

Deep Learning approach for classification and interpretation of Autism Spectrum Disorder

Pindi Krishna Chandra Prasad*
krishna.chandra@research.iit.ac.in
IIT Hyderabad

Yash Khare*
yash.khare@research.iit.ac.in
IIT Hyderabad

Kamalaker Dadi
kamalaker.dadi@research.iit.ac.in
IIT Hyderabad

P. K. Vinod
vinod.pk@iit.ac.in
IIT Hyderabad

Bapi Raju Surampudi
bapi.raju@iit.ac.in
IIT Hyderabad

Abstract—Autism spectrum disorder (ASD) is a neurodevelopmental disorder predominantly found in children. The current behavior-based diagnosis of ASD is arduous and requires expertise. Therefore, it is appealing to develop an accurate computer-aided tool for diagnosing ASD. Although resting-state functional magnetic resonance imaging (rsfMRI) has proven to be successful in capturing the neural organization of the brain, automated detection of ASD using rsfMRI scans is a challenging task due to heterogeneity in the dataset and limited sample size. This paper proposes a Multilayer Perceptron (MLP) based classification model with autoencoder pretraining for classifying ASD from Typically Developing (TD) using rsfMRI scans obtained from the ABIDE-1 dataset. Our model achieves new state-of-the-art performance on the ABIDE-1 dataset with a 10-fold cross-validation accuracy of 74.82%. Further, we use the Integrated Gradients (IG) and DeepLIFT techniques to identify the correlations between brain regions that contribute most to the classification task. Our analysis identifies the following regions, Left Lingual Gyrus, Right Insula Lobe, Right Cuneus, Right Middle Frontal Gyrus, Left Superior Temporal Gyrus to be associated with ASD. Interestingly, these regions in the brain are primarily responsible for social cognition, language, attention, decision making and visual processing, which are known to be altered in ASD.

Index Terms—Autism Classification, Neural Network, Autoencoder, Pretraining, rsfMRI, ABIDE, ASD Brain Biomarkers

I. INTRODUCTION & RELATED WORK

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that occurs in children. It is caused due to early deviations in brain development, and neural rearrangement [1]. The children with this disorder suffer from social communication deficits, monotonous and abnormal sensory-motor behaviors. ASD, as the very name suggests, is a spectrum disorder with a broad range of types, severities, and symptoms [2]. Centers for disease control and prevention reported that 1 out of 54 children in the United States and 1 out of 160 children worldwide are diagnosed with ASD. All socioeconomic and ethnic groups have been affected by ASD. The prevalence of ASD has been increasing worldwide, and the underlying cause is unclear. Diagnosis of ASD is typically made by

observing behavioral patterns and cognitive developmental delays because there are no well-grounded biomarkers [1].

Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R) are the “gold standard” assessment measures for the evaluation of ASD. However, diagnosis using these procedures suffers from subjectivity, lack of accuracy and also requires experts, which often results in a significant economic burden on the families of ASD patients [3]. Early detection of ASD is crucial, and it can significantly improve the quality of life of individuals with ASD, their careers, and families, as mentioned in the clinical study by Elder et al. [4]. Therefore, developing a computer-aided tool for the accurate diagnosis of ASD will make the diagnosis objective, swift, and economical.

Brain disorders are nothing but abnormalities in one or more brain regions or atypical connectivity between different brain regions [5]. Atypical Functional Connectivity (FC) between different brain regions is widely used for the study of various neurodegenerative and psychiatric disorders like ASD, Mild cognitive impairment, Bipolar disorder, Attention deficit hyperactivity disorders, and Schizophrenia [6]. FC is estimated from the temporal correlation of spontaneous BOLD signal among two or more anatomically distinct brain regions. If the signals from two or more brain regions show synchronized fluctuations, then the regions are said to be functionally correlated (connected). FC in the resting human brain was first identified by Biswal et al. [7].

With the increasing popularity of deep learning methods, many studies have combined FC and deep neural network architectures to classify children with ASD from the typically developing (TD). Heinsfeld et al. [8] used stacked denoising autoencoders (SDA) at the pretraining stage and then used a multilayer perceptron (MLP) as the classifier. They reported an overall 10-fold cross-validation accuracy of 70% on the ABIDE-1 dataset. Parisot et al. [9] have used a graph-based approach to build a population graph and then trained a Graph Convolution Network to classify the nodes as autistic or typical control. They achieved a classification accuracy of 70.4% on the ABIDE-1 dataset. Eslami et al. [10] proposed a joint learning procedure using an autoencoder and a single

*The first two authors contributed equally to this work

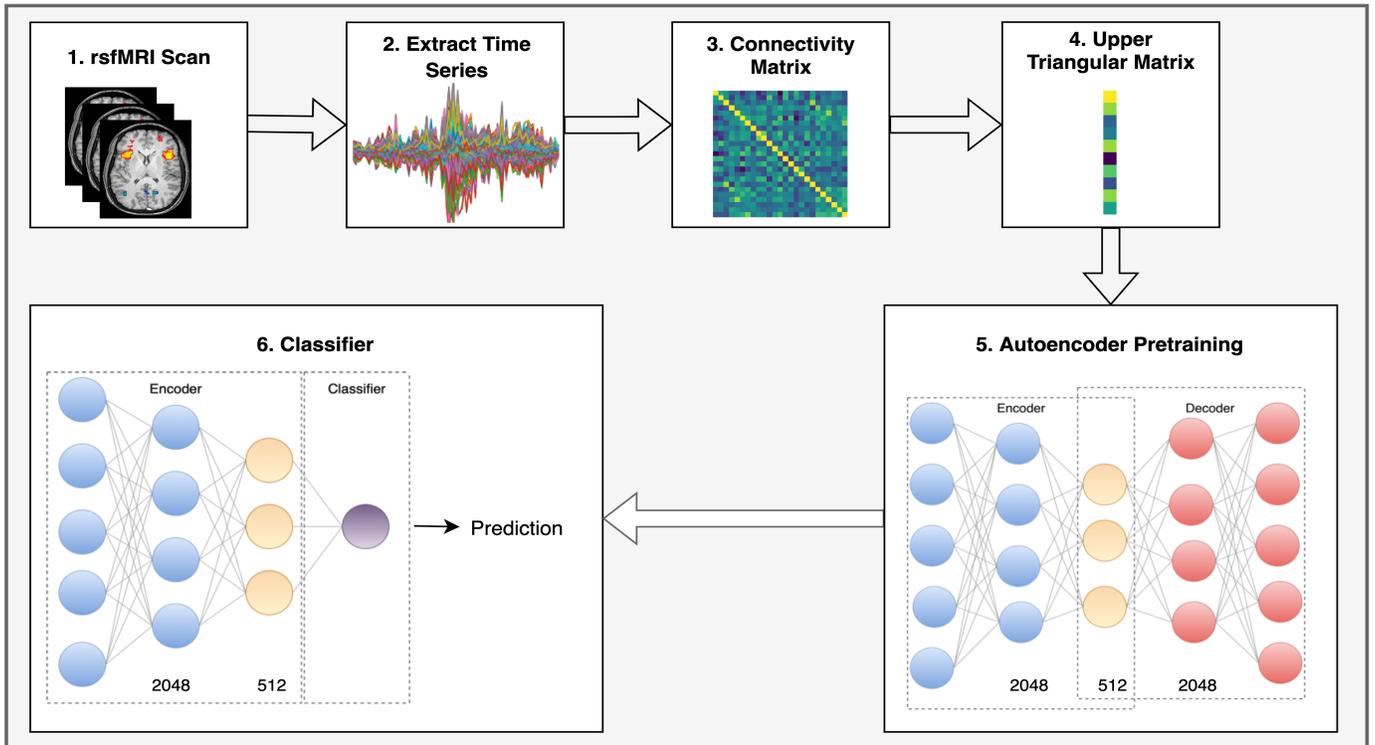


Fig. 1. Classification Pipeline. 1. Pre-processing the rsfMRI scan. 2. Extracting the time-series data using pre-defined anatomical or functional brain parcellations. 3. Extracting the static functional connectivity matrix. 4. Flattening the Upper triangular part of functional connectivity matrix and passing as input to the auto-encoder. 5. Training the auto-encoder. 6. Fine-tuning the classifier with pre-trained auto encoder weights.

layer perceptron for the ASD classification task, called *ASD-Diagnet*. Their model achieved 70.3% classification accuracy on the ABIDE-1 dataset. Sherkatghanad et al. [11] used a Convolutional Neural Network (CNN) based classification model and reported an accuracy of 70.22% on the ABIDE-1 dataset. Wang et al. [12] proposed a multi-atlas deep feature representation and ensemble learning method based on SDA and MLP for the ASD identification task. They reported an accuracy of 74.52% on the ABIDE-1 dataset.

The above mentioned approaches lack model interpretability i.e. identifying the brain regions that contribute most to the ASD classification task. This study proposes a two hidden layer neural network with autoencoder pre-trained weights and model interpretation using Integrated Gradients (IG) [13] and DeepLIFT [14]. Our model achieves new state-of-the-art performance with a 10-fold cross-validation accuracy of 74.82% on the ABIDE-1 dataset in 10x less training time as compared to the previous best method [12]. The main contributions of this paper are:

- 1) Proposed a two hidden layer feed-forward neural network with autoencoder pretraining.
- 2) Applied IG and DeepLIFT, prominent feature attribution methods, to interpret brain biomarkers in ASD subjects.
- 3) Revealed the impact of heterogeneity in the dataset and compared the effect of different preprocessing pipelines, brain parcellation schemes on the classification performance.

II. METHODS

A. Feature Extraction

Static Functional Connectivity (sFC) is measured by calculating the pair-wise correlations between every pair of brain regions. Mostly these correlations are linear, which are captured using the Pearson Correlation Coefficient (PCC). The PCC, ρ_{xy} for two signals, x and y each of length T and mean \hat{x} and \hat{y} respectively, can be computed using the following equation.

$$\rho_{xy} = \frac{\sum_{t=1}^T (x_t - \hat{x})(y_t - \hat{y})}{\sqrt{\sum_{t=1}^T (x_t - \hat{x})^2} \sqrt{\sum_{t=1}^T (y_t - \hat{y})^2}} \quad (1)$$

Given n brain regions, we obtain an $n \times n$, sFC symmetric matrix where each $(i, j)^{th}$ entry represents the PCC between i^{th} and j^{th} regions. We extract the upper triangular values of the sFC matrix and use them as an input to our model.

B. Autoencoder and Classification Method

An Autoencoder [15] is a type of neural network trained in an unsupervised way to estimate the latent representation of the original input. It consists of two parts – an encoder, which encodes the input vector x into a latent vector z , and a decoder which then reconstructs the original input vector \hat{x} from the latent vector z . Both the encoder and decoder are neural networks that are trained end-to-end by minimizing the reconstruction loss between the original input vector x , and the

TABLE I

SITE-WISE PHENOTYPIC INFORMATION OF THE ABIDE-1 DATASET. ASD AND TD COUNT REPRESENT THE NUMBER OF AUTISM AND TYPICALLY DEVELOPING SAMPLES IN EACH SITE, RESPECTIVELY. AVERAGE MEAN FRAME-WISE DISPLACEMENT (MFD) REPRESENTS THE AVERAGE MEAN FRAME-WISE DISPLACEMENT OF ALL SCANS IN THE SAME SITE MEASURED IN MILLIMETERS (MM). TIMESERIES REPRESENT THE NUMBER OF FRAMES IN EACH SCAN ALONG THE TEMPORAL DIMENSION. DURATION REPRESENTS THE TIME TAKEN FOR EACH SCAN IN SECONDS(S). REPETITION TIME (TR) REPRESENTS THE AMOUNT OF TIME BETWEEN SUCCESSIVE PULSE SEQUENCES APPLIED TO THE SAME SLICE IN SECONDS (S)

SITE	ASD count	TD count	Average Age (years)	Male count	Female count	Average MFD (mm)	Time-series	Duration (s)	TR (s)
CALTECH	14	18	27.36	25	7	0.0701	146	292.0	2
CMU	13	13	26.69	20	6	0.2957	236/316	632.0	2
KKI	18	28	10.02	34	12	0.14	152	380	2.5
LEUVEN	20	34	17.81	46	8	0.0924	246	410	1.6
MAX MUN	17	28	25.42	42	3	0.1441	116/196	588	3
NYU	60	100	15.51	125	35	0.07	176	352	2
OHSU	9	14	10.87	23	0	0.09	78	195	2.5
OLIN	18	15	16.67	29	4	0.1901	206	309	1.5
PITT	24	27	18.89	45	6	0.1564	196	294	1.5
SBL	12	15	32.85	27	0	0.1528	196	431	2.2
SDSU	9	22	14.32	24	7	0.0954	176	352	2
STANFORD	17	20	9.89	30	7	0.1063	176	352	2
TRINITY	19	25	16.90	44	0	0.1104	146	292	2
UCLA	46	44	12.90	79	11	0.1915	116	348	3
UM	58	74	14.09	106	26	0.1608	296	592	2
USM	42	25	22.74	67	0	0.1496	236	472	2
YALE	23	28	12.79	37	14	0.1137	196	392	2

decoder output \hat{x} , to make the reconstruction as close to the original input vector as possible. Usually, the reconstruction loss is computed using the Mean Squared Error (MSE).

$$Loss(x, \hat{x}) = \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (2)$$

In this study, we use the autoencoder in the pretraining stage. The encoder architecture consists of two hidden layers of size 2048 and 512 respectively, as shown in Fig.1. We use \tanh as the nonlinear activation function for both the layers. The decoder architecture mirrors that of the encoder, with two hidden layers of size 512 and 2048, respectively.

For each fold, first, we train the autoencoder and then remove the decoder part and add the classifier, which has a single output layer as shown in Fig.1. The model comprising the encoder and classifier, are then fine-tuned for the classification task by minimizing the binary cross-entropy loss H between the original class label, y_i and the model prediction, \hat{y}_i for all the N subjects.

$$H = \frac{-1}{N} \sum_{i=1}^N (y_i * \log \hat{y}_i + (1 - y_i) * \log(1 - \hat{y}_i)) \quad (3)$$

C. Model Interpretation

Many studies have developed classification pipelines to detect ASD from TD with high accuracy. However, identifying the brain bio-markers in ASD subjects is less explored. In this study, we have attempted to find out the regions associated with ASD using Integrated Gradients (IG) [13] and DeepLIFT [14]. IG and DeepLIFT are considered to best attribution methods so far as it satisfies the two fundamental axioms, Sensitivity, and Implementation Invariance, that

all attribution methods must satisfy [13]. Therefore, IG and DeepLIFT were used to determine which input features to blame or credit for predicting the corresponding classification label.

IG and DeepLIFT assign a weight to a feature relative to the baseline (where that feature is absent). Passing the input with null feature values reveals how the model performs when no information is provided and is neutral to the model's prediction as recommended in [16]. Therefore, initializing the baseline as input of zeros would be appropriate in this study.

IG aggregates the gradients along the inputs that fall on the straight line between the baseline and the input. The integrated gradient along the i^{th} dimension for an input x and baseline x' is defined as follows. Here, $\partial F(x)/\partial x_i$ is the gradient of $F(x)$ along the i^{th} dimension.

$$IG_i(x) = (x_i - x'_i) \times \int_{\alpha=0}^1 \frac{\partial F(x' + \alpha \times (x - x'))}{\partial x_i} d\alpha \quad (4)$$

DeepLIFT method computes feature importance based on explicating the difference in output from some 'reference' output in terms of the difference of the input from some 'reference' input. Though the gradient is zero, information can be propagated using this approach based on differences from the reference values [14].

III. DATASET & PREPROCESSING

This paper utilized the dataset from the preprocessed Autism Brain Imaging Data Exchange (ABIDE-1) initiative [17]. The dataset consists of 1,112 rsfMRI scans collected from 17 different sites. There exist 505 ASD and 530 Typical controls (henceforth, referred to as typically developing (TD) cohort), out of which 86 ASD scans have been discarded due to their missing time series [12]. There is significant noise due

to patient-specific difficulties such as repetitive head motion, body trembling, keeping eyes open/closed during the scan, and heterogeneity due to high variation in demographics, scan parameters, and age groups of participants collected from different sites as shown in table I.

The dataset was experimented with four different pipelines such as Connectome Computation System (CCS) [18], Configurable Pipeline for the Analysis of Connectomes (C-PAC) [19], Data Processing Assistant for Resting-State fMRI (DPARF) [20], Neuro Imaging Analysis Kit (NIAK) [21], are publicly available. Basic preprocessing steps include dropping first n columns, slice timing correction, motion realignment, intensity normalization, nuisance signal removal, bandpass filtering (0.01 - 0.1 Hz), and global signal regression. The above-mentioned preprocessing pipelines include all the basic preprocessing steps, but their software implementations, specific algorithms, and parameters for various steps are different.

It is essential to parcellate the brain into structurally and functionally well-defined regions and obtain the time-series data to understand the correlation between regions in the brain [22]. Timeseries data obtained using different brain parcellation schemes like Havard-Oxford (HO), Automated Anatomical Labeling (AAL) [23], Dosenbach 160 [24], Craddock 200 (CC200) [25] on preprocessed rsfMRI scans are publicly available and have been used directly in our work.

IV. EXPERIMENTS AND RESULTS

A. Experimental Setup

This study evaluates the proposed method using 10-fold cross-validation [26] on the ABIDE-1 dataset with 949 samples. In each fold, 20% of the training set is used as a validation set for hyper-parameter tuning. The upper triangular part of the sFC matrix is flattened and passed as input to train the autoencoder part (pretraining stage) for 50 epochs and the classifier part for 50 epochs using Adam optimizer with a learning rate of 10^{-4} and weight decay of 0.1. The encoder part of the classifier is initialized with pretrained autoencoder weights and the whole classifier has been trained. All the hyper-parameters were tuned using cross-validation. The code is implemented in pytorch [27] and made available here¹. Computations are performed on a computing system with 16GB RAM and NVidia K80 GPUs for the entire training stage.

B. Classification

We report the accuracy, sensitivity, and specificity of our model. Our model achieves a new state-of-the-art performance with an accuracy of 74.82%, sensitivity of 67.33%, specificity of 80.75% on the ABIDE-1 dataset. Table II reports the performance of the proposed method in comparison with previous studies for the ASD classification task.

¹<https://github.com/pindi-krishna/Classification-and-Interpretation-of-ASD.git>

C. Impact of different Preprocessing Pipelines

In this section, we discuss the impact of using different preprocessing pipelines on the performance of our proposed method. We train our model on data obtained using four preprocessing pipelines, CPAC, NIAK, DPARF, and CCS. Table III shows the classification results using these pipelines. We get the best accuracy of 74.82% for the CPAC pipeline followed by 71.21% for DPARF, 71.04% for CCS, and 64.55% for NIAK. As we can see, there is a 13% difference in the accuracy when using the CPAC pipeline compared to the NIAK pipeline, which suggests that choosing the correct preprocessing pipeline for the classification task plays a significant role in the overall results.

TABLE II
OUR METHOD OUTPERFORMS ALL THE PREVIOUS METHODS IN OVERALL ACCURACY.

Method	Accuracy(%)	Sensitivity(%)	Specificity(%)
Heinsfeld [8]	70.0	74.0	63.0
Pariset [9]	70.4	-	-
ASD-Diagnet [10]	70.3	68.3	72.20
CNN [11]	70.22	77.46	61.82
AIMAFE [12]	74.52	80.69	66.71
Proposed Method	74.82	67.33	80.75

TABLE III
RESULTS USING DIFFERENT PREPROCESSING PIPELINES ON ABIDE-1 DATASET USING CC200 BRAIN PARCELLATION. CPAC PIPELINE HAS YIELDED THE BEST PERFORMANCE

Pipeline	Accuracy(%)	Sensitivity(%)	Specificity(%)
CPAC	74.82	67.33	80.75
NIAK	64.55	46.78	78.6
DPARF	71.21	64.46	76.57
CCS	71.04	66.02	75.02

TABLE IV
RESULTS USING DIFFERENT ATLASSES ON ABIDE-1 DATASET PREPROCESSED USING CPAC PIPELINE

Atlas	Accuracy(%)	Sensitivity(%)	Specificity(%)
CC200	74.82	67.33	80.75
HO	72.29	66.06	77.21
Dosenbach	70.92	64.69	75.85
AAL	69.95	60.90	77.09

D. Impact of different parcellation schemes

This section presents the results of a comparative study using four different brain parcellations – CC200, Dosenbach, AAL, and HO – which parcellate the brain into 200, 160, 116, and 111 regions, respectively. As the number of regions of interest (ROIs) is different for each atlas, the input vector size also changes accordingly, affecting the model's overall performance. Table IV shows the classification results using the above mentioned atlases. We get the best accuracy of 74.82% using the CC200 atlas.

TABLE V
LEAVE-ONE-SITE-OUT RESULTS ON ABIDE-1 DATASET PREPROCESSED
USING CPAC PIPELINE AND CC200 BRAIN PARCELLATION SCHEME

Site	Accuracy(%)	Sensitivity(%)	Specificity(%)
CALTECH	71.25	64.29	76.67
CMU	71.54	58.46	84.62
KKI	73.91	65.56	79.28
LEUVEN	74.07	33.0	98.24
MAXMUN	57.34	48.24	62.86
NYU	78.75	72.33	82.60
OHSU	60	44.44	70
OLIN	74.9	66.67	82.22
PITT	73.73	63.33	82.96
SBL	54.82	18.34	84.0
SDSU	76.13	40.0	90.91
STANFORD	58.38	78.82	41.0
TRINITY	65.0	65.26	64.8
UCLA	76.0	71.31	80.91
UM	75.76	75.86	75.67
USM	85.08	84.29	86.4
YALE	84.31	89.56	80.0
Mean	71.23	61.16	77.83

E. Leave-One-Site-Out Results

One major challenge in the ASD classification task using the ABIDE-1 dataset is the inter-site variability as it contains rsfMRI scans from 17 different sites. To evaluate the performance of our proposed method on new sites, we train our model using the leave-one-site-out approach. In this approach, we keep the data from each site separately for testing, and we train our model on the data from the remaining sites. This approach allows us to estimate the model’s generalizability to new, unseen sites. Table V shows the leave-one-site-out results. Our model achieves the best accuracy of 85.08% on scans collected from the University of Southern Mississippi (USM) site, followed by 84.31% for Yale and 78.75% for NYU.

F. Interpretation

Attribution methods have been applied on the model trained using the data obtained from CPAC pre-processing pipeline and CC200 brain parcellation scheme as this combination yielded the best result. The steps for finding the associated regions using IG and DeepLIFT are as follows: A) Group all the correctly predicted autism samples from the test set. B) Apply attribution method on these autism samples to find the attributions for each feature passed as input to the model with a zero embedded vector as the baseline. C) Replace the attributions in the top one percentile with 1 and remaining values with 0. D) Construct a 2D matrix of size 200×200 with attribution vector contributing to the upper triangular matrix. E) Calculate the row-wise sum and pick those ROIs with the maximum value. F) Repeat steps A-E on each fold and find out the most repeated, common ROIs in all the folds.

As we are interested in regions that led to the prediction of autism, we have considered only correctly predicted autism samples for applying attribution methods. The results are qualitatively similar when the attributions in top k percentile ($k = 0.5, 2, \text{ and } 3$) replaced with 1. These attribution methods have been implemented using *Captum*, a model interpretability

library for PyTorch [28]. IG and DeepLIFT analysis identify the following regions: Left Lingual Gyrus (LLG), Right Insula Lobe (RIL), Right Cuneus (RC), Right Middle Frontal Gyrus (RMFG), Left Superior Temporal Gyrus (LSTG) to be associated with ASD classification (see Fig.2). Primary function and co-ordinates of these regions based on CC200 parcellation are shown in Table VI. We conclude that feature attribution methods such as IG and DeepLIFT enable accurate identification of brain regions whose activation seems to be altered in ASD as compared to TD.

TABLE VI
ROIS THAT CONTRIBUTED MOST TO THE ASD CLASSIFICATION TASK
BASED ON CC200 BRAIN PARCELLATION

ROI name	ROI number	Center of mass	Primary Function
LLG	177	(-14.3;-74.2;-10.1)	Face/Object Recognition
RIL	59	(36.7;17.2;3.6)	Decision making
RC	142	(17.8;-89.5;22.7)	Visual Processing
RMFG	106	(28.6;34.6;42.0)	Attention
LSTG	200	(-41.9;-31.5;15.2)	Language Comprehension

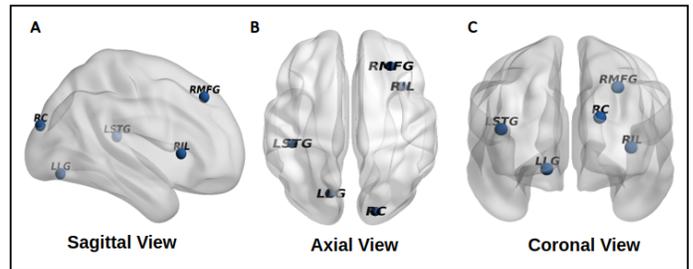


Fig. 2. Visualization of regions associated with ASD in 3 different views using Brain Net Viewer [29]. A. Sagittal view B. Axial View C. Coronal View. LLG : Left Lingual Gyrus, RIL : Right Insula Lobe, RC : Right Cuneus, RMFG : Right Middle Frontal Gyrus, LSTG : Left Superior Temporal Gyrus

V. DISCUSSION AND CONCLUSION

This work shows that the proposed method demonstrates effective results on the ASD classification task. Among all the studies reported in Table II, our method achieved the best classification accuracy of 74.82%. The entire training time of our model is less than an hour which is approximately 10x faster than the previous state-of-the-art method. This suggests that the autoencoder pretraining weights acted as an excellent initialization point for training the classifier, which helped in faster convergence of the model.

As the scans are collected from various sites, and each site uses a different set of scan parameters and protocols, the data have significant heterogeneity. The variations in the leave-one-site-out results as shown in Table V highlight the impact of the heterogeneity on the classification task. Although there is no consensus and conceptual clarity on how different preprocessing pipelines and brain parcellations affect the ASD classification results, we observe a significant variation in the results as shown in Table III and IV, respectively. These

variations may occur due to the implementation of different algorithms and parameters used in preprocessing pipelines and different ROIs extracted from each parcellation scheme. The empirical results obtained in this study reveal that the combination of CPAC preprocessing pipeline and CC200 parcellation yield the best accuracy for the ASD classification task. Going forward, these results might be useful in defining a common benchmark dataset and specifications to enable a fair and viable comparison of methods being proposed for ASD classification.

This study identifies LLG, RIL, RC, RMFG, LSTG associated with ASD. A lingual gyrus plays a crucial role in vision processing, primarily related to letters. Chandran et al. [30] found a significant association between the left lingual gyrus cortical thickness and the right lateral occipital cortex surface area among autism subjects. The greater volume and gyrification of the lingual gyrus and lateral occipital cortex may cause abnormal visual processing in individuals with higher autistic symptoms. Yamada et al. study [31] on 36 ASD subjects and 38 TD subjects found a significant change in the anterior sector of the left insula and the middle ventral sub-region of the right insula in the ASD brain. They noticed a notable volumetric increase in the ASD brain compared with the TD brain in the middle ventral sub-region, in the right insula. The right cuneus is responsible for visual processing. Stock et al. study [32] on 66 adults with high-functioning autism and 66 TD investigated gray matter abnormalities in the two groups. They found increased gray-matter volume in frontal brain regions, including the medial prefrontal cortex, superior and inferior frontal gyri, and middle temporal gyrus, and reduced gray-matter volume in posterior brain regions, including the posterior hippocampus, cuneus, in individuals with ASD in relative to TD. Middle Frontal Gyrus and Face processing seem quite relevant as ASD patients have difficulty with face processing and facial emotion identification. In ASD subjects, significantly reduced activity in the middle frontal gyrus when involving face processing tasks and middle temporal gyrus during nonface social tasks is found when compared to TD subjects [33]. The superior temporal gyrus (STG) is responsible for language comprehension. Bigler et al. study [34] on 30 autistic children and 39 controls of similar age, education, and head circumference, investigated the link between autism and intellectual-language-based abilities. Clinical Evaluation of Language Fundamentals–Third Edition (CELF–3) [35], divided into three index scores: expressive, receptive, and total, has been used to measure language ability. This study observed a positive correlation between receptive language scores and STG volume in control subjects and zero correlation in autistic subjects.

Although the BOLD time series signal is a time varying signal over the duration of the rsfMRI scan, in the proposed approach we calculated the grand mean FC value over the entire time series, thus yielding static functional connectivity (sFC). Features from the sFC matrix were used in the proposed approach to obtain the best results. Thus, there is still scope for improvement using various measures to capture the brain

dynamics using the dynamic functional connectivity (dFC) information in place of sFC. In future, efforts can focus on integrating phenotypic and demographic information with dynamic functional connectivity to attempt superior classification performance and possible biomarkers for the diagnosis of ASD.

ACKNOWLEDGMENT

We thank IHub data, International Institute of Information Technology for financial support.

REFERENCES

- [1] C. Lord, M. Elsabbagh, G. Baird, and J. Veenstra-Vanderweele, "Autism spectrum disorder," *The Lancet*, vol. 392, no. 10146, pp. 508–520, 2018.
- [2] C. Lord, T. S. Brugha, T. Charman, J. Cusack, G. Dumas, T. Frazier et al., "Autism spectrum disorder," *Nature Reviews Disease Primers*, vol. 6, no. 1, pp. 1–23, 2020.
- [3] A. Frigaux, R. Evrard, and J. Lighezzolo-Alnot, "Adi-r and ados and the differential diagnosis of autism spectrum disorders: Interests, limits and openings," *L'encephale*, vol. 45, no. 5, pp. 441–448, 2019.
- [4] J. H. Elder, C. M. Kreider, S. N. Brasher, and M. Ansell, "Clinical impact of early diagnosis of autism on the prognosis and parent–child relationships." *Psychology research and behavior management*, 2017.
- [5] Y. Du, Z. Fu, and V. D. Calhoun, "Classification and prediction of brain disorders using functional connectivity: promising but challenging," *Frontiers in neuroscience*, vol. 12, p. 525, 2018.
- [6] E. Canario, D. Chen, and B. Biswal, "A review of resting-state fmri and its use to examine psychiatric disorders," *Psychoradiology*, vol. 1, no. 1, pp. 42–53, 2021.
- [7] B. Biswal, F. Zerrin Yetkin, V. M. Haughton, and J. S. Hyde, "Functional connectivity in the motor cortex of resting human brain using echo-planar mri," *Magnetic resonance in medicine*, vol. 34, no. 4, pp. 537–541, 1995.
- [8] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the abide dataset," *NeuroImage: Clinical*, vol. 17, pp. 16–23, 2018.
- [9] S. Parisot, S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker et al., "Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease," *Medical image analysis*, vol. 48, pp. 117–130, 2018.
- [10] T. Eslami, V. Mirjalili, A. Fong, A. R. Laird, and F. Saeed, "Asd-diagnet: a hybrid learning approach for detection of autism spectrum disorder using fmri data," *Frontiers in neuroinformatics*, vol. 13, p. 70, 2019.
- [11] Z. Sherkatghanad, M. Akhondzadeh, S. Salari, M. Zomorodi-Moghadam, M. Abdar, U. R. Acharya et al., "Automated detection of autism spectrum disorder using a convolutional neural network," *Frontiers in neuroscience*, vol. 13, p. 1325, 2020.
- [12] Y. Wang, J. Wang, F.-X. Wu, R. Hayrat, and J. Liu, "Aimafe: autism spectrum disorder identification with multi-atlas deep feature representation and ensemble learning," *Journal of Neuroscience Methods*, p. 108840, 2020.
- [13] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *International Conference on Machine Learning*. PMLR, 2017, pp. 3319–3328.
- [14] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *International Conference on Machine Learning*. PMLR, 2017, pp. 3145–3153.
- [15] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proceedings of ICML workshop on unsupervised and transfer learning*. JMLR Workshop and Conference Proceedings, 2012, pp. 37–49.
- [16] P. Sturmfels, S. Lundberg, and S.-I. Lee, "Visualizing the impact of feature attribution baselines," *Distill*, vol. 5, no. 1, p. e22, 2020.
- [17] A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts et al., "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism," *Molecular psychiatry*, vol. 19, no. 6, pp. 659–667, 2014.
- [18] T. Xu, Z. Yang, L. Jiang, X.-X. Xing, and X.-N. Zuo, "A connectome computation system for discovery science of brain," *Science Bulletin*, vol. 60, no. 1, pp. 86–95, 2015.

- [19] 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*, vol. 42, pp. 10–3389, 2013.
- [20] C. Yan and Y. Zang, “Dparsf: a matlab toolbox for” pipeline” data analysis of resting-state fmri,” *Frontiers in systems neuroscience*, vol. 4, p. 13, 2010.
- [21] P. Bellec, F. Carbonnell, V. Perlberg, C. Lepage, O. Lyttelton, V. Fonov, A. Janke, J. Tohka, and A. Evans, “A neuroimaging analyses kit for matlab and octave,” in *Human Brain Mapping HBM 2011 17th Annual Meeting of the Organization on Human Brain Mapping, Quebec City, Canada, June 26-30, 2011*. Organization on Human Brain Mapping, 2011, pp. 1–5.
- [22] S. B. Eickhoff, B. T. Yeo, and S. Genon, “Imaging-based parcellations of the human brain,” *Nature Reviews Neuroscience*, vol. 19, no. 11, pp. 672–686, 2018.
- [23] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, “Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain,” *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.
- [24] N. U. Dosenbach, B. Nardos, A. L. Cohen, D. A. Fair, J. D. Power, J. A. Church, S. M. Nelson, G. S. Wig, A. C. Vogel, C. N. Lessov-Schlaggar *et al.*, “Prediction of individual brain maturity using fmri,” *Science*, vol. 329, no. 5997, pp. 1358–1361, 2010.
- [25] R. C. Craddock, G. A. James, P. E. Holtzheimer III, X. P. Hu, and H. S. Mayberg, “A whole brain fmri atlas generated via spatially constrained spectral clustering,” *Human brain mapping*, vol. 33, no. 8, pp. 1914–1928, 2012.
- [26] 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,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
- [27] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan *et al.*, “Pytorch: An imperative style, high-performance deep learning library,” *Advances in neural information processing systems*, vol. 32, pp. 8026–8037, 2019.
- [28] N. Kokhlikyan, V. Miglani, M. Martin, E. Wang, B. Alsallakh, J. Reynolds, A. Melnikov, N. Kliushkina, C. Araya, S. Yan *et al.*, “Cap-tum: A unified and generic model interpretability library for pytorch,” *arXiv preprint arXiv:2009.07896*, 2020.
- [29] M. Xia, J. Wang, and Y. He, “Brainnet viewer: a network visualization tool for human brain connectomics,” *PLoS one*, vol. 8, no. 7, p. e68910, 2013.
- [30] V. A. Chandran, C. Pliatsikas, J. Neufeld, G. O’Connell, A. Haffey, V. DeLuca, and B. Chakrabarti, “Brain structural correlates of autistic traits across the diagnostic divide: A grey matter and white matter microstructure study,” *NeuroImage: Clinical*, vol. 32, p. 102897, 2021.
- [31] T. Yamada, T. Itahashi, M. Nakamura, H. Watanabe, M. Kuroda, H. Ohta, C. Kanai, N. Kato, and R.-i. Hashimoto, “Altered functional organization within the insular cortex in adult males with high-functioning autism spectrum disorder: evidence from connectivity-based parcellation,” *Molecular autism*, vol. 7, no. 1, pp. 1–15, 2016.
- [32] T. Eilam-Stock, T. Wu, A. Spagna, L. J. Egan, and J. Fan, “Neuroanatomical alterations in high-functioning adults with autism spectrum disorder,” *Frontiers in neuroscience*, vol. 10, p. 237, 2016.
- [33] M. A. Patriquin, T. DeRamus, L. E. Libero, A. Laird, and R. K. Kana, “Neuroanatomical and neurofunctional markers of social cognition in autism spectrum disorder,” *Human brain mapping*, vol. 37, no. 11, pp. 3957–3978, 2016.
- [34] E. D. Bigler, S. Mortensen, E. S. Neeley, S. Ozonoff, L. Krasny, M. Johnson, J. Lu, S. L. Provençal, W. McMahon, and J. E. Lainhart, “Superior temporal gyrus, language function, and autism,” *Developmental neuropsychology*, vol. 31, no. 2, pp. 217–238, 2007.
- [35] E. Semel, E. Wiig, and W. Secord, “Clinical evaluation of language fundamentals—third edition (celf-3) psychological corporation,” *San Antonio, TX*, 1995.