Physica Medica 65 (2019) 99–105

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Original paper

Functional connectivity-based classification of autism and control using SVM-RFECV on rs-fMRI data

Canhua Wang[a,b], Zhiyong Xiao[c], Jianhua Wu[d,*]

[a] School of Mechatronics Engineering, Nanchang University, Nanchang 330031, China
[b] School of Computer, Jiangxi University of Traditional Chinese Medicine, Nanchang 330004, China
[c] School of Software, Jiangxi Agricultural University, Nanchang 330045, China
[d] School of Information Engineering, Nanchang University, Nanchang 330031, China

ARTICLE INFO

Keywords:
Autism
fMRI
Feature selection
Machine learning
Classification

ABSTRACT

Considering the unsatisfactory classification accuracy of autism due to unsuitable features selected in current studies, a functional connectivity (FC)-based algorithm for classifying autism and control using support vector machine-recursive feature elimination (SVM-RFE) is proposed in this paper. The goal is to find the optimal features based on FC and improve the classification accuracy on a large sample of data. We chose 35 regions of interest based on the social motivation hypothesis to construct the FC matrix and searched for informative features in the complex high-dimensional FC dataset by the SVM-RFE with a stratified-4-fold cross-validation strategy. The selected features were then entered into an SVM with a Gaussian kernel for classification. A total of 255 subjects with autism and 276 subjects with typical development from 10 sites were involved in the study. For the data of global sites, the proposed classification algorithm could identify the two groups with an accuracy of 90.60% (sensitivity 90.62%, specificity 90.58%). For the leave-one-site-out test, the proposed algorithm achieved a classification accuracy of 75.00%–95.23% for data from different sites. These promising results demonstrate that the proposed classification algorithm performs better than those in recent similar studies in that the importance of features can be measured accurately and only the most discriminative feature subset is selected.

1. Introduction

Autism spectrum disorder (ASD) is a life-long neuro-developmental condition, currently estimated to affect 1 in 59 children in the USA and 1%–1.5% of children and adults worldwide according to the most recent investigation of the Centers for Disease Control and Prevention [1]. It is regrettable that the etiology and pathogenesis of ASD are rarely known. Diagnosis is typically made in early childhood based on clinical interviews and observation of behavior. It is necessary to search for objective biomarkers to guide the clinical diagnosis and treatment of ASD. In recent years, researchers have used the method of machine learning (ML) to extract messages from the resting-state functional magnetic resonance imaging (rs-fMRI) data of ASD for classification or prediction. The ML method has become a popular trend in the study of ASD.

One of the most replicated abnormalities in brain imaging studies of autism has been underconnectivity of distributed brain networks [2–7]. The functional connectivity (FC) from rs-fMRI data has been extensively applied to classify ASD and control due to the hypo- and hyper-connectivity in the brain. The previous studies that utilized ML attained a relatively high classification accuracy of 70%–96.3%, but the subjects studied came from a single research institute and the number of subjects investigated was very small; for example, 40 subjects with ASD and 40 typical development (TD) subjects in [8], 13 ASD and 14 TD subjects in [9], and 20 ASD and 20 TD subjects in [10]. In response, many researchers began to investigate large data samples from the Autism Brain Imaging Data Exchange (ABIDE) dataset [11]. One study, which used the ABIDE database to calculate the FC measurements from a lattice of 7266 ROI covering the entire grey matter (26.4 million connections), attained an accuracy of only 60% [12]. In another study, the classification procedure attained an accuracy of 76.67% for 89 ASD and 89 TD [13]. By building participant-specific connectomes from functionally defined brain areas, Abraham et al. achieved a classification accuracy of 67% on the full ABIDE dataset [14]. Heinsfeld et al. investigated patterns of FC and achieved an accuracy of classification of 70% on the full ABIDE dataset by using the method of deep learning

* Corresponding author.
E-mail address: jhwu@ncu.edu.cn (J. Wu).

https://doi.org/10.1016/j.ejmp.2019.08.010
Received 27 April 2019; Received in revised form 8 August 2019; Accepted 13 August 2019
Available online 22 August 2019

[15]. Of course, a higher classification accuracy could be obtained by using a large sample of data. One study used the probabilistic neural network and intrinsic connectivity analysis, obtaining an accuracy of classification of 86.9% between 312 patients with ASD and 328 TD by using 10-fold cross-validation [16]. However, the authors did not investigate the differences in classification from different sites. Classification accuracy drops significantly in the context of larger population samples from different sites [12].

The FC measurements mentioned above are generally a high-dimensional matrix constructed by calculating the Pearson correlation coefficients between time series from regions of interest (ROI). A poor classification accuracy might be due to the fact that the most discriminative feature subset not being well selected from the original dataset with high-dimensional features. Hence, a sophisticated feature selection is critical for classifying smaller population samples in the high-dimensional space. The support vector machine-recursive feature elimination (SVM-RFE) algorithm is an efficient feature selection technique that has been widely applied in many scenarios [17–19], but rarely in the analysis of brain FC measurements. For a specific application, it is quite difficult to predetermine how many top-ranked features should be selected from the feature rank of SVM-RFE.

In order to address the above mentioned issues in ASD classification, in this paper, we chose 35 ROI based on the social motivation hypothesis [20] to construct an FC matrix for avoiding a primitive feature set with too-high dimensions, and defined the number of informative features from the initial FC dataset by the support vector machine-recursive feature elimination with a stratified-4-fold cross-validation (SVM-RFECV) strategy. The selected features were then entered into an SVM with a Gaussian kernel for classification. A total of 255 subjects with autism and 276 subjects with TD from 10 sites were involved in the study. For the data of global sites, the proposed classification algorithm could identify the two groups with an accuracy of 90.60% (sensitivity 90.62%, specificity 90.58%). For the leave-one-site-out test, the proposed algorithm achieved a classification accuracy of 75.00%–95.23% for data from different sites. These promising results demonstrate that the proposed classification algorithm performs better than those in recent similar studies in that the importance of features can be measured accurately and only the most discriminative feature subset is selected.

2. Materials and methods

2.1. Participants and data preprocessing

This study was carried out using rs-fMRI data downloaded from the ABIDE database [11]. We downloaded the rs-fMRI data from 10 sites in the ABIDE, in which the NYU data consisted of two parts from ABIDE I and ABIDE II. All the resting-state functional images were preprocessed by the Data Processing Assistant for Resting-State fMRI (DPARSF) software [21,22]. The preprocessing procedures are as follows: (1) removing the first 10 time points, (2) slice timing correction, (3) head motion realignment, (4) T1-weighted individual structural images co-registration to the mean functional images using a 6-degree-of-freedom linear transformation, (5) segmentation, (6) nuisance covariate regression (global signal regression (GSR) was not performed because of concerns about increasing negative correlations [23,24]), (7) normalization by DARTEL, and (8) temporal filtering. After preprocessing, we selected the subjects whose structural images were covered completely, with good registration, and a mean frame-wise displacement (FD Jenkinson) less than 0.2 mm [25]. Thus, 531 subjects (255 ASD patients and 276 TD controls) were selected, as listed in Table 1. The names and abbreviations of the institutes and scanning parameters are given in Table 2. All fMRI images were obtained with informed consent according to procedures established by human subject research boards.

Table 1
Phenotype summary.

Site	ASD			TC		
	Age Avg	Sex	Size	Age Avg	Sex	Size
CALTECH	26.9	M 9, F 4	13	28.8	M 15, F 4	19
LEU	19.09	M 13, F 1	14	20.02	M 16, F 1	17
NYU I	15.22	M 60, F 10	70	16.30	M 70, F 25	95
NYU II	10.19	M 38, F 4	42	9.68	M 26, F 1	27
OHSU	11.61	M 12, F 0	12	9.96	M 12, F 0	12
OLIN	17.90	M 8, F 3	11	18.10	M 8, F 2	10
PITT	19.21	M 13, F 3	16	19.69	M 10, F 2	12
SBL	32.80	M 5, F 0	5	29.75	M 4, F 0	4
SDSU	15.04	M 9, F 0	9	14.39	M 13, F 5	18
USM	23.63	M 44, F 0	44	22.18	M 40, F 0	40
YALE	13.20	M 14, F 5	19	12.67	M 16, F6	22

NYU I: downloaded from ABIDE I; NYU II: downloaded from ABIDE II; M: Male; F: Female.

Table 2
Scanning parameters and experimental settings in different sites.

Institute	MRI vendor	TR (msec)	TE (msec)	FA (deg)	Volumes
CALTECH	Siemens	2000	30	75	150
LEU	Phillips	1667	33	90	250
NYU	Siemens	2000	15	90	180
OHSU	Siemens	2500	30	90	82
OLIN	Siemens	1500	27	60	210
PITT	Siemens	1500	25	70	200
SBL	Phillips	2200	30	80	200
SDSU	GE	2000	30	90	180
USM	Siemens	2000	28	90	240
YALE	Siemens	2000	25	60	200

CALTECH, California Institute of Technology; LEU, University of Leuven; NYU, NYU Langone Medical Center; OHSU, Oregon Health and Science University; OLIN, Institute of Living at Hartford Hospital; PITT, University of Pittsburgh School of Medicine; SBL, Social Brain Lab; SDSU, San Diego State University; USM, University of Utah School of Medicine; YALE, Yale Child Study Center; FA, Flip angle.

2.2. Connectivity measures and feature matrices

In this section, we used 35 ROI summarized by Clements et al. [20], as listed in Table 3. We defined 35 spheres with the MNI coordinates of these regions as the center and 5 mm as the radius. Time courses were extracted from each of the 35 spheres and averaged within each region. Pearson's correlation coefficients (r) were computed between these average time courses. Individuals' r values were normalized to z values using Fisher's z transformation. The z-transformed correlation coefficients were represented in a 35×35 matrix (1225 cells), which was symmetric with regard to the diagonal. The values within upper triangles and the main diagonal of the matrix were removed. Therefore, the number of remaining effective cells in the lower triangle of the matrix was 595. We flattened the remaining triangle (i.e., collapsed it into a one-dimensional vector) to retrieve a vector of features, with the purpose of using it for classification.

2.3. Feature selection: SVM-RFECV

In this study, the smallest data sample from the SBL site included just 9 subjects (5 with ASD and 4 with TD), and most of the other samples included only dozens of subjects. However, the FC mentioned above has as many as 595 features. Therefore, it was summarized by the expression "high dimension and small sample." It was critical to accurately measure the feature importance and select the most discriminative feature subset. The SVM-RFE is a kind of backward elimination method that starts with a full set of all features and then removes the most irrelevant features one by one. The top-ranked

Table 3
35 ROI for building FC.

Number	Region	Hemisphere	MNI Coordinates
1	Anterior cingulate gyrus, caudate (L)	R, L	6, 6, 20
2	Caudate	R	22, 24, 12
3	Central opercular cortex, insula	R	42, −6, 18
4	Cerebellum	R	32, −52, −26
5	Cerebellum	L	−34, −84, −36
6	Frontal pole	R	26, 54, −4
7	Inferior frontal gyrus, pars triangularis	R	34, 28, 18
8	Lateral occipital cortex (inferior)	L	−42, −86, −18
9	Middle frontal gyrus	R	24, 14, 34
10	Nucleus accumbens, subcallosal cortex	L	−2, 6, −10
11	Occipital fusiform gyrus, occipital pole, Cerebellum	L	−10, −88, −28
12	Occipital pole, lingual gyrus, cerebellum	R	2, −90, −20
13	Parietal operculum cortex	R	30, −40, 16
14	Precentral gyrus	R	52, −2, 30
15	Precentral gyrus	L	−58, −4, 36
16	Precuneus cortex	L	−22, −42, 10
17	Subcallosal cortex	R, L	2, 16, −6
18	Thalamus	R	2, −26, 16
19	Thalamus	L	−12, −34, 16
20	Anterior cingulate gyrus	R	6, 30, 8
21	Frontal pole	L	−22, 52, 18
22	Lateral occipital cortex (inferior)	R	58, −66, 12
23	Lateral occipital cortex (inferior), angular gyrus	R	52, −66, 16
24	Lateral occipital cortex (superior)	R	16, −80, 40
25	Occipital fusiform gyrus, temporal occipital fusiform cortex	R	36, −58, −10
26	Parahippocampal gyrus (anterior)	R	28, −6, −30
27	Planum temporale, central opercular cortex	R	60, −16, 10
28	Planum temporale, superior temporal gyrus	L	−56, −26, 6
29	Precuneus cortex	L	−2, −72, 42
30	Putamen, amygdala	R	24, −2, −10
31	Putamen, insula	L	−30, 0, 10
32	Superior frontal gyrus	L	−12, 28, 58
33	Superior temporal gyrus	L	−48, −22, −4
34	Superior temporal gyrus (posterior)	L	−36, −34, 4
35	Insula	R	34, 14, −10

features removed in the last iteration of SVM-RFE are the most important, while the bottom-ranked ones are the least informative and are removed in the first iteration. In detail, the SVM-RFE algorithm can be expressed as below:

Let $X = \{x^{(1)}, x^{(2)}, ...,x^{(N)}\}$ be the dataset containing N initial features and let class labels be denoted by $Y = \{y^{(1)}, y^{(2)}, ...,y^{(N)}\}$, $y^{(i)} \in \{-1, 1\} and\ i \in \{1, 2, ...,N\}$.

Step 1: Train an SVM on the initial features set:

$$L = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y^{(i)}y^{(j)}\left[k\left(x^{(i)}, x^{(j)}\right) + \lambda\delta_{ij}\right] - \sum_{i=1}^{N}\alpha_i \tag{1}$$

$$\sum_{i=1}^{N} a_i y^{(i)} = 0\, and\, 0 \leqslant \alpha_i \leqslant C,\ (i = 1, 2, ...,N) \tag{2}$$

Eq. (1) is to be minimized subject to Eq. (2). In Eqs. (1) and (2), $k(x^{(i)}, x^{(j)})$ is a kernel function, δ_{ij} is the Kronecker symbol ($\delta_{ij} = 1 \text{if } i = j$ and 0 otherwise), and $\alpha = \{\alpha_1, \alpha_2, ...,\alpha_N\}$ are the parameters to be determined. λ and C are positive constants that ensure convergence even when the problem is non-linearly separable or poorly conditioned.

Step 2: Compute the weight vector and the ranking criteria according to Eq. (3) and Eq. (4).

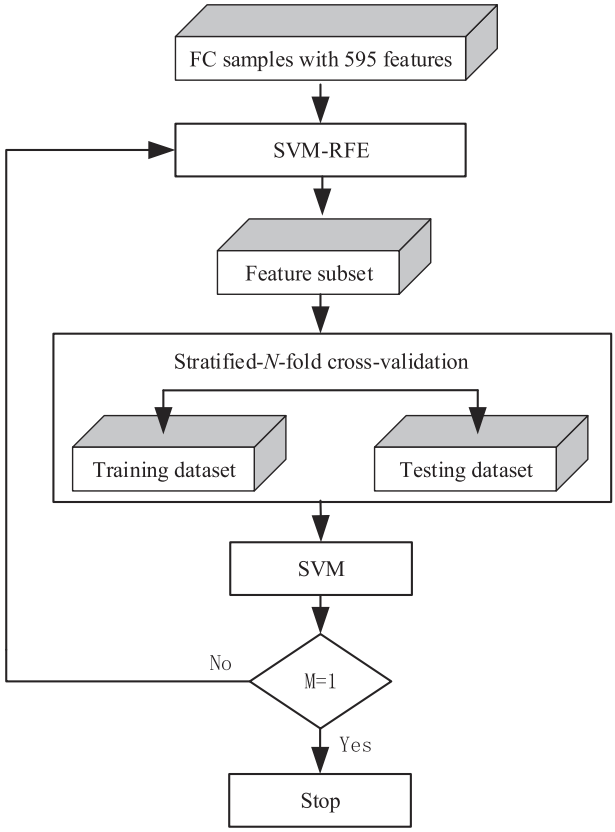

Fig. 1. Flowchart of feature selection based on SVM-RFECV.

$$w_i = \sum_{i=1}^{N}\alpha_i y^{(i)}x^{(i)} \tag{3}$$

$$c_i = w_i^2 \tag{4}$$

Step 3: Find the feature with the smallest ranking criterion and eliminate it.

Step 4: Update the features dataset and $N = N$-1.

Step 5: Repeat Steps 1–4 until the features set is empty.

Notably, when the features dataset is not empty, the iteration can also stop. The stopping criterion could be the desired number of features users want to retain. Meanwhile, the rank of features is also obtained.

It was critical to decide how many top-ranked features should be selected from the feature rank of SVM-RFE. To make a trade-off between accuracy and robustness, we adopt the linear SVM-RFE with a stratified-N-fold cross-validation strategy to find the minimum features dataset on which the SVM can get the highest accuracy. The flowchart of the feature selection by SVM-RFECV is shown in Fig. 1.

The following steps explain the scheme in detail.

Step 1: Using all the features of the data samples from 531 subjects as the initial features dataset, we can obtain the rank of features and a feature subset; the bottom-ranked feature is removed by the SVM-RFE algorithm. Here, let the numbers of initial features and selected feature subsets be M and $M - 1$, respectively. In addition, the kernel function of SVM is linear.

Step 2: The feature subset selected is split into a training dataset and testing dataset by a stratified-N-fold strategy. The classification accuracy is computed by the same SVM that was applied for recursive feature elimination.

Step 3: Let $M = M - 1$and update the feature subset, which is then entered into the SVM-RFE.

Step 4: Repeat Steps 1–3 until the feature subset is empty.

Step 5: The minimum features dataset is determined on which the

SVM can get the highest score.

We can predetermine how many top-ranked features should be selected by SVM-RFE with a stratified-*N*-fold cross-validation strategy. For example, let $N = 4$. Whenever the bottom-ranked feature was removed by SVM-RFE, the resulting feature subset was evaluated by the stratified-four-fold cross-validation. The highest score was achieved based on the feature subset consisting of the top 148 ranked features. Similarly, by letting $N = 10$, the feature subset including the top 97 ranked features becomes optimal. Through the evaluation of a specialized classifier, the optimal number of top-ranked features was determined to be between 148 and 97.

2.4. Classification method and performance evaluation

Although the linear SVM was used to evaluate the feature subset, the highest classification accuracy was less than 60%. The classification accuracy was poor because the feature subset was not normalized and the parameters of the linear SVM used as the classifier were not optimized. In order to find an optimal classifier, the linear SVM, extreme gradient boosting (XGB), and logistic regression (LR) were applied to identify the features selected by SVM-RFECV, respectively. The pattern recognition scheme employed in the study is shown in Fig. 2.

The classification quality was assessed by the following performance indices:

$$Accuracy = (TP + TN)/(TP + FN + TN + FP) \quad (5)$$

$$Sensitivity = TP/(TP + FN) \quad (6)$$

$$Specificity = TN/(TN + FP) \quad (7)$$

$$PPV = TP/(TP + FP) \quad (8)$$

$$NPV = TN/(TN + FN) \quad (9)$$

where TP, FN, TN, and FP denote true positive, false negative, true negative, and false positive, respectively. These are defined, respectively, as the number of ASD correctly classified, the number of ASD predicted to be TD, the number of TD correctly classified, and the number of TD predicted to be ASD. The optimal classifier would be selected from the XGB, LR, and SVM according to the accuracy, sensitivity, specificity, PPV, and NPV.

3. Results and discussion

In this section, the results of recognition for ASD and TD were evaluated in terms of different classifiers and the number of features selected by SVM-RFE in order to select the optimal pattern recognition scheme. Meanwhile, the most discriminative features were found. Furthermore, a leave-one-site-out test was carried out for evaluating classifier performance across sites and the proposed approach was compared with those in recent studies about the classification of autism. The SVM-RFE with a stratified-*N*-fold cross-validation strategy and classification algorithms were implemented using Scikit-learn [26].

Table 4
Results of classification by different number of features.

	$N = 3$	$N = 4$	$N = 5$	$N = 6$	$N = 7$	$N = 8$	$N = 9$	$N = 10$
Number of features	135	148	51	112	21	5	17	97
Accuracy	87.18	90.60	77.57	84.73	70.05	60.11	70.43	80.99

3.1. Results of classification by different number of features

For the SVM-RFE with a stratified-*N*-fold cross-validation, the number of features selected varies with *N*. Features were extracted from 531 subjects (255 ASD patients and 276 TD controls) from all sites. Table 4 shows the classification results by different numbers of features using the linear SVM.

As shown in Table 4, different *N* resulted in different numbers of features and different accuracies of classification. The highest accuracy, 90.60%, came from the 148 features selected by SVM-RFE with a stratified-4-fold cross-validation. The other feature sets, which were only a subset of the 148 feature sets, might not well represent ASD and TD. Although the 5 features selected by SVM-RFE with a stratified-8-fold cross-validation resulted in the lowest accuracy, these features may be the most discriminative.

3.2. Results of classification by different classifiers

According to the above analysis, the feature subset consisting of the top 148 ranked features was optimal. The SVM, XGB, and LR were applied to identify these features, separately. We used the GridSearchCV strategy to search for the optimal parameters for the three classifiers. The results of classification by a stratified-10-fold cross-validation are given in Table 5.

As shown in Table 5, the SVM achieved a mean classification accuracy of 90.60% (sensitivity 90.62%, specificity 90.58%) from the stratified-10-fold cross-validation. Based on the existing literature, this is the highest classification achieved so far. The LR classifier achieved a mean accuracy of 85.67% (sensitivity 88.73%, specificity 82.35%); meanwhile, the XGB classifier achieved a mean accuracy of 72.56% (sensitivity 66.61%, specificity 60.80%). To assess how our model would behave in a real clinical world, we calculated PPV and NPV. The SVM attained a relatively high PPV of 91.65% and an NPV of 90.19%. The results show that the SVM is superior to the other two classifiers on the multi-site ABIDE data.

In order to test if a classification score is significant, we performed a permutation analysis: we first randomly reassigned subject labels, and then performed the 10-fold cross-validation classification. This procedure was repeated for 1000 iterations. The *p*-value (0.00099) was then given by the percentage of runs for which the score obtained was greater than the classification score obtained.

As shown in Fig. 3, the permutation test demonstrates that the features selected from the relatively large datasets are of statistical significance.

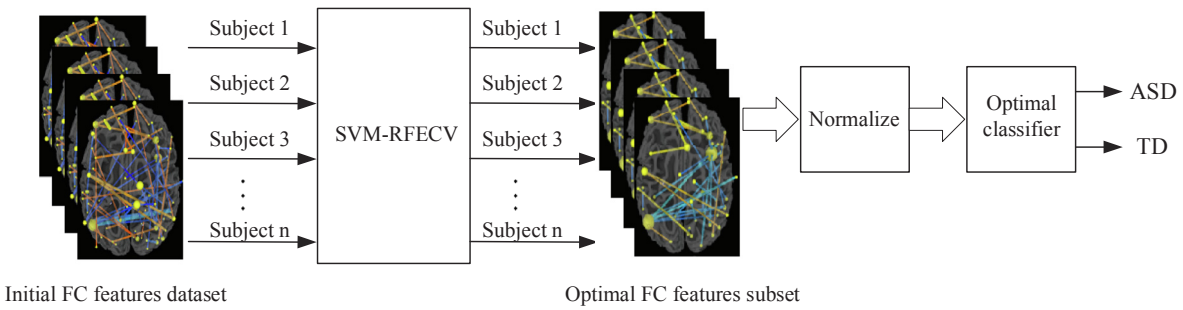

Fig. 2. Pattern recognition scheme for ASD and TD.

Table 5
Results of classification for the data of all sites.

Classifier	Accuracy	Sensitivity	Specificity	PPV	NPV
XGB	72.56	66.61	60.80	64.91	62.86
LR	85.67	88.73	82.35	84.92	87.33
SVM	90.60	90.62	90.58	91.65	90.19

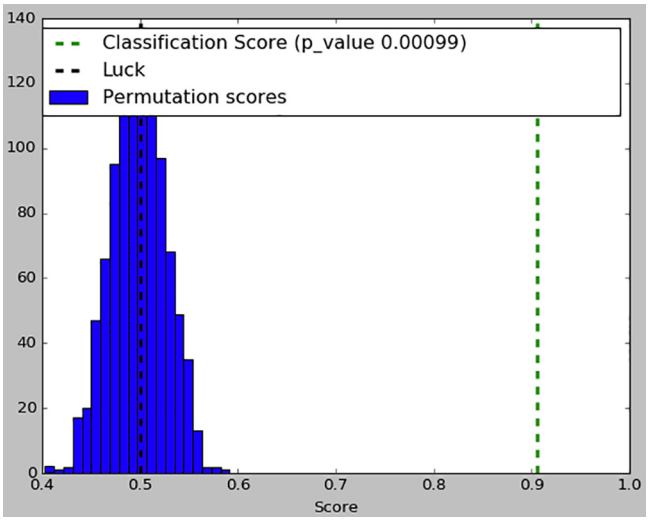

Fig. 3. Histogram of permutation scores.

3.3. Leave-one-site-out classification

To evaluate the classifier performance across different sites, we performed a leave-one-site-out cross-validation (LOSOCV) process. This process involves using one site's data as the test set, and the remaining data as the training set. This process is then repeated with each site's data used exactly once as the test data. Using the 148 features and the linear SVM, the LOSOCV results are as reported in Table 6.

Three sites, SBL, OHSU and OLIN, with relatively small sample sizes (Table 1) showed significantly higher accuracies than the global result. The highest accuracy of classification from OLIN reached 95.23%. However, the site with the largest number of subjects, NYU I, did not show the lowest accuracy of all sites. The worst identification came from CALTECH, which has a small sample size less than that of NYU I. Table 6 shows that the sensitivity and NPV are relatively low in those sites with a higher proportion of females, such as YALE, SDSU, NYU I, and CALTECH, except for OLIN.

3.4. Neural patterns: connectivity in the autistic brain

The FC feature matrix was collapsed into a one-dimensional vector before feeding the classifier. In order to find the location of the features selected, we restored and located them by the support file generated by the SVM-RFECV. The FC areas of the 148 features selected and the 5 most discriminative features are shown in Figs. 4 and 5, respectively. The red regions denote the selected features.

Table 6
LOSOCV results using the linear SVM.

Site	Accuracy	Sensitivity	Specificity	PPV	NPV
CALTECH	75.00	73.68	76.92	83.25	66.67
LEU	80.64	82.35	78.57	82.35	78.57
NYU I	76.36	71.57	82.86	85.00	68.23
NYU II	84.05	96.30	76.19	72.22	96.97
OHSU	91.67	100	83.33	85.71	100
OLIN	95.23	100	90.90	90.90	100
PITT	82.14	83.33	81.25	76.92	86.67
SBL	88.88	100	80.00	80.00	100
SDSU	81.48	72.22	100	100	64.29
USM	83.33	80.00	86.36	84.21	82.61
YALE	78.05	77.27	78.95	80.95	75.00

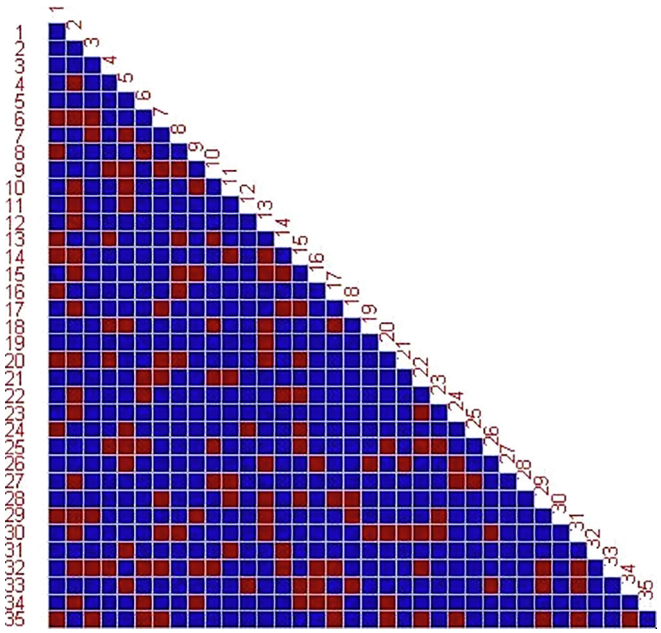

Fig. 4. FC areas of the 148 selected features.

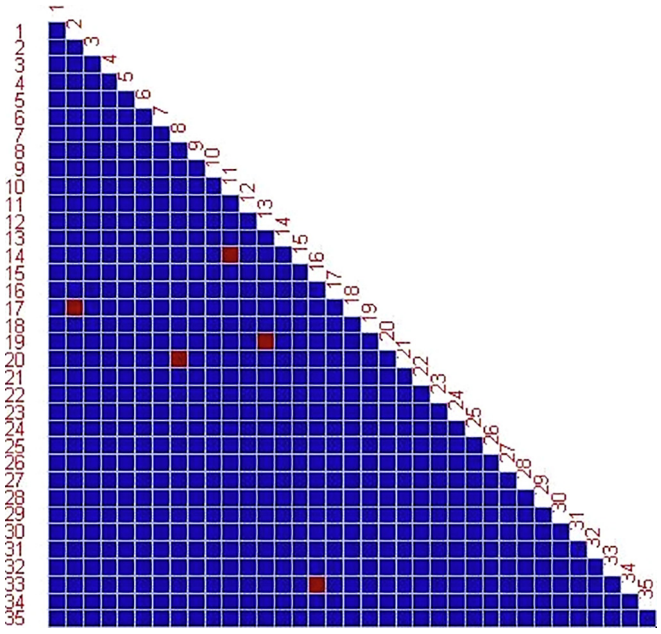

Fig. 5. The 5 most discriminative features selected.

As shown in Fig. 4, most of these selected features come from FC areas related to the caudate, parietal operculum cortex, superior frontal gyrus, precentral gyrus (L), anterior cingulate gyrus, and insula.

As shown in Fig. 5, the five most discriminative features selected represent the FC from the precentral gyrus (R) and occipital fusiform gyrus, occipital pole, cerebellum (L), subcallosal cortex and caudate (R), thalamus (L) and parietal operculum cortex (R), anterior cingulate gyrus (R) and lateral occipital cortex (inferior) (L), and superior temporal gyrus (L) and Precuneus cortex (L). We chose the ROI based on the social motivation hypothesis to construct a functional connectivity

matrix, which achieved a very high classification accuracy and supported the viewpoint on aberrant neural processing of social, nonsocial, and potentially restricted interest rewards in individuals with ASD.

3.5. Comparison with recent studies about the classification of autism

In recent years, researchers have generally used a large data sample to classify ASD and TD based on the FC. Nielsen et al., Plitt et al., Abraham et al., Heinsfeld et al., and Iidaka, as mentioned in the introduction of this article, attained accuracies of 60%, 76.67%, 67%, 70%, and 86.9%, respectively [12–16]. In addition, H. Chen et al. achieved an accuracy of 79.17% using the SVM to classify 112 ASD and 128 TD subjects [27]. To evaluate classifier performance across sites, researchers have generally performed LOSOCV. H. Chen et al. tested the LOSOCV for 6 sites and achieved an accuracy of 60%–77% [27]. Heinsfeld et al. attained an accuracy of 63%–68% for the test of LOSOCV across 17 sites [15].

In this paper, we employed feature selection using the SVM-RFECV and the SVM with a Gaussian kernel to identify ASD and TD with an accuracy of 90.60% based on a large sample of data. Furthermore, the accuracy of the LOSOCV across 10 sites reached 75%–95.23%. The classification accuracy across different sites does not drop significantly in the context of a larger population sample from different sites. It is proved that the proposed method has advantages for eliminating the impact of cross-sites generated from different equipment and demographics. Although the dataset is not exactly the same as the recent similar studies, our method possesses a satisfactory classification performance. The reasons are as follows. Firstly, we used a sophisticated preprocessing approach for the rs-fMRI data including selecting the subjects whose structural images were covered completely and head motion followed the strict frame-wise displacement. Subjects with a poor fMRI data quality were excluded. In this study, we selected 531 subjects with high-quality data from 708 subjects involved from 10 sites. Compared with the studies including 964 subjects from 16 sites [12], including 871 subjects from 17 sites [14], and including 1035 subjects from 17 sites [15], the proposed method achieved a higher classification accuracy based on relatively a smaller sample of data. Compared with the studies including 178 subjects from 3 sites [13] and including 368 subjects from 6 sites [27], the proposed method also achieved a higher classification accuracy based on a relatively larger sample of data. The classification accuracy of a study including 640 subjects from 12 sites [13] is close to ours. Large multi-site datasets increase sample size at the cost of uncontrolled heterogeneity [14]. The heterogeneity results in some disturbances to data samples as well as some loss in classification accuracy. The data sample from 10 sites may be considered as a moderate choice. Moreover, we used the 35 ROI based on the social motivation hypothesis to build an FC matrix. Compared with other ROI, these ROI might well represent the features of autism. Moreover, it was critical that the SVM-RFECV could accurately measure the feature importance and select the most discriminative feature subset.

4. Conclusion

In conclusion, the FC-based algorithm for classifying autism and control using SVM-RFECV not only achieved a high classification accuracy on the global dataset but also on the across-sites' dataset. The importance of features could be measured accurately and the most discriminative feature subset could be selected by the strategy. In addition, large multi-site datasets increased sample size at the cost of uncontrolled heterogeneity. The data sample from 10 sites might be considered as a moderate choice. Although abnormal functional connectivity networks are not yet objective biomarkers of autism [13], this proposed method could be used as an important reference for autism diagnosis, and the 148 FC features found in this study might be a significant part of ASD biomarkers. In this study, only the index of FC was used, and some other features such as grey matter volume will be introduced in the future; this may lead to better results.

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.

Acknowledgment

This work is partly supported by the National Natural Science Foundation of China (Grant No. 61662047). The authors would like to thank researchers and funding agencies that have contributed to ABIDE.

References

[1] https://www.cdc.gov/ncbddd/autism/data.html.
[2] Belmonte MK, Allen G, Beckel-Mitchener A, Boulanger LM, Carper RA, Webb SJ. Autism and abnormal development of brain connectivity. J Neurosci 2004;24(42):9228–31. https://doi.org/10.1523/JNEUROSCI.3340-04.2004.
[3] Just MA, Cherkassky VL, Keller TA, Minshew NJ. Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. Brain 2004;127(Pt 8):1811–21. https://doi.org/10.1093/brain/awh199.
[4] Geschwind DH, Levitt P. Autism spectrum disorders: developmental disconnection syndromes. Curr Opin Neurobiol 2007;17(1):103–11. https://doi.org/10.1016/j.conb.2007.01.009.
[5] Casanova M, Trippe J. Radial cytoarchitecture and patterns of cortical connectivity in autism. Philos Trans R Soc Lond B Biol Sci 2009;364(1522):1433–6. https://doi.org/10.1098/rstb.2008.0331.
[6] Muller RA, Shih P, Keehn B, Deyoe JR, Leyden KM, Shukla DK. Underconnected, but how? A survey of functional connectivity MRI studies in autism spectrum disorders. Cereb Cortex 2011;21(10):2233–43. https://doi.org/10.1093/cercor/bhq296.
[7] Di Martino A, Yan CG, Li Q, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry 2014;19(6):659–67. https://doi.org/10.1038/mp.2013.78.
[8] Anderson JS, Nielsen JA, Froehlich AL, et al. Functional connectivity magnetic resonance imaging classification of autism. Brain 2011;134(Pt 12). https://doi.org/10.1093/brain/awr263.
[9] Murdaugh DL, Shinkareva SV, Deshpande HR, Wang J, Pennick MR, Kana RK. Differential deactivation during mentalizing and classification of autism based on default mode network connectivity. PLoS One 2012;7(11):e50064https://doi.org/10.1371/journal.pone.0050064.
[10] Uddin LQ, Supekar K, Lynch CJ, et al. Salience network–based classification and prediction of symptom severity in children with autism. JAMA Psychiatry 2013;70(8):869–79. https://doi.org/10.1001/jamapsychiatry.2013.104.
[11] http://fcon_1000.projects.nitrc.org/indi/abide/index.html.
[12] Nielsen JA, Zielinski BA, Fletcher PT, et al. Multisite functional connectivity MRI classification of autism: ABIDE results. Front Hum Neurosci 2013;7:599. https://doi.org/10.3389/fnhum.2013.00599.
[13] Plitt M, Barnes KA, Martin A. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker. Neuroimage Clin 2015;7:359–66. https://doi.org/10.1016/j.nicl.2014.12.013.
[14] Abraham A, Milham M, Martino AD, et al. Deriving reproducible biomarkers from multi-site resting-state data: an Autism-based example. Neuro Image 2017;147:736–45. https://doi.org/10.1016/j.neuroimage.2016.10.045.
[15] Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. Neuroimage Clin 2018;17(Suppl 3):16–23. https://doi.org/10.1016/j.nicl.2017.08.017.
[16] Iidaka T. Resting state functional magnetic resonance imaging and neural network classified autism and control. Cortex 2015;63:55–67. https://doi.org/10.1016/j.cortex.2014.08.011.
[17] Guyon I, Weston J, Barnhill S, Vapnik VJML. Gene selection for cancer classification using support vector machines. Mach Learn 2002;46(1–3):389–422. https://doi.org/10.1023/a:1012487302797.
[18] Lin X, Li C, Zhang Y, Su B, Fan M, Wei H. Selecting feature subsets based on SVM-RFE and the overlapping ratio with applications in bioinformatics. Molecules 2017;23(1). https://doi.org/10.3390/molecules23010052.
[19] Ding X, Yang Y, Stein EA, Ross J. Multivariate classification of smokers and nonsmokers using SVM-RFE on structural MRI images. Hum Brain Mapp 2015;36(12):4869–79. https://doi.org/10.1002/hbm.22956.
[20] Clements CC, Zoltowski AR, Yankowitz LD, Yerys BE, Schultz RT, Herrington JDJJP. Evaluation of the social motivation hypothesis of autism: a systematic review and meta-analysis. JAMA Psychiatry 2018;75:797–808. https://doi.org/10.1001/jamapsychiatry.2018.1100.
[21] Chao-Gan Y, Yu-Feng Z. DPARSF: A MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. Front Syst Neurosci 2010;4:13. https://doi.org/10.3389/fnsys.2010.00013.
[22] Yan CG, Wang XD, Zuo XN, Zang YFJN. DPABI: data processing & analysis for

(resting-state) brain imaging. Neuroinformatics 2016;14(3):339–51. https://doi.org/10.1007/s12021-016-9299-4.
[23] Murphy K, Birn RM, Handwerker DA, Jones TB, Bandettini PA. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced? Neuroimage 2009;44(3):893–905. https://doi.org/10.1016/j.neuroimage.2008.09.036.
[24] Weissenbacher A, Kasess C, Gerstl F, Lanzenberger R, Moser E, Windischberger C. Correlations and anticorrelations in resting-state functional connectivity MRI: a quantitative comparison of preprocessing strategies. Neuroimage 2009;47:1408–16. https://doi.org/10.1016/j.neuroimage.2009.05.005.
[25] Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 2002;17(2):825–41. https://doi.org/10.1006/nimg.2002.1132.
[26] Pedregosa F, Gramfort A, Michel V, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2013;12(10):2825–30. https://doi.org/10.1524/auto.2011.0951.
[27] Chen H, Xujun Duan, et al. Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivity-A multicenter study. Prog Neuro Psychoph 2016;64:1–9. https://doi.org/10.1016/j.pnpbp.2015.06.014.