Neurocomputing 324 (2019) 63–68

Contents lists available at ScienceDirect

Neurocomputing

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

Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier

Yazhou Kong[a], Jianliang Gao[a], Yunpei Xu[a], Yi Pan[a,b], Jianxin Wang[a], Jin Liu[a,*]

[a] School of Information Science and Engineering, Central South University, Changsha 410083, China
[b] Department of Computer Science, Georgia State University, Atlanta, GA 30303, USA

ARTICLE INFO

Article history:
Received 31 December 2017
Revised 10 March 2018
Accepted 3 April 2018
Available online 24 May 2018

Keywords:
ASD
Morphological features
Individual network
Deep neural network
Classification

ABSTRACT

Autism spectrum disorder (ASD) is a common neurodevelopmental disorder that seriously affects communication and sociality of patients. It is crucial to accurately identify patients with ASD from typical controls (TC). Conventional methods for the classification of ASD/TC mainly extract morphological features independently at different regions of interest (ROIs), rarely considering the connectivity between these ROIs. In this study, we construct an individual brain network as feature representation, and use a deep neural network (DNN) classifier to perform ASD/TC classification. Firstly, we construct an individual brain network for each subject, and extract connectivity features between each pair of ROIs. Secondly, the connectivity features are ranked in descending order using *F*-score, and the top ranked features are selected. Finally, the selected 3000 top features are used to perform ASD/TC classification via a DNN classifier. An evaluation of the proposed method has been conducted with T1-weighted MRI images from the Autism Brain Imaging Data Exchange I (ABIDE I) by using ten-fold cross validation. Experimental results show that our proposed method can achieve the accuracy of 90.39% and the area under receiver operating characteristic curve (AUC) of 0.9738 for ASD/TC classification. Comparison of experimental results illustrates that our proposed method outperforms some state-of-the-art methods in ASD/TC classification.

1. Introduction

Autism spectrum disorder (ASD) is a common complex neurodevelopmental disorder. It is associated with several comorbid disorders, including intellectual impairment, seizures and anxiety [1,2]. The 2013 report showed that 1 in 55 children aged 6–17 years had been identified with ASD [3]. These who have ASD behave mild to severe impairments in social interaction and communication along with restricted, repetitive of behaviors and interests [4]. It is a crucial step to accurately identify patients with ASD from typical controls (TC).

At present, neuroimaging technology has widely been used in the study of various brain diseases, such as ASD [5,6], Alzheimer's disease [7,8] and schizophrenia classification [9,10]. Magnetic resonance imaging (MRI) is a powerful and safe technique, which provides high quality three-dimensional (3D) images of brain structures and detailed structural information. Morphological studies based on MRI images have been applied to the related diseases, and have achieved good results. For example, Akshoomoff et al. [11] used six preselected brain volume-based features to perform ASD classification, and the results indicated that variability in cerebellar and cerebral size is correlated with ASD. Jiao et al. [12] used regional cortical thickness extracted from surface-based morphology to perform ASD classification and achieved good classification performance. Wee et al. [13] extracted morphological features of different ROIs (including mean cortical thickness, regional cortical volume, etc.) to perform alzheimer's disease classification. Han et al. [14] extracted the morphological features of MRI data to classify schizophrenia and showed that the gray matter density of frontal lobes in the schizophrenic patients is lower than that in the normal control groups [15,16]. In this study, we use the morphological features derived from T1w MRI to investigate the predictive power in ASD/TC classification.

Traditional machine learning methods are widely used in disease classification [17–21]. For example, Ecker et al. [18] used SVM classifier to investigate the predictive value of whole-brain structural volumetric changes in ASD. Xiao et al. [21] applied three popular machine learning classifiers including RF, NB and SVM to perform ASD classification. Due to the deep feature representation can be extracted from the deep neural network, the DNN framework has been widely used for medical image classification [22,23]. For example, Hinton et al. [24] proposed the first deep autoencoder

* Corresponding author.
E-mail address: liujin06@csu.edu.cn (J. Liu).

https://doi.org/10.1016/j.neucom.2018.04.080
0925-2312/

Table 1
Demographic information of 182 subjects from NYU Langone Medical Center.

Type	Number	Age	Gender (M/F)
ASD	78	14.54 ± 5.29	68/10
TC	104	15.87 ± 5.04	78/26

M: Male; F: Female. The values are denoted as mean ± standard deviation.

network to effectively address some of most challenging problems in computer vision learning. Cruz-Roa et al. [25] proposed convolutional autoencoder neural network architecture for histopathological image representation learning. It can be seen that the combination of deep learning technology and MRI images will become a research trend [26]. In this study, we propose a deep neural network classifier based on stacked autoencoder to perform ASD classification. The DNN classifier consists of two autoencoders and a softmax function. First of all, we construct an individual brain network to extract cortical gray matter volume (CGMV) from the structural MRI image. Then, all features are ranked by *F*-score in descending order, and the top ranked features are selected. Finally, the selected 3000 top features are used to perform ASD/TC classification via the DNN classifier.

The rest of the article is organized as follows. First, we describe the details of our proposed ASD/TC classification method in Section 2. Then, we illustrate the experiments of ASD/TC classification and discuss comparison results in Section 3. Finally, we draw conclusions in Section 4.

2. Materials and methods

2.1. Image acquisition and preprocessing

A subset of the T1w MRI images from the Autism Brain Imaging Data Exchange I (ABIDE I) site: NYU Langone Medical Center is used to evaluate our proposed method. This subset includes 182 subjects with T1w MRI images, which are composed of 78 subjects with ASD and 104 TC subjects. The demographic information of 182 subjects is shown in Table 1. It is worth mentioning that the average age of subjects with ASD and TC subjects in this study are about 15 years old, and male subjects are significantly more than female subjects. For more details with the ABIDE I, please see http://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html.

All T1w MRI images are preprocessed by using FreeSurfer image analysis suite [27]. The preprocessing procedures in this study mainly include motion correction, intensity normalization, skull stripping and cerebellum removal.

2.2. Construction of individual networks

In order to parcellate the whole brain, Destrieux atlas proposed by Destrieux et al. [28] is used in this study. This atlas parcellates the brain into 148 cortical regions (74 in each hemisphere). Details of the Destrieux atlas are shown in Table 2. For more details with Destrieux atlas, please see http://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation.

Since brain network can unravel the extraordinary complexity of neuronal connectivity, it plays an important role in the studies of brain disease [29,30]. In this study, we propose to construct an individual network based on cortical regions of Destrieux atlas, denoted G_{CGMV}. For G_{CGMV}, the gray matter volume of each ROI is defined as the node of the individual network. The correlation between each pair of ROIs is defined as the edge of the individual network. The correlation is calculated as follows:

$$c(a,b) = \frac{1}{d(a,b)+1}, (a \neq b) \tag{1}$$

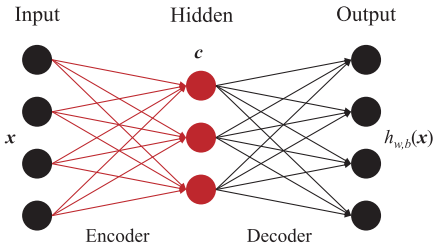

Fig. 1. An example of an autoencoder.

$$d(a,b) = |t(a) - t(b)|^2 \tag{2}$$

where $d(a, b)$ is defined as the difference between ROIs a and b, and $t(a)$ and $t(b)$ are the gray matter volume of ROIs a and b, respectively. Finally, each subject can be represented by a $148 \times (148-1)/2 = 10,878$ dimensional vector.

2.3. Feature ranking

Solving classification problems with data of high dimensionality is a challenging task due to the curse of dimensionality. It is particularly obvious for neuroimaging classification problems. With the presence of uninformative, irrelevant or redundant features, learning models tend to overfit and become less generalizable. Feature ranking is a useful and important means to identify relevant features for dimensionality reduction and improving generalization performance [31,32]. In this study, we use the *F*-score [33] to rank all features extracted above.

F-score is a simple technique which measures the discrimination of two sets of real numbers. Given training vectors $\boldsymbol{x}_k$, $k = 1, \ldots, m$, if the number of positive instances and negative instances are n_+ and n_-, respectively, then the *F*-score of the ith feature is defined as:

$$F(i) = \frac{\left(\bar{\boldsymbol{x}}_i^{(+)} - \bar{\boldsymbol{x}}_i\right)^2 + \left(\bar{\boldsymbol{x}}_i^{(-)} - \bar{\boldsymbol{x}}_i\right)^2}{\frac{1}{n_+-1}\sum_{k=1}^{n_+}\left(\boldsymbol{x}_{k,i}^{(+)} - \bar{\boldsymbol{x}}_i^{(+)}\right)^2 + \frac{1}{n_--1}\sum_{k=1}^{n_-}\left(\boldsymbol{x}_{k,i}^{(-)} - \bar{\boldsymbol{x}}_i^{(-)}\right)^2} \tag{3}$$

where $\bar{\boldsymbol{x}}_i$, $\bar{\boldsymbol{x}}_i^{(+)}$, $\bar{\boldsymbol{x}}_i^{(-)}$ are the average of the ith feature of the whole, positive, and negative data sets, respectively; $\boldsymbol{x}_{k,i}^{(+)}$ is the ith feature of the kth positive instance, and $\boldsymbol{x}_{k,i}^{(-)}$ is the ith feature of the kth negative instance. The numerator indicates the discrimination between the positive and negative sets, and the denominator indicates the one within each of the two sets. The greater the *F*-score value, the more the discernment of this feature may be.

2.4. Deep neural network classifier

An autoencoder (AE) is a simple neural network which can reconstruct the input as much as possible, that is, tries to learning a function model of $h_{w,b}(\boldsymbol{x}) \approx \boldsymbol{x}$. As is shown in Fig. 1, an autoencoder is composed with an encoder and a decoder. Firstly, the raw signal $\boldsymbol{x}$ is input to the encoder, which corresponds to a lower dimensional hidden layer representation $\boldsymbol{c}$. Then, the representation $\boldsymbol{c}$ is input into the decoder, and accordingly there is an output $h_{w,b}(\boldsymbol{x})$. Finally, we continue to adjust the parameters of model by back-propagation to minimize the error between the input $\boldsymbol{x}$ and the output $h_{w,b}(\boldsymbol{x})$. There are many kinds of autoencoder, such as sparse autoencoder [34], denoising autoencoder [35], contractive autoencoder [36], etc. Sparse autoencoder often can better learn the representation of the raw data. The loss function of the sparse

Table 2
Cortical regions defined in this study.

Cortical regions (74 in each hemisphere)			
Regions	*ID*	*Regions*	*ID*
G_and_S_frontomargin	1	G_temporal_middle	38
G_and_S_occipital_inf	2	Lat_Fis-ant-Horizont	39
G_and_S_paracentral	3	Lat_Fis-ant-Vertical	40
G_and_S_subcentral	4	Lat_Fis-post	41
G_and_S_transv_frontopol	5	Pole_occipital	42
G_and_S_cingul-Ant	6	Pole_temporal	43
G_and_S_cingul-Mid-Ant	7	S_calcarine	44
G_and_S_cingul-Mid-Post	8	S_central	45
G_cingul-Post-dorsal	9	S_cingul-Marginalis	46
G_cingul-Post-ventral	10	S_circular_insula_ant	47
G_cuneus	11	S_circular_insula_inf	48
G_front_inf-Opercular	12	S_circular_insula_sup	49
G_front_inf-Orbital	13	S_collat_transv_ant	50
G_front_inf-Triangul	14	S_collat_transv_post	51
G_front_middle	15	S_front_inf	52
G_front_sup	16	S_front_middle	53
G_Ins_lg_and_S_cent_ins	17	S_front_sup	54
G_insular_short	18	S_interm_prim-Jensen	55
G_occipital_middle	19	S_intrapariet_and_P_trans	56
G_occipital_sup	20	S_oc_middle_and_Lunatus	57
G_oc-temp_lat-fusifor	21	S_oc_sup_and_transversal	58
G_oc-temp_med-Lingual	22	S_occipital_ant	59
G_oc-temp_med-Parahip	23	S_oc-temp_lat	60
G_orbital	24	S_oc-temp_med_and_Lingual	61
G_pariet_inf-Angular	25	S_orbital_lateral	62
G_pariet_inf-Supramar	26	S_orbital_med-olfact	63
G_parietal_sup	27	S_orbital-H_Shaped	64
G_postcentral	28	S_parieto_occipital	65
G_precentral	29	S_pericallosal	66
G_precuneus	30	S_postcentral	67
G_rectus	31	S_precentral-inf-part	68
G_subcallosal	32	S_precentral-sup-part	69
G_temp_sup-G_T_transv	33	S_suborbital	70
G_temp_sup-Lateral	34	S_subparietal	71
G_temp_sup-Plan_polar	35	S_temporal_inf	72
G_temp_sup-Plan_tempo	36	S_temporal_sup	73
G_temporal_inf	37	S_temporal_transverse	74

autoencoder is defined as follow:

$$J_{sparse}(W,b) = J(W,b) + \beta * \sum_{j}^{n} KL(\rho \parallel \widehat{\rho}_j) + \frac{\lambda}{2} \sum_{k}^{2} \sum_{i}^{k} \sum_{j}^{k+1} \left(w_{ij}^{(k)}\right)^2 \tag{4}$$

$$J(W,b) = \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{2} \left\| h_{w,b}(x^{(i)}) - y^{(i)} \right\|^2 \right) \tag{5}$$

$$KL(\rho \parallel \widehat{\rho}_j) = \rho \log \frac{\rho}{\widehat{\rho}_j} + (1-\rho) \log \frac{1-\rho}{1-\widehat{\rho}_j} \tag{6}$$

$$\widehat{\rho}_j = \frac{1}{m} \sum_{i=1}^{m} [a_j^{(2)}(x^{(i)})] \tag{7}$$

where $J_{sparse}(W, b)$ represents the loss function of the input and reconstructed output, $J(W, b)$ is the mean square error, m is the number of samples, $h_{w,\,h}(x^{(i)})$ denotes the output of the ith sample, $x^{(i)}$ denotes the input of the ith sample, $KL(\rho \parallel \widehat{\rho}_j)$ a sparse penalty term, which can learn relatively sparse features, ρ denotes a sparse penalty parameter, $\widehat{\rho}_j$ denotes the average activation level of hidden layer neuron j, $a_j^{(2)}(x^{(i)})$ denotes the output of the jth neuron in the hidden layer of the ith sample, $w_{ij}^{(k)}$ denotes the weight decay term.

Compared to single autoencoder, stacked autoencoder has stronger learning ability [37]. Therefore, stacked autoencoder is widely used in complicated problems. For example, Wei et al.

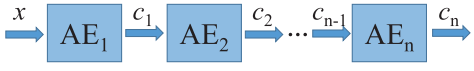

Fig. 2. An example of a stacked autoencoder.

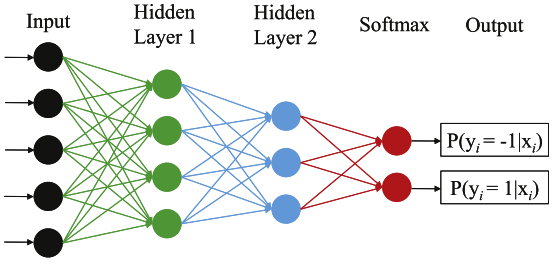

Fig. 3. Our proposed deep neural network classifier.

[38] proposed a deep learning based predictor by using stacked autoencoder. As shown in Fig. 2, a stacked autoencoder neural network is stacked by n autoencoders. For the raw data $\boldsymbol{x}$, we obtain the hidden layer representation $\boldsymbol{c}_1$ by AE_1, then put the representation $\boldsymbol{c}_1$ as the raw data of AE_2, until to the last layer. The whole network is trained through greedy layer-wise method. Based on the above analysis, we propose to use two sparse autoencoder and a softmax layer to form a deep neural network classifier as shown in Fig. 3. The input represents the top ranked features. The two hidden layer represent two autoencoder, which are trained in an unsupervised manner to obtain the hidden features and to

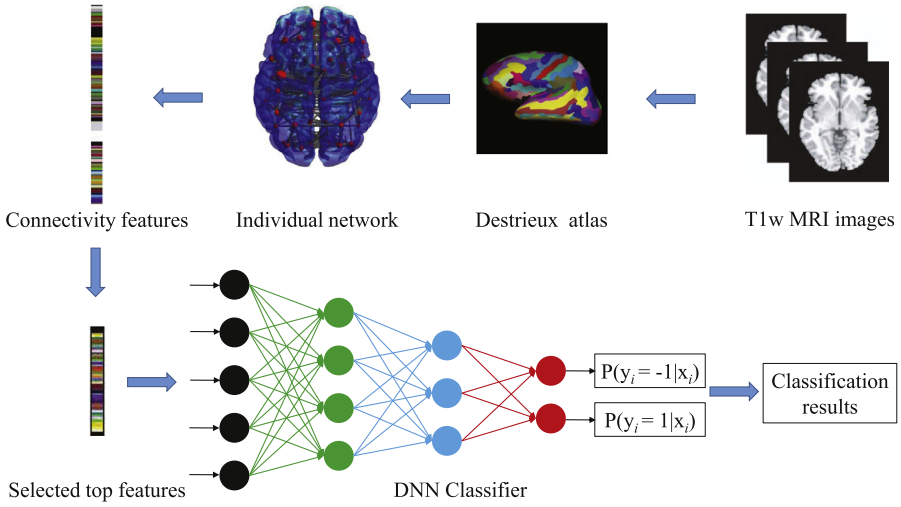

Fig. 4. An overall flowchart for ASD/TC classification in this study.

decrease the features dimension. The softmax layer is trained in a supervised manner with the features obtained from the hidden layer and label to perform ASD/TC classification. The output represents the class identified by the DNN classifier.

2.5. Evaluation

An overview of the proposed ASD/TC classification framework is depicted in Fig. 4. Firstly, we construct an individual network for each subject, and extract connectivity features of each pair of ROIs. Next, we use *F*-score to rank all features in descending order and select top features to identify subjects with ASD from TC subjects via a DNN classifier.

Four indicators including accuracy (ACC), sensitivity (SEN), specificity (SPE), and the area under receiver operating characteristic curve (AUC) are calculated to evaluate the performance of our proposed ASD/TC classification method. The accuracy, sensitivity and specificity can be formulated as follows:

$$Accuracy = \frac{TN + TP}{TN + FN + TP + FP} \tag{8}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{9}$$

$$Specificity = \frac{FN}{TN + FP} \tag{10}$$

where TP, FP, TN and FN are the number of true positive subjects, false positive subjects, true negative subjects and false negative subjects, respectively. The larger the AUC, the better the classification performance.

3. Results and discussion

3.1. Classification performance

We use a ten-fold cross-validation strategy and repeat twenty times to evaluate our proposed method. To avoid overfitting, firstly, all subjects for each classification are randomly equally partitioned into 10 subjects $\{S1, S2, \ldots, S10\}$ and $S1$ is selected to as testing set. Then, $\{S2, \ldots, S10\}$ as a whole data set is further randomly equally into 10 subsets, and one subset is randomly selected as validation set and the other 9 subsets are used to train a classifier. We have done a series of experiments based on different numbers of top ranked features as shown in Fig. 5. As can be seen from

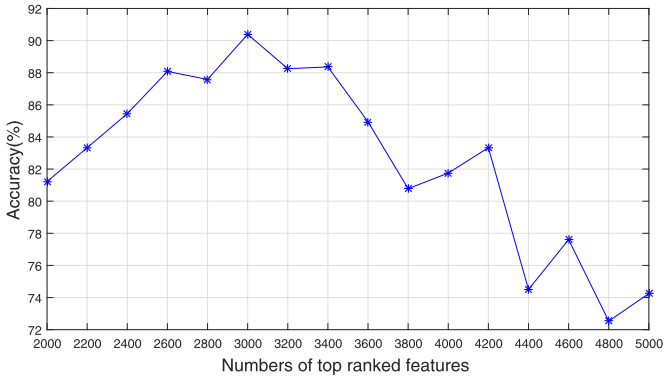

Fig. 5. The classification accuracy of different numbers of top ranked features.

Table 3
Comparison of different morphological representation for ASD/TC classification.

Type	ACC (%)	SEN (%)	SPE (%)	AUC
CT	67.58	56.66	75.46	0.7150
CTstd	69.17	57.94	77.33	0.7832
CGMV	90.39	84.37	95.88	0.9738

Fig. 5, when we select 3000 top features, our proposed method achieves the highest classification accuracy, and its accuracy, sensitivity, specificity and AUC are 90.39%, 84.37%, 95.88% and 0.9738, respectively.

3.2. Comparison with different morphological representation

In this section, we investigate the performance of different morphological features (CGMV, cortical thickness (CT) and standard deviation of cortical thickness (CTstd)) based on our proposed framework for ASD/TC classification. The experimental results are shown in Table 3. As can be seen from Table 3, CGMV outperforms other morphological features (CT and CTstd) in terms of ACC, SEN, SPE, AUC. Therefore, it is more reasonable to use CGMV for ASD/TC classification.

3.3. Comparison with existing classification methods

Recently, some existing methods have achieved relatively good results for ASD/TC classification. Katuwal et al. [39] investigated

Table 4
Comparison to existing methods using NYU dataset for ASD/TC classification.

Type	ACC (%)	SEN (%)	SPE (%)	AUC
Xiao et al. [21]	59.56	72.68	42.05	0.6226
Katuwal et al. [39]	74.46	63.97	82.62	0.8389
Our proposed	90.39	84.37	95.88	0.9738

Table 5
Comparison to existing methods using KKI dataset for ASD/TC classification.

Type	ACC (%)	SEN (%)	SPE (%)	AUC
Xiao et al. [21]	52.35	75.61	17.56	0.4626
Katuwal et al. [39]	71.60	38.88	95.25	0.8309
Our proposed	86.70	67.76	100.00	0.9803

the predictive power in ASD classification with brain morphological features and random forest. Xiao et al. [21] proposed to generate relatively stable predictive model based on anatomical brain features to perform ASD/TC classification. To demonstrate the superiority of our proposed method in ASD/TC classification, we compare the above two methods with the same dataset. The experimental results are shown in Table 4. As can be seen in Table 4, our proposed method is superior to the other two methods.

To further prove the superiority of our proposed method, we repeat the same experiment with T1w images from ABIDE I: Kennedy Krieger Institute (KKI) site for ASD/TC classification as shown in Table 5. As can be seen from Table 5, our proposed method is also better than the other two existing methods.

4. Conclusion

In this study, we propose an ASD aided diagnosis method based on deep neural network classifier. We firstly construct an individual network for each subject and extract gray matter volume as features. Then, we use *F*-score algorithm to rank all features in descend order. Afterwards, the top ranked features are applied to the DNN classifier to obtain the best classification accuracy. The experimental results illustration that our proposed method is an effective assistant diagnostic strategy in ASD/TC classification.

Acknowledgments

The authors would like to express their gratitude for the support from the National Natural Science Foundation of China under Grant Nos. 61672536, 61420106009 and 61622213.

References

[1] D.G. Amaral, C.M. Schumann, C.W. Nordahl, Neuroanatomy of autism, Trends Neurosci. 31 (3) (2008) 137–145.
[2] K.C. Turner, L. Frost, D. Linsenbardt, J.R. McIlroy, R.-A. Müller, Atypically diffuse functional connectivity between caudate nuclei and cerebral cortex in autism, Behav. Brain Funct. 2 (1) (2006) 34.
[3] S.J. Blumberg, M.D. Bramlett, M.D. Kogan, L.A. Schieve, J.R. Jones, M.C. Lu, Changes in Prevalence of Parent-Reported Autism Spectrum Disorder in School-Aged US Children: 2007 to 2011–2012, no. 65, US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2013.
[4] M. Langen, S. Durston, W.G. Staal, S.J. Palmen, H. van Engeland, Caudate nucleus is enlarged in high-functioning medication-naive subjects with autism, Biol. Psychiatry 62 (3) (2007) 262–266.
[5] C. Ecker, A. Marquand, J. Mourão-Miranda, P. Johnston, E.M. Daly, M.J. Brammer, S. Maltezos, C.M. Murphy, D. Robertson, S.C. Williams, et al., Describing the brain in autism in five dimensionsmagnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach, J. Neurosci. 30 (32) (2010) 10612–10623.
[6] Autism and Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators, Prevalence of autism spectrum disorder among children aged 8 years-Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2010, in: MMWR Surveill Summ., 63(2), Centers for Disease Control and Prevention (CDC), 2014, pp. 1–21.
[7] J. Liu, M. Li, W. Lan, F.-X. Wu, Y. Pan, J. Wang, Classification of Alzheimer's disease using whole brain hierarchical network, IEEE/ACM Trans. Comput. Biol. Bioinform. 15 (2) (2018) 624–632, doi:10.1109/TCBB.2016.2635144.
[8] J. Liu, J. Wang, Z. Tang, B. Hu, F.-X. Wu, Y. Pan, Improving Alzheimeres disease classification by combining multiple measures, IEEE/ACM Trans. Comput. Biol. Bioinform. (2017), doi:10.1109/TCBB.2017.2731849.
[9] J. Liu, M. Li, Y. Pan, F.-X. Wu, X. Chen, J. Wang, Classification of schizophrenia based on individual hierarchical brain networks constructed from structural MRI images, IEEE Trans. NanoBiosci. 16 (7) (2017) 600–608.
[10] J. Liu, X. Wang, X. Zhang, Y. Pan, X. Wang, J. Wang, MMM: classification of schizophrenia using multi-modality multi-atlas feature representation and multi-kernel learning, Multimed. Tools Appl. (2017), doi:10.1007/s11042-017-5470-7.
[11] N. Akshoomoff, C. Lord, A.J. Lincoln, R.Y. Courchesne, R.A. Carper, J. Townsend, E. Courchesne, Outcome classification of preschool children with autism spectrum disorders using MRI brain measures, J. Am. Acad. Child Adolesc. Psychiatry 43 (3) (2004) 349–357.
[12] Y. Jiao, R. Chen, X. Ke, K. Chu, Z. Lu, E.H. Herskovits, Predictive models of autism spectrum disorder based on brain regional cortical thickness, Neuroimage 50 (2) (2010) 589–599.
[13] C.-Y. Wee, P.-T. Yap, D. Shen, Prediction of Alzheimer's disease and mild cognitive impairment using cortical morphological patterns, Hum. Brain Mapp. 34 (12) (2013) 3411–3425.
[14] X. Han, J. Jovicich, D. Salat, A. van der Kouwe, B. Quinn, S. Czanner, E. Busa, J. Pacheco, M. Albert, R. Killiany, et al., Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer, Neuroimage 32 (1) (2006) 180–194.
[15] M. Nieuwenhuis, N.E. van Haren, H.E.H. Pol, W. Cahn, R.S. Kahn, H.G. Schnack, Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples, Neuroimage 61 (3) (2012) 606–612.
[16] M.N.I. Qureshi, J. Oh, D. Cho, H.J. Jo, B. Lee, Multimodal discrimination of schizophrenia using hybrid weighted feature concatenation of brain functional connectivity and anatomical features with an extreme learning machine, Front. Neuroinform. 11 (2017) 59.
[17] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297.
[18] C. Ecker, V. Rocha-Rego, P. Johnston, J. Mourao-Miranda, A. Marquand, E.M. Daly, M.J. Brammer, C. Murphy, D.G. Murphy, M.A. Consortium, et al., Investigating the predictive value of whole-brain structural mr scans in autism: a pattern classification approach, Neuroimage 49 (1) (2010) 44–56.
[19] W. Lan, M. Li, K. Zhao, J. Liu, F.-X. Wu, Y. Pan, J. Wang, LDAP: a web server for lncRNA-disease association prediction, Bioinformatics 33 (3) (2016) 458–460.
[20] W. Lan, J. Wang, M. Li, J. Liu, F.-X. Wu, Y. Pan, Predicting microRNA-disease associations based on improved microRNA and disease similarities, IEEE/ACM Trans. Comput. Biol. Bioinform. (2016), doi:10.1109/TCBB.2016.2586190.
[21] X. Xiao, H. Fang, J. Wu, C. Xiao, T. Xiao, L. Qian, F. Liang, Z. Xiao, K.K. Chu, X. Ke, Diagnostic model generated by MRI-derived brain features in toddlers with autism spectrum disorder, Autism Res. 10 (4) (2017) 620–630.
[22] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: Proceedings of Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
[23] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, A. Madabhushi, Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images, IEEE Trans. Med. Imaging 35 (1) (2016) 119–130.
[24] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[25] A.A. Cruz-Roa, J.E.A. Ovalle, A. Madabhushi, F.A.G. Osorio, A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2013, pp. 403–410.
[26] J. Liu, Y. Pan, M. Li, Z. Chen, L. Tang, C. Lu, J. Wang, Applications of deep learning to MRI images: a survey, Big Data Min. Anal. 1 (1) (2018) 1–18.
[27] B. Fischl, Freesurfer, Neuroimage 62 (2) (2012) 774–781.
[28] C. Destrieux, B. Fischl, A. Dale, E. Halgren, Automatic parcellation of human cortical Gyri and Sulci using standard anatomical nomenclature, Neuroimage 53 (1) (2010) 1–15.
[29] U. Braun, S.F. Muldoon, D.S. Bassett, On human brain networks in health and disease, in: eLS, Wiley, 2015, doi:10.1002/9780470015902.a0025783.
[30] J. Liu, M. Li, Y. Pan, W. Lan, R. Zheng, F.-X. Wu, J. Wang, Complex brain network analysis and its applications to brain disorders: a survey, Complexity (2017) 8362741, doi:10.1155/2017/8362741.
[31] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (Mar) (2003) 1157–1182.
[32] H. Liu, L. Yu, Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17 (4) (2005) 491–502.
[33] Y.-W. Chen, C.-J. Lin, Combining SVMs with various feature selection strategies, in: Feature Extraction, Springer, 2006, pp. 315–324.
[34] A. Ng, Sparse autoencoder, CS294A Lecture Notes, 72(2011), Stanford University, 2011, pp. 1–19.

[35] X. Lu, Y. Tsao, S. Matsuda, C. Hori, Speech enhancement based on deep denoising autoencoder., in: Proceedings of Interspeech, 2013, pp. 436–440.
[36] S. Rifai, P. Vincent, X. Muller, X. Glorot, Y. Bengio, Contractive auto-encoders: Explicit invariance during feature extraction, in: Proceedings of the 28th International Conference on International Conference on Machine Learning, Omnipress, 2011, pp. 833–840.
[37] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, Greedy layer-wise training of deep networks, in: Proceedings of Advances in Neural Information Processing Systems, 2007, pp. 153–160.
[38] L. Wei, Y. Ding, R. Su, J. Tang, Q. Zou, Prediction of human protein subcellular localization using deep learning, J. Parallel Distrib. Comput. 117 (2017) 212–217.
[39] G.J. Katuwal, N.D. Cahill, S.A. Baum, A.M. Michael, The predictive power of structural MRI in autism diagnosis, in: Proceedings of 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015, pp. 4270–4273.

Yazhou Kong is a MS candidate in School of Information Science and Engineering, Central South University, Changsha, Hunan, PR China. His current research interests include medical image analysis and machine learning.

Jianliang Gao received the Ph.D. degree in Computer System Structure from Institute of Computing Technology, Chinese Academy of Sciences, China, in 2011. He is currently an associate professor at the School of Information Science and Engineering, Central South University, Changsha, Hunan, PR China. His main research interests include big data process technology, machine learning and large scale image data process.

Yunpei Xu is a MS candidate in School of Information Science and Engineering, Central South University, Changsha, Hunan, PR China. His current research interests include medical image analysis and machine learning.

Yi Pan is a Regents' Professor of Computer Science and an Interim Associate Dean and Chair of Biology at Georgia State University, USA. Dr. Pan joined Georgia State University in 2000 and was promoted to Full Professor in 2004, named a Distinguished University Professor in 2013 and designated a Regents' Professor (the highest recognition given to a faculty member by the University System of Georgia) in 2015. He served as the Chair of Computer Science Department from 2005–2013. He is also a visiting Changjiang Chair Professor at Central South University, China. Dr. Pan received his B.Eng. and M.Eng. degrees in computer engineering from Tsinghua University, China, in 1982 and 1984, respectively, and his Ph.D. degree in computer science from the University of Pittsburgh, USA, in 1991. His profile has been featured as a distinguished alumnus in both Tsinghua Alumni Newsletter and University of Pittsburgh CS Alumni Newsletter. Dr. Pan's research interests include parallel and cloud computing, wire-less networks, and bioinformatics. Dr. Pan has published more than 330 papers including over 180 SCI journal papers and 60 IEEE/ACM Transactions papers. In addition, he has edited/authored 40 books. His work has been cited more than 8800 times. Dr. Pan has served as an editor-in-chief or editorial board member for 15 journals including 7 IEEE Transactions. He is the recipient of many awards including IEEE Transactions Best Paper Award, 4 other international conference or journal Best Paper Awards, 4 IBM Faculty Awards, 2 JSPS Senior Invitation Fellowships, IEEE BIBE Outstanding Achievement Award, NSF Research Opportunity Award, and AFOSR Summer Faculty Research Fellowship. He has organized many international conferences and delivered keynote speeches at over 50 international conferences around the world.

Jianxin Wang received the BEng and MEng degrees in computer engineering from Central South University, China, in 1992 and 1996, respectively, and the Ph.D. degree in computer science from Central South University, China, in 2001. He is the Chair of and a Professor in Department of Computer Science, Central South University, Changsha, Hunan, PR China. His current research interests include algorithm analysis and optimization, parameterized algorithm, Bioinformatics and computer network.

Jin Liu received the Ph.D. degree in computer science from Central South University, China, in 2017. He is currently a Lecturer at the School of Information Science and Engineering, Central South University, Changsha, Hunan, PR China. His current research interests include medical image analysis, bioinformatics and machine learning.