{L} \in \mathbb{R}^{N \times N}$ is a low-dimensional matrix enforcing a low-rank constraint on $S$. $\|\cdot\|_F$ denotes the Frobenius norm, $\mathbf{I}_N \in \mathbb{R}^{N \times N}$ and $\mathbf{I}_c \in \mathbb{R}^{N \times N}$ are identity matrices, and $N_c$ is the cluster number. Following (Jiao et al., 2020), we set the cluster number $N_c = 3$ empirically in the experiments.

In Eq. (2), the first term encourages that the similarity between two subjects should be small if their kernel-based distance is large, and the second term is used to prevent $S$ from becoming an identity matrix. The third term and the constraint on $\mathbf{L}$ enforces a low-rank structure of $S$. The last term imposes constraints on the kernel weights to avoid the selection of a single kernel. Through an alternating optimization algorithm, each variable in Eq. (2) can be optimized while fixing others until convergence. Specifically, with fixed $\mathbf{L}$ and $\mathbf{w}$, we can update $\mathbf{S}$ via the following:

$$\min_{\mathbf{S}} \sum_{i=1}^N \sum_{j=1}^N \sum_{g=1}^G -w_g K_g(x^i, x^j) S_{ij} + \|S\|_F^2 + \|\mathbf{t}(\mathbf{L}^T (\mathbf{I}_N - S) \mathbf{L})\|_1$$

s.t. $\sum_{j=1}^N S_{ij} = 1, S_{ij} \geq 0.$

(3)

Then, with fixed $\mathbf{S}$ and $\mathbf{L}$, we can update $\mathbf{w}$ via the following:

$$\min_{\mathbf{w}} \sum_{g=1}^G \sum_{i=1}^N \sum_{j=1}^N -w_g K_g(x^i, x^j) S_{ij} + \sum_{g=1}^G w_g \log w_g$$

s.t. $\sum_{g=1}^G w_g = 1, w_g \geq 0.$

(5)

Such an alternating optimization algorithm is performed by repeatedly optimizing one variable with the other two fixed until convergence. Based on the similarity matrix $S$ calculated within a specific group (e.g., ASD), we can perform subpopulation clustering by extracting each of $N_c$ blocks gathered by similar brain FC networks.

3.2.3. Nested feature extraction

Although multi-site clustering is helpful to detect subpopulations in the ASD/HC group, there may still be heterogeneity in the data distribution of different subpopulations. In addition, subjects in each subpopulation still have high-dimensional (i.e., 2,016) FC features with significant redundancy. Previous studies (Mourao-Miranda et al., 2005; Vinjamuri et al., 2009; Arribas et al., 2010; Tak et al., 2011; Guo et al., 2020) have shown that singular value decomposition (SVD) is effective in dimensionality reduction and extracting important features based on several top singular values. Therefore, we design a nested SVD strategy for multi-site brain FC network feature extraction.

Specifically, within each group (e.g., ASD or HC), we first apply the 1st-level SVD on each cluster/subpopulation to learn local cluster-shared features. We denote original features of the $c$-th ($c = 1, \ldots, N_c$) subpopulations of all training ASD patients as $\mathbf{A'}$, and the $c$-th subpopulations of HC group as $\mathbf{H'}$. In the 1st-level SVD, we aim to obtain the new features (e.g., $\mathbf{A''}$ and $\mathbf{H''}$) by keeping those top $T_1$ largest singular values for each of $N_c$ subpopula-
tions in two groups via the following:

\[
\hat{\mathbf{A}}^c = \text{SVD}(\hat{\mathbf{A}}^c), \\
\hat{\mathbf{H}}^c = \text{SVD}(\hat{\mathbf{H}}^c),
\]

which is expected to extract local (i.e., cluster/subpopulation-level) FC features for the ASD and HC groups, respectively.

Based on the learned new representation (i.e., \(\hat{\mathbf{A}}\) and \(\hat{\mathbf{H}}\) for all \(N_c\) clusters in the ASD/HC group, we then apply a 2nd-level SVD on all training subjects (represented by newly-learned local features) to learn more global (i.e., group-level) category-shared features across the ASD and HC groups by keeping those top \(T_f\) largest singular values. Here, \(T_f\) denotes the dimension of finally learned FC features, and \(T_f\) is empirically set as 78 in the experiments. This helps model the global group-level differences between these two categories. We denote local features of the ASD group as \(\hat{\mathbf{A}} = [\hat{\mathbf{A}}^1, \ldots, \hat{\mathbf{A}}^{N_c}]\), and those of the HC group as \(\hat{\mathbf{H}} = [\hat{\mathbf{H}}^1, \ldots, \hat{\mathbf{H}}^{N_c}]\). The to-be-extracted global FC network features (denoted as \(\hat{\mathbf{A}} \in \mathbb{R}^{T_f}\) for the ASD group and \(\hat{\mathbf{H}} \in \mathbb{R}^{T_f}\) for the HC group) can be obtained via the 2nd-level SVD as follows:

\[
\hat{\mathbf{A}} = \text{SVD}(\hat{\mathbf{A}}), \\
\hat{\mathbf{H}} = \text{SVD}(\hat{\mathbf{H}}).
\]

Using such a nested SVD strategy for FC network feature extraction, it is expected to reduce feature dimension and, more importantly, to alleviate inter-site heterogeneity by learning local cluster-shared features across sites within each category and global category-shared features within each specific group (i.e., ASD or HC).

3.2.4. Biomarker identification and ASD diagnosis

In the training stage, based on the learned group-specific FC features, we can select the most informative functional connectivities in the ASD and HC groups by mapping the new features to the original FC space. More details on locating discriminative functional connectivities can be found in Section 3 of the Supplementary Materials. To facilitate automated disease diagnosis, we further construct a linear support vector machine (SVM) based on the learned new \(T_f\)-dimensional FC features of training subjects (ASD and HC) for ASD detection.

In the test stage, we first input the FC network of an unseen test subject into the proposed framework to extract features. Then we feed the features into trained SVM for generating group/class-specific probability scores. Finally, we assign the class label with a higher probability score to the test subject.

4. Experiment

4.1. Experimental setting

A leave-one-out cross-validation (LOO-CV) strategy is used in the experiments. Following (Jiao et al., 2020), in the proposed MC-NFE model, we empirically set the number of clusters \(N_c = 3\) for ASD and HC groups, respectively, and set the feature dimension as \(T_f = 1.100\) and \(T_f = 78\) in nested SVD. We have also studied their influence by varying the values of \(N_c\) and \(T_f\) in Section 5. Five metrics are used to evaluate the performance of ASD vs. HC classification, including (1) accuracy (ACC), (2) sensitivity (SEN), (3) specificity (SPE), (4) balanced accuracy (BAC), and (4) the area under the ROC curve (AUC).

4.2. Competing method

We compare our MC-NFE with three baseline methods and two state-of-the-art methods, with details introduced below.

1. Principal component analysis (PCA): PCA is used to reduce feature dimension and select the most discriminative features. Due to the difference between sites, the dimension of selected features ranges from 20 to 100 in PCA. Similar to our MC-NFE, a linear SVM with a default penalty coefficient (C = 1) is used for classification in this method.

2. Support vector machine recursive feature elimination (SVM-RFE) (Guyon et al., 2002): This is an embedded feature selection algorithm, which uses criteria derived from the coefficients in SVMs to assess features and recursively removes features that have small criteria. The nonlinear SVM-RFE with an RBF kernel is used for comparison, considering its advantage over the linear version (Yan and Zhang, 2015). In this method, the penalty coefficient is set to \(C = 1\) and the bandwidth parameter of the RBF kernel is set to \(\gamma = 2^{-6}\) in SVM.

3. Gaussian random projection (GRP): This method reduces feature dimension by projecting FC features in the original input space onto a randomly generated matrix. For GRP, we first retain the reduced feature dimension as 110, and then select the most representative features from these features via cross-validation (CV). A linear SVM with a default penalty coefficient (C = 1) is used for classification.

4. Brain network atlas-guided feature selection (NAG-FS) (Mhiri and Rekik, 2020): This method first divides the population into HC and ASD groups, and then creates multiple cluster-specific network atlases via SIMLR for each group. Within each group, it also estimates the average local network at the center of each subpopulation via the similarity network fusion (SNF) technique, thus generating an ASD-related atlas and an HC-related atlas. With the estimated ASD and HC atlases, it computes the absolute difference between two network atlases, where the non-zero features with the highest discrepancy are selected as the discriminate feature for SVM training. For this method, we use the same parameters as the corresponding paper (Mhiri and Rekik, 2020).

5. Invertible network (IN-Net) (Zhuang et al., 2019): This method first maps the FC matrix to a feature domain, followed by a fully-connected layer to classify ASD from HC group. It determines decision boundary as a high-dimensional plane and projects data points onto the decision boundary. The difference between a data point and its projection onto the decision boundary can be viewed as the explanation. An importance measure is then defined as the explanation weighted by the gradient of prediction with respect to the input, and this importance measure is finally used to identify biomarkers. We use the same parameter settings as the original paper (Zhuang et al., 2019).

Note that four methods (i.e., PCA, SVM-RFE, GRP, and IN-Net) do not consider the data heterogeneity issue, since they combine multi-site data for feature extraction/selection and classification. The NAG-FS method first generates a group-averaged network atlas for ASD and HC groups via SIMLR and then performs feature selection by computing the absolute difference between two atlases, without learning local cluster-shared FC network features. In contrast, our MC-NFE method helps alleviate inter-site heterogeneity by learning (1) local cluster-shared features across multiple sites via SIMLR and the 1st-level SVD, and (2) global category-shared features across ASD and HC groups via the 2nd-level SVD.

4.3. Results on single-site data

To evaluate the generalization ability of different methods in ASD detection based on rs-fMRI data without significant inter-site heterogeneity, we report the results of six different methods in ASD vs. HC classification on 6 single sites from ABIDE in Table 2, and the demographic information of these single-site data are given in Supplementary Materials. More results on the remaining 15 single sites are reported in Supplementary Materials. It can be seen from Table 2 that the proposed MC-NFE produces the over-
all best results on six single sites. For instance, on Site 2 (with 21 HC and 13 ASD subjects), our method achieves very good results (AUC=78.75%; ACC=79.78%). On Site 14 with 19 subjects, our method yields a promising AUC result (AUC=88.09%), with an improvement of at least 10.31% when compared with those of the competing methods. On Site 16, our approach can achieve the best results in terms of ACC (77.42%), SEN (73.33%), SPE (81.25%), BAC (77.29%), and AUC (83.33%), respectively. On Site 32, our MC-NFE method still obtains good performances in terms of five metrics. These results imply that our method can reliably identify ASD patients from HCs, even with limited training samples from each single site. Thus, our MC-NFE is practical with considerably better adaptability.

### 4.4. Results on multi-site data

We then evaluate our MC-NFE method and the competing methods on the challenging multi-site ASD detection problem, by using data from all 21 sites in ABIDE. Note that these multi-site data exist significant inter-site heterogeneity due to several factors such as age and gender. The results of ASD vs. HC classification achieved by six different methods are reported in Table 3. It can be seen from Table 3 that MC-NFE yields consistently better results in terms of ACC, SEN, SPE, BAC and AUC. The underlying reason is that the MC-NFE explicitly models the inter-site heterogeneity (via multi-site clustering) and mitigates the feature differences between sites (via nested SVD), but four competing methods do not consider their heterogeneity by combining multi-site data. These results further confirm the necessity of normalizing rs-fMRI data in multi-site studies.

### 4.5. Comparison with State-Of-The-Arts

We also compare our MC-NFE method with five state-of-the-art (SOTA) methods for rs-fMRI based ASD detection using subjects from ABIDE: (1) **MLSCG** (Ktena et al., 2018) that uses 871 subjects including 403 ASD patients and 468 HCs recruited from 20 sites in ABIDE; (2) **DTLN** (Li et al., 2018a) that uses 149 ASD patients and 161 HCs; (3) **IASDDL** (Dvornek et al., 2017) that mixes data from 17 sites while keeping the proportions between sites, including 539 ASD and 573 healthy controls; (4) **MT-GCN** (Yao et al., 2019) that chooses 1,029 subjects from 34 imaging sites in ABIDE-I and ABIDE-II, including 485 ASD and 544 HCs; and (5) **TSST** (Zhao et al., 2019) that employs 303 ASD patients and 390 HC subjects from ABIDE-II with 10 imaging sites. The comparison between our method and five SOTA methods is summarized in Table 4. Note that results in Table 4 are not fully comparable due to the use of different numbers of subjects from

---

**Table 2**

<table>
<thead>
<tr>
<th>Site</th>
<th>Metric</th>
<th>PCA</th>
<th>SVM-RFE</th>
<th>GRP</th>
<th>NAG-FS</th>
<th>IN-Net</th>
<th>MC-NFE (Ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC</td>
<td>82.35±4.74</td>
<td>79.41±4.94</td>
<td>76.47±3.55</td>
<td>76.47±4.13</td>
<td>75.66±3.17</td>
<td>79.78±4.38</td>
<td></td>
</tr>
<tr>
<td>SEN</td>
<td>50.00±5.01</td>
<td>25.00±6.37</td>
<td>50.00±5.89</td>
<td>37.50±6.20</td>
<td>61.12±2.92</td>
<td>61.54±2.88</td>
<td></td>
</tr>
<tr>
<td>SPE</td>
<td>92.71±5.54</td>
<td>96.15±6.06</td>
<td>91.65±6.20</td>
<td>94.66±5.04</td>
<td>92.22±5.03</td>
<td>90.48±5.12</td>
<td></td>
</tr>
<tr>
<td>BAC</td>
<td>70.36±5.87</td>
<td>60.58±5.26</td>
<td>67.31±3.14</td>
<td>62.98±4.01</td>
<td>72.00±2.67</td>
<td>76.01±4.17</td>
<td></td>
</tr>
<tr>
<td>AUC</td>
<td>71.63±4.04</td>
<td>66.48±5.63</td>
<td>70.19±4.73</td>
<td>70.67±3.12</td>
<td>71.63±2.79</td>
<td>78.75±4.21</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3**

<table>
<thead>
<tr>
<th>Method</th>
<th>ACC</th>
<th>SEN</th>
<th>SPE</th>
<th>BAC</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>53.8±4.89</td>
<td>50.00±3.76</td>
<td>55.56±5.24</td>
<td>52.78±4.95</td>
<td>51.85±5.27</td>
</tr>
<tr>
<td>SVM-RFE</td>
<td>54.17±2.48</td>
<td>51.71±5.86</td>
<td>58.33±5.78</td>
<td>55.02±5.13</td>
<td>56.94±5.14</td>
</tr>
<tr>
<td>GRP</td>
<td>59.38±3.25</td>
<td>60.91±1.18</td>
<td>57.14±1.39</td>
<td>59.13±3.56</td>
<td>61.11±1.02</td>
</tr>
<tr>
<td>NAG-FS</td>
<td>61.76±6.10</td>
<td>58.82±5.87</td>
<td>60.70±5.80</td>
<td>59.76±5.68</td>
<td>58.96±6.48</td>
</tr>
<tr>
<td>IN-Net</td>
<td>65.27±2.59</td>
<td>60.77±3.58</td>
<td>58.14±1.29</td>
<td>59.46±2.55</td>
<td>61.36±1.62</td>
</tr>
<tr>
<td>MC-NFE (Ours)</td>
<td>68.42±3.75</td>
<td>70.05±5.71</td>
<td>63.64±2.98</td>
<td>68.85±3.65</td>
<td>69.31±4.52</td>
</tr>
</tbody>
</table>
Table 4
Comparison with state-of-the-art methods in ASD vs. HC classification with rs-fMRI data from ABIDE, with best results shown in bold.

<table>
<thead>
<tr>
<th>Method</th>
<th>MLGSC (Xtena et al., 2018)</th>
<th>DTLNN (Li et al., 2018a)</th>
<th>IASDDL (Dovneck et al., 2017)</th>
<th>MTGCN (Yao et al., 2019)</th>
<th>TSST (Zhao et al., 2019)</th>
<th>MC-NFE (Ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(ASD/HC)</td>
<td>(403/468)</td>
<td>(149/161)</td>
<td>(539/573)</td>
<td>(485/544)</td>
<td>(303/390)</td>
<td>(280/329)</td>
</tr>
<tr>
<td>ACC (%)</td>
<td>62.90</td>
<td>65.00</td>
<td>65.00</td>
<td>67.30</td>
<td>65.30</td>
<td>68.42</td>
</tr>
<tr>
<td>SEN (%)</td>
<td>-</td>
<td>68.90</td>
<td>69.00</td>
<td>64.20</td>
<td>-</td>
<td>70.05</td>
</tr>
<tr>
<td>SPE (%)</td>
<td>-</td>
<td>67.00</td>
<td>62.00</td>
<td>70.00</td>
<td>-</td>
<td>63.64</td>
</tr>
</tbody>
</table>

Fig. 2. Top 5 and top 30 discriminative FCs identified by our MC-NFE method in ASD vs. HC classification, where the thickest solid line denotes the most discriminative functional connectivity, and the largest circle denotes the most discriminative regions-of-interest (ROIs). R: right; L: left.

ABIDE in different methods. Through this rough comparison, it can be seen that our conventional machine learning based method (i.e., MC-NFE) achieves the best ACC and SEN results (ACC=69.79%, SEN=70.05%), compared with five SOTA methods. Although the MTGCN method achieves the best SPE value (SPE=70.00%), this method is based on deep neural networks, so it may be challenging for users to interpret their learned features from rs-fMRI data. As shown in Section 5.1, using our method, we can visually show discriminative functional connectivities for ASD detection to aid diagnosis, while deep learning methods lack interpretability. This means that our method would be very helpful in clinical applications by helping early detect ASD-associated brain FC dysfunction.

5. Discussion

In this section, we first visually show the discriminative functional connectivities identified by our method, and study the influence of several key parameters. We finally list several limitations of the current work and possible future research directions. More experimental results on the influence of brain templates for ROI partition, unbalanced data, age, and gender are reported in the Supplementary Materials.

5.1. Identified discriminative FCs

To better understand brain functional connectivity (FC) network variations between ASD and HC groups, we visualize the most discriminative FC features identified by our MC-NFE method, with top 5 and top 30 FCs shown in Fig. 2. The regions-of-interest (ROIs) corresponding to those top 30 FCs are reported in Table SVI in the Supplementary Materials. In this figure, the thickest solid line denotes the most discriminative functional connectivity, and the largest circle denotes the most discriminative ROIs. It clearly shows the dysfunction between posterior cingulate cortex and precuneus, which can be treated as functional hubs in the default mode network (DMN) used to process information regarding the self. Note that DMN dysfunction is related to ASD (Padmanabhan et al., 2017; Washington et al., 2014; Menon, 2018). Our identified discriminative FCs also include the salience network and cerebellum region, and these regions are also shown to be closely related to ASD (Green et al., 2016; Wang et al., 2014). These results verify the reliability of our MC-NFE in detecting informative functional connectivity for ASD identification.

5.2. Influence of nested SVD

To visually evaluate the influence of the nested SVD, we use t-SNE (Van der Maaten and Hinton, 2008) to reduce the FC features to 3 dimensions. The distributions of original FC features and those learned by our method using only the 1st-level SVD and the nested SVD on 21 sites are reported in Fig. 3. This figure suggests that our MC-NFE with nested SVD can significantly increase the distribution gap between ASD and HC groups, compared with 1st-level SVD. As we can see from Fig. 3 (a)-(b), compared with the distribution of the original functional connectivity features, using 1st-level SVD can enhance the gap between ASD and HC groups. From Fig. 3 (c), we can see that, with the nested SVD, the distribution gap between the ASD and HC groups has further increased, making features of the two groups more separable.

We further report the results of MC-NFE with 1st-level SVD and nested SVD on six single sites (i.e., Site 2, Site 14, Site 16, Site 17, Site 22 and Site 32) in Fig. 4, with more detailed results given in Table SVII of the Supplementary Materials. Fig. 4 suggests that MC-NFE using nested SVD produces the overall better performance, compared with that using only the 1st-level SVD on six single sites. The main reason could be that the nested SVD strategy used in MC-NFE helps extract both local features of each cluster and global features shared by different clusters between sites, while using only the 1st-level SVD cannot capture global FC network representations.

5.3. Influence of feature extraction strategy

To evaluate the influence of the proposed nested feature extraction strategy (via two-layer/nested SVD) in MC-NFE, we fur-
Finally, we evaluate our MC-NFE method using the original functional connectivity features and the new features learned by MC-NFE with and without the nested SVD, as well as the new features learned by SIMLR+Nested-SVM-RFE, SIMLR+SVM-RFE, and SIMLR+PCA, respectively. These results are shown in Fig. 3 and Fig. 4. For comparison, we also report the results of the conventional feature extraction methods, including FC and t-SNE. The results are shown in Table SVIII of the Supplementary Materials.

5.4. Influence of cluster number

To evaluate the influence of the number of clusters within ASD and HC groups, we vary the value of $N_c$ within the range of [2, 3, 4, 5] and report the corresponding results of MC-NFE on

---

Fig. 3. Distributions of (a) original FC features, (b) new features learned by MC-NFE using only the 1st-level SVD, and (c) new features learned by MC-NFE with nested SVD on all 21 sites. To facilitate visualization, these features are reduced to 3 dimensions through t-SNE projection (Van der Maaten and Hinton, 2008). Yellow and black denote autism spectrum disorder (ASD) and healthy control (HC) subjects, respectively.

Fig. 4. Results (%) of our MC-NFE using only the 1st-level SVD and the nested SVD strategies on 6 single sites (i.e., S2, S14, S16, S17, S22, and S32).

Fig. 5. Results (%) of the proposed MC-NFE method and its four variants (i.e., SIMLR+PCA, SIMLR+SVM-RFE, SIMLR+Nested-PCA, and SIMLR+Nested-SVM-RFE) with different feature extraction strategies in ASD vs HC classification on multi-site data.
all sites in Table 5. This table shows that, with \( N_r = 3 \), the MC-NFE achieves the overall best performances (e.g., ACC=68.42% and AUC=69.71%). When \( N_r = 2 \) and \( N_r = 5 \), the classification results achieved by MC-NFE are not good. For instance, with \( N_r = 2 \), the results of ACC and AUC are 55.36% and 53.13%, respectively. This is consistent with previous findings (Jiao et al., 2020). Also, these results imply that the number of clusters is an important parameter when clustering subjects within each group into several subgroups to capture heterogeneous distributions of multi-site data.

5.5. Influence of feature dimension

We further analyze the influence of different feature dimensions learned by nested SVD on the performance of the MC-NFE method. Specifically, we vary the feature dimension \( T_f \) from the range of [5, 30, 78, 100, 110] and record the results in Table 6. It can be seen from Table 6 that, with a very low or high feature dimension (e.g., \( T_f = 5 \) and \( T_f = 110 \)), our method cannot yield good results. With \( T_f = 78 \) and \( T_f = 100 \), we can obtain relatively better results. This implies that, for the original 2,016-dimensional FC network features, it is reasonable to set the reduced feature dimension between [50,100] when using the proposed MC-NFE for feature extraction and selection.

5.6. Limitation and future work

Several issues need to be considered in the future. First, we used nested SVD for extracting features in the current work, and the improved performance of our method indicated that nested SVD could produce informative features for autism detection. It is desired to apply more advanced deep learning techniques to further improve the performance of ASD diagnosis. Second, we use only functional imaging modality to discover potential biomarkers for ASD detection, without considering the structural imaging modalities that may also be related to brain states. In the future, we will integrate both functional and structural modalities to train our model for ASD identification. In addition, we use multi-site clustering and nested feature extraction for identifying ASD, and rely on full category labels of training subjects. Self-supervised methods (Mak et al., 2019; Kalanderian and Nasrallah, 2019) could be used to further reduce the demand for category labels, which will also be our future work.

6. Conclusion

In this work, we propose a multi-site clustering and nested feature extraction (MC-NFE) method for fMRI-based ASD detection, which can help mitigate inter-site heterogeneity and handle high-dimensional fMRI data. The proposed MC-NFE shows good performance in ASD detection on the heterogeneous ABIDE database with 21 sites. The most discriminative functional connectivities identified by MC-NFE involve the default mode network, salience network, and cerebellum region, which could be used as potential biomarkers for ASD detection.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary material


CRediT authorship contribution statement

Nan Wang: Methodology. Software, Writing – original draft. Dongren Yao: Data curation. Lizhuang Ma: Writing – review & editing. Supervision. Mingxia Liu: Conceptualization, Validation, Writing – review & editing, Supervision.

References


