DIAGNOSING AUTISM SPECTRUM DISORDER USING ENSEMBLE 3D-CNN: A PRELIMINARY STUDY

Jingsheng Deng[1] *Md Rakibul Hasan*[2] *Minhaz Mahmud*[3]
Md Mahbub Hasan[4] *Khandaker Asif Ahmed*[5] *Md Zakir Hossain*[1,6,7]

[1]School of Computing, Australian National University, Canberra, Australia
[2]BRAC University, Dhaka, Bangladesh
[3]Bangladesh Army International University of Science and Technology, Cumilla, Bangladesh
[4]Khulna University of Engineering & Technology, Khulna, Bangladesh
[5]L&W CSIRO, Black Mountain, Canberra, Australia
[6]Biological Data Science Institute, Australian National University, Canberra, Australia
[7]A&F CSIRO, Black Mountain, Canberra, Australia
{zakir.hossain}@anu.edu.au

2022 IEEE International Conference on Image Processing (ICIP) | 978-1-6654-9620-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIP46576.2022.9897628

ABSTRACT

Autism spectrum disorder (ASD) is a neuro-developmental disorder that results in behavioural retardation in verbal communications and social interactions. Traditional ASD detection methods involve assessing patients' behavioural patterns by medical practitioners, which often lack credibility and precision. The contribution of the current study involves a 3D-CNN (convolutional neural network) model to diagnose ASD patients from healthy individuals using functional magnetic resonance imaging (fMRI) of the brain. We utilised a publicly available dataset, Autism Brain Imaging Data Exchange (ABIDE I), and tested different CNN-based models in individual and combined brain parcellations. Our model showed a better outcome (74.53% accuracy, 69.98% sensitivity, and 76.00% specificity) for combined parcellations than individuals. Further, we compared our model with several state-of-the-art models and discussed the suitability of our model for future prospects. The current model would be a predecessor of future prognosis models or behavioural patterns-based multi-modal models for early detection of ASD.

Index Terms— Autism, fMRI, Functional connectivity, CNN, ABIDE I

1. INTRODUCTION

Autism spectrum disorder (ASD) is a common neurodevelopmental disease that manifests as communication and social barriers and progresses towards repetitive and uncontrollable behaviours in daily life. According to the World Health Organisation (WHO), ASD is affecting one in 160 children over the world [1]. Since ASD mostly appears in early childhood between 6 and 17 years old, it is crucial to diagnose at early stage to avoid complications [2]. Modern clinics detect autism based on some codes of conduct, such as the autism diagnostic observation schedule and autism diagnostic interview-revised. These methods lack objectivity with a higher rate of misdiagnosis. Moreover, behaviours are complex and should not be utilised as decisive evidence to determine a disease [3]. To save medical resources, achieve early diagnosis, and enhance the higher quality of life of ASD people and their families, it is an urge to create a more objective and accurate ASD diagnosis method.

Magnetic resonance imaging (MRI) is a low-cost diagnostic tool in the computational medical field [4]. It helps diagnose brain disorders including epilepsy, Alzheimer's, schizophrenia, and autism by analysing functional brain structure [5]. Since functional MRI (fMRI) can detect correlated fluctuations by measuring blood oxygen levels in the brain, recent studies often aim to explore ASD biomarkers using fMRI [6]. Functional connectivity has been widely employed in the progression of brain disorders (e.g., ASD) that injure ambiguous connections among several brain regions [6]. By observing the intensity changes of fMRI images onto a time series, we can quantify the brain activity by a predefined atlas or parcellation technique. Note that a parcellation technique defines the boundaries of different regions of interest (ROI) and restricts the spatial scale, which becomes easy to understand and interpret [7]. In this context, a well-known dataset that contains neuroimaging and phenotypic information, for example Autism Brain Imaging Data Exchange I (ABIDE I), can be used for data-driven autism diagnosis method [8]. Overall, we present our analysis on ABIDE I dataset to diagnose ASD from control groups.

Several works have shown a proof of concept to use deep learning in ASD [4, 9, 10, 11]. Authors in [9] used multi-atlas deep feature representation with ensemble learning and

achieved 74.52% accuracy. Graph convolutional networks (GCNs) achieved 70.40% accuracy to predict autism in [10]. However, their model must be retrained from scratch if a new subject is added to the analysis. Several works used fMRI images from ABIDE datasets to classify individuals with ASD from healthy controls [4, 11]. Authors in [4] used only Craddock 200 (CC200) atlas in Extra-Trees and support vector machine (SVM) algorithms. Authors in [11] used Harvard-Oxford atlas and nine single summary measures independently as input to a 3D-CNN (convolutional neural network) ensemble approach. We propose an optimized ASD identification model using three atlases in a 3D-CNN configuration.

2. MATERIALS & METHODS

2.1. Dataset & Data Preprocessing

ABIDE I dataset consists of resting-state fMRI images of 1,112 individuals (539 ASD patients and 573 healthy controls) collected from 17 different geo-locations around the world [8]. The dataset was manually inspected by neurology experts, ensuring the authenticity of the overall dataset. From several pre-processed ABIDE I datasets released by preprocessed connectomes project[1], we chose "Configurable Pipeline for the Analysis of Connectomes" (C-PAC) pipeline since it covers the operation of slice timing correction, motion correction, and voxel intensity normalisation [12]. To remove fMRI artifacts, resulting from different physiological (e.g., head motion, respiration, cardiac pulsation) or mechanical issues (e.g., low-frequency drifts, global signal), a band-pass filter of 0.01–0.1Hz was used to preprocess the dataset, which resulted in 860 rs-fMRI with proper phenotypic annotations. Overall the final dataset consists of fMRI of 440 ASD individuals (366 males and 74 females) and 420 controls (361 males and 59 females).

Different studies utilised different atlases to detect ASD. Authors in [11] and [4] showed Craddock 200 (CC200) and Craddock 400 (CC400) atlases are more suitable, while authors in [13] reported better performance with the automated anatomical labeling 116 (AAL116) atlas. As the performance can differ based on models and hyperparameters, we experimented on all three atlases—AAL116, CC200, and CC400—to be consistent across the study. We experimented on individual atlases and a combination of all three atlases together.

2.2. Architecture of 3D-CNN Model

Figure 1 depicts the elements involved in our proposed method. We first extracted functional connectivity matrices from fMRI by computing the connectivity among different time blocks between each pair of regions of the three brain atlases. These are upper triangular symmetric matrices, with all the values on the diagonal being one. Each cell in the matrix is a Pearson correlation coefficient, where 1 represents the highest correlation between the two areas of the brain, and -1 represents the lowest. These matrices were the original feature representations used as CNN models' input.

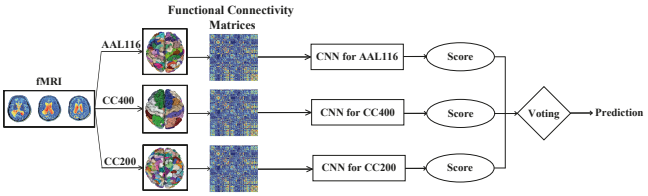

Fig. 1. Overall arrangement for ASD diagnosis from fMRI images.

In a similar manner applied in [13], we employed a stochastic sampling method named Poisson disk sampling, where samples are drawn at a minimum distance apart computed by spatial proximity. The Gray matter was then divided roughly into equal-sized parcels to prevent them from crossing hemisphere boundaries for the given parcellation scheme. Since the remaining voxels were assigned to the nearest regional centre, we introduced randomness in the centre of the ROI. We used the MNI152 template at 3 mm resolution to process both fMRI data and parcellations combined with the grey matter tissue with subcortical structures. Furthermore, we used the cortical mantle mask to construct functional connectivity profiles.

Fully connected layers connect all units from one layer to the next, while convolutional layers only connect a unit to its spatially connected units. The weights of these connections are shared, which facilitates learning with fewer parameters. CNN uses filters to detect image features. Pooling layers reduce feature maps' size. One of the widely used pooling methods is max pooling. It selects maximum features in prefixed local neighbourhoods to represent them.

We constructed CNN architectures with multiple layers to deal with high-dimensional images. Our experiment utilised three 3D CNN models for three different parcellations. Though [13] inspired our data processing, we tuned three distinct CNNs for three atlases, whereas they separately trained a single model for different atlases. We systematically changed hyperparameters and tracked performances for each setting, which guided us to decide on optimized CNNs for each atlas. We reduced the number of layers and finally decreased computation cost by controlling the number of parameters to around 300k. Before trying our model, we also experimented with some popular computer-vision architectures in 3D versions, such as ResNet-50, VGGNet, and their variants. However, we obtained poor performance, proving

[1] http://preprocessed-connectomes-project.org/abide/

that deep and bulky neural networks do not fit this dataset and the task to some extent.

The input to each CNN model was given by calculating 3D-correlation maps among different atlases that represent three multi-channel 3D matrices. We created 30 stochastic parcellations by taking cross-sections through different coordinates and directions, and then we calculated their mean accuracy as a benchmark. Like [13], we used both atlas-based and stochastic brain parcellation methods to define target ROIs in our implementation.

Figure 2 illustrates three CNNs for the three parcellations. AAL116 CNN consists of average pooling of size 2 and stride 2 in the first layer with the same border mode, which acts as a downsampling function. It has two convolutional layers with exponential linear unit (ELU) activation functions. The first convolutional layer had 64 filters of size 3 in three dimensions, and the second one had 16 filters of the same size kernel. A max-pooling layer of size 2 followed the convolutional layers in 3 dimensions. The output was flattened and fed to a dense layer with 16 nodes, and again ELU activation with L2 (0.005) kernel regularisation was applied. The last layer was fully connected with only one node and sigmoid activation for final classification for AAL116 atlas.

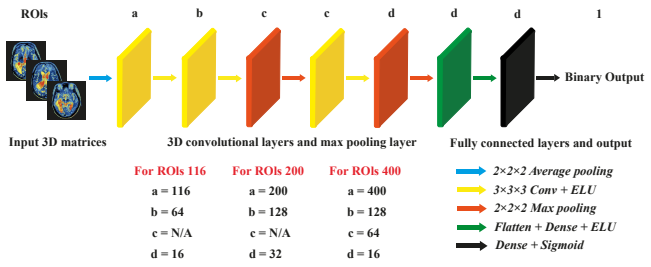

Fig. 2. Three CNN architectures for AAL116, CC200, CC400 atlases.

We only changed the first convolutional layer's filter size to 128 and the second layer's filter size to 32 in the CC200 CNN model because of its more complicated brain map. More filters made it possible to learn more precisely and concisely. CC400 CNN was based on CC200 CNN, but one more layer was added between two previous convolutional layers. It had 64 filters followed by a max-pooling layer of size 2.

We used binary cross-entropy validation loss and mini-batch training with 64 batch size. Stochastic gradient descent (SGD) was the optimiser having 0.001 as the learning rate and 0.9 as the momentum parameter. We ran 50 epochs for each parcellation. We found the models were overfitting as the training accuracy approached 100% so fast that when it got to 25 epochs, it had already reached 98%-99% accuracy. Then we added regularisation having a rate of 0.005 and dropout in all CNN models to avoid overfitting. In AAL116 CNN, only one spatial dropout was applied between two convolutional layers. However, for CC200 and CC400 CNNs, another dropout was applied between two dense fully-connected layers along with the spatial dropout. All dropout rates were set to 0.5 after optimising.

Furthermore, we implemented SVM and RF (random forest) algorithms on the three parcellation schemes as benchmarks. We used the "sigmoid" kernel for the SVM classifier, and we chose "sqrt" and "balanced" for max features and class weight for the RF classifier.

2.3. Evaluation of Model Performance

We used accuracy, specificity, and sensitivity metrics with 5-fold cross-validation to evaluate our proposed method's performance. All 860 subjects were divided into five subsets: one subset was the test set, the others were training sets. Even though accuracy could be interpreted as the only criteria for classification tasks, specificity and sensitivity are also significant to show different aspects of results in clinical expectation.

3. RESULTS & DISCUSSION

To classify ASD and healthy control, we designed three CNN architectures for three different parcellations with the pre-processed ABIDE I dataset. We further combined the three CNNs to achieve the best accuracy. The total execution time was about 35h using the GeForce GTX 1050 Ti GPU.

Table 1 shows cross-validated test score of each 3D CNN for three distinct atlases and the combination of them (we recorded the trained parameters for each CNN model and started a new 5-fold cross-validation for the combination method). We further calculated another evaluation score—area under the receiver operating characteristic curve—that was 0.74 for the ensemble CNN model.

Table 1. 5-fold cross-validated results of CNN models on parcellations.

	Accuracy % ($\pm se$)	**Sensitivity** % ($\pm se$)	**Specificity** % ($\pm se$)
AAL116	69.53 ($\pm$0.05)	67.04 ($\pm$0.04)	73.27 ($\pm$0.03)
CC200	70.91 ($\pm$0.04)	69.83 ($\pm$0.02)	73.76 ($\pm$0.04)
CC400	71.94 ($\pm$0.03)	69.50 ($\pm$0.03)	74.01 ($\pm$0.02)
Combined	74.53 ($\pm$0.04)	69.98 ($\pm$0.04)	76.00 ($\pm$0.08)

The results confirm the suitability of our design concept. The combined parcellation scheme consistently performed better than individual alas-based models. Every atlas has its outstanding ideas and meaning, and thus, ensemble models—considering them all and combining their advantages—is a wise and safer choice for supervised machine learning on connectomes. Different atlases have their own "tended" network architectures, for which we designed different CNNs to make them exert their greatest strength. The 3D CNNs

stored all information from the fMRI data for learning and training. Specifically, the spatial convolutions captured data's structural and topographic features and valuable information relevant for classification.

According to our analysis, summarising the connectome data using a single atlas is often sub-optimal for training machine learning models. An ensemble technique—averaging various models trained with diverse parcellation schemes—can produce significantly more accurate predictions.

We also evaluated the performance of SVM and RF classifiers on the same preprocessed dataset for each atlas. After hyperparameter tuning, the average accuracy was 69% and 64%, respectively. The results are presented in Table 2 (all values for each atlas are the mean values of 5-fold validation). CNN-based architectures proved their selection by outperforming these two traditional machine learning classifiers even after optimisation.

Table 2. 5-fold cross-validated accuracy (%) on baseline methods.

Parcellation	SVM	RF
AAL116	65.32	63.71
CC200	71.48	61.44
CC400	69.22	65.78

We found that the accuracy fluctuates significantly (high variation) during training. We analysed the low accuracy folds and found that the test accuracy decreased remarkably when the test set included data from several specific sources (Stanford, UM, Leuven, OHSU, USM, and Yale). Then we tested our ensemble model on the "filtered" (discarded the above sites) data set with only 554 subjects; the mean accuracy indeed increased a lot to around 80% with 5-fold cross-validation. By observing the contents of these data, surprisingly, we found that their magnitudes are either too smaller or too larger than others, which we called "outliers" in the aspect of the dataset. Thus, these outliers have reduced the overall accuracy.

3.1. Comparative Analysis

To demonstrate the high quality of our proposed method, we illustrate previously comparative studies in Table 3. Hinted by the study of [13], we explored a different direction with various spatial scales and parcellation techniques. Their model used seven parcellations and the same CNN model to predict, while our model consists of three different CNN architectures with dropout rates. We used only ABIDE I dataset and achieved 74.53% accuracy whereas they used both I and II datasets and achieved only 72.80% accuracy. Our result overperformed many previous studies. The most plausible reason is that we employed three distinct CNN models for three different atlases, whereas other researchers used either a single atlas or a single prediction model on multiple atlases. This area still requires significant effort to reach a more acceptable accuracy in ASD prediction to help real-life medical diagnoses.

Table 3. Comparison with previous studies on ASD prediction.

Reference	ABIDE Dataset	Preprocessing Method	Algorithm	Accuracy (%)
[14]	I	C-PAC	DNN	70
[15]	I	CCS	LSTM	70.10
[16]	I	C-PAC	CNN	70.22
[10]	I	C-PAC	GCN	70.40
[17]	I	C-PAC	Ensemble GCN	70.86
[4]	I	C-PAC	Extra-Trees	72.20
[9]	I	C-PAC	Ensemble MLP	74.52
Proposed	**I**	**C-PAC**	**3D-CNN**	**74.53**
[5]	I subset (size: 182)	C-PAC	Autoencoder-DNN	90.39

Authors in [5] achieved 90.39% accuracy using a subset of the ABIDE I dataset having only 182 subjects from a single source (NYU Langone Medical Center). Having data from a single source inherently limits ASD symptoms' heterogeneity that would have been solved if data from diverse sources had been considered (we considered all 17 sources). As discussed previously, we got higher accuracy after excluding several sources—proving less diversity indeed gives higher classification accuracy—which justifies [5]'s higher accuracy. However, an efficient ASD identification tool should be trained from diversified data, making it robust and suitable enough for a broader area in practical scenarios.

4. CONCLUSION AND FUTURE WORK

To diagnose ASD, we present a detailed analysis of applying three suitable CNN models for three different atlases on rs-fMRI data from the ABIDE I dataset. We first calculated three functional connectivity matrices based on three different brain atlases. We applied ensemble learning, where the individual atlases' CNN results voted for the final prediction, which we then compared with the ground truth to calculate our method's performance. We achieved 74.53% accuracy, outperforming state-of-the-results on ASD prediction considering the full-scale ABIDE dataset. In future, we would attempt to extend our method to more atlases to check its robustness and applicability. We would evaluate different combinations of atlases and deep-learning models for a robust and trustworthy ASD diagnosis and prognosis system. Moreover, as we found that excluding several sites brings better results, we will apply more rules and discipline among site, age, and sex dimensions while selecting data.

5. REFERENCES

[1] James D Lee and Hedda Meadan, "Parent-mediated interventions for children with asd in low-resource settings: a scoping review," *Review Journal of Autism and Developmental Disorders*, vol. 8, no. 3, pp. 285–298, 2021.

[2] Catherine Lord, Susan Risi, Linda Lambrecht, Edwin H Cook, Bennett L Leventhal, Pamela C DiLavore, Andrew Pickles, and Michael Rutter, "The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits associated with the spectrum of autism," *Journal of autism and developmental disorders*, vol. 30, no. 3, pp. 205–223, 2000.

[3] Sami Timimi, Damian Milton, Virginia Bovell, Steven Kapp, and Ginny Russell, "Deconstructing diagnosis: Four commentaries on a diagnostic tool to assess individuals for autism spectrum disorders," *Autonomy (Birmingham, England)*, vol. 1, no. 6, 2019.

[4] Yaya Liu, Lingyu Xu, Jun Li, Jie Yu, and Xuan Yu, "Attentional connectivity-based prediction of autism using heterogeneous rs-fmri data from cc200 atlas," *Experimental neurobiology*, vol. 29, no. 1, pp. 27–37, 2020.

[5] Yazhou Kong, Jianliang Gao, Yunpei Xu, Yi Pan, Jianxin Wang, and Jin Liu, "Classification of autism spectrum disorder by combining brain connectivity and deep neural network classifier," *Neurocomputing*, vol. 324, pp. 63–68, 2019.

[6] Christopher S Monk, Scott J Peltier, Jillian Lee Wiggins, Shih-Jen Weng, Melisa Carrasco, Susan Risi, and Catherine Lord, "Abnormalities of intrinsic functional connectivity in autism spectrum disorders," *Neuroimage*, vol. 47, no. 2, pp. 764–772, 2009.

[7] Alexandre Abraham, Michael P Milham, Adriana Di Martino, R Cameron Craddock, Dimitris Samaras, Bertrand Thirion, and Gael Varoquaux, "Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example," *NeuroImage*, vol. 147, pp. 736–745, 2017.

[8] Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, 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.

[9] Yufei Wang, Jianxin Wang, Fang-Xiang Wu, Rahmatjan Hayrat, and Jin Liu, "Aimafe: Autism spectrum disorder identification with multi-atlas deep feature representation and ensemble learning," *Journal of Neuroscience Methods*, vol. 343, pp. 108840, 2020.

[10] Sarah Parisot, Sofia Ira Ktena, Enzo Ferrante, Matthew Lee, Ricardo Guerrero, Ben Glocker, and Daniel Rueckert, "Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease," *Medical image analysis*, vol. 48, pp. 117–130, 2018.

[11] Rajat Mani Thomas, Selene Gallo, Leonardo Cerliani, Paul Zhutovsky, Ahmed El-Gazzar, and Guido van Wingen, "Classifying autism spectrum disorder using the temporal statistics of resting-state functional mri data with 3d convolutional neural networks," *Frontiers in psychiatry*, vol. 11, 2020.

[12] Cameron Craddock, Yassine Benhajali, Carlton Chu, Francois Chouinard, Alan Evans, András Jakab, Budhachandra Singh Khundrakpam, John David Lewis, Qingyang Li, Michael Milham, et al., "The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives," *Frontiers in Neuroinformatics*, vol. 7, 2013.

[13] Meenakshi Khosla, Keith Jamison, Amy Kuceyeski, and Mert R Sabuncu, "Ensemble learning with 3d convolutional neural networks for functional connectome-based prediction," *NeuroImage*, vol. 199, pp. 651–662, 2019.

[14] Anibal Sólon Heinsfeld, Alexandre Rosa Franco, R Cameron Craddock, Augusto Buchweitz, and Felipe Meneguzzi, "Identification of autism spectrum disorder using deep learning and the abide dataset," *NeuroImage: Clinical*, vol. 17, pp. 16–23, 2018.

[15] Nicha C Dvornek, Pamela Ventola, and James S Duncan, "Combining phenotypic and resting-state fmri data for autism classification with recurrent neural networks," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 725–728.

[16] Zeinab Sherkatghanad, Mohammadsadegh Akhondzadeh, Soorena Salari, Mariam Zomorodi-Moghadam, Moloud Abdar, U Rajendra Acharya, Reza Khosrowabadi, and Vahid Salari, "Automated detection of autism spectrum disorder using a convolutional neural network," *Frontiers in neuroscience*, vol. 13, 2020.

[17] Rushil Anirudh and Jayaraman J Thiagarajan, "Bootstrapping graph convolutional neural networks for autism spectrum disorder classification," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 3197–3201.