Functional connectivity magnetic resonance imaging classification of autism spectrum disorder using the multisite ABIDE dataset

1st Xin Yang
Mathematics & Computer Science
Southern Arkansas University
Magnolia, USA
XinYang@saumag.edu

2nd Mohammad Samiul Islam
Electrical & Computer Engineering
North South University
Dhaka, Bangladesh
samiul.islam08@northsouth.edu

3rd A M Arefin Khaled
Mathematics & Computer Science
Southern Arkansas University
Magnolia, USA
akhaled9024@muleriders.saumag.edu

***Abstract*—The goal of this paper is to apply machine learning algorithms to classify autism spectrum disorder (ASD) patients and typically developing (TD) participants using resting-state functional MRI (rs-fMRI) data from a large multisite data repository ABIDE(Autism Brain Imaging Data Exchange) and identify the important brain connectivity features. In this study, we implemented a data-driven approach to classify ASD patients and TD participants by using the rs-fcMRI features extracted from rs-fMRI data. We applied several classical machine learning classifiers such as support vector machines, logistic regression, and ridge.**

Our contribution has mainly three parts: (1) We used a cross-validation grid search method to find the optimal parameters for each classifier. By using the optimal parameters, the best accuracy we achieved is 71.98%, which is slightly higher than the present best accuracy 70% using deep learning and the same data from the multisite repository ABIDE. We also obtained satisfactory recall and precision results. (2) We implemented the same experiments for seven different brain atlas data in ABIDE, and we identified the most promising brain atlas is Craddock 400 (CC400). (3) We identified the top five highly correlated and anti-correlated region of interests (ROIs) from the brain atlas CC400 for the ASD group and TD group.

***Index Terms*—rs-fMRI, functional connectivity, autism spectrum disorder (ASD), classification, ABIDE, ROIs**

I. Introduction

Autism spectrum disorder (ASD) is a brain disorder that is characterized by social and communication impairments as well as restricted interests and repetitive behaviors. According to recent research studies, one in every 68 children in the United States is affected by ASD [1]. Early diagnosis of ASD is critical for the implementation of the early intervention and providing a proper treatment plan. Although ASD has been identified since the early 1960s, its exact cause is still unknown. Generally, the symptom-based diagnosis of ASD requires a very significant amount of time behavioral assessments under the guidance of a highly experienced multidisciplinary team. However, symptom-based diagnosis often results in poor treatment due to lack of knowledge of neuropathology. In the past few years, an increasing number of neuroscience research studies have used machine learning and deep learning to implement data-driven diagnosis of ASD, which would lead to more effective treatment outcomes. One promising candidate for the data-driven diagnosis is resting-state functional connectivity MRI (rs-fcMRI) data [2]. In past research, extensive brain imaging studies have reported that ASD is associated with brain connectivity [3]–[5] Despite extensive research evidence that ASD is a brain connectivity disorder, it lacks a distributed framework of brain abnormalities [6]. It is still unclear whether brain abnormalities are associated with specific brain regions in ASD [3]. In this study, we implemented a data-driven approach to classify ASD patients and typically developing (TD) participants by using the rs-fcMRI features extracted from resting-state functional MRI (rs-fMRI) data.

The goal of this paper is to apply machine learning algorithms to classify ASD patients and TD participants using the rs-fMRI data from a large multisite data repository ABIDE(Autism Brain Imaging Data Exchange) and identify the important brain connectivity features. We performed the functional connectivity classification of ASD patients and TD participants by applying several classical machine learning classifiers such as support vector machines, logistic regression, and ridge. We used a cross-validation grid search method to find the optimal parameters for each classifier. By using the optimal parameters, each classifier achieved its best performance. The best accuracy 71.98% is achieved with the classifier Ridge and the brain atlas CC400. Our best accuracy result is slightly higher than the present best accuracy 70% using deep learning and the same data from the multisite repository ABIDE [1]. In terms of time efficiency, our training process takes much less time than the deep learning process in [1]. In addition to accuracy, we also got satisfactory recall and precision results for all classifiers. In the study of fMRI data, brain atlas is critical to identify the pathology of ASD. We implemented the same experiments for seven different brain atlas data in ABIDE, and we identified the most promising brain atlas is Craddock 400 (CC400). Finally, we identified the top five highly correlated and anti-correlated region of interests (ROIs) from the brain atlas CC400 for the ASD group and TD group.

978-1-7281-0848-3/19/$31.00 ©2019 IEEE

All the experimental results are displayed in Table II and Table III.

II. Material and Methods

Since the rs-fMRI data can reflect the functional connectivity relationship between the area of the brain [7], in this paper, we implemented a data-driven approach to classify ASD patients and TD participants by using the rs-fcMRI features extracted from the rs-fMRI data.

A. ABIDE dataset

The preprocessed rs-fMRI data with ASD and TD are downloaded from a large multisite data repository ABIDE(Autism Brain Imaging Data Exchange). ABIDE is a multisite platform that has aggregated functional and structural brain imaging data collected from 17 different laboratories around the world. The preprocessed connectomes project (PCP) from the ABIDE has openly released 539 individuals who have ASD and 573 TD to the public [8]. These 1112 datasets consist of structural and preprocessed resting state functional MRI data along with phenotypic information. The rs-fMRI data are slice time corrected, motion corrected and normalized. In this study, all rs-fMRI data are selected from the CPAC preprocessing pipeline and band-pass filtered (0.01-0.1Hz). From these 1112 subjects, 1035 subjects are screened for our study since only 1035 subjects have been given the corresponding completed phenotypic information, which is essential for our further study. In these 1035 subjects, there are 505 ASD and 530 TD, 157 females and 878 males. The summary information of the screened subjects is displayed in Table I. Table I contains the key phenotypical information of ASD and TD such as gender, age and lab site ID.

TABLE I
ABIDE data Phenotypical Information

site	*ASD*	*TD*	*M*	*F*	*Age Range*
CALTECH	19	18	29	8	17∼56
CMU	14	13	21	6	19∼40
KKI	20	28	36	12	8∼13
LEUVEN	29	34	55	8	12∼32
MAX_MUN	24	28	48	4	7∼58
NYU	75	100	139	36	6∼39
OHSU	12	14	26	0	8∼15
OLIN	19	15	29	5	10∼24
PITT	29	27	48	8	9∼35
SBL	15	15	30	0	20∼64
SDSU	14	22	29	7	9∼17
STANFORD	19	20	31	8	8∼13
TRINITY	22	25	47	0	12∼26
UCLA	54	44	86	12	8∼18
UM	66	74	113	27	8∼29
USM	46	25	71	0	9∼50
YALE	28	28	40	16	7∼18
TOTAL	505	530	157	878	6∼64

B. Feature selection

Since the functional connectivity is a manifestation of the co-activation level of the brain regions [3]–[5], in this study, we use functional connectivity of ROIs to classify ASD patients and TD participants. The rs-fcMRI features are extracted from rs-fMRI data. In order to extract the functional connectivity of ROIs, a correlation is calculated between two ROIs based on the average of the time series of the rs-fMRI data. The correlation coefficient ranges from 1 to -1. 1 indicates that two ROIs are highly correlated, -1 indicates that two ROIs are anti-correlated. The feature selection step is completed by using the python package nilearn. Nilearn is a python toolbox for statistical learning on neuroimaging data.

There are seven various brain parcellations in ABIDE: Automated Anatomical Labeling (AAL, 116 ROIs), Dosenbach (160 ROIs), Eickoff-Zilles (EZ, 116 ROIs), Harvard-Oxford Atlas (HOA,110 ROIs), Talairach-Tournoux (TT, 110 ROIs), Craddock 200 (CC200, 200 ROIs), Craddock 400 (CC400, 400 ROIs) [9]. We have compared all these seven brain atlas in our study, CC400 achieves the best accuracy result. Fig. 1 shows an example of rs-fMRI time series of the brain atlas CC400. Fig. 2 shows an example of the connectivity matrix for the brain atlas HOA.

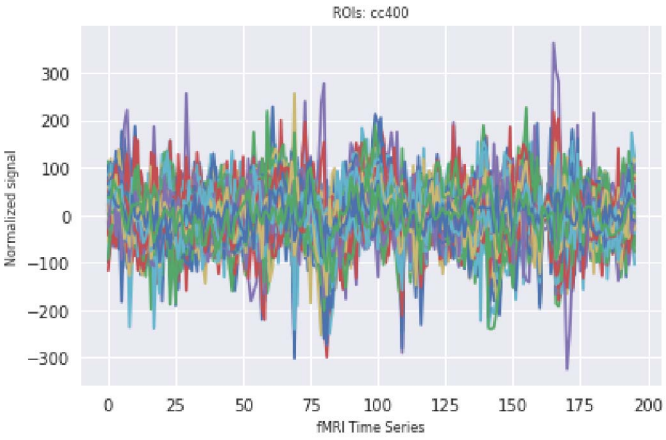

Fig. 1. rs-fMRI time series example

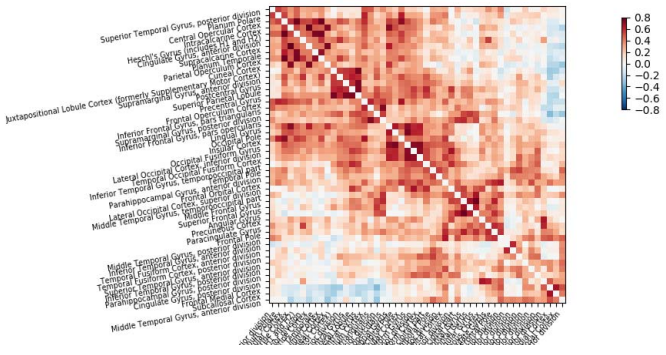

Fig. 2. functional connectivity of HOA example

C. Classification methods

In the past few years, a growing number of studies have shown that machine learning (ML) classifiers can be used for

the analysis of fMRI data [10]–[13]. Most studies applied supervised learning methods for the classification of ASD and TD, such as support vector machines(SVM), logistic regression (LR), Ridge, random forest (RF). Along with the growth in interest and breadth of applications, ML classifiers can achieve up to 97% accuracy for classification of ASD patients and TD participants for a single site [14]. Generally, the studies with a single site and a small number of subjects could achieve high classification accuracy. However, the classification accuracy drops significantly when the multi-site data and a large number of subjects are studied. 76.67% classification accuracy was achieved through rs-fMRI with 178 subjects from ABIDE [15]. As high as 60% accuracy was obtained with 964 subjects and 16 separate international sites [16]. Recently, deep learning has gained increasing attention in the field of fMRI. The authors improved the state-of-art by achieving 70% with 1035 subjects from 17 international sites by using a deep neural network (DNN) [1]. In this study, we aim to improve that highest accuracy with 1035 subjects from 17 international sites using supervised learning classifiers with optimal parameters.

III. Results

In this study, we applied and compared four different traditional machine learning classifiers for the classification of ASD and TD. The four classifiers are logistic regression, Ridge, linear support vector machine with l_2 penalty and support vector machine with a Gaussian kernel. All experiments are implemented with 5-fold cross-validation. According to our research, the present studies did not report that the classification analysis is based on the optimal parameters of the traditional classifiers. The classifiers can produce significantly different results using different parameters. In order to achieve the best performance, we need to adjust the parameters for each classifier. Therefore, the purpose of this paper is to improve the accuracy of traditional classifiers by using optimal parameters. In order to find out the optimal parameter for each classifier, we applied the grid-search method. We implemented the same experiments for all seven brain atlas: CC400, CC200, AAL, HOA, TT, EZ, Dosenbach. In addition to accuracy, we also report the recall and precision for each classifier and each brain atlas. All cross-validation results are displayed in Table II. From Table II, we can see that the best accuracy 71.98% was achieved by using Ridge classifier and CC400 atlas. Our best accuracy is slightly higher than the present best accuracy 70%, which is achieved by using the DNN and the same data from the multisite repository ABIDE [1]. In general, deep learning would produce better results than traditional machine learning in terms of big data. However, the advantage of deep learning is not obvious for the ABIDE dataset, since the data sample is not big.

In the study of fMRI data, brain atlas is critical; appropriate brain atlas can achieve better performance. From Table II, we can conclude that CC400 is the most promising brain atlas for the future study. This conclusion also confirms that parcellation containing a larger number of ROIs can represent functional connectivity patterns at the voxel level and more accurate [9]. In Table III, we listed the top five highly correlated and anti-correlated ROIs from the CC400 atlas for the ASD group and TD group. These selected ROIs are selected according to the mean of the connectivity matrix for each group.

ASD is a developmental brain disorder that affect social interaction, verbal and nonverbal communication. Many research work have highlighted ASD with a deficit in the integration of cognitive mechanisms. Recent fMRI studies have showed that ASD are characterized by reduced functional connectivity throughout the cortical language systems [17], [18]. Our studies also supported this view. The highly correlated Left Occipital Fusiform Gyrus and Left Lingual Gyrus ROIs are only found in ASD. The highly correlated Left Cingulate Gyrus and Right Precuneous Cortex are only found in TD participants. For anti-correlation ROIs, the Right Lateral Occipital Cortex and Right Central Opercular Cortex are not found in ASD group, which are found TD.

The brain connectome associated with the correlated and anti-correlated ROIs of ASD and TD is shown in Fig. 3 and Fig. 4. In Fig. 3 and Fig. 4, the red color brain connectivity indicates correlated, and the blue color brain connectivity indicates anti-correlated. All the experiment results are reproducible and all the source code can be found in github: https://github.com/XinYangSAU/ABIDE.

TABLE II
Experimental Results

Atlas	*Classifier*	*Accuracy*	*Recall*	*Precision*
CC400	Ridge	**71.98%**	**70.89%**	**71.53%**
	LR	71.79%	70.69%	71.29%
	linearSVC-l_2	71.40%	70.10%	70.93%
	SVC-rbf	71.40%	69.90%	71.12%
CC200	Ridge	69.28%	65.94%	69.61%
	LR	69.08%	66.73%	69.04%
	linearSVC-l_2	68.60%	66.14%	68.56%
	SVC-rbf	68.12%	62.57%	69.31%
AAL	Ridge	65.99%	62.57%	66.13%
	LR	66.18%	63.56%	66.05%
	linearSVC-l_2	65.41%	61.39%	65.68%
	SVC-rbf	65.60%	63.96%	65.10%
HOA	Ridge	67.92%	66.53%	67.47%
	LR	67.73%	67.33%	66.92%
	linearSVC-l_2	67.34%	66.14%	66.78%
	SVC-rbf	67.34%	66.34%	66.79%
EZ	Ridge	66.09%	62.18%	66.37%
	LR	66.09%	60.99%	66.69%
	linearSVC-l_2	66.57%	61.19%	67.38%
	SVC-rbf	65.31%	64.75%	64.63%
TT	Ridge	67.54%	65.15%	67.30%
	LR	66.96%	64.75%	66.62%
	linearSVC-l_2	67.73%	65.15%	67.60%
	SVC-rbf	66.47%	61.39%	67.10%
Dosenbach	Ridge	63.00%	61.78%	62.22%
	LR	63.19%	61.58%	62.45%
	linearSVC-l_2	63.48%	62.18%	62.74%
	SVC-rbf	63.09%	62.57%	62.16%

TABLE III
ROIs ANALYSIS FOR CC400

Top 5 Correlation Connectivity ROIs	
ASD	Left Intracalcarine Cortex ↔ Right Intracalcarine Cortex
	Left Occipital Pole ↔ Right Occipital Pole
	Left Cuneal Cortex ↔ Right Cuneal Cortex
	Left Precuneous Cortex ↔ Right Precuneous Cortex
	Left Occipital Fusiform Gyrus ↔ Left Lingual Gyrus
TD	Left Intracalcarine Cortex ↔ Right Intracalcarine Cortex
	Left Precuneous Cortex ↔ Right Precuneous Cortex
	Left Cingulate Gyrus ↔ Right Precuneous Cortex
	Left Occipital Pole ↔ Right Occipital Pole
	Left Cuneal Cortex ↔ Right Cuneal Cortex
Top 5 Anti-Correlation Connectivity ROIs	
ASD	Left Lateral Occipital Cortex ↔ Right Frontal Operculum Cortex
	Right Frontal Medial Cortex ↔ Right Frontal Operculum Cortex
	Left Middle Temporal Gyrus ↔ Right Frontal Operculum Cortex
	Right Precuneous Cortex ↔ Right Frontal Operculum Cortex
	Left Precuneous Corte ↔ Right Frontal Operculum Cortex
TD	Left Lateral Occipital Cortex ↔ Right Frontal Operculum Cortex
	Left Precuneous Cortex ↔ Right Frontal Operculum Cortex
	Right Frontal Medial Cortex ↔ Right Frontal Operculum Cortex
	Right Lateral Occipital Cortex ↔ Right Frontal Operculum Cortex
	Right Lateral Occipital Cortex ↔ Right Central Opercular Cortex

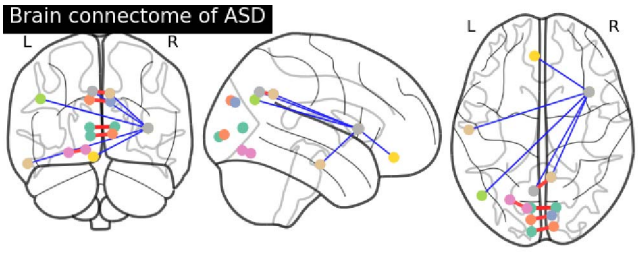

Fig. 3. Top 5 correlated and anti-correlated ROIs of ASD

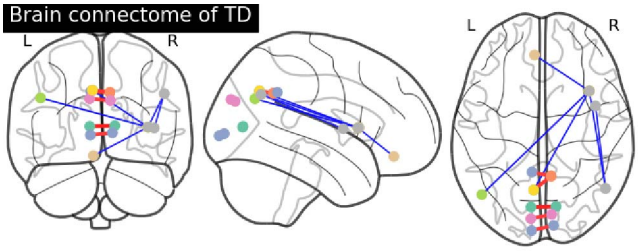

Fig. 4. Top 5 correlated and anti-correlated ROIs of TD

IV. DISCUSSION

In this study, we have implemented four classical machine learning classifiers for the classification of ASD patients and TD participants from a large multisite data repository ABIDE. We have applied the algorithms to seven different brain atlas: CC400, CC200, AAL, EZ, HOA, TT, Dosenbach. The experiment results show that we can achieve up to 71.98% accuracy by using Ridge classifier and CC400 atlas. Although the machine learning classifier cannot be directly applied to the clinical diagnosis, our work demonstrates that the classical machine learning classifiers can find some correlated ROIs to classify fMRI data of ASD and TD from some extend. We anticipate that the combination of machine learning classifiers with other ASD clinical features will improve the accuracy of the ASD diagnosis.

REFERENCES

[1] Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage: Clinical, 17, 16-23.

[2] Yamada, T., Hashimoto, R. I., Yahata, N., Ichikawa, N., Yoshihara, Y., Okamoto, Y., ... & Kawato, M. (2017). Resting-state functional connectivity-based biomarkers and functional MRI-based neurofeedback for psychiatric disorders: a challenge for developing theranostic biomarkers. International Journal of Neuropsychopharmacology, 20(10), 769-781.

[3] Anderson, J. S., Nielsen, J. A., Froehlich, A. L., DuBray, M. B., Druzgal, T. J., Cariello, A. N., ... & Alexander, A. L. (2011). Functional connectivity magnetic resonance imaging classification of autism. Brain, 134(12), 3742-3754.

[4] Belmonte, M. K., Allen, G., Beckel-Mitchener, A., Boulanger, L. M., Carper, R. A., & Webb, S. J. (2004). Autism and abnormal development of brain connectivity. Journal of Neuroscience, 24(42), 9228-9231.

[5] Just, M. A., Cherkassky, V. L., Keller, T. A., & Minshew, N. J. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. Brain, 127(8), 1811-1821.

[6] Mller, R. A., Shih, P., Keehn, B., Deyoe, J. R., Leyden, K. M., & Shukla, D. K. (2011). Underconnected, but how? A survey of functional connectivity MRI studies in autism spectrum disorders. Cerebral cortex, 21(10), 2233-2243.

[7] Franco, A. R., Mannell, M. V., Calhoun, V. D., & Mayer, A. R. (2013). Impact of analysis methods on the reproducibility and reliability of resting-state networks. Brain connectivity, 3(4), 363-374.

[8] Craddock, C., Benhajali, Y., Chu, C., Chouinard, F., Evans, A., Jakab, A., Khundrakpam, B.S., Lewis, J.D., Li, Q., Milham, M. and Yan, C. (2013). The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. Neuroinformatics.

[9] Craddock, R. C., James, G. A., Holtzheimer III, P. E., Hu, X. P., & Mayberg, H. S. (2012). A whole brain fMRI atlas generated via spatially constrained spectral clustering. Human brain mapping, 33(8), 1914-1928.

[10] Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. Neuroimage, 45(1), S199-S209.

[11] Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends in cognitive sciences, 10(9), 424-430.

[12] Haynes, J. D., & Rees, G. (2006). Neuroimaging: decoding mental states from brain activity in humans. Nature Reviews Neuroscience, 7(7), 523.

[13] O'Toole, A. J., Jiang, F., Abdi, H., Pnard, N., Dunlop, J. P., & Parent, M. A. (2007). Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data. Journal of cognitive neuroscience, 19(11), 1735-1752.

[14] Just, M. A., Cherkassky, V. L., Buchweitz, A., Keller, T. A., & Mitchell, T. M. (2014). Identifying autism from neural representations of social interactions: neurocognitive markers of autism. PloS one, 9(12), e113879.

[15] Plitt, M., Barnes, K. A., & Martin, A. (2015). Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. NeuroImage: Clinical, 7, 359-366.

[16] Nielsen, J. A., Zielinski, B. A., Fletcher, P. T., Alexander, A. L., Lange, N., Bigler, E. D., ... & Anderson, J. S. (2013). Multisite functional connectivity MRI classification of autism: ABIDE results. Frontiers in human neuroscience, 7, 599.

[17] Uhlhaas, P. J., & Singer, W. (2006). Neural synchrony in brain disorders: relevance for cognitive dysfunctions and pathophysiology. neuron, 52(1), 155-168.

[18] Hill, E. L., & Frith, U. (2003). Understanding autism: insights from mind and brain. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358(1430), 281-289.