IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 11, 2019, accepted August 16, 2019, date of publication August 21, 2019, date of current version September 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2936639

Identification of Autism Based on SVM-RFE and Stacked Sparse Auto-Encoder

CANHUA WANG[1,2], ZHIYONG XIAO[3], BAOYU WANG[4], AND JIANHUA WU[4]

[1]School of Mechatronics Engineering, Nanchang University, Nanchang 330031, China
[2]School of Computer, Jiangxi University of Traditional Chinese Medicine, Nanchang 330004, China
[3]School of Software, Jiangxi Agricultural University, Nanchang 330045, China
[4]School of Information Engineering, Nanchang University, Nanchang 330031, China

Corresponding author: Jianhua Wu (jhwu@ncu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61662047.

ABSTRACT In order to improve the classification accuracy of patients with autism based on the full Autism Brain Imaging Data Exchange dataset, a total of 501 subjects with autism and 553 subjects with typical control across 17 sites were involved in the study. Firstly, we applied the resting-state functional magnetic resonance imaging data to calculate the functional connectivity (FC) based on the automated anatomical labeling atlas with 116 brain regions. Secondly, we adopted the support vector machine-recursive feature elimination algorithm to select top 1000 features from the primitive FC features. Thirdly, we trained a stacked sparse auto-encoder with two hidden layers to extract the high-level latent and complicated features from the 1000 features. Finally, the optimal features obtained were fed into the softmax classifier. Experimental results demonstrate that the proposed classification algorithm is able to identify the autism with a state-of-the-art accuracy of 93.59% (sensitivity 92.52%, specificity 94.56%).

INDEX TERMS Autism, fMRI, deep learning, SVM-RFE, classification.

I. INTRODUCTION

Autism spectrum disorders (ASD) are a cluster of neurodevelopmental conditions associated with core deficits in social communication, social interaction, and restricted and repetitive behaviors [1]. The prevalence of ASD has dramatically increased recently, reaching estimates of 1 in 59 children in the USA and 1% to 1.5% of children and adults worldwide according to the most recent investigation of the Centers for Disease Control and Prevention. It is regrettable that the exact etiopathogenesis of ASD remains unclear. Currently, diagnosis for ASD is solely behavior-based and relies on symptom-based clinical criteria which cannot well distinguish patients from typical controls (TC).

Recently, machine learning (ML) method has been wildly applied to extract biological information from magnetic resonance imaging (MRI) data, classify individuals with ASD and TC and further predict the tendency of the disease. The previous studies that utilized ML attained a relatively high classification accuracy of 65%-96.27% on the small dataset with less than 200 samples [2]–[11]. The classification algorithms in the related studies include support vector machine (SVM), logistic regression, random forests, linear discriminant analysis and deep belief network. Among numerous ML algorithms, SVM was the most prevalent classification method and obtained the highest accuracy in ASD classification (accuracy = 96.27%) based on a small dataset with only 117 samples [7]. One common limitation of these studies is the aggregation of a large number of features on an under-powered sample size, leading to models overfitting to the small dataset and generalizing poorly to a large data sample [12].

The Autism Brain Imaging Data Exchange (ABIDE) [13] consortium collected the MRI data consisting of 539 ASD and 573 matched controls across 17 independent sites, providing an unprecedented opportunity for a large-scale investigation of ASD. In [14], Abraham et al. demonstrated the feasibility of inter-site classification of neuropsychiatric status, with an application to the ABIDE database. By building participant-specific connectomes from functionally-defined brain areas, they achieved a classification accuracy of 67% in the full ABIDE dataset. In [15], the ABIDE dataset was investigated for functional connectivity of the regions of the brain as specified by the Craddock 200 template, Heinsfeld et al. applied the method of deep learning to obtain an accuracy of classification of 70% based on the whole ABIDE dataset.

The associate editor coordinating the review of this article and approving it for publication was Longzhi Yang.

However, large multi-site datasets increase sample size at the cost of uncontrolled heterogeneity. The heterogeneity means some disturbances to data samples as well as some loss in classification accuracy.

In order to improve the classification accuracy on the full ABIDE dataset, in this paper, we proposed a novel ML framework to distinguish autistic patients from normal controls. Firstly, we applied the resting-state functional magnetic resonance imaging (rs-fMRI) data to calculate the functional connectivity (FC) based on the automated anatomical labeling (AAL) atlas with 116 brain regions. Secondly, we adopted the support vector machine-recursive feature elimination (SVM-RFE) algorithm to select top 1000 features from the primitive FC features. Thirdly, we trained a stacked sparse auto-encoder (SSAE) with two hidden layers to extract the high-level latent and complicated features from the 1000 features. Finally, the optimal features obtained were fed into the softmax classifier. Experimental results demonstrate that the proposed classification algorithm is able to identify the autism with a state-of-the-art accuracy of 93.59% (sensitivity 92.52%, specificity 94.56%), which is an inspiring result for auxiliary clinical diagnosis of patients with ASD.

This paper is organized as follows: related work is introduced in Section II and the dataset used, the feature extraction and the proposed method are described in Section III; Section IV describes the main results from this study and a comparison with recent similar studies, followed by conclusion in Section V.

II. RELATED WORK

A. SVM-RFE

The SVM-RFE is a kind of backward elimination methods which starts with a full set of all features and then removes the most irrelevant features one by one [16]–[19]. The top ranked features removed in the last iteration of SVM-RFE are the most important, while the bottom ranked ones are the least informative and removed in the first iteration. In detail, the SVM-RFE algorithm can be expressed as follows:

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

Step 1: Training of an SVM on the initial features set:

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

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

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

Step 2: Computation of the weight vector and the ranking criteria according to:

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

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

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

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

Step 5: Repetition of Steps 1-4 until the features set is empty.

The pseudo code of SVM-RFE algorithm can be expressed as Table 1.

TABLE 1. SVM-RFE algorithm.

The pseudo code of SVM-RFE
Input: training dataset
Output: the ranked features dataset
Begin
F={all input features}
While (F is not empty) do
Train F by a linear SVM
Compute the weight vector w_i
Rank the features in F by w_i^2
Find the bottom ranked feature
$F = F -$ {the bottom ranked feature}
Return the ranked features dataset
End

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

B. AUTO-ENCODER

An auto-encoder (AE) neural network is an unsupervised learning framework that applies back propagation, setting the target values to be equal to the inputs. For a set of m unlabeled training examples$\{x^{(i)}|x^{(i)} \in R^{n+1}, i = 1, \ldots, m\}$, with $x_0^{(i)} = 1$ related to the bias input, the AE tries to learn an approximation function $h_{W,b}(x) \approx x$ such that the output is close to input, as depicted in Fig. 1.

The overall cost function of a sparse AE (SAE) is defined as

$$J_{sparse}(W, b) = J(W, b) + \beta\sum_{j=1}^{s_2} KL\left(\rho \,\|\, \hat{\rho}_j\right) \tag{5}$$

$$J(W, b) = \left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|h_{W,b}(x^{(i)}) - x^{(i)}\right\|^2\right)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^{(l)}\right)^2 \tag{6}$$

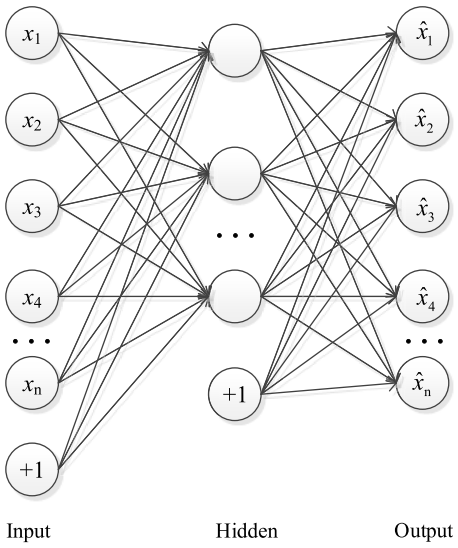


FIGURE 1. The structure of an AE.

$$KL(\rho\|\hat{\rho}_j) = \rho\, log\frac{\rho}{\hat{\rho}_j} + (1-\rho)log\frac{1-\rho}{1-\hat{\rho}_j} \tag{7}$$

$$\hat{\rho}_j = \frac{1}{m}\sum_{i=1}^{m} a_j^{(2)}(x^{(i)}) \tag{8}$$

where the parameter β in Eq. (5) controls the weight of the sparse penalty term; the first term in Eq. (6) is an average sum-of-square error term; the second term in Eq. (6) is a regularization term that tends to decrease the magnitude of the weights and helps prevent overfitting; the parameter λ controls the relative importance of the two terms; m is the number of training samples; n_l is the number of layers in the network; s_l is the number of nodes of layer l; $\hat{\rho}_j$ denotes the average activation of the hidden unit j; ρ is a sparse parameter. The target is to minimize Eq. (5) by resorting to an iterative optimization algorithm such as Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) [20], [21].

C. SOFTMAX REGRESSION

Let $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots .(x^{(m)}, y^{(m)})\}$ be a set of m labeled training examples and $x^{(i)} \in R^n$, $y^{(i)} \in \{1, 2, \ldots, k\}$. We want to estimate the probability of the class label taking on each of the k different possible values. The hypothesis $h_\theta(x)$ of softmax takes the form:

$$h_\theta\left(x^{(i)}\right) = \begin{bmatrix} p\left(y^{(i)}=1\left|x^{(i)};\theta\right.\right) \\ p\left(y^{(i)}=2\left|x^{(i)};\theta\right.\right) \\ \vdots \\ p\left(y^{(i)}=k\left|x^{(i)};\theta\right.\right) \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \tag{9}$$

here $\theta = (\theta_1^T, \theta_2^T, \ldots, \theta_k^T)^T$ denotes the parameter set of the model and is usually represented as a $k \times (n+1)$ matrix obtained by stacking up $\theta_j, j = 1, \cdots, k$ with $\theta_j \in R^{n+1}$ in rows.

The cost function is

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\left\{y^{(i)}=j\right\} \log\frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}\right] + \frac{\lambda_s}{2}\sum_{i=1}^{k}\sum_{j=0}^{n}\theta_{ij}^2 \tag{10}$$

where $1\{\cdot\}$ is the indicator function, the weight decay term $\frac{\lambda_s}{2}\sum_{i=1}^{k}\sum_{j=0}^{n}\theta_{ij}^2$ is used to penalize large values of the parameters and make $J(\theta)$ strictly convex.

III. MATERIALS AND METHODS

A. PARTICIPANTS AND DATA PREPROCESSING

The original fMRI and demographic data were collected from the ABIDE which allows unrestricted usage for non-commercial research purpose. The scanning parameters of these institutes and composition of subjects are listed in Table 2. The Preprocessed Connectomes Project (PCP) opened sharing of preprocessed neuroimaging data from ABIDE [22]. All data in the study were preprocessed by the DPARSF [23] pipeline in the PCP. Specifically, after a dropout of the first 4 volumes, slice timing correction, head-motion realigned and intensity normalized. Nuisance signals removal was performed to clean confounding variation due to physiological processes (heart beat and respiration),

TABLE 2. Scanning parameters and composition of subjects in different sites.

Institute	MRI vendor	TR (msec.)	Volumes	ASD	TC
CIT	Siemens	2000	150	19	19
CMU	Siemens	2000	240	6	8
KKI	Phillips	2500	156	21	33
LEU	Phillips	1667	250	29	33
LMUM	Siemens	3000	120	24	32
NYU	Siemens	2000	180	79	103
OHSU	Siemens	2500	82	13	15
OLIN	Siemens	1500	210	19	16
PITT	Siemens	1500	200	20	25
SBL	Phillips	2200	200	7	10
SDSU	GE	2000	180	14	22
STAN	GE	2000	180	20	20
TRIN	Phillips	2000	150	24	25
UCLA	Siemens	3000	120	54	45
UM	GE	2000	300	66	77
USM	Siemens	2000	240	58	42
YALE	Siemens	2000	200	28	28

CIT, California Institute of Technology; CMU, Carnegie Mellon University; KKI, Kennedy Krieger Institute; LEU, University of Leuven; LMUM, Ludwig Maximilians University Munich; 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; STAN, Stanford University; TRIN, Trinity Centre for Health Sciences; UCLA, University of California Los Angeles; UM, University of Michigan; USM, University of Utah School of Medicine; YALE, Yale Child Study Center.

head motion, and low frequency scanner drifts. A band-pass filtering (0.01 – 0.1 Hz) was applied after nuisance variable regression. A transform from original space to template (MNI152) space was calculated for each dataset by a combination of functional-to-anatomical and anatomical-to-template transforms. According to literature [24], the head movement had no significant effect on the classification accuracy of the autism. Therefore, we only excluded the subjects whose structural images were not covered completely after preprocessing. Thus, 501 ASD and 553 TC across 17 sites were involved in the study.

B. CONNECTIVITY MEASURES AND FEATURE MATRICES

The AAL atlas [25], which divides the brain into 116 regions, is a standard brain template for creating intrinsic connectivity. Time courses were extracted from each of the 116 regions and averaged within each region. Pearson's correlation coefficients ($R(i,j)$) were computed between these average time courses according to:

$$R(i,j) = \frac{C(i,j)}{\sqrt{C(i,i)C(j,j)}} \tag{11}$$

where, $C(i,j)$ denotes the correlation coefficient between time-courses i and j.

In order to make it more statistically significant, Fisher's z transformation was implemented according to:

$$Z(i,j) = \frac{1}{2} \times \log\left[\frac{(1+R(i,j)}{(1-R(i,j)}\right] \tag{12}$$

So, the Fisher's z-transformed correlation coefficients were represented in a 116×116 matrix (13,456 features), 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 6670. 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.

C. FEATURE SELECTION BASED ON SVM-RFE

The FC mentioned above resulted in 6670 features while the sample size was 1054. So, this was summarized in the expression ''high dimensions, small samples''. It was difficult to learn the most discriminative feature that the 6670 features were directly fed to SAE in the case of small samples. SVM-RFE could pre-eliminate some meaningless features. So, we applied the SVM-RFE to select the top 1000 ones in 6670 initial features in order to make the SAE learning high-level features well. The top 1000 features were selected by the SVM-RFE in 6670 FC set, as shown in Fig. 2.

D. FEATURE SELF-TAUGHT LEARNING AND CLASSIFICATION

In order to extract the high-level latent and complicated features of FC, we constructed a feature self-taught learning network consisting of two layers of SAE in which the

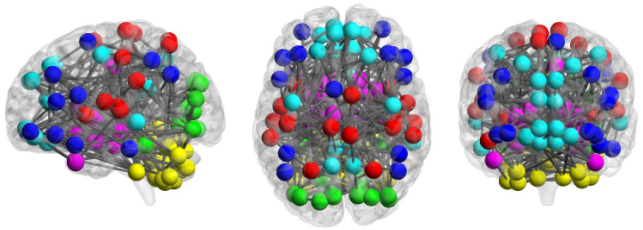


FIGURE 2. The top 1000 FC features selected by SVM-RFE.

outputs of the first layer are wired to the inputs of the second layer. Let $W^{(k,1)}$, $W^{(k,2)}$,$b^{(k,1)}$,$b^{(k,2)}$ denote the weight parameters and bias for the kth SAE, respectively. A good way to obtain good parameters for an SAE is to use a greedy layer-wise training [26]. To do this, the first layer is trained firstly on raw input to learn parameters $W^{(1,1)}$, $W^{(1,2)}$, $b^{(1,1)}$ and $b^{(1,2)}$. The first layer transforms the raw input into a vector consisting of activations of the hidden units. The second layer is then trained on this vector to learn parameters $W^{(2,1)}$,$W^{(2,2)}$,$b^{(2,1)}$ and $b^{(2,2)}$. This training process is repeated for subsequent layers by using the output of each layer as input for the subsequent layer. The network constructed in this way by two or more SAE is called stacked sparse auto-encoder (SSAE).

As shown in Fig. 3, the feature self-taught learning system based on SSAE was configured as 1000-200-100. The top 1000 features selected by the SVM-RFE were fed into the first SAE with 200 hidden units in which 200 high-level features were learned. The 100 sophisticated features were learned in turn by the second SAE. The features self-taught learning extracted from the two successive SAE are shown in Fig. 4 and Fig. 5 respectively.

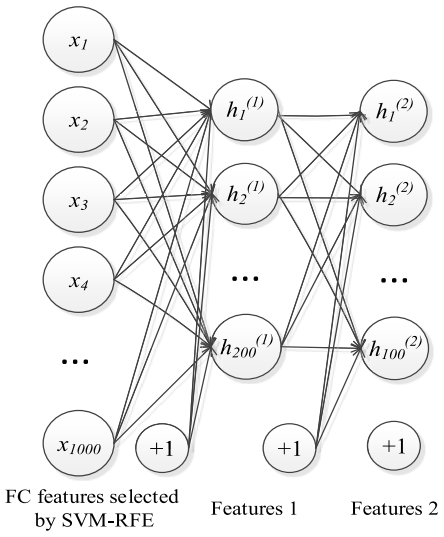


FIGURE 3. The structure of SSAE consisting of two SAE.

The outputs of the SSAE, used as features, were fed into the softmax classifier generalized by the logistic regression [27]. The total pattern recognition scheme is showed in Fig. 6.

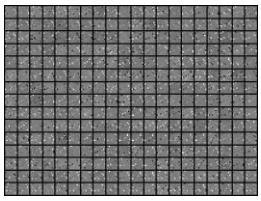


FIGURE 4. The features learned firstly.

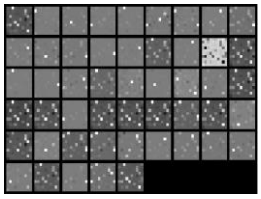


FIGURE 5. The features learned again.

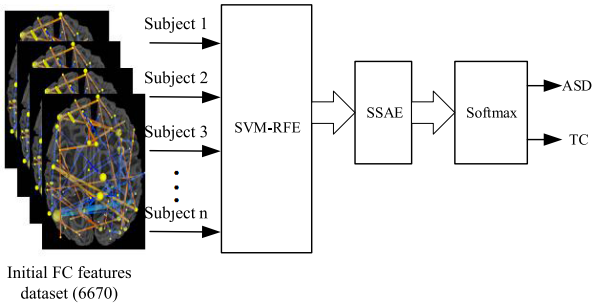


FIGURE 6. Pattern recognition schemes for ASD and TC.

For the SSAE, the sparse parameter ρ, the weight decay parameter λ, the weight of the sparse penalty terms β and the number of iterations, are set to be 0.08, 1e-4, 2 and 200, respectively. For the softmax classifier, the weight decay parameter λ_s is 1e-4 and the number of iterations is 100. In addition, we made use of fine-tuning to improve the performance of the SSAE after pre-training.

IV. RESULTS AND DISCUSSION

The study was carried out on the full ABIDE dataset. The dataset of 1054 samples (501 ASD, 553 TC) were randomly split with k equally sized subsets and fed into the proposed pattern recognition network. The classification quality was assessed by the following performance indices:

$$Accuracy = (TP + TN)/(TP + FN + TN + FP) \quad (13)$$

$$Sensitivity = TP/(TP + FN) \quad (14)$$

$$Specificity = TN/(TN + FP) \quad (15)$$

$$PPV = TP/(TP + FP) \quad (16)$$

$$NPV = TN/(TN + FN) \quad (17)$$

where, true positive (TP), false negative (FN), true negative (TN), and false positive (FP) denote, respectively, the number of ASD correctly classified, the number of ASD predicted to be TC, the number of TC correctly classified, and the number of TC predicted to be ASD. The results of recognition for ASD and TC are listed in Table 3 in terms of different fold cross-validation (CV). The values of various indicators such as accuracy are very stable even if k is different. The average classification accuracy achieves up to 93.59%.

TABLE 3. Results of classification with different CV.

	5-fold	10-fold	20-fold	30-fold	Average
Accuracy	91.65%	93.17%	94.67%	94.87%	93.59%
Sensitivity	90.62%	91.21%	94.22%	94.02%	92.52%
Specificity	92.58%	94.93%	95.10%	95.64%	94.56%
PPV	91.76%	94.29%	94.71%	95.31%	94.01%
NPV	91.60%	92.32%	94.92%	94.81%	93.41%

The sensitivity measured the proportion of ASDs that were correctly identified and specificity measured the proportion of TCs that were correctly identified. To assess a realistic prospect of how our model would behave in the real clinical world, we calculated PPV and NPV respectively. The proposed method attained a relatively high PPV of 94.01% and an NPV of 93.41%. Based on the literature, this is the best result of classification achieved so far.

Furthermore, if the SSAE was configured 6670-200-100 or 6670-3072-1024-512-100, that is, original FC features were entered into the SSAE without selection by the SVM-RFE, the classification accuracy was only 63.37% and 64.01% respectively. Although the number of hidden layer of SSAE was increased, it was difficult for the SSAE to extract the most discriminatively features in the case of ''high dimensions, small samples''. In addition, if the 1000 or 100 FC features selected by the SVM-RFE were directly entered into the softmax classifier without the SSAE, the classification accuracy was only 67.25% and 63.12% respectively.

The result of identification was dissatisfactory only relying on deep learning since the sample was small relative to the size of FC features even if the sample came from the whole ABIDE dataset. Also, the classification accuracy was unsatisfactory only relying on the SVM-RFE without the SSAE. The issue about ''high dimensions, small samples'' was addressed well by combining the SVM-RFE with SSAE. Using the SVM-RFE to pre-eliminate some meaningless features could reduce some noise for the SSAE. Some high-level latent features could be learned by the SSAE based on the features selected by the SVM-RFE. At a higher abstract level, these transformations created representations that were used for the classification task.

Compared with the recent similar studies, the proposed method performs better than literature [14] and [15] based on the same large multi-site dataset. The comparison of classification accuracy by different methods is shown in Table 4.

In addition, the receiver operating characteristic (ROC) curve of the proposed method, the method of the literature 14 and the method of the literature 15 are shown in Fig.7, from which it can be seen that for all classifiers the true positive rate increases with the increase of the false positive rate,

TABLE 4. Results of classification with different methods.

	The proposed method	The method of literature 14	The method of literature 15
Accuracy	93.59%	67.57%	70.00%
Sensitivity	92.52%	62.63%	74.00%
Specificity	94.56%	72.32%	63.00%

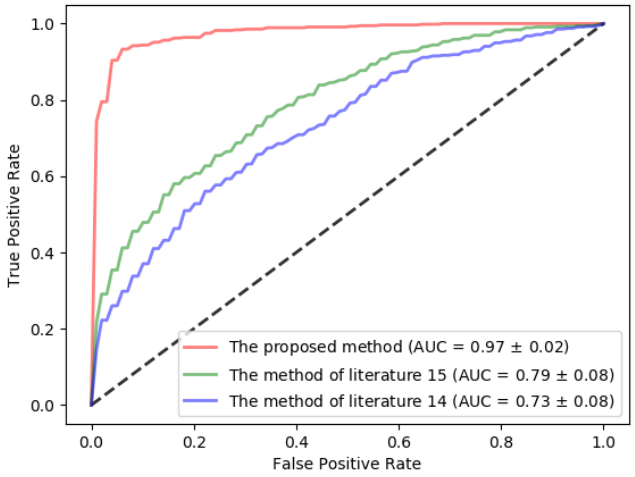


FIGURE 7. The ROC curves of different algorithms.

while the proposed method achieves the largest area under the curve (AUC).

The proposed method is completely data-driven without any prior knowledge. The initial FC features in this study come from the regions of total brain rather than local regions such as the default mode network. Some selected regions are not really linked with ASD according to the existing literatures [28]–[33], but it is possible that these regions play some unrecognized roles in autism. It is perhaps the reason that the data-driven method can address well the uncontrolled heterogeneity generated by different equipment and demographics in the large multi-site dataset. As a result, doing more conveying on unrecognized regions would be essential for further development of ASD diagnosis.

V. CONCLUSION

In this study, a deep learning method combined with SVM-RFE was proposed to improve the classification accuracy of ASD based on the whole ABIDE dataset. The proposed method used SVM-RFE to pre-eliminate some meaningless features in order to enable the SSAE to extract well the sophisticated features in the case of "high dimensions, small samples". A total of 501 subjects with autism and 553 subjects with typical control across 17 sites were involved in the study. The state-of-the-art average accuracy of 93.59%, ROC and AUC are all better than that in recent similar studies on the largest dataset of autism. The experimental results demonstrate that some selected ROIs by the proposed method might play some unrecognized roles in autism although these ones are not really linked with ASD according to the existing literatures. It does show that the data-driven method can address well the uncontrolled heterogeneity generated by different equipment and demographics in the large multi-site dataset. It is an inspiring result for auxiliary clinical diagnosis of patients with ASD.

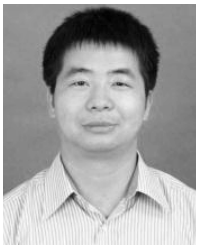
Further research is necessary to combine multiple indices of brain imaging such as grey matter volume and cortex thickness with the FC, which might lead to better results of identification.

REFERENCES

[1] E. P. K. Pua, S. C. Bowden, and M. L. Seal, "Autism spectrum disorders: Neuroimaging findings from systematic reviews," *Res. Autism Spectr. Disorders*, vol. 34, pp. 28–33, Feb. 2017.
[2] J. S. Anderson, J. A. Nielsen, A. L. Froehlich, M. B. DuBray, T. J. Druzgal, A. N. Cariello, B. Cooperrider, Z. Caitlin, R. P. Fletcher, A. Alexander, E. Bigler, N. Lange, and J. Lainhart, "Functional connectivity magnetic resonance imaging classification of autism," *Brain*, vol. 134, pp. 3742–3754, Dec. 2011.
[3] D. L. Murdaugh, S. V. Shinkareva, H. R. Deshpande, J. Wang, M. R. Pennick, and R. K. Kana, "Differential deactivation during mentalizing and classification of autism based on default mode network connectivity," *PLoS ONE*, vol. 7, no. 11, 2012, Art. no. e50064.
[4] L. Q. Uddin, K. Supekar, C. J. Lynch, A. Khouzam, J. Phillips, C. Feinstein, S. Ryali, and V. Menon, "Salience network-based classification and prediction of symptom severity in children with autism," *JAMA Psychiatry*, vol. 70, pp. 869–879, Aug. 2013.
[5] X.-A. Bi, Y. Wang, Q. Shu, Q. Sun, and Q. Xu, "Classification of autism spectrum disorder using random support vector machine cluster," *Frontiers Genet.*, vol. 9, p. 18, Feb. 2018.
[6] 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 Clin.*, vol. 7, pp. 359–366, 2015.
[7] C.-Y. Wee, L. Wang, F. Shi, P.-T. Yap, and D. Shen, "Diagnosis of autism spectrum disorders using regional and interregional morphological features," *Hum. Brain Mapping*, vol. 35, pp. 3414–3430, Jul. 2014.
[8] L. Wang, C.-Y. Wee, X. Tang, P.-T. Yap, and D. Shen, "Multi-task feature selection via supervised canonical graph matching for diagnosis of autism spectrum disorder," *Brain Imag. Behav.*, vol. 10, pp. 33–40, Mar. 2016.
[9] X. Xiao, H. Fang, J. Wu, C. Xiao, T. Xiao, L. Qian, F. Liang, Z. Xiao, K. K. Chu, and X. Ke, "Diagnostic model generated by MRI-derived brain features in toddlers with autism spectrum disorder," *Autism Res.*, vol. 10, pp. 620–630, Apr. 2017.
[10] A. Chaddad, C. Desrosiers, L. Hassan, and C. Tanougast, "Hippocampus and amygdala radiomic biomarkers for the study of autism spectrum disorder," *BMC Neurosci.*, vol. 18, p. 52, Jul. 2017.
[11] M. A. Aghdam, A. Sharifi, and M. M. Pedram, "Combination of rs-fMRI and sMRI data to discriminate autism spectrum disorders in young children using deep belief network," *J. Digit. Imag.*, vol. 31, no. 6, pp. 895–903, Dec. 2018.
[12] A. M. Pagnozzi, E. Conti, S. Calderoni, J. Fripp, and S. E. Rose, "A systematic review of structural MRI biomarkers in autism spectrum disorder: A machine learning perspective," *Int. J. Develop. Neurosci.*, vol. 71, pp. 68–82, Dec. 2018.
[13] [Online]. Available: http://fcon_1000.projects.nitrc.org/indi/abide
[14] 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, Feb. 2017.
[15] 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, Clin.*, vol. 17, pp. 16–23, Jan. 2018.
[16] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, nos. 1–3, pp. 389–422, 2002.
[17] X. Lin, C. Li, Y. Zhang, B. Su, M. Fan, and H. Wei, "Selecting feature subsets based on SVM-RFE and the overlapping ratio with applications in bioinformatics," *Molecules*, vol. 23, p. 52, Dec. 2017.

[18] J. Ding, J. Shi, and F.-X. Wu, "SVM-RFE based feature selection for tandem mass spectrum quality assessment," *Int. J. Data Mining Bioinf.*, vol. 5, pp. 73–88, Feb. 2011.
[19] X. Ding, Y. Yang, E. A. Stein, and T. J. Ross, "Multivariate classification of smokers and nonsmokers using SVM-RFE on structural MRI images," *Hum. Brain Mapping*, vol. 36, pp. 4869–4879, Dec. 2015.
[20] M.'A. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, "Efficient learning of sparse representations with an energy-based model," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19, 2006, pp. 1137–1144,
[21] I. J. Goodfellow, Q. V. Le, A. M. Saxe, H. Lee, and A. Y. Ng, "Measuring invariances in deep networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2009, pp. 646–654.
[22] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. S. Khundrakpam, J. D. Lewis, Q. Li, M. Milham, C. Yan, and P. Bellec, "The Neuro bureau preprocessing initiative: Open sharing of preprocessed neuroimaging data and derivatives," *Frontiers Neuroinform.*, to be published.
[23] Y. Chao-Gan and Z. Yu-Feng, "DPARSF: A MATLAB toolbox for 'Pipeline' data analysis of resting-state fMRI," *Frontiers Syst. Neurosci.*, vol. 4, p. 13, May 2010.
[24] T. Iidaka, "Resting state functional magnetic resonance imaging and neural network classified autism and control," *Cortex*, vol. 63, pp. 55–67, Feb. 2015.
[25] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage*, vol. 15, pp. 273–289, Jan. 2002.
[26] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
[27] (Apr. 7, 2013). *UFLDL Tutorial*. [Online]. Available: http://ufldl.stanford.edu/wiki/index.php/Softmax_Regression
[28] M. A. Just, V. L. Cherkassky, T. A. Keller, and N. J. Minshew, "Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity," *Brain*, vol. 127, pp. 1811–1821, Aug. 2004.
[29] D. H. Geschwind and P. Levitt, "Autism spectrum disorders: Developmental disconnection syndromes," *Current Opinion Neurobiol.*, vol. 17, pp. 103–111, Feb. 2007.
[30] M. Casanova and J. Trippe, "Radial cytoarchitecture and patterns of cortical connectivity in autism," *Philos. Trans. Roy. Soc. B, Biol. Sci.*, vol. 364, pp. 1433–1436, May 2009.
[31] R.-A. Müller, P. Shih, B. Keehn, J. R. Deyoe, K. M. Leyden, and D. K. Shukla, "Underconnected, but How? A survey of functional connectivity MRI studies in autism spectrum disorders," *Cerebral Cortex*, vol. 21, pp. 2233–2243, Oct. 2011.
[32] 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, B. Deen, S. Delmonte, I. Dinstein, B. Ertl-Wagner, D. A. Fair, L. Gallagher, D. P. Kennedy, C. L. Keown, C. Keysers, and J. E. Lainhart, "The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism," *Mol. Psychiatry*, vol. 19, no. 6, pp. 659–667, 2014.
[33] M. K. Belmonte, G. Allen, A. Beckel-Mitchener, L. M. Boulanger, R. A. Carper, and S. J. Webb, "Autism and abnormal development of brain connectivity," *J. Neurosci.*, vol. 24, pp. 9228–9231, Oct. 2004.

CANHUA WANG was born in Pengze, Jiangxi, China, in 1978. He received the B.S. degree in electrical engineering and the M.S. degree in communication and information system from Nanchang University, Nanchang, China, in 2001 and 2007, respectively, where he is currently pursuing the Ph.D. degree in mechatronics engineering. His research interests include machine learning, signal processing, and computer vision.



ZHIYONG XIAO was born in Xingan, Jiangxi, China, in 1978. He received the master's degree in control theory and engineering from the East China University of Science and Technology, Shanghai, China, in 2008. He is currently pursuing the Ph.D. degree with the School of Mechatronic Engineering, Nanchang University. His research interests include image processing, computer vision, and pattern recognition.

BAOYU WANG received the B.E. degree in Internet-of-Things engineering from Jiangxi Science and Technology Normal University, Nanchang, China, in 2017. He is currently pursuing the M.E. degree in electronics and communication engineering with Nanchang University. His current research interests include image processing, computer vision, object detection, and automatic optical inspection.

JIANHUA WU was born in Jinxian, Jiangxi, China. He received the bachelor's degree in information engineering from the Harbin Institute of Technology, in 1982, the master's degree in science with a major in communication and electronic systems from the South China University of Technology, Guangzhou, China, in 1985, and the Ph.D. degree in image and signal processing from the University of Poitiers, Poitiers, France, in 2005.

He is currently a Professor with the Department of Electronic Information Engineering, Nanchang University. He has published more than 20 articles in academic journals, such as *Optics Communications* and *Optics & Laser Technology*. His research interests include image and signal processing, power quality signal analysis, and pattern recognition.