Assignment 3: Reimplementation of ASD-DiagNet: A Hybrid Learning Approach for Detection of Autism Spectrum Disorder Using fMRI Data

Shivam Kumar Sharma
shivam.krsharma44@gmail.com

Abstract

Autism Spectrum Disorder (ASD) is a lifelong neurodevelopmental brain disorder that impacts many children and adults. The current psychiatric diagnostic process is based purely on the behavioral observation of symptomology (DSM-5/ICD-10). There is no quantitative test that can be prescribed to patients that may lead to a definite diagnosis. The diagnosis of such disorders is currently made by experts who employ standard questionnaires and look for specific behavioral markers through manual observation. Such methods for diagnosis are subjective, difficult to repeat, and costly. This assignment aims to replicate the work done in this field by [3]. Code used in this assignment is available at GitHub and the report can be accessed from Overleaf.

1. Introduction

ASD is a complex developmental condition that impacts the nervous system and affects an individual's overall cognitive, emotional, social and physical health. The overlapping of symptoms with other mental disorders makes it difficult to diagnose by doctors. Quantitative analysis of brain imaging data can provide valuable biomarkers that result in a more accurate diagnosis of brain diseases. So the goal of the work by [3] was to improve upon the work of [5] that aimed to classify subjects suffering from ASD from healthy control subjects using fMRI data.

Various studies have been performed in this field. Authors in [8],[1] have used Pearson's Correlation as a feature for their model, similar to what used for this reimplementation task. [8] uses CNN to aggregate the features. The architecture of[1] is very similar to the ASD-DiagNet except that they have used a deep neural network for later classification. [1] claims that they obtained better results than DiagNet and have the current best accuracy score if the dataset is ABIDE.

[4],[7],[6] have used behavioral markers for classification instead of using just the time-varying BOLD signals. Behavioral markers include pose estimation like calculating pitch, yaw, roll, response time and speed of movement of head. [7] used self-simulatory behaviors like repetitive movement of body parts such as arm flapping, headbanging and spinning for classification.

2. Methodology

The methodology can be divided into three basic steps:

1. Feature extraction: In this step, Pearson's Correlation Coefficient is calculated for the atlas. [3] used CC200 parcellation, so there are (199*200) /2 = 19900 unique features available for each subject(due to symmetry, only the upper right triangle was used). Also, out of 19900 features, only 1/4th highest and 1/4th lowest values were extracted. This was done by creating a mask from the average correlation matrix (computed taking all the samples into account).

2. Augmentation: For augmentation Synthetic Minority Over-sampling Technique was used. SMOTE generates synthetic data in feature space by using the nearest neighbors of a sample. Five nearest neighbors were considered for this. [3] have used a similarity metric called Extended Frobenius Norm (EROS) for finding the nearest neighbors. Also, augmentation was performed only on the training data after the split and not on the testing data.

3. Training: The model architecture consists of an autoencoder and a Single layer Perceptron for classification. The autoencoder is trained to minimize its reconstruction error, computed as the Mean Squared Error between the input and its reconstruction. In comparison, the SLP network is trained by minimizing the Binary Cross Entropy loss. The autoencoder and the SLP classifier are trained simultaneously. This can potentially result in obtaining low dimensional features that may contain discriminatory information for the classification task.

3. Dataset

The data used in the experiment is preprocessed using CPAC pipeline. The dataset used in [3] has records of about 1035 patients obtained from ABIDE-1, while the same dataset had records of 1102 patients when used in this reimplementation assignment. The phenotypic file available on the ABIDE website has few FILE_ID columns mentioned

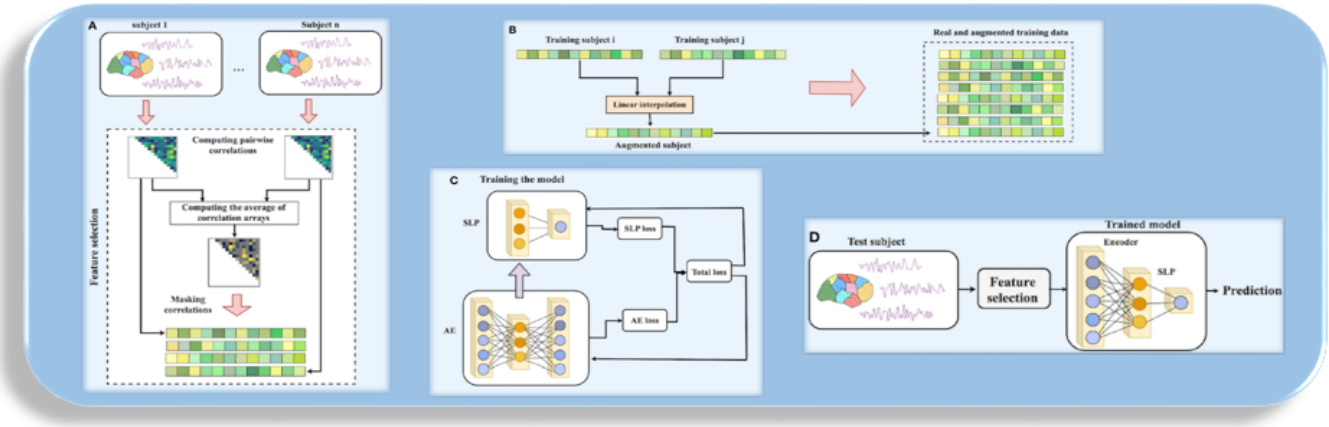

Figure 1. Workflow of ASD-DiagNet: (A) Feature Extraction. (B) Augmentation. (C) Joint training of Autoencoder and SLP. (D) For a test subject, the features are extracted using the mask generated in part A, followed by passing the features through the encoder part of the autoencoder, and finally predicting its label using the trained SLP.

Atlas	My Results	Eslami et al.[3]	Heinsfeld et at.[5]
CC200(+)	70.8	70.3	-
CC200(-)	70.3	69.4	65.4
AAL(+)	68.0	67.5	-
AAL(-)	66.2	64.5	63.5

Table 1. Accuracy Scores of SLP classifier. '+' indicates with augmentation and '-' indicates without augmentation.

Atlas	Accuracy Scores
CC200(+)	68.4
CC200(-)	66.7

Table 2. Results when SLP was used only after autoencoder layer was fully trained.

as 'no_filename'. These 'no_filename' rows were renamed according to their respective 'SITE_ID' and 'SUB_ID' in my reimplementation task which accounts for the difference in the number of patients. Table 1 and Table 2 contain results on the dataset that has renamed subjects and for the rest results were obtained without renaming the subjects.

4. Results

The results obtained for different atlases can be seen in Table 1. There is a marginal increase in accuracy in all the atlases because the dataset that I took into account contains 67 more subjects than the one used by [3]. This reinforces the fact that if more training examples are available, we may get a better accuracy score.

Usually, an autoencoder is fully trained such that its reconstruction error is minimized, then the features from the bottleneck layer are used as input for training the SLP classifier separately. In contrast, here, the autoencoder and the SLP classifier were trained simultaneously. So to verify this, the code was modified a bit to use SLP only after the autoencoder layer was fully trained. The results obtained for this are displayed in Table 2. We see a drop in accuracy in this case and the results indeed are in line with author's assumption.

Accuracy score on the same dataset as [3] (i.e., containing 1035 subjects) for CC200 atlas with augmentation was found out to be 70.3% and without augmentation was 69.7%.

Table 3 shows intrasite accuracy scores of five low-performing sites and three high-accuracy sites. It contains results from data preprocessed under CPAC and NIAK pipeline. Looking at these eight sites, we can say NIAK preprocessed data on average performs better on sites where CPAC gives low accuracy and vice versa. Further investigation is needed to determine the exact reason behind this as there are many different parameters taken into account in these preprocessing pipelines. Two subsets of the original dataset were created by removing the first five low accuracy score sites and the following three high accuracy score sites (see Table 3). Classification results of both the subsets are displayed in table 4. From this, we can deduce that reducing the dataset not necessarily mean a reduction in accuracy.

Also, leave one site out validation was performed on a subset of ABIDE data containing ten sites. It was found that the average accuracy score stayed around 61%.

5. Discussion

The accuracy score is relatively low compared to scores in other domains. One of the reasons behind low accuracy scores may be amount of training data and features used.

Sr. No. and Site $--> $	1. MaxMun	2. SBL	3. Caltech	4. Trinity	5. SDSU	6. KKI	7. UCLA	8. OHSU
Eslami et al. (CPAC)	48.6	51.6	52.8	54.1	63	69.5	73.2	82
My Result (CPAC)	48.3	41.6	47.6	56.8	60.6	69.2	67.2	80.1
My Result (NIAK)	49.5	53	61.9	67.6	55.1	56.6	55.5	46.9

Table 3. Site wise classification scores of dataset pre-processed with different pipeline.

Subset	Size	Accuracy
Low Removed (1 to 5)	833	71.9
High Removed (6 to 8)	863	69.1

Table 4. Results of subset of original dataset.

If we compare this with a typical architecture that uses an image as input of, let's say, size 512*512, the input feature in use is close to 2.6 lakh and thousands of these images are taken as input for training the network. But here, only around 10 thousand features with almost 1000 examples are available for training and testing. Another Reason might be that there is no standard atlas, although there are several atlases available like Cracdok's CC200 or CC400, Gordon atlas and AAL atlas but there is no common consensus on which one is the best.

Most of papers like [3],[8],[1],[2] did not make much change in architecture, kept the model not too deep, and used simple architecture. One of the reasons to explain this would be that the number of features and examples used here is so low that a complex or deep architecture will not be able to learn weights to be good enough for classification. So simple architecture is predominant when using Pearson's correlation coefficient as features. Also, it is evident from such studies that the accuracy score saturates around 71% when using these features as input. So other features and methods should be explored, such as the ones used in [4],[6],[7] which uses behavioral data like head movement and facial expression to obtain a robust prediction model.

6. Conclusion

In [3], authors have designed and developed method for classifying brain scans that exhibit ASD from healthy controls scans. Other methods like [7],[6] have taken into account features that are close to what doctors consider while doing their diagnostics, which make these models more robust to misclassification. One thing that can be done in addition to that is to find a way to combine phenotypic information as well like in [2]. Development in these direction can possibly lead to better understanding of the neurobiological underpinning of the Autism Spectrum Disorder.

References

[1] F. Almuqhim and F. Saeed. Asd-saenet: A sparse autoencoder, and deep-neural network model for detecting autism spectrum disorder (asd) using fmri data. *Frontiers in Computational Neuroscience*, 15:27, 2021.

[2] N. C. Dvornek, P. Ventola, and J. 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)*, pages 725–728. IEEE, 2018.

[3] 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*, 13:70, 2019.

[4] J. Hashemi, M. Tepper, T. Vallin Spina, A. Esler, V. Morellas, N. Papanikolopoulos, H. Egger, G. Dawson, and G. Sapiro. Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants. *Autism research and treatment*, 2014, 2014.

[5] 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*, 17:16–23, 2018.

[6] S. Jaiswal, M. F. Valstar, A. Gillott, and D. Daley. Automatic detection of adhd and asd from expressive behaviour in rgbd data. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pages 762–769. IEEE, 2017.

[7] S. S. Rajagopalan and R. Goecke. Detecting self-stimulatory behaviours for autism diagnosis. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 1470–1474. IEEE, 2014.

[8] Z. Sherkatghanad, M. Akhondzadeh, S. Salari, M. Zomorodi-Moghadam, M. Abdar, U. R. Acharya, R. Khosrowabadi, and V. Salari. Automated detection of autism spectrum disorder using a convolutional neural network. *Frontiers in neuroscience*, 13:1325, 2020.