Neurocomputing xxx (xxxx) xxx

Contents lists available at ScienceDirect

Neurocomputing

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

MAGE: Automatic diagnosis of autism spectrum disorders using multi-atlas graph convolutional networks and ensemble learning

Yufei Wang [a,c], Jin Liu [a,*], Yizhen Xiang [a], Jianxin Wang [a], Qingyong Chen [b,c], Jing Chong [b,c]

[a] Hunan Provincial Key Lab on Bioinformatics, School of Computer Science and Engineering, Central South University, Changsha 410083, China
[b] China mobile(Chengdu)Industrial Research Institute, Chengdu610041, China
[c] Mobile Health Ministry of Education-China Mobile Joint Laboratory, Changsha 410008, China

ARTICLE INFO

Article history:
Received 23 February 2020
Revised 14 June 2020
Accepted 14 June 2020
Available online xxxx

Keywords:
Autism spectrum disorders
Automatic diagnosis
Functional connectivity
Multi-atlas
Graph convolutional networks
Ensemble learning

ABSTRACT

Currently, it is still a great challenge in clinical practice to accurately diagnose autism spectrum disorder (ASD). To address this challenge, in this study we propose a method for automatic diagnosis of ASD based on multi-atlas graph convolutional networks and ensemble learning. Firstly, we extract multiple feature representations based on functional connectivity (FC) of different brain atlases from fMRI data of each subject. Then, to obtain the features that are more helpful for ASD automatic diagnosis, we propose a multi-atlas graph convolutional network method (MAGCN). Finally, to combine different feature representations, we propose a stacking ensemble learning method to perform the final ASD automatic diagnostic task. Our proposed method is evaluated on 949 subjects (including 419 subjects with ASD and 530 subjects with typical control (TC)) from the Autism Brain Imaging Data Exchange (ABIDE). Experimental results show that our proposed method achieves an accuracy of 75.86% and an area under the receiver operating characteristic curve (AUC) of 0.8314 for automatic diagnosis of ASD. In addition, compared with some methods published in recent years, our proposed method obtains the best performance of ASD diagnosis. Overall, our proposed method is effective and promising for automatic diagnosis of ASD in clinical practice.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

Autism Spectrum Disorder (ASD) [1] is a pervasive developmental disorder of the central nervous system, which is caused by the stimulation of various environmental factors and genetic factors. People with ASD have the following symptoms: severe loneliness, lack of emotional response, language development disorders and repetitive behaviors. These symptoms can seriously affect the physical and mental health of patients, which brings a heavy burden on their entire family. Nowadays, diagnosing ASD mainly relies on the patient's behavioral symptoms. This approach requires professional doctors to diagnose and the diagnosing performance heavily depends on the level and ability of the medical profession. Therefore, it is necessary to develop an objective and accurate ASD diagnosis method [2].

Nowadays, applying magnetic resonance imaging (MRI) to brain research has become a hot topic in various brain disease diagnosis [3], such as ASD [4,5], schizophrenia [6,7] and Alzheimer's disease [8–10]. Such technology includes structural MRI (sMRI) and functional MRI (fMRI). Compared with sMRI, since fMRI could capture both time and space resolving information, it has gained more attention in the study of brain dysfunction disorders [11,12].

In the past decade, traditional machine learning methods have been widely used to identify patients and reveal discriminant biomarkers based on fMRI data. Since functional connectivity (FC) reflects the interrelationship and temporal connectivity between different brain regions, it is usually extracted as features in ASD classification tasks. FC can be calculated via the time-series of regions of interest (ROIs) from manual segmentation or existing brain atlases. Abraham et al. [13] extracted FC as features from a data-driven atlas based on multi-subject dictionary learning (MSDL) strategy and then applied support vector machine (SVM) as a classifier to perform ASD classification. They also compared another two atlases (i.e., Harvard-Oxford atlas, Yeo atlas) and three data-driven atlas strategies (i.e., Ward strategy, ICA strategy and K-means strategy). Chen et al. [14] calculated FC in the two frequency bands (slow-4: 0.027–0.073 Hz and slow-5: 0.01–0.027 Hz) based on Dosenbach atlas to diagnose ASD. Yahata et al. [15] used FC based on the Brainvisa Sulci Atlas [16] as features and L1-norm regularized sparse canonical correlation analysis (L1-SCCA) [17] as a classifier to perform ASD identification. Jahedi et al. [18]

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

https://doi.org/10.1016/j.neucom.2020.06.152
0925-2312/© 2021 Elsevier B.V. All rights reserved.

Please cite this article as: Y. Wang, J. Liu, Y. Xiang et al., MAGE: Automatic diagnosis of autism spectrum disorders using multi-atlas graph convolutional networks and ensemble learning, Neurocomputing, https://doi.org/10.1016/j.neucom.2020.06.152

extracted FC based on the brain atlas of Power et al. [19] for ASD classification.

In recent years, deep learning methods have become popular and show great potential in medical domain. Compared with traditional machine learning methods, deep learning methods could get more hidden high level features via neural network. Heinsfeld et al. [20] proposed a method that used stacked denoising autoencoder to do pretraining and a fully connected neural network to learn deep feature representations from Harvard-Oxford (HO) atlas FC for ASD identification. Sherkatghanad [21] extracted FC based on CC400 functional parcellation atlas as features and used convolutional neural networks to detect ASD. In addition, to explore interactions between subjects, an increasing research associated with graph convolutional network (GCN) [22] has attracted great attention in ASD identification task. GCN provides a more powerful ability of aggregating information and extracting deep feature representation based on graph. Parisot et al. [23] introduced GCN to the medical domain which utilized HO atlas FC and non-imaging features to construct graph for ASD classification. Kazi et al. [24] proposed an InceptionGCN model which aggregated different filter kernel sizes for ASD prediction. Li et al. [25] applied the graph embedding and infomax loss to detect ASD and typical control (TC). However, these studies were only based on single brain atlas. Yao et al. [26] suggested that since there were great discrepancies between different brain atlases, these different brain atlases may provide complementary information with each other. So far, although some results have been achieved for ASD/TC classification, how to fully make use of complementary information of multiple brain atlases to improve the performance of ASD/TC classification is still a challenge problem.

Based on the above analysis, to improve the performance of ASD automatic diagnosis, in this study we propose a new method based on multi-atlas graph convolutional networks and ensemble learning, which is denoted as MAGE. We first extract multiple feature representations based on FC of different brain atlases from fMRI data of each subject. After that, recursive feature elimination (RFE) is used to perform feature selection. Then, we propose a multi-atlas graph convolutional network method (MAGCN) to obtain features that are more helpful for ASD automatic diagnosis. Afterward, we propose a stacking ensemble learning method to combine different feature representations. Finally, we apply a Ridge classifier to perform the final ASD automatic diagnostic task. Our proposed method is evaluated on 949 subjects (including 419 subjects with ASD and 530 subjects with TC) from the Autism Brain Imaging Data Exchange (ABIDE) [27].

2. Methods

The overall flowchart of our proposed method for automatic diagnosis of ASD is shown in Fig. 1. As can be seen from Fig. 1, our proposed method consists of three main steps: 1) image preprocessing and feature representations, 2) multi-atlas GCNs, and 3) classification. Next, we will describe our proposed method in detail.

2.1. Image preprocessing and feature representations

As can be seen from Fig. 1, the fMRI image data of each experimental subject used in this study is provided by the Autism Brain Imaging Data Exchange (ABIDE) [27]http://fcon_1000.projects.nitrc.org/indi/abide/. For the fMRI data of each subject in the ABIDE, the Preprocessed Connectomes Project (PCP) has preprocessed the fMRI data of each subject into time-series data of each brain region from different brain atlases. For more details about the PCP, please visit the link:http://preprocessed-connectomes-project.org/abide/. In this study, we select time-series data of brain regions as our image experimental data using Configurable Pipeline for the Analysis of Connectomes(CPAC) [28].

The ABIDE contains 1112 subjects including 539 subjects with ASD and 573 subjects with TC. We exclude the subjects with missing time-series of brain regions after preprocessing by CPAC tool. In this study, we select 949 subjects from 17 international sites as experimental subjects which includes 419 subjects with ASD and 530 subjects with TC. Demographic information of these subjects in this study is summarized in Table 1.

Next, based on the time-series data of each brain region, we propose to calculate the functional connectivity (FC) between paired brain regions using Pearson correlation coefficient (PCC) as follows:

$$FC(r_i, r_j) = PCC(r_i, r_j) = \frac{E(r_i r_j) - E(r_i)E(r_j)}{\sqrt{E(r_i^2) - E^2(r_i)}\sqrt{E(r_j^2) - E^2(r_j)}} \quad (1)$$

where r_i and r_j are the time series of brain regions i and j, respectively; $E(\cdot)$ is the mathematical expectation. For example, the Automated Anatomical Labeling (AAL) [29] atlas has 116 brain regions. We calculate the FC between each pair of brain regions based on the time series of each subject. After this step, we can obtain a 116×116 matrix for each subject. Then the lower triangle of the matrix for each subject is used as feature representation. Since the feature representation obtained above is high-dimensional, which may contain a lot of redundant or unrelated features for ASD diagnosis, we propose to adopt support vector machine-recursive feature elimination (SVM-RFE) method [30] to perform feature selection on the feature representation. Following the work of Parisot et al. [23], we also finally select the top 2000 features as feature representation of each subject.

In this study, in addition to obtaining feature representation of each subject based on AAL atlas, we adopt other five commonly used brain atlases to extract FC between paired brain regions. These brain atlases include Eickhoff-Zilles (EZ), Harvard-Oxford (HO), Talaraich and Tournoux (TT), Craddock 200 (CC200) and Dosenbach (DOH). For more details about the above six brain atlases, please visit this link:http://preprocessed-connectomes-project.org/abide/Pipelines.html.

By performing the above step, we can obtain six selected feature representations for each experimental subject, which are denoted as $F_{AAL}, F_{EZ}, F_{HO}, F_{TT}, F_{CC}$ and F_{DOH}, respectively.

2.2. Multi-atlas graph convolutional networks

Currently, graph-based mining technology has been widely used in feature selection [31–34]. To further obtain features that are more helpful for ASD diagnosis from the above six feature representations, we propose a multi-atlas graph convolutional network method (MAGCN) as shown in Fig. 1. This method first uses the relationship between pairs of subjects to construct multiple graphs, and then applies the graph convolution operation to these graphs. The details of this method are as follows.

2.2.1. Graph construction

In general, a graph can be defined as $G = (V, E, W)$, where V is the set of vertices, E is the set of edges (i.e., connectivity of paired vertices), and W is an adjacency matrix describing the graph's connectivity. In this study, we construct six graphs. The vertices of these six graphs are composed of all subjects, and the adjacency matrix of these six graphs is the same. The only difference among these six graphs is that the vertices of each graph are represented differently.

In a recently published study on the diagnosis of ASD using GCN model, Parisot et al. [23] proposed an adjacency matrix calculation

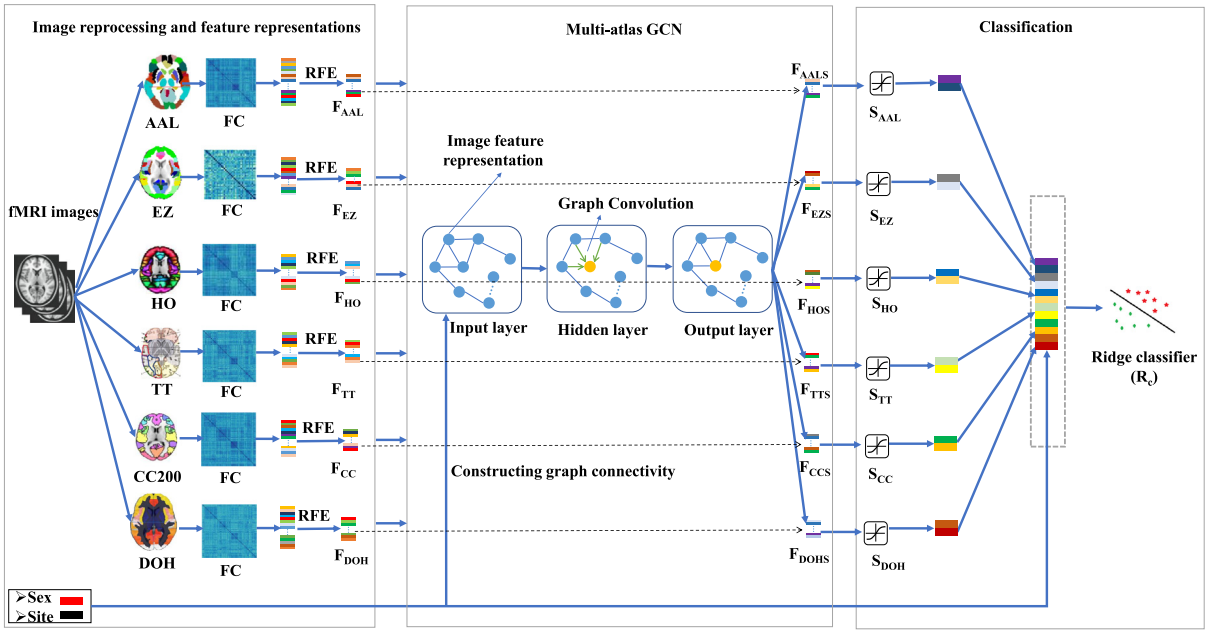

Fig. 1. An overall flowchart for ASD automatic diagnosis based on multi-atlas GCNs and ensemble learning.

method based on image and non-image feature representations. The adjacency matrix $W \in \mathbb{R}^{N\times N}$, where N is the number of subjects, was defined as follows:

$$W_{a,b} = exp(-\frac{\|L_a - L_b\|_2^2}{2\sigma^2})\sum_{h=1}^{H} f_h(X_h(a), X_h(b)) \tag{2}$$

where $W_{a,b}$ is the weight of the connectivity between two subjects a and b; L_a and L_b are the image feature representations of subjects a and b, respectively; σ is the width of the Gaussian kernel; H is the number of non-image feature representations (such as sex and site); $X_h(a)$ and $X_h(b)$ are the h-th non-image feature representation of subjects a and b, respectively; $f_h(X_h(a), X_h(b))$ is a distance measurement function between subjects a and b in the h-th non-image feature representation. It is worth mentioning that in the work of Parisot et al. [23], when the h-th non-image feature representation is sex or site, $f_h(X_h(a), X_h(b))$ is the Kronecker delta function ($\delta(\cdot,\cdot)$) as follows:

$$f_h(X_h(a), X_h(b)) = \delta(X(a), X(b)) = \begin{cases} 1 \ if \ X(a) = X(b) \\ 0 \ if \ X(a) \neq X(b) \end{cases} \tag{3}$$

As can be seen from Eqs. (2) and (3), the connectivity of the graph mainly depends on the distance measurement between subjects based on non-image feature representations (such as sex and site). Following the work of Kipf et al. [35], we can learn that the convolution operation relies on the structure of graph and does not require any feature representation of vertices. Therefore, to reduce the computational cost, in this study we propose to use only non-image feature representations to compute the adjacency matrix as follows:

$$W_{a,b} = \sum_{h=1}^{H} f_h(X_h(a), X_h(b)) \tag{4}$$

After the above step, we can obtain six different graphs with different vertex feature representations (i.e., $F_{AAL}, F_{EZ}, F_{HO}, F_{TT}, F_{CC}$ and F_{DOH}) and the same adjacency matrix, which are denoted as $G_{AAL}, G_{EZ}, G_{HO}, G_{TT}, G_{CC}$ and G_{DOH}, respectively. It is worth mentioning that since Parisot et al. [23] used the gender and site of each subject to construct the graph to obtain the best ASD diagnostic accuracy, in this study we only use the gender and site of each subject to construct graphs.

2.2.2. GCN models

In order to further obtain more useful feature representations of each subject for ASD diagnosis, in this study we apply graph convolution network (GCN) model to each of the above six graphs (i.e., $G_{AAL}, G_{EZ}, G_{HO}, G_{TT}, G_{CC}$ and G_{DOH}). The GCN model is briefly described as follows.

The key concept of GCN is to define graph convolution based on Laplacian matrix and graph Fourier transform in the spectral domain. Suppose there is a graph G with N vertices whose adjacency matrix is $W \in \mathbb{R}^{N\times N}$ and its diagonal degree matrix is D. The normalized Laplacian matrix (L) of the graph can be defined as follows:

$$L = I - D^{-\frac{1}{2}}WD^{-\frac{1}{2}} \tag{5}$$

where I is an identity matrix. As L is a real symmetric positive semidefinite matrix, it can be decomposed as follows:

$$L = UVU^{\mathrm{T}} \tag{6}$$

where U is an orthogonal matrix, and V is a diagonal matrix of eigenvalues. Suppose x is a feature representation defined on the vertices of the graph G, the graph convolution can be defined as follows:

$$y = g_\theta(L)x = g_\theta(UVU^{\mathrm{T}})x = Ug_\theta(V)U^{\mathrm{T}}x \tag{7}$$

where y is the feature representation corresponding to $x, g_\theta(\cdot)$ is a filter which can be regarded as a function of V, and θ is a vector of Fourier coefficients to be learned. Recently, to address the computational complexity problem, Kipf and Welling [35] proposed a simplified filter, which is approximated by a truncated expansion in terms of Chebyshev polynomials. In this study, the graph convolution is defined as follows:

$$H^{l+1} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{W}\tilde{D}^{-\frac{1}{2}}H^l\Theta^l) \tag{8}$$

where $\tilde{W} = W+I$, H^l is the feature representation of the l- hidden

layer, and Θ^l is the weight matrix of the l-th hidden layer.

After performing the GCN models based on the above six vertex feature representations, we can obtain six deep feature representations for each subject, which are denoted as $F_{AALS}, F_{EZS}, F_{HOS}, F_{TTS}, F_{CCS}$ and F_{DOHS}, respectively.

2.3. Classification

In order to use the above six deep feature representations more effectively to perform ASD automatic diagnosis, we propose an ensemble learning method based on a stacking strategy [36] as shown in Fig. 1. From Fig. 1, we can observe that the stacking ensemble learning method includes two-step ASD/TC classification, and is described in detail below.

Step 1 : We propose to use a softmax regression method [37] to perform the ASD/TC classification task. In this step, $F_{AALS}, F_{EZS}, F_{HOS}, F_{TTS}, F_{CCS}$ and F_{DOHS} are used as the inputs to the corresponding softmax classifiers (i.e., $S_{AAL}, S_{EZ}, S_{HO}, S_{TT}, S_{CC}$ and S_{DOH}) to perform ASD/TC classification task, respectively, as shown in Fig. 1. After this step, for each classifier, we can not only obtain the ASD/TC classification performance of the feature representation of each corresponding input, but also obtain a 2-dimensional feature representation of each subject, which consists of the probability that this subject is positive and the probability that it is negative.

Step 2 : We first propose to use a stacking strategy to concatenate the six 2-dimensional feature representations of each subject obtained in step 1 to a 12-dimensional feature representation. Then, considering the role of non-image features (such as sex and site) in the diagnosis of ASD, we again use a stacking strategy to concatenate the 12-dimensional feature representation obtained above and the non-image feature representation for each subject. From Fig. 1, we can observe that in this study we only use the two non-image features of each subject's sex and site. Finally, a 14-dimensional feature representation of each subject can be obtained and these features are used as the input of a Ridge classifier [38] (i.e., R_c) to perform the final ASD/TC classification task.

3. Experiments and results

3.1. Experimental settings

In order to get an unbiased estimation and more robust performance for ASD diagnosis, in this study we adopt a 10-fold cross validation strategy [39] and repeat 5 experiments. Mean classification accuracy (ACC), sensitivity (SEN), and specificity (SPE) of the five experiments for ASD diagnosis are used as metrics, which are formulated as

$$ACC = \frac{1}{5}\sum_{i=1}^{n}\frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \tag{9}$$

$$SEN = \frac{1}{5}\sum_{i=1}^{n}\frac{TP_i}{TP_i + FN_i} \tag{10}$$

$$SPE = \frac{1}{5}\sum_{i=1}^{n}\frac{TN_i}{TN_i + FP_i} \tag{11}$$

where TP_i, FP_i, TN_i and FN_i are the number of correctly classified true positive subjects, the number of incorrectly classified false positive subjects, the number of correctly classified true negative subjects, and the number of incorrectly classified false negative subjects in the i-th experiment, respectively. In addition, we calculate the area under receiver operating characteristic curve (AUC) to evaluate the overall performance of our proposed method in the diagnosis of ASD.

In addition, the hyperparameters in our proposed method are set as follows: the number of selected features based on RFE is 2000, the dropout rate is 0.3, the learning rate is 0.005, the number of hidden layer is 1, the number of hidden neurons is 16, the number of epochs is 200, the weight decay is 0.00005, the loss function is cross entropy error function. It is worth mentioning that the experiments in this study are done on a GeForce RTX 2080 with 128 RAM server.

3.2. Experimental results

In this section, to verify the effectiveness of our proposed method, we compare the result of R_c using ensemble learning method with another six results of single feature sets. We present the performance of different softmax classifiers (i.e., $S_{AAL}, S_{EZ}, S_{HO}, S_{TT}, S_{CC}, S_{DOH}$ and R_c) in our proposed method for ASD diagnosis as shown in Table 2.

From Table 2, we observe that the final classifier R_c in our proposed method achieves 75.86% ACC, 79.24% SEN, 71.53% SPE and 0.8314 AUC for ASD diagnosis. The results show that our proposed method is effective for ASD automatic diagnosis. Also, we observe that the performance (including ACC, SEN, SPE and AUC) of R_c based on multi-atlas feature representation is better than that of other classifiers (i.e., $S_{AAL}, S_{EZ}, S_{HO}, S_{TT}, S_{CC}$ and S_{DOH}) based on single-atlas feature representation. This comparative result shows that our proposed stacking ensemble learning method can effectively integrate different feature representations from different

Table 1
Demographic information of these subjects in this study.

Site	ASD/TC	Age (Mean(STD))	Male/Female	Time-series length
CALTECH	14/18	27.36(10.62)	25/7	146
CMU	13/13	26.69(5.77)	20/6	236/316
KKI	18/28	10.02(1.25)	34/12	152
LEUVEN	20/34	17.81(4.74)	46/8	246
MAX MUN	17/28	25.42(11.13)	42/3	116/196
NYU	60/100	15.51(6.66)	125/35	176
OHSU	9/14	10.87(1.80)	23/0	78
OLIN	18/15	16.67(3.49)	29/4	206
PITT	24/27	18.89(6.91)	45/6	196
SBL	12/15	32.85(6.14)	27/0	196
SDSU	9/22	14.32(1.88)	24/7	176
STANFORD	17/20	9.89(1.59)	30/7	176
TRINITY	19/25	16.90(3.49)	44/0	146
UCLA	46/44	12.90(2.12)	79/11	116
UM	58/74	14.09(3.25)	106/26	296
USM	42/25	22.74(8.57)	67/0	236
YALE	23/28	12.79(2.89)	37/14	196
Total	419/530	16.88(7.76)	803/146	–

Table 2
The performance of different classifiers in our proposed method for ASD diagnosis.

Classifiers	ACC(%)	SEN(%)	SPE(%)	AUC
S_{AAL}	70.71	73.21	67.54	0.7821
S_{EZ}	71.97	74.15	69.21	0.7795
S_{HO}	71.76	74.72	68.02	0.7794
S_{TT}	71.75	77.74	64.20	0.7773
S_{CC}	72.50	76.60	67.30	0.7935
S_{DOH}	70.18	70.00	70.41	0.7614
R_c	**75.86**	**79.24**	**71.53**	**0.8314**

The significance of bold in the Tables indicates that the result in the corresponding metric is the best.

Table 3
The performance of different adjacency matrix calculation methods for ASD diagnosis.

Adjacency matrix calculation	ACC(%)	SEN(%)	SPE(%)	AUC
Eq. (2)	75.80	79.05	71.63	0.8274
Eq. (4)	75.86	79.24	71.53	0.8314

brain atlases. It is worth noting that the AUC of our proposed method is greater than 0.8, which shows that our proposed method has good robustness for ASD automatics diagnosis.

In order to prove the fact that our proposed method is related to the graph structure, we also have done a comparative experiment with the adjacency matrix calculation method of Eq. (2). The comparative experimental results are shown in Table 3.

As can be seen from Table 3, the diagnostic performance of ASD obtained by using these two adjacency matrix calculation methods in our proposed method is similar. This comparative result proves that our proposed method is more affected by the structure of the graph than the weight of the graph, and it effectively reduces the computational cost of our proposed method.

4. Discussion

4.1. Comparison with different classifiers

In this subsection, in order to demonstrate the superiority of Ridge classifier used in our proposed method for ASD/TC classification, we compare several typical classifiers including linear support vector machine (L-SVM) [40], Gaussian kernel support vector machine (G-SVM) [40], Gradient Boosting Decision Tree (GBDT) [41], XGboost [42] and Multi-layer Perceptron (MLP) [43]. For this purpose, we have done a series of experiments using these five different classifiers in our proposed method. It is worth mentioning that in general, before using SVM as a classifier for classification, a feature selection method needs to be used to select the optimal subset as the input of SVM to obtain better classification results. Therefore, before using L-SVM and G-SVM in our proposed method for ASD/TC classification, we use minimum Redundancy Maximum Relevance (mRMR) [44,45] to select the optimal feature subset as the input of these two classifiers. In this study, these five classifiers with the default hyperparameters are implemented by the Scikit-learn library [46]. Experimental results using our proposed method with different classifiers for ASD/TC classification are shown in Table 4.

As can be seen from Table 4, the ACC, SPE and AUC are obtained by using Ridge classifier in our proposed framework for ASD/TC classification. Although the best SEN is not obtained by using Ridge classifier (79.24%) in our proposed framework for ASD/TC classification, it is also close to the best SEN (79.51%). The comparative results further demonstrate that our proposed method is effective for ASD/TC classification.

4.2. Comparison with existing methods

In this subsection, in order to verify the superiority of our proposed method, we compare some existing state-of-the-art diagnostic methods for ASD. For this purpose, we have done a series of experiments based on five recently published diagnostic methods for ASD including three traditional machine learning methods [14,13,18] and two deep learning methods [23,20]. These five methods have briefly been described in Section Introduction. For fair comparisons, the experimental data of these five methods and our proposed method is the same. The experimental results of these five existing methods and our proposed method for ASD/TC classification are shown in Table 5.

As can be seen from Table 5, our proposed method achieves the best classification performance which verifies the superiority of our proposed method to some extent. The comparative results further demonstrate that our proposed method is effective and has certain advantages for ASD automatic diagnosis.

From Table 5, we also observe that deep learning methods (i.e., [23,20] and MAGE) are generally superior to traditional machine learning methods (i.e., [14,13,18]) for ASD/TC classification. Although deep learning methods use a large number of parameters and require superior computing equipment, compared with traditional machine learning methods, deep learning methods obtain the best ASD classification performance. Furthermore, with the continuous improvement and optimization of computing equipment performance, the computing time of deep learning model with a large number of parameters will also become shorter.

4.3. Important brain regions

In this study, in addition to using our proposed method to perform ASD diagnosis, we also analyze the above six feature representations (i.e., $F_{AAL}, F_{EZ}, F_{HO}, F_{TT}, F_{CC}$ and F_{DOH}) aimed at finding brain regions that are highly related with ASD. We consider these brain regions as important brain regions in the diagnosis of ASD.

To achieve this, we map the above six feature representations to the brain regions of the corresponding brain atlas. After this step, we find that these six feature representations have four common brain regions including angular gyrus, precentral gyrus, precuneus and thalamus. Through literature reading, we found that these four brain regions have been published by researchers and related to ASD. For example, Monk et al. [47] suggested that angular gyrus is highly associated with ASD symptoms; Nebel et al. [48] found

Table 4
The performance of different classifiers in our proposed method for ASD/TC classification.

Classifiers	ACC(%)	SEN(%)	SPE(%)	AUC
L-SVM	75.80	**79.51**	71.49	0.8221
G-SVM	69.93	75.53	62.80	0.7681
GBDT	73.97	79.43	66.99	0.7321
XGboost	70.49	76.60	62.74	0.6967
MLP	75.18	79.30	67.47	0.8223
Ridge	**75.86**	79.24	**71.53**	**0.8314**

The significance of bold in the Tables indicates that the result in the corresponding metric is the best.

Table 5
The performance of different methods for ASD/TC classification.

Method	ACC(%)	SEN(%)	SPE(%)	AUC	No.parameters
Chen et al., 2016 [14]	69.77	72.16	63.72	0.6639	< 100
Abraham et al., 2017 [13]	65.96	68.55	63.17	0.6717	< 100
Jahedi et al., 2017 [18]	66.90	71.43	61.78	0.6724	< 100
Parisot et al., 2018 [23]	71.86	75.55	67.43	0.7864	~1,700,000
Heinsfeld et al., 2018 [20]	71.32	78.17	62.59	0.7641	~1,600,000
MAGE	**75.86**	**79.24**	**71.53**	**0.8314**	~2,000,000

The significance of bold in the Tables indicates that the result in the corresponding metric is the best.

that precentral gyrus is related to the severity of ASD traits. Jiang et al. [49] observed some significant interactions with left precuneus in subjects with ASD; Schuetze et al. [50] suggested that there are some abnormalities in the thalamus between ASD and TC. Although these existing studies have found that the four brain regions mentioned above are related to ASD, these four brain regions were found separately.

In this study, we find four important brain regions related to ASD at the same time. The result of this analysis once again proves that our proposed method is effective. In a word, our proposed method can not only obtain an accuracy of greater than 75%, but also find some imaging markers related to ASD.

5. Conclusion

In this study, we propose an automatic diagnosis of autism spectrum disorders method using multi-atlas graph convolutional networks and ensemble learning (MAGE). Experimental results demonstrate that our proposed framework is effective for ASD automatic diagnosis. Also, we find four brain regions highly related to ASD. To sum up, our proposed method is not only effective and promising for automatic diagnosis of ASD in clinical practice, but also paves the way to discriminative imaging markers for automatic diagnosis of ASD.

CRediT authorship contribution statement

Yufei Wang: Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Writing - original draft, Writing - review & editing. **Jin Liu:** Conceptualization, Methodology, Writing - original draft. **Yizhen Xiang:** Validation. **Jianxin Wang:** Conceptualization, Methodology. **Qingyong Chen:** Supervision. **Jing Chong:** Supervision.

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.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China under Grant No.61802442, No.61877059, the Natural Science Foundation of Hunan Province under Grant No.2019JJ50775, No.2018JJ2534, the 111 Project (No.B18059), the Hunan Provincial Science and Technology Program (2018WK4001), the Mobile Health Ministry of Education-China Mobile Joint Laboratory (2020MHL02003), the Fundamental Research Funds for the Central Universities of Central South University(2019zzts590).

References

[1] A.P. Association, et al., Diagnostic and statistical manual of mental disorders (DSM-5), American Psychiatric Pub, 2013..
[2] B.E. Yerys, B.F. Pennington, How do we establish a biological marker for a behaviorally defined disorder? autism as a test case, Autism Research 4 (4) (2011) 239–241.
[3] 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 Mining and Analytics 1 (1) (2018) 1–18.
[4] Y. Kong, J. Gao, Y. Xu, Y. Pan, J. Wang, J. Liu, Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier, Neurocomputing 324 (2019) 63–68.
[5] L. Xu, Y. Liu, J. Yu, X. Li, X. Yu, H. Cheng, J. Li, Characterizing autism spectrum disorder by deep learning spontaneous brain activity from functional near-infrared spectroscopy, Journal of Neuroscience Methods 331 (2020) 108538.
[6] 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, Multimedia Tools and Applications 77 (22) (2018) 29651–29667.
[7] Y. Xiang, J. Wang, G. Tan, F.-X. Wu, J. Liu, Schizophrenia identification using multi-view graph measures of functional brain networks, Frontiers in Bioengineering and Biotechnology 7 (2020) 479.
[8] 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 Transactions on Computational Biology and Bioinformatics 15 (2) (2018) 624–632.
[9] J. Liu, J. Wang, Z. Tang, B. Hu, F.-X. Wu, Y. Pan, Improving alzheimer's disease classification by combining multiple measures, IEEE/ACM Transactions on Computational Biology and Bioinformatics 15 (5) (2018) 1649–1659.
[10] J. Liu, Y. Pan, F.-X. Wu, J. Wang, Enhancing the feature representation of multi-modal mri data by combining multi-view information for mci classification, Neurocomputing 400 (2020) 322–332.
[11] B.H. Silverstein, S.L. Bressler, V.A. Diwadkar, Inferring the dysconnection syndrome in schizophrenia: Interpretational considerations on methods for the network analyses of fmri data, Frontiers in Psychiatry 7 (2016) 132.

[12] A. Khazaee, A. Ebrahimzadeh, A. Babajani-Feremi, Identifying patients with alzheimer's disease using resting-state fmri and graph theory, Clinical Neurophysiology 126 (11) (2015) 2132–2141.
[13] A. Abraham, M.P. Milham, A. Di Martino, R.C. Craddock, D. Samaras, B. Thirion, G. Varoquaux, Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example, NeuroImage 147 (2017) 736–745.
[14] H. Chen, X. Duan, F. Liu, F. Lu, X. Ma, Y. Zhang, L.Q. Uddin, H. Chen, Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivity?a multi-center study, Progress in Neuro-Psychopharmacology and Biological Psychiatry 64 (2016) 1–9.
[15] N. Yahata, J. Morimoto, R. Hashimoto, G. Lisi, K. Shibata, Y. Kawakubo, H. Kuwabara, M. Kuroda, T. Yamada, F. Megumi, et al., A small number of abnormal brain connections predicts adult autism spectrum disorder, Nature Communications 7 (1) (2016) 1–12.
[16] M. Perrot, D. Rivière, J.-F. Mangin, Cortical sulci recognition and spatial normalization, Medical Image Analysis 15 (4) (2011) 529–550.
[17] D.M. Witten, R. Tibshirani, T. Hastie, A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis, Biostatistics 10 (3) (2009) 515–534.
[18] A. Jahedi, C.A. Nasamran, B. Faires, J. Fan, R.-A. Müller, Distributed intrinsic functional connectivity patterns predict diagnostic status in large autism cohort, Brain Connectivity 7 (8) (2017) 515–525.
[19] J.D. Power, A.L. Cohen, S.M. Nelson, G.S. Wig, K.A. Barnes, J.A. Church, A.C. Vogel, T.O. Laumann, F.M. Miezin, B.L. Schlaggar, et al., Functional network organization of the human brain, Neuron 72 (4) (2011) 665–678.
[20] A.S. Heinsfeld, A.R. Franco, R.C. Craddock, A. Buchweitz, F. Meneguzzi, Identification of autism spectrum disorder using deep learning and the abide dataset, NeuroImage: Clinical 17 (2018) 16–23.
[21] Z. Sherkatghanad, M. Akhondzadeh, S. Salari, M. Zomorodi-Moghadam, M. Abdar, U.R. Acharya, R. Khosrowabadi, V. Salari, Automated detection of autism spectrum disorder using a convolutional neural network, Frontiers in Neuroscience 13..
[22] M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, in: Advances in neural information processing systems, 2016, pp. 3844–3852..
[23] S. Parisot, S.I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, D. Rueckert, Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer?s disease, Medical Image Analysis 48 (2018) 117–130.
[24] A. Kazi, S. Shekarforoush, S.A. Krishna, H. Burwinkel, G. Vivar, K. Kortüm, S.-A. Ahmadi, S. Albarqouni, N. Navab, Inceptiongcn: receptive field aware graph convolutional network for disease prediction, in: International Conference on Information Processing in Medical Imaging, Springer, 2019, pp. 73–85..
[25] X. Li, N.C. Dvornek, J. Zhuang, P. Ventola, J. Duncana, Graph embedding using infomax for asd classification and brain functional difference detection, arXiv preprint arXiv:1908.04769..
[26] Z. Yao, B. Hu, Y. Xie, P. Moore, J. Zheng, A review of structural and functional brain networks: small world and atlas, Brain Informatics 2 (1) (2015) 45–52.
[27] A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F.X. Castellanos, K. Alaerts, J.S. Anderson, M. Assaf, S.Y. Bookheimer, M. Dapretto, et al., The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism, Molecular Psychiatry 19 (6) (2014) 659–667.
[28] C. Craddock, S. Sikka, B. Cheung, R. Khanuja, S.S. Ghosh, C. Yan, Q. Li, D. Lurie, J. Vogelstein, R. Burns, et al., Towards automated analysis of connectomes: The configurable pipeline for the analysis of connectomes (c-pac), Front Neuroinform 42..
[29] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain, Neuroimage 15 (1) (2002) 273–289.
[30] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Machine Learning 46 (1–3) (2002) 389–422.
[31] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun, Graph neural networks: A review of methods and applications, arXiv preprint arXiv:1812.08434..
[32] 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 Transactions on Computational Biology and Bioinformatics 15 (6) (2018) 1774–1782.
[33] T.B. Mudiyanselage, Y. Zhang, Feature selection with graph mining technology, Big Data Mining and Analytics 2 (2) (2019) 73–82.
[34] J. Liu, D. Zeng, M. Lu, J. Wang, Mild cognitive impairment identification based on multi-view graph convolutional networks, in: 2019 Seventh International Conference on Advanced Cloud and Big Data (CBD), IEEE, 2019, pp. 309–314.
[35] T.N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, arXiv preprint arXiv:1609.02907..
[36] C. Zhang, Y. Ma, Ensemble Machine Learning: Methods and Applications, Springer, 2012.
[37] W. Liu, Y. Wen, Z. Yu, M. Yang, Large-margin softmax loss for convolutional neural networks., in: ICML, vol. 2, 2016, p. 7..
[38] A.E. Hoerl, R.W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, Technometrics 12 (1) (1970) 55–67.
[39] J. Liu, J. Wang, B. Hu, F.-X. Wu, Y. Pan, Alzheimer's disease classification based on individual hierarchical networks constructed with 3-d texture features, IEEE Transactions on Nanobioscience 16 (6) (2017) 428–437.
[40] L. Wang, Support Vector Machines: Theory and Applications, vol. 177, Springer Science & Business Media, 2005.
[41] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, Lightgbm: A highly efficient gradient boosting decision tree, in: Advances in Neural Information Processing Systems, 2017, pp. 3146–3154..
[42] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[43] B.W. Suter, The multilayer perceptron as an approximation to a bayes optimal discriminant function, IEEE Transactions on Neural Networks 1 (4) (1990) 291.
[44] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (8) (2005) 1226–1238.
[45] A. Alkhateeb, I. Rezaeian, S. Singireddy, D. Cavallo-Medved, L.A. Porter, L. Rueda, Transcriptomics signature from next-generation sequencing data reveals new transcriptomic biomarkers related to prostate cancer, Cancer Informatics 18 (2019), 1176935119835522.
[46] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al., Scikit-learn: Machine learning in python, Journal of Machine Learning Research 12 (Oct) (2011) 2825–2830.
[47] C.S. Monk, S.J. Peltier, J.L. Wiggins, S.-J. Weng, M. Carrasco, S. Risi, C. Lord, Abnormalities of intrinsic functional connectivity in autism spectrum disorders, Neuroimage 47 (2) (2009) 764–772.
[48] M.B. Nebel, A. Eloyan, A.D. Barber, S.H. Mostofsky, Precentral gyrus functional connectivity signatures of autism, Frontiers in Systems Neuroscience 8 (2014) 80.
[49] L. Jiang, X.-H. Hou, N. Yang, Z. Yang, X.-N. Zuo, Examination of local functional homogeneity in autism, BioMed Research International (2015).
[50] M. Schuetze, M.T.M. Park, I.Y. Cho, F.P. MacMaster, M.M. Chakravarty, S.L. Bray, Morphological alterations in the thalamus, striatum, and pallidum in autism spectrum disorder, Neuropsychopharmacology 41 (11) (2016) 2627–2637.

Yufei Wang is a MS Candidate in School of Computer Science and Engineering, Central South University, Changsha, Hunan, P.R. China. His current research interests include medical image analysis and machine learning.

Jin Liu Jin Liu received the PhD degree in computer science from Central South University, China, in 2017. He is currently an associate professor at the School of Computer Science and Engineering, Central South University, Changsha, Hunan, P.R. China. His current research interests include medical image analysis and machine learning. His research has been cited over 1000 times (Google Scholar) with H-index = 15.

Yizhen Xiang is a MS Candidate in School of Computer Science and Engineering, Central South University, Changsha, Hunan, P.R. China. His current research interests include medical image analysis and machine learning.

Jianxin Wang received his BEng and MEng degrees in computer science from Central South University of Technology, China, in 1992 and 1996, respectively, and the PhD degree in computer science from Central South University, China, in 2001. He is a professor and Dean in the School of Computer Science and Engineering at Central South University, Changsha, Hunan, P.R. China. His current research interests include bioinformatics and computer algorithm. He published over 200 papers in the field of computational biology and bioinformatics in journals including Bioinformatics, TCBB and conferences including ISMB and BIBM. His research has been cited over 8800 times (Google Scholar) with H-index = 42. He is a senior member of the IEEE, the Chair of ACM Sigbio China and a senior member of China Computer Federation. He serves as editor of TCBB, IJBRA, IJDMB, NMSHIB, IJDSN, CPPS, PPL, CBIO and guest editor of BMC Bioinformatics, TCBB. He served as program chair, co-chair or committee in several international conferences including FAW, ISBRA and BIBM.

Qingyong Chen, expert of 5 g Industry Alliance in Sichuan Province, director of industry application center of China Mobile (Chengdu) Industry Research Institute. Vice president of Sichuan intelligent terminal industry alliance and expert of Sichuan cultural tourism public service platform construction participated in various 5 g academic research exchanges across the country for many times, providing 5 g industry development consultation and review for several governments. He is the author of 5 g + medical care: how new technology changes the business model of medical industry, presided over the preparation of technical development strategy and top-level design report in the field of education, technical development strategy and top-level design report in the field of medical care, etc., have more than 30 national invention patents submitted and participated as the first inventor.

Jing Chong, Master of Software Engineering, 18 years of professional experience, Expert member in vertical field of Technical Advisory Committee of China Mobile Group, is Deputy General Manager of R&D Dept. No.1 of China Mobile (Chengdu) Industrial Research Institute. Former Principal Researcher of China Mobile Group Research Institute. His research interests include 5 g/ NB-IoT/ 4G/ 3G telecom network and service technology, computer science and software engineering. Project Leader of 5G industrial solutions strategic projects in Telemedicine & Intelligent education of China Mobile Group. Many innovation projects which he hosted and participated in have won science & technology awards of academic organization of CIC and CIE, and awards of CMCC. He has applied for more than 30 china patents, 15 authorizations.