North South University

Department of Electrical and Computer Engineering

Senior Design Project

Autism Detection Using Machine Learning and Deep Learning

By

Nawal Ayesha Khan	**ID # 1911301042**
Sarah Jasim	**ID # 1912260042**
Seam E Nur Sami	**ID # 1912489042**

Under the supervision of

Dr. Atiqur Rahman

Associate Professor

Department of Electrical and Computer Engineering

North South University

CSE 499 - Senior Design

Fall 2022

Acknowledgement

Firstly, we would like to thank our course instructor, Dr. Atiqur Rahman, for his direction, management, and coordinated effort in helping us accomplish our Senior Design Project with the dedication, focus, enthusiasm, and resolution needed. We appreciate his ceaseless help, supervision, advice, and recommendation that allowed us to finish this report.

We are also grateful to the Electrical and Computer Engineering (ECE) department of North South University for providing this course and assisting us in developing this project. Finally, we would like to thank our family, friends, and everyone else who has supported us in this journey, as their constant guidance and support assisted us throughout this whole period.

Abstract

Autism Spectrum Disorder (ASD) is a developmental disability caused by differences in the brain. People with ASD often have problems with social communication and interaction, and restricted or repetitive behaviors or interests. They may also have different ways of learning, moving, or paying attention. Typically, ASD patients are analyzed by doctors based on history and behavior observation without the ability to diagnose instantaneously.

In recent years, machine learning based intelligent diagnosis has been evolved to complement the traditional clinical methods which can be time consuming and expensive. In this project, we design and compare multiple ASD (Autism Spectrum Disorder) classification systems using different Machine Learning and Deep Learning algorithms and identify the best performing classifier suitable for the obtained ASD dataset. fMRI (Functional Magnetic Resonance Imaging) images from the Autism Brain Imaging Data Exchange (ABIDE) dataset was used to research the brain imaging of ASD patients. Based on the results, the Feedforward neural network at 68%, followed by the SVM and Logistic Regression classifiers at 67%, while the Naïve Bayes classifier was the least best performing at an overall accuracy of 56%.

Approval

This is to certify that the CSE499 report entitled *Autism Detection Using Machine Learning and Deep Learning*, submitted by Nawal Ayesha Khan (ID: 1911301042), Sarah Jasim (ID: 1912260042), and Seam E Nur (ID: 1912489042) are undergraduate students of the Department of Electrical and Computer Engineering, North South University. This report partially fulfills the requirements for the degree of Bachelor of Science in Computer Science and Engineering on February 15th, 2023, and has been accepted as satisfactory.

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation.

Dr. Atiqur Rahman
Associate Professor
Department of Electrical and Computer Engineering
North South University, Dhaka
Bangladesh

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation.

Dr. Rajesh Palit
Professor and Chair
Department of Electrical and Computer Engineering
North South University, Dhaka
Bangladesh

Declaration

It is hereby acknowledged that:

No illegitimate procedure has been practiced during the preparation of this document. This document does not contain any previously published material without proper citation. This document represents our own accomplishment while being Undergraduate students in North South University. We declare that this CSE499 report entitled *Autism Detection Using Machine Learning and Deep Learning* has not been accepted for any degree and is not currently submitted in candidature for any other degree. We would like to request you to accept this report as a partial fulfilment of the Bachelor of Science degree under the Electrical and Computer Engineering Department of North South University.

Sincerely,

Nawal Ayesha Khan
1911301042

Sarah Jasim
1912260042

Seam E Nur Sami
1912489042

Table of Contents

Chapter 1: Introduction

1.1 Introduction

Autism Spectrum Disorder (ASD), a neuro-developmental disorder, is often accompanied by sensory issues such as an over or under-sensitivity to sights, sounds, smells, or touch. It is a lifelong condition and cannot be cured. Although its main cause is genetic in nature, early detection and treatment can help improve the conditions. The symptoms and signs appear when a child is very young, and early detection of ASD can help both patient and healthcare service providers by prescribing proper therapy and/or medication needed, thus reducing the long-term costs and effects associated with delayed diagnosis. A study conducted by Wiggin et al. found that 33% of children with difficulties other than ASD have some ASD symptoms while not meeting the full classification criteria [4]. Typically, ASD patients are analyzed by doctors based on history and behavior observation without the ability to diagnose instantaneously. This research intends to study the ASD biomarker based on neuroimaging functional Magnetic Resonance Imaging (fMRI) images, which can aid doctors in diagnosing ASD.

Early screening and diagnosis are crucial. Parents, teachers, and other people without special training or qualifications can screen children for autism. Identification of young children with developmental challenges or youngsters at risk of language and communication problems is the goal of screening and monitoring. Early detection will result in early action. The Autism Spectrum Quotient (AQ), the Social Communication Questionnaire (SCQ), and the Modified Checklist for Autism in Toddler are a few of the tools used to assess children's development (M-CHAT). A health care provider can utilize the data acquired throughout the screening and monitoring procedure to make a diagnosis and gain a thorough picture of the social skills or behavior of the kid. However, a proper diagnosis can only be performed by trained health care professionals. In diagnosing patients with autism, health care professionals will use standardized diagnostic instruments. However, behavioral indices like psychological assessment could not get rid of some limitations like subjectivity and reporter-dependency. These methods are error prone which may cause harmful side effects due to overprescribing drugs [7]. Patients with mild expressions of the disease, also known as High Functioning individuals, do not have noticeable speech impairments or cognitive deficits may be misdiagnosed with other personality disorders such as Schizophrenia or the diagnosis may be missed altogether when tested with the traditional diagnostic instruments [20]. Misdiagnosis or missed diagnosis upon evaluation from healthcare professionals is prevalent among high functioning women [21] or children who display behavioral issues as a response to trauma [22]. Misdiagnosis can lead to harmful side effects from the prescription of the wrong medication, and delayed or missed diagnosis can have detrimental effects as well, as ASD requires lifelong treatment to manage the symptoms.
Currently, there are 3 attested, standardized diagnostic instruments often used in diagnosing autism, namely the Autism Diagnostic Observation Schedule (ADOS), Autism Diagnostic Interview-Revised (ADI-R), and Diagnostic and Statistical Manual of Mental Disorders 5 (DSM-5) [17, 18].

Although standardized diagnostic tools have frequently been used by clinicians for ASD diagnostics, a major disadvantage of their usage is that administering them requires a long duration of time to conduct the assessments necessary to interpret the scores. The method of machine learning has been proposed as a solution to this issue. The main aim of machine learning research in the diagnosis of ASD is to minimize the diagnostic time with increased accuracy, with the added benefit of recognizing the best ranked ASD features and diminishing the dimensionality of the input dataset [19]. By reducing the diagnostic time,

ASD patients can receive immediate intervention and the assistance they need to handle their condition.

Hence, researchers have been looking into relatively stable biomarkers of ASD as additional diagnostic evidence. The identification of patterns of activation for ASD and the association of the patterns with neural and psychological components contribute to the understanding of the etiology of mental disorders, and the use of Deep Learning and Machine Learning algorithms has helped substantially in these regards. Most papers have looked into using more computationally intensive algorithms like Support Vector Machines, K-means clustering, Random Forest Classifiers, or Neural Networks.

Images from various layers of the brain are often provided by MRI and fMRI procedures, which are frequently utilized as the input of machine learning algorithms. The brain is split into tiny cubic components called voxels for fMRI data, and a time series representing each voxel's activity across space and time is retrieved. The correlation between two voxels' time series values is used to characterize the statistical relationship between them, sometimes referred to as their functional connectivity [3]. Using only fMRI data for classifying ASD vs TD subjects allows clinicians to assess the subject's condition without being biased in decision making process with other demographic information. Although adding additional data may improve prediction accuracy, our objective is to just use brain fMRI data to identify ASD disorder, which is a more difficult task. We were intrigued in considering functional connectivity as our model features, since alterations in functional connectivity may cause different brain disorders such as Alzheimer's, Schizophrenia, and other disorders.

The focus of this paper is to determine the best performing classifiers for these binary datasets based on fMRI data alone by using the evaluation metrics Recall, Precision, F-measures and Accuracy for each model, as well as by looking at the ROC curves of the machine learning models and the loss curves of the deep learning models. We pre-process our dataset by removing the attributes with missing values and also those that do not offer any benefit during the analyses. The Autism Brain Imaging Data Exchange (ABIDE) dataset was used to research the brain imaging of ASD patients. We compare the state-of-the-art classification methods and identify the best performing classifier that is suitable for our dataset. These findings may inform future development of machine-learning models for psychiatric disorders and may improve the accuracy and speed of clinical assessments.

1.2 Objective

The main objectives of this study have been detailed below:

1. Analyzing the fMRI images of patients from an ASD dataset of the ABIDE database.
2. Exploring various Machine Learning and Deep Learning models in correctly classifying ASD and TD patients.
3. Comparing the state-of-the-art classification methods to determine the best for identifying positive cases of ASD based on fMRI images alone.

Chapter 2: Literature Review

2.1 Related Works

One study [1] applied a deep learning method from Convolutional Neural Network (CNN) variants to detect whether the patients were ASD or non-ASD and extracted the characteristics from neuro-images in fMRI images. The model interprets the accuracy performance of pre-processed images to classify the neural patterns. The Autism Brain Imaging Data Exchange (ABIDE) dataset was used to research the brain imaging of ASD patients. The results achieved using CNN models namely VGG-16 and ResNet-50 are 63.4% and 87.0% accuracy, respectively [See Table 1A and 1B]. This method assists doctors in detecting Autism from a quantifiable method that is not dependent on the behavioral observations of suspected autistic children.

Recall	Precision	Accuracy
$\text{Recall} = \frac{26}{26+11}$	$\text{Precision} = \frac{26}{26+4}$	$\text{Accuracy} = \frac{26}{26+11+4}$
$= 0.703 * 100$	$= 0.867 * 100$	$= 0.634 * 100$
$= 70.3\%$	$= 86.7\%$	$= 63.4\%$

Table 1A: The performance of VGG-16

Recall	Precision	Accuracy
$\text{Recall} = \frac{40}{40+4}$	$\text{Precision} = \frac{40}{40+2}$	$\text{Accuracy} = \frac{40}{40+4+2}$
$= 0.909 * 100$	$= 0.952 * 100$	$= 0.87 * 100$
$= 90.9\%$	$= 95.2\%$	$= 87.0\%$

Table 1B: The performance of ResNet-50

In the case of high-dimensional datasets such as fMRI data, deep learning models are much more efficient than traditional machine learning method. Hence, [2] investigated patterns of functional connectivity that objectively identify ASD participants from functional brain imaging data, and attempted to unveil the neural patterns that emerged from the classification. DL algorithms were applied to identify ASD patients from a large brain imaging dataset (ABIDE dataset), based solely on the patients' brain activation patterns. Data was selected from the C-PAC preprocessing pipeline. The fMRI data was slice time corrected, motion corrected, and the voxel intensity was normalized. Denoising auto-encoders were used to train the predictive model for better generalization. The deep neural network achieved a mean classification accuracy of 70% (sensitivity 74%, specificity 63%) from cross-validation folds, and a range of accuracy of 66% to 71% in individual folds. The SVM classifier achieved mean accuracy of 65% (sensitivity 68%, specificity 62%); while the Random Forest classifier achieved mean accuracy of 63% (sensitivity 69%, specificity 58%) [See Table 2].

Method	Accuracy	Sensitivity	Specificity	Time
SVM	0.65	0.68	0.62	1 m 37 s
RF	0.63	0.69	0.58	20 m 55 s
DNN	0.70	0.74	0.63	32h 52 m 36 s

Table 2: Comparison of DNN, RF, and SVM trained on the entire dataset

In [3], they investigated the discriminative power of features that have been extracted using MLP by feeding them to an SVM classifier on a dataset consisting of only fMRI images. The dataset was from the ABIDE database and used C-PAC pipeline (including motion correction, slice timing correction, nuisance signal removal, low frequency drifts, and voxel intensity normalization), similar to [1], & 5-fold Cross-validation, then compared the average accuracy, sensitivity and specificity of each method. A technique called SMOTE was used for performing data augmentation to avoid overfitting, and a tool called ATM was used for hyper-parameter tuning. The entire method is called Auto-ASD-Network, and it was able to achieve higher accuracies for 4 different datasets [See Table 3] and significantly improved the results for SVM classifiers.

Site	Method	Accuracy	Sensitivity	Specificity
OHSU	SVM	54	0	100
	SVM-ATM	**72.3**	56.6	83.3
NYU	SVM	57.1	0	100
	SVM-ATM	**69.1**	53.3	81
USM	SVM	64.7	100	0
	SVM-ATM	**69.6**	84.3	42
UCLA	SVM	55.1	100	0
	SVM-ATM	**72.2**	83.8	57

Table 3: Performance comparison of traditional SVM and SVM optimized with ATM (SVM-ATM)

Researchers of [4] attempted to generate predictive models for toddlers using RF, NB and SVM classifiers. RF showed the best accuracy with 80.9% compared to 80% for NB and SVM [See Table 4A]. They also conducted an evaluation about the predicted performance of models based on different cortical features such as thickness (CT), surface area (SA) and regional average cortical volume (CV). They collected their own data from 85 participants aged 18-37 months, from which 46 were diagnosed with ASD. Their work confirmed that classification is significantly more accurate for CT compared to CV and SA. Models based on CT had 75% accuracy while CV and SA had 68% and 72% respectively. [See Table 4A and 4B] This may indicate that cortical thickness might be the most prominent feature of abnormal cortex in ASD.

	ACC (%)	SEN (%)	SPE (%)	AUC	MCC
RF	80.9 ± 1.5	81.3 ± 1.2	81 ± 4	0.88 ± 0.01	0.62 ± 0.03
NB	80 ± 2	86 ± 3	73 ± 3	0.79 ± 0.02	0.59 ± 0.04
SVM	80 ± 4	84.8 ± 1.5	74 ± 7	0.79 ± 0.04	0.59 ± 0.07

Table 4A: Performance Metrics Comparison of Three Predictive Model

	ACC (%)	SEN (%)	SPE (%)	AUC	MCC	F1 (%)
Model based on CV	68 ± 2	70 ± 6	66 ± 5	0.72 ± 0.01	0.35 ± 0.05	70 ± 3
Model based on CT	75 ± 2	80 ± 3	70 ± 3	0.80 ± 0.02	0.50 ± 0.04	77.8 ± 1.9
Model based on SA	72 ± 2	77 ± 4	66 ± 4	0.77 ± 0.04	0.44 ± 0.05	75 ± 2

Table 4B: Performance Metrics for Three Features Generated by RF

In [5], they conducted research comparing the different preprocessing pipelines used in the ABIDE dataset to identify the most promising preprocessing pipeline for deep learning models. MLP models of 4 different configurations (2 layers of 128-64, 256-128, etc.) were built using the TensorFlow framework to classify ASD from TD using the rs-fMRI data, among which 128-64 DNN had the highest accuracy (75.27%) and recall (74%), whereas 256-128 DNN had the highest precision (78.37%). The results have been summarized in Figure: 1 below. Motivated by this work, it is expected that a combination of topological analysis and DNN using the right preprocessing pipeline (CPAC) would be a promising direction to identify the biomarker ROIs for the ASD group.

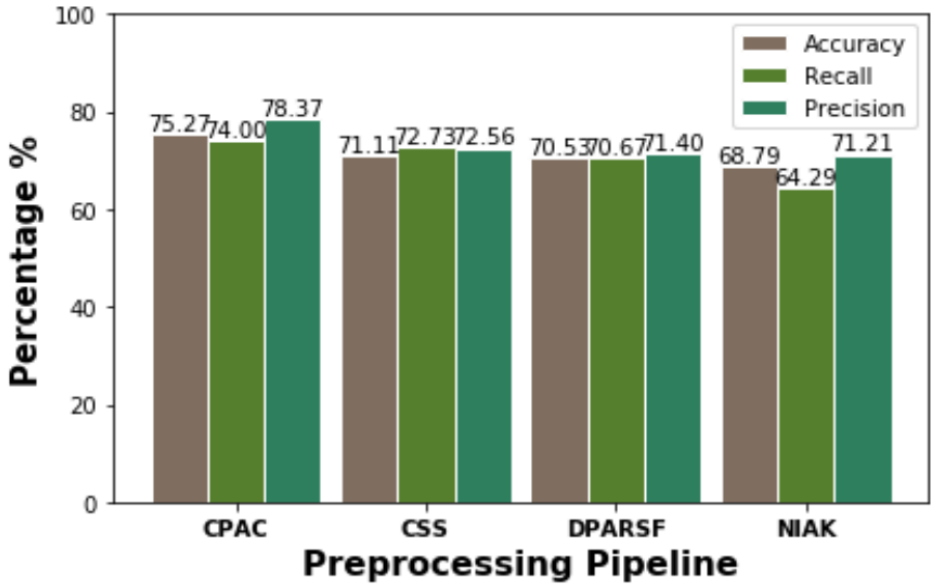

Figure 1: Results of all 4 pipelines by Accuracy, Recall, and Precision.

In paper [6] they used a cross-species machine learning framework that used connectome-based features learned from a primate genetic model of autism and then built a classifier for diagnostic utility in humans. Nine core regions predominantly distributed in frontal and temporal cortices [see Figure 2] were identified in monkeys using group lasso algorithm and those were used as templates to construct the monkey-derived classifier that was used

in the classification of human autism. The classifier based on these core regions achieved an accuracy of 61.31% on the human brain. Similarly, in the ABIDE-I data set, this group lasso algorithm identified four core regions. The classifier based on the four core regions achieved an accuracy of 60.40% (95% CI=52.04, 68.21), with a sensitivity of 53.33% (95% CI=40.10, 66.14), a specificity of 65.17% (95% CI=54.26, 74.76) in the classification.

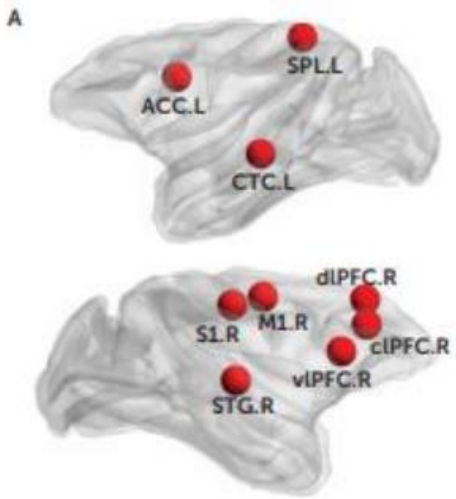

Figure 2: Nine identified core regions out of 94 nodes of the monkey brain

In [16], the authors investigated the predictive power of sMRI (Structural Magnetic Resonance Imaging) in ASD using 373 ASD and 361 TDC male subjects from the ABIDE dataset. The brain morphometric features were derived using FreeSurfer, a brain imaging software package for the analysis and visualization of structural, functional, and diffusion of neuroimaging data, and classification was performed using RF, SVM & GBM. The paper also investigated the relationship between ADOS and autism class probability, an autism score produced by a classifier. Seven different morphometric properties: surface area, volume, Gaussian curvature, mean curvature, folding index, thickness, thickness standard deviation of cortical and subcortical structures were derived using the recon-all workflow of FreeSurfer. Many features were found to be common among the sites, as shown in Figure 3. 3 experiments were carried out, where subjects across all sites were used and classification was performed with and without selection of optimal features. The AUC metric for each classifier based on each morphometric feature has been displayed in Figure 4 below. Based on their findings, the accuracy across all sites improved to 67% when IQ and age information were added to morphometric features. The highest RF classification accuracy of 60% (sensitivity = 57%, specificity = 64) was achieved when mtry = 25 under LOOCV. This study suggests that high-performance diagnostic models can be built with very few features from sMRI data in a small well-matched dataset at the cost of overfitting.

	AUC | *p*-value		
	RF	SVM	GBM
Volume	0.59 | 3E-5	0.58 | 8.1E-5	0.58 | 3E-4
Area	0.59 | 4.3E-5	0.59 | 1.1E-5	0.57 | 8.3E-4
Thickness	0.51 | 0.64	0.57 | 4.5 E-4	0.54 | 0.046
Folding Index	0.57 | 2E-3	0.52 | 0.43	0.57 | 1.6E-3
Mean Curvature	0.56 | 2.6E-3	0.56 | 3.6E-3	0.57 | 1.6E-3
Gauss Curvature	0.57 | 6.4E-4	0.53 | 0.12	0.57 | 2E-3
All	0.60 | 8.3E-6	0.59 | E-5	0.60 | 6.6E-6

Figure 3: Discriminative power of morphometric properties based on AUC metric

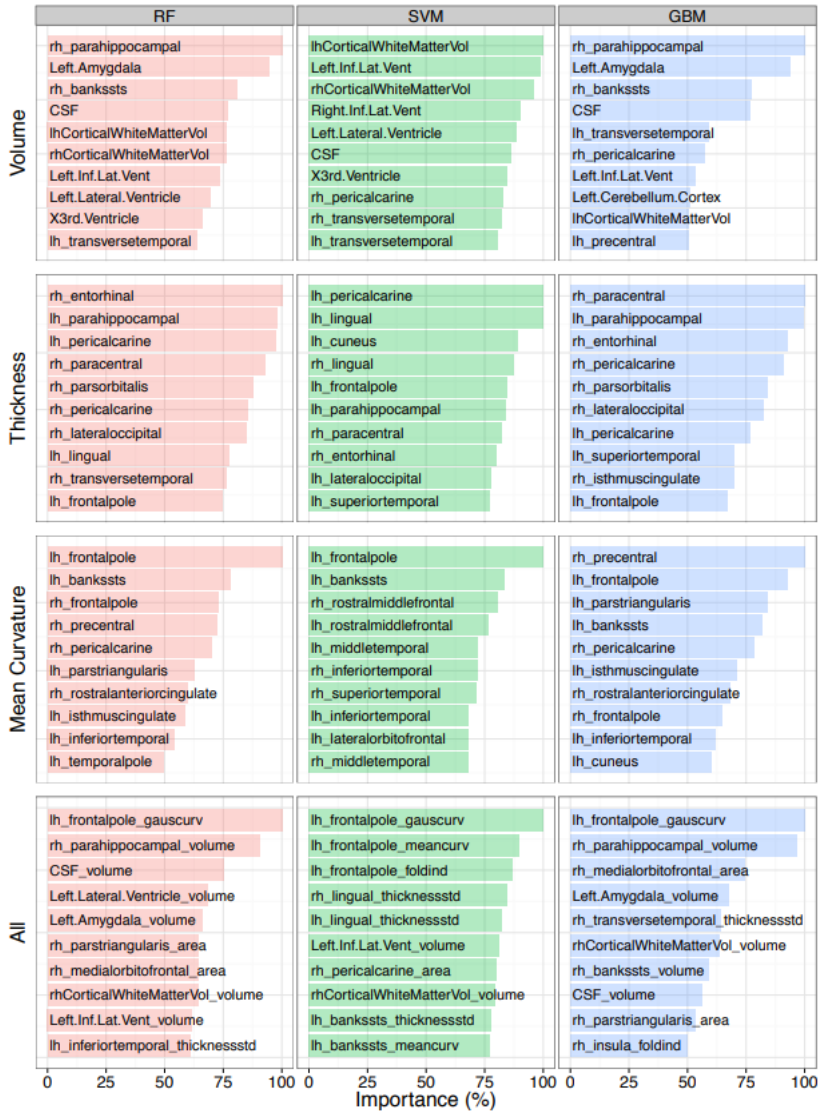

Figure 4: Top 10 features for each morphometric property for ASD vs. TDC classification

Chapter 3: Methodology

We followed the standard procedure by performing data pre-processing and Exploratory Data Analysis (EDA) on the dataset. The dataset was imported and pre-processed to get rid of any null or incorrect values, and columns that were not required were omitted completely. The data was then visualized using EDA (Exploratory Data Analysis) in Jupyter Notebook to get a better understanding of our data set.

3.1 Dataset

The dataset used in this project was from the ABIDE I database. It includes resting state fMRI data from 16 international imaging sites that have aggregated and are openly sharing neuroimaging data from 539 individuals suffering from ASD and 573 typical controls. These 1112 datasets are composed of structural and resting state functional MRI data along with an extensive array of phenotypic information, which has been displayed in Figure 5.

Autism Brain Imaging Data Exchange (ABIDE) Dataset

Site	Count			Age Range
	ASD	Control	Total	
Caltech	5	10	15	17–56
CMU	6	4	10	19–40
KKI	12	20	32	8–13
LEUVEN	26	30	56	12–32
MAX_MUN	19	27	46	7–58
NYU	74	98	172	6–39
OHSU	12	13	25	8–15
OLIN	14	14	28	10–24
PITT	24	26	50	9–35
SBL	12	14	26	20–64
SDSU	8	18	26	9–17
Stanford	12	13	25	8–13
Trinity	19	25	44	12–26
UCLA	48	37	85	8–18
UM	46	73	119	8–29
USM	43	24	67	9–50
YALE	22	18	40	8–18
TOTAL	402	464	866	6–64

Figure 5: Phenotypic information summary of the participants from the ABIDE dataset.

Only the fMRI data was used without including other demographic information, such as age, gender and IQ. This is because the paper's goal was to solely rely on brain fMRI data for detecting ASD without being biased with other demographic information, similar to [16]. The dataset before pre-processing has been shown below in Figure 6.

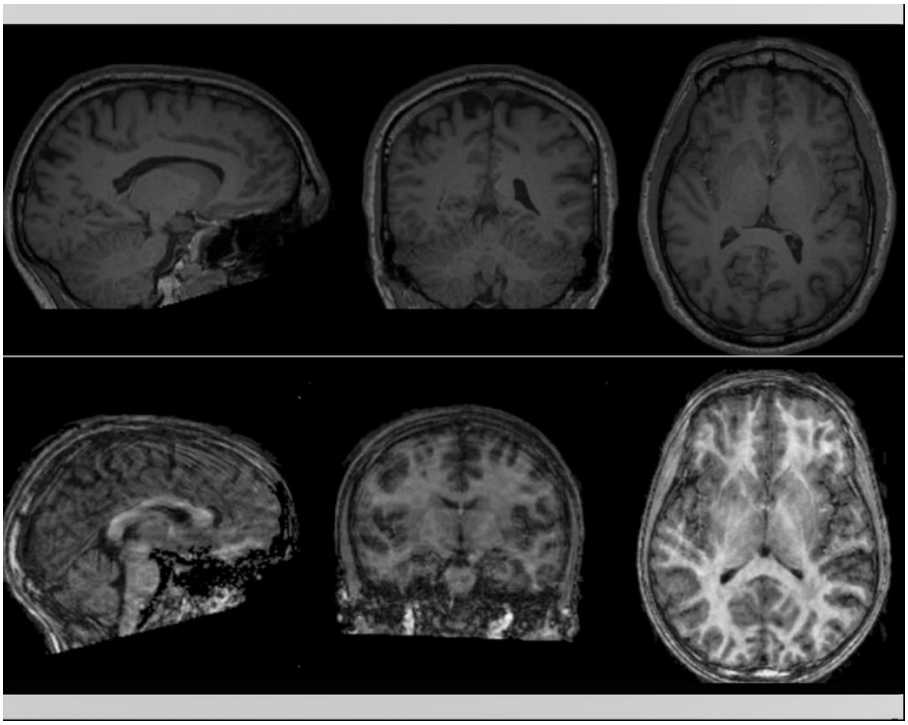

Figure 6: Example of brain scan images used in dataset. Top panel shows scans of a Neuro-divergent/Autistic brain (Class: 0). Bottom panel shows scans of a Neuro-typical/Typically Developing brain (Class: 1).

The results of the EDA have been displayed in the figures below. A pie chart has been plotted to show the distribution of classification between ASD and TD diagnosed patients, shown in Figure 7.

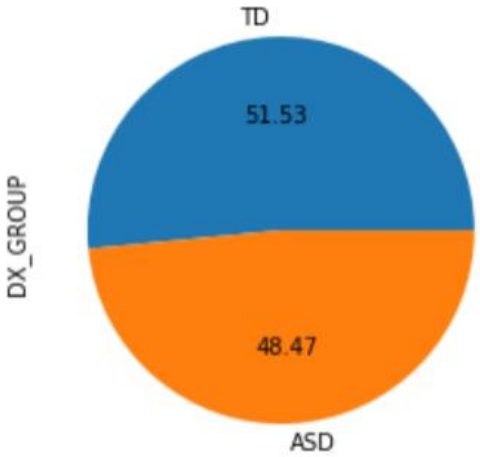

Figure 7: Pie chart of ASD and TD patients

For pre-processing the data, we used the C-PAC pipeline, which includes steps for the fMRI data to be time slice corrected, motion corrected and normalized. These steps were necessary as the dataset consists of 16 different data sites that each had different parameters for taking the MRI scans. Parameters such as keeping the eyes open during scan, the repetition time and time to echo were different for each of the sites. After screening, our final dataset had 871 subjects, of which 403 had ASD and 468 were TD.

3.2 . Feature Selection

For classification between ASD and TD, we used the functional connectivity between regions of interests (ROIs) as the main feature. Functional connectivity refers to the correlation of activity levels between two ROIs based on the fMRI data. The range is from 1, which indicates high correlation between the regions, to -1 indicating the regions are anti-correlated. We used nilearn, which is a package from python, to select and calculate the correlations.

We used the BASC Multiscale 2015 brain parcellation to get the fMRI data on the labelled regions of interests before finding the correlations. The parcellation contained 64 different regions, and as we used all the correlations between each pair of regions, we got $\frac{N(N-1)}{2}$ distinct correlations, where N is the number of regions. Thus, we had a total of 2,016 features for each brain image, as displayed in Figure 8.

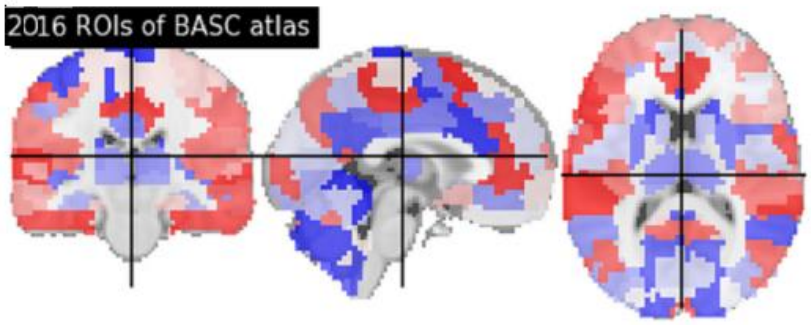

Figure 8: Regions of Interest (ROIs) for each brain image of the BASC atlas

3.3 . Machine Learning Algorithms Used

For the Machine learning models, the dataset was first split for training and testing samples, and the models were defined, fitted, and then trained and tested on each respective data portion. The results are then evaluated by plotting a confusion matrix and calculating the accuracy, after which 5-fold cross validation or Leave One Out (LOO) cross validation is applied, and finally, ROC curves are plotted for all models (except for SVM model). The Machine Learning algorithms implemented have been listed below.

i) SVM Classifier: Support vector machines (SVMs) are a set of supervised learning methods used for classification, regression, and detection of outliers. SVMs are known for their effectiveness in high dimensional spaces, and in cases where the number of dimensions is greater than the number of samples. It constructs a hyperplane, or a set thereof, within a high-dimensional space, which can be utilized for classification [9]

ii) Decision Tree Classifier: Decision tree classifiers are supervised machine learning models that work by splitting data into a series of binary decisions, which allow us to traverse down the tree based on these decisions until we end at a leaf node, which will return the predicted classification. They're generally faster to train than other algorithms such as neural networks, and can handle high dimensional data with high degrees of accuracy.[8]

iii) Random Forest Classifier: It is an ensemble of decision tree classifiers trained on various sub-samples of the dataset and averaged to gain a better accuracy than any individual tree

trained on the whole dataset. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree. [10]

iv) Logistic Regression: Logistic Regression is a linear model for classification rather than regression. In this model, the probabilities describing the possible outcomes of a single trial are modeled using a logistic function. It can handle both dense and sparse input. While it may technically not qualify as a classification method, it represents a discrete choice model, and we therefore use it as such. [11]

v) Naïve Bayes: Naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong independence assumptions between the features, thus vastly simplifying the computational space. They are among the simplest Bayesian network models, but coupled with kernel density estimation, they can achieve high accuracy levels. [12]

vi) Gradient Boosting: Gradient boosting is a method standing out for its prediction speed and accuracy, particularly with large and complex datasets. It helps reduce the bias error of the model [13]. This algorithm builds an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions [14].

3.4 . Deep Learning models used

For the Deep learning models, the dataset was first split for training and testing samples, and then scaled and converted to tensors. Data loaders are used to manage and iterate over the batches. Subsequently, the models were defined with the necessary functions for forward passing, optimization, and loss computation. The models are then trained and tested on each respective batch. The results are then evaluated by plotting a confusion matrix and calculating the accuracy. The Deep Learning algorithms implemented have been listed below.

i) Multi-Layer Perceptron (MLP): A multilayer perceptron is a special case of a feedforward neural network where every layer is a fully connected layer, and in some definitions, the number of nodes in each layer is the same. Furthermore, in many definitions the activation function across hidden layers is the same. MLPs are preferable when one knows very little about the structure of the problem. Using fully connected layers only, which defines an MLP, is a way of learning structure rather than imposing it [15].

The layers in the MLP defined in this study consist of 2 hidden layers, consisting of 256 and then 50 nodes respectively. The ReLu and Sigmoid activation functions and dropout regularization are also applied on the network. Back-propagation is used to update the parameters, and Adam optimizer has been applied to act as a replacement optimizer for the gradient descent. The configuration of layers in the MLP has been displayed in Figure 9.

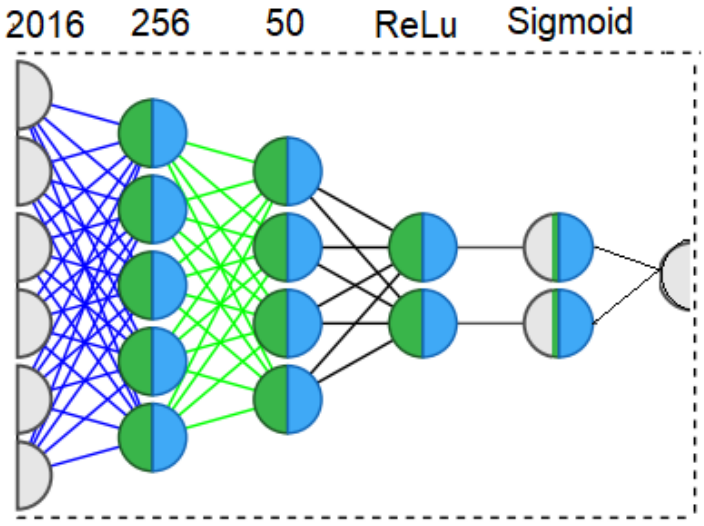

Figure 9: Configuration of layers in MLP classifier

ii) Feed-Forward Neural Network: A feedforward neural network involves sequential layers of function compositions. Each layer outputs a set of vectors that serve as input to the next layer, which is a set of functions. There are three types of layers: an Input layer for the raw input data, Hidden layer(s) to apply functions to either inputs or outputs of previous hidden layers, and an Output layer for the final function or set of functions [15].
The layers in the FFNN defined in this study, similar to the MLP, consist of 2 hidden layers, consisting of 256 and then 50 nodes respectively. The ReLu activation function and dropout regularization are also applied on the network. Back-propagation and Adam optimizer have been applied too. The configuration of layers in the FFNN has been displayed in Figure 10.

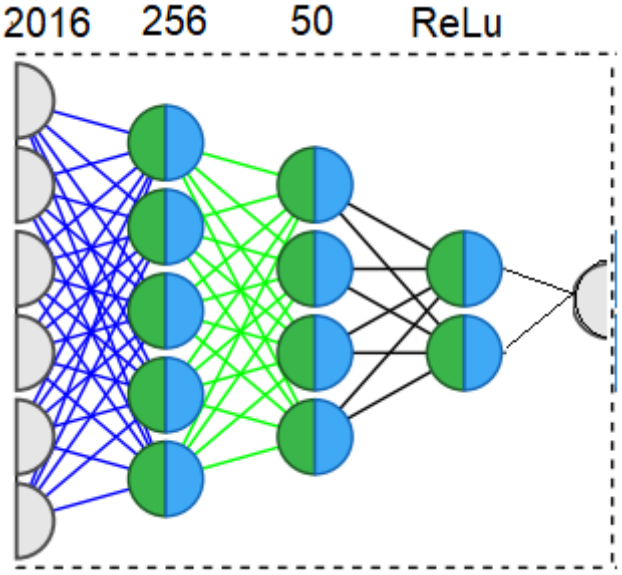

Figure 10: Configuration of layers in Feedforward Neural Network

While multiple papers such as [1], [3], [5], [6], [16] have utilized Convolutional Neural Networks (CNN), due to their extensively long training duration and large amount of computational power needed, both of which are not available to the authors of this study, we have excluded their usage in the DL models for our paper.

Chapter 4: Results

The results of our study have been summarized below in Table: 5 with their respective precision, recall, f1-score, and overall accuracy scores.

MODEL	PRECISION		RECALL		F1-SCORE		ACCURACY
	ASD	*TD*	*ASD*	*TD*	*ASD*	*TD*	*Overall*
SVM Classifier	0.66	0.69	0.62	0.72	0.64	0.70	0.67
Logistic Regression	0.63	0.70	0.68	0.66	0.65	0.68	0.67
Gradient Boosting	0.63	0.62	0.51	0.73	0.56	0.67	0.62
Decision Tree	0.58	0.63	0.56	0.65	0.57	0.64	0.61
Random Forest	0.64	0.63	0.47	0.78	0.54	0.70	0.60
Naïve Bayes	0.59	0.65	0.60	0.64	0.60	0.65	0.56
MLP Classifier	0.66	0.66	0.60	0.71	0.63	0.68	0.66
Feedforward Neural Network	0.64	0.73	0.70	0.67	0.67	0.70	0.68

Table 5: Summary of evaluation metrics for each ML and DL model

4.1. Metrics for Evaluation

The feasibility of the predictions made by our models will be cross-validated and critiqued through standard Precision, Recall, and F1-scores, as well as Recipient Operating Classification (ROC) curves for the ML models. The definition of these metrics are as follows:

i) Accuracy: It is a machine learning classification model performance metric that is defined as the ratio of true positives and true negatives to all positive and negative observations. In other words, accuracy tells us how often we can expect our machine learning model will correctly predict an outcome out of the total number of times it made predictions.

$$\textbf{\textit{Accuracy Score}} = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative}$$

ii) Precision: It measures the proportion of positively predicted labels that are actually correct. The precision score is a useful measure of the success of prediction when the classes are very imbalanced.

$$\textbf{\textit{Precision Score}} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

iii) Recall: It represents the model's ability to correctly predict the positives out of actual positives. A high recall score indicates that the model is good at identifying positive examples. Recall score can be used in the scenario where the labels are not equally divided among classes.

$$\textbf{\textit{Recall Score}} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

iv) F1-Score: An F1 score is the harmonic mean of Precision and Recall; it therefore is commonly utilized as a classification evaluation metric due to weighing each metric evenly. This is a useful measure of the model in the scenarios where one tries to optimize either of precision or recall score and as a result, the model performance suffers.

$$\textbf{\textit{F1-Score}} = \frac{2 * Precision\ Score * Recall\ Score}{Precision\ Score + Recall\ Score}$$

v) K-fold Cross-validation: It is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the

other used to validate the model. In *k*-fold cross-validation, the data is first partitioned into *k* equally (or nearly equally) sized segments or folds. Subsequently *k* iterations of training and validation are performed such that within each iteration a different fold of the data is held-out for validation while the remaining $k - 1$ folds are used for learning. For our ML models, we have used 5-fold cross validation for optimum results. Data partitioning using 5-fold cross-validation has been displayed in Figure 11.

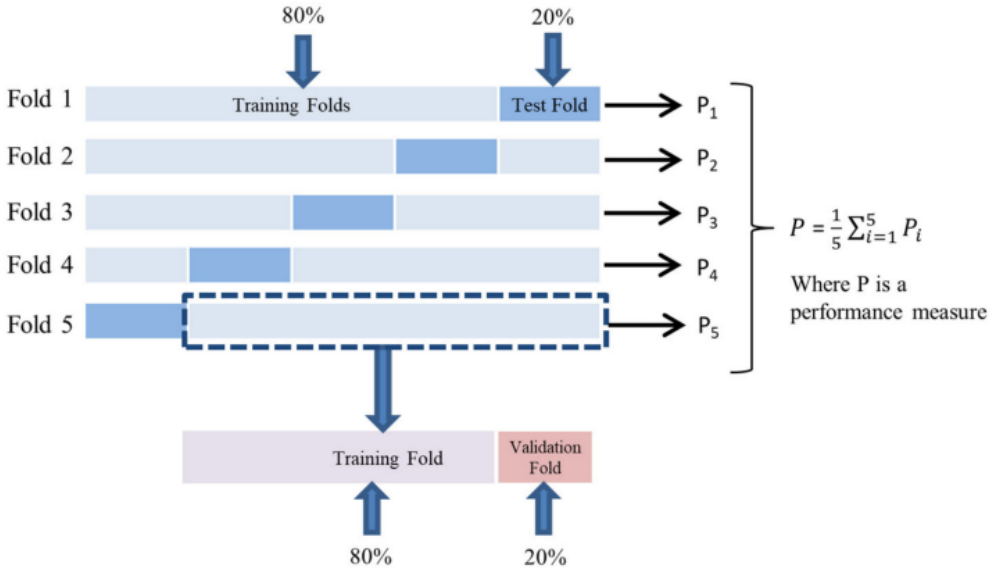

Figure 11: Data partitioning using 5-fold cross-validation

4.2. Results of ML model simulations

The SVM classifier model, along with the Logistic Regression model, had the highest overall accuracy among the ML models at 67%. It correctly predicted 50 ASD cases out of 81 instances, and 68 cases of TD out of 94. It had a precision value of 0.66 (66%), a recall value of 0.62 (62%), and an overall f1-score of 0.64 (64%). After 5-fold cross validation, the average CV score came to 0.63 (63%). The confusion matrix for the SVM classifier has been displayed in Figure 12 below.

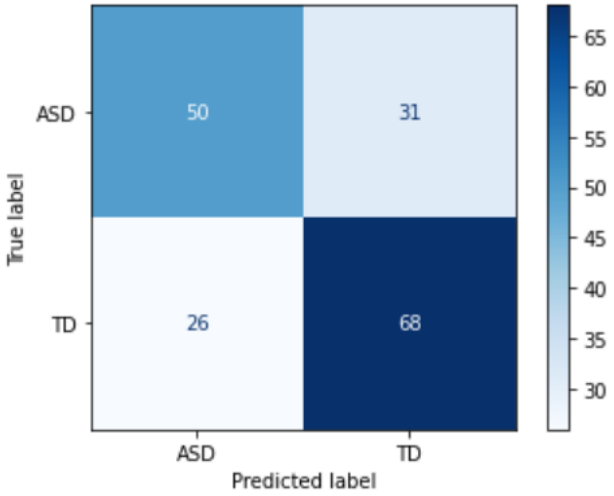

Figure 12: Confusion Matrix for SVM model

Along with the SVM, the Logistic Regression model also had the highest overall accuracy of 67%. It correctly predicted 55 ASD cases from 81, and 62 cases of TD out of 94. It had a recall value of 0.68 (68%), a precision value of 0.63 (63%), and an overall f1-score of 0.65 (65%). After 5-fold Leave One Out cross validation, the average CV score came to 0.61 (61%). The confusion matrix and ROC (Receiver Operating Characteristic Curve) graph for the LR classifier have been displayed in Figures 13 and 14 below.

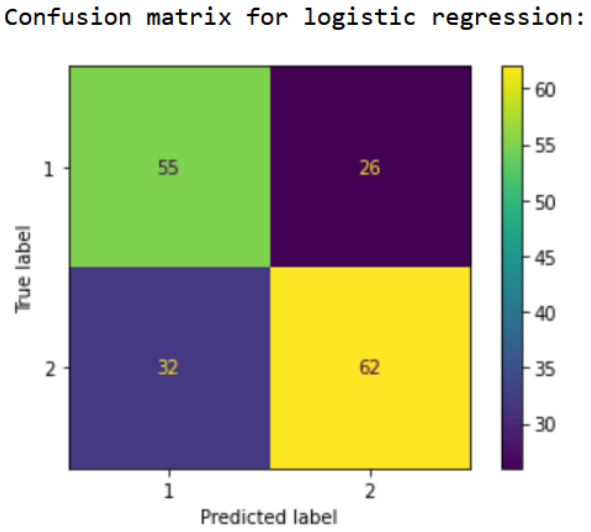

Figure 13: Confusion Matrix for LR model

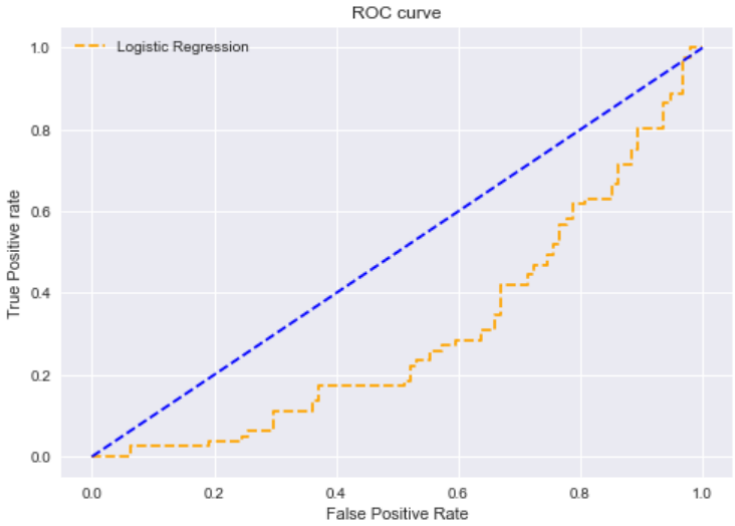

Figure 14: ROC graph for LR model

The Gradient Boosting had the second highest overall accuracy of 62%. It correctly predicted 38 ASD cases from 83, and 73 cases of TD out of 92. It had a precision value of 0.63 (63%), a recall value of 0.51 (51%), and an overall f1-score of 0.56 (56%). After 5-fold cross validation, the average CV score came to 0. 56 (56%). The confusion matrix and ROC graph for the GB classifier have been displayed in Figures 15 and 16 below.

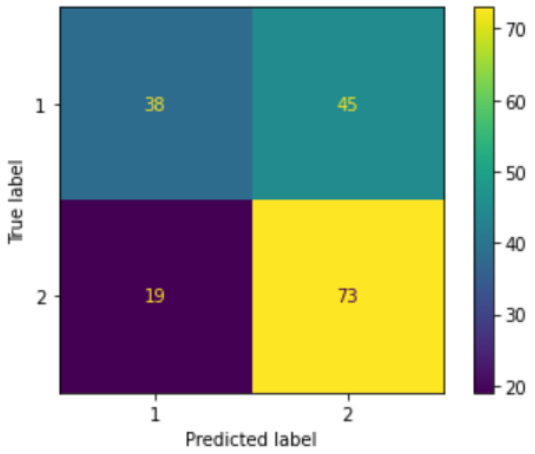

Figure 15: Confusion Matrix for GB model

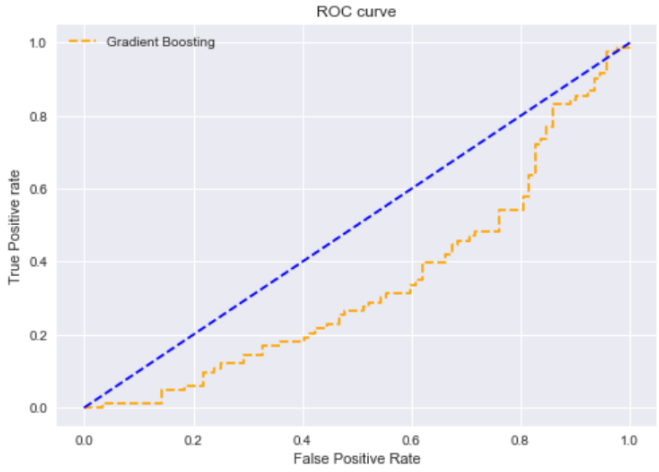

Figure 16: ROC graph for GB model

The Decision Tree classifier had an overall accuracy of 61%. It correctly predicted 45 ASD cases from 81, and 61 cases of TD out of 94. It had a recall value of 0.56 (56%), a precision value of 0.58 (58%), and an overall f1-score of 0.57 (57%). After 5-fold Leave One Out cross validation, the average CV score came to 0. 45 (45%). The confusion matrix and ROC graph for the DT classifier have been displayed in Figures 17 and 18 below.

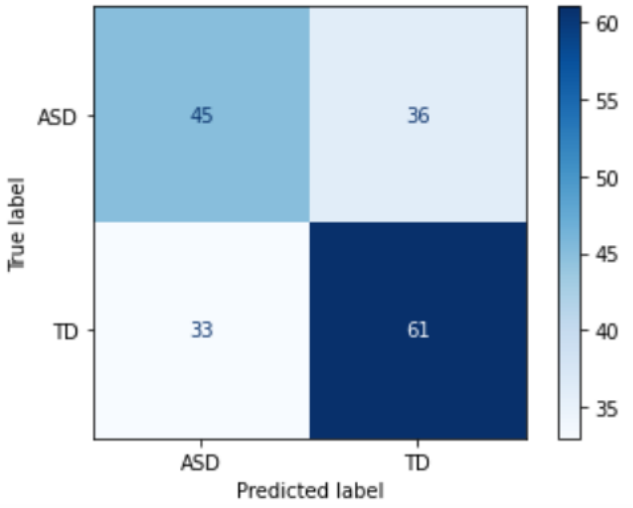

Figure 17: Confusion Matrix for DT model

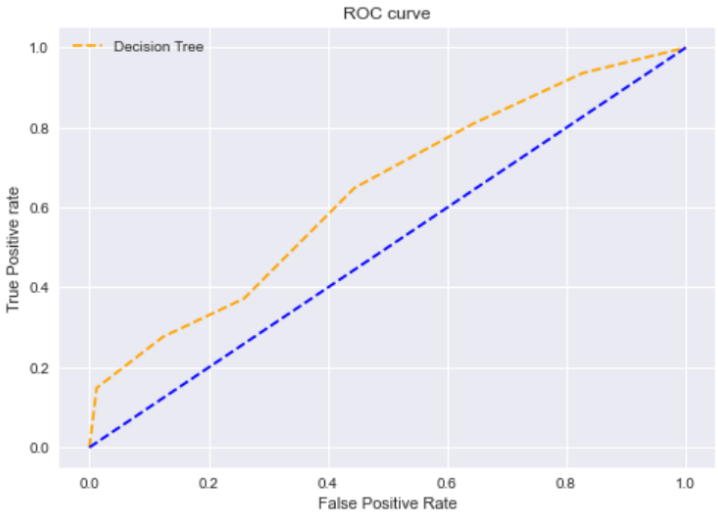

Figure 18: ROC for DT model

The Random Forest classifier had an overall accuracy of 60%. It correctly predicted 38 ASD cases from 81, and 73 cases of TD out of 94. It had a recall value of 0.47 (47%), a precision value of 0.64 (64%), and an overall f1-score of 0.54 (54%). After 5-fold cross validation, the average CV score came to 0.61 (61%). The confusion matrix and ROC graph for the RF classifier have been displayed in Figures 19 and 20 below.

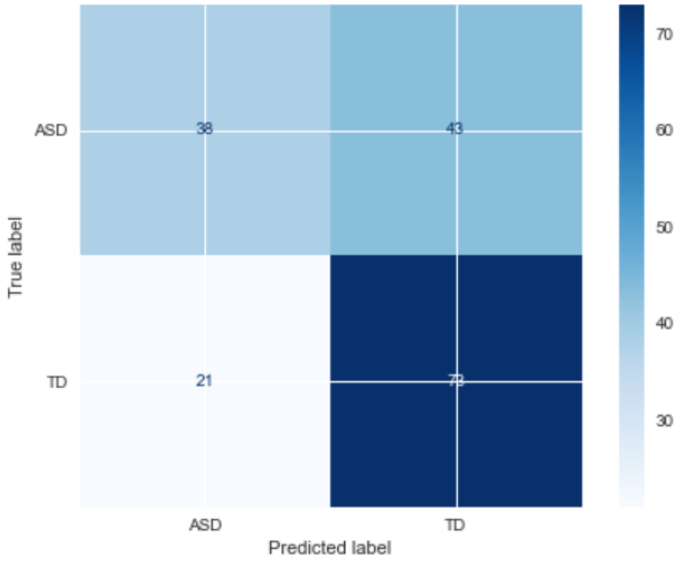

Figure 19: Confusion Matrix for RF model

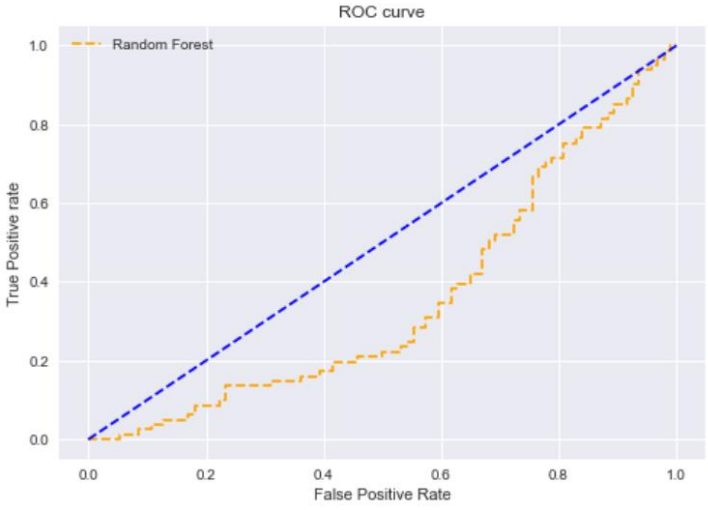

Figure 20: ROC for RF model

The Naïve Bayes model had the lowest overall accuracy of 56%. It correctly predicted 49 ASD cases from 81, and 60 cases of TD out of 94. It had a recall value of 0.51 (51%), a precision value of 0.59 (59%), and an overall f1-score of 0.60 (60%). After 5-fold Leave One Out cross validation, the average CV score came to 0. 58 (58%). The confusion matrix and ROC graph for the NB classifier have been displayed in Figures 21 and 22 below.

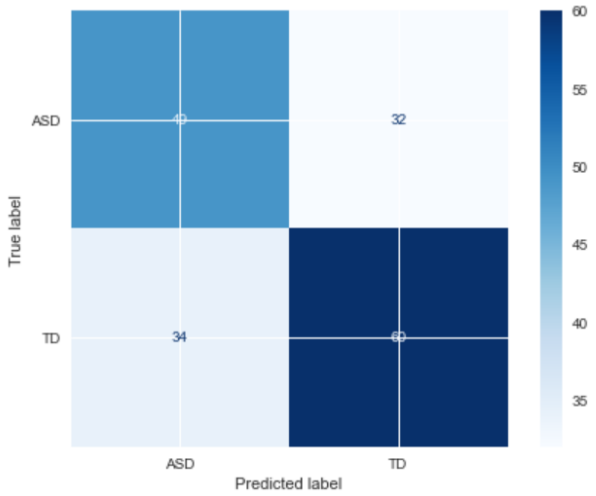

Figure 21: Confusion Matrix for NB model

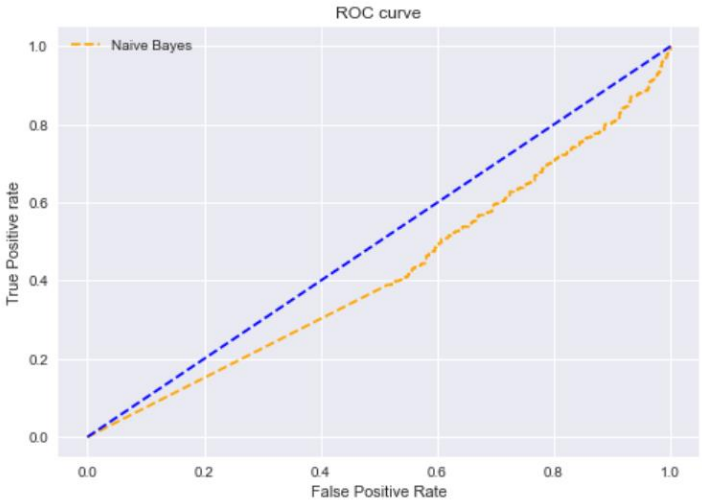

Figure 22: ROC for NB model

4.3. Results of DL model simulations

Among the DL models, the Feedforward Neural network achieved the highest overall accuracy at 68%. It had a precision value of 0.61 (61%), a recall value of 0.67 (67%), and an overall f1-score of 0.64 (64%). The training and validation loss curves for the FFNN classifier have been displayed in Figures 23 and 24 below.

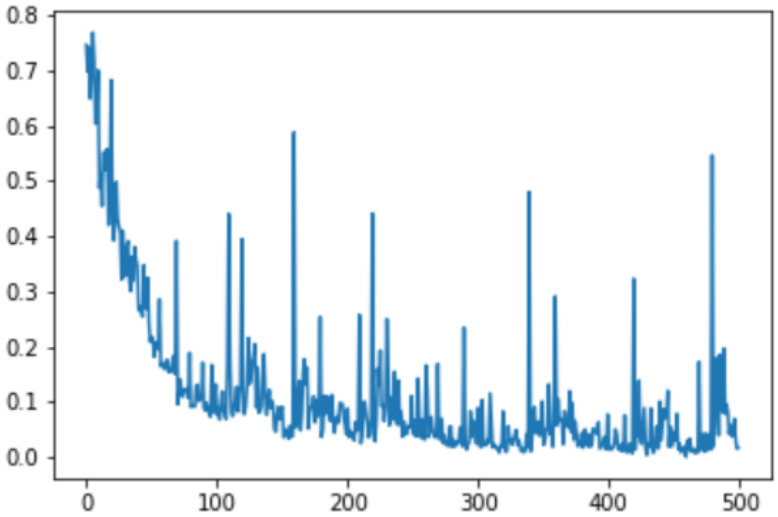

Figure 23: Training loss curve of FFNN

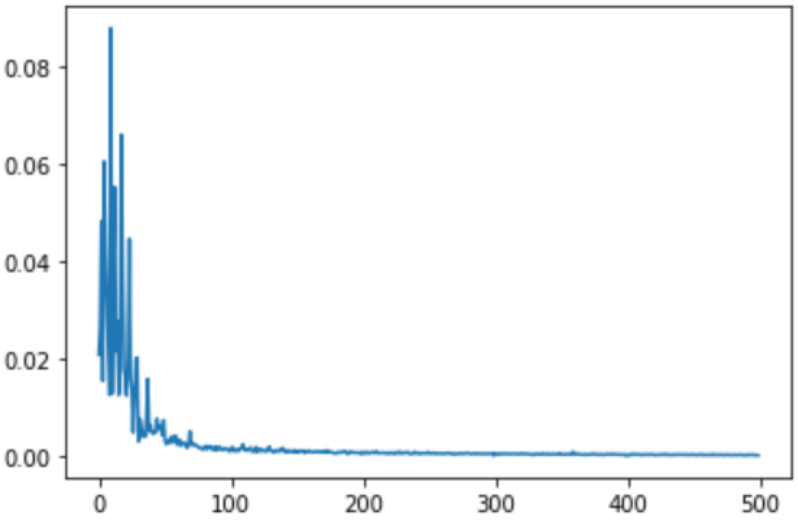

Figure 24: Validation loss curve of FFNN

On the other hand, the MLP classifier achieved an overall accuracy at 66%. It had a precision value of 0.66 (66%), a recall value of 0.60 (60%), and an overall f1-score of 0.63 (63%). The training and validation loss curves for the MLP classifier have been displayed in Figures 25 and 26 below.

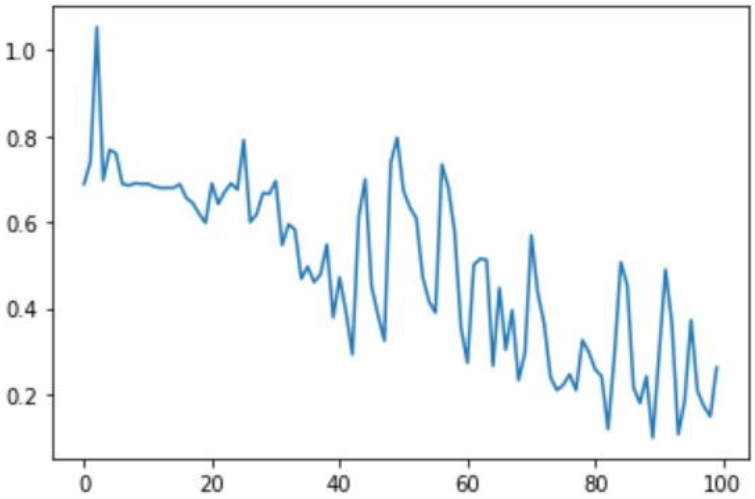

Figure 25: Training loss curve of MLP

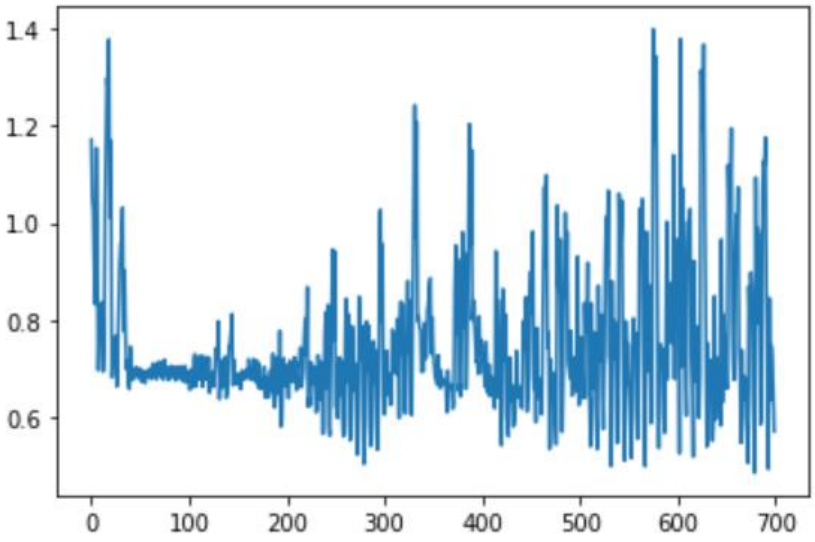

Figure 26: Validation loss curve of MLP

Chapter 5: Conclusion and Future Work

5.1.Conclusion

Individuals who fall under the Autism Disorder Spectrum display various behavioral patterns and characteristics, which is why it may be difficult to accurately detect it in many individuals just by using medical history, psychological evaluation, and clinical observation. The use of machine learning and deep learning models to identify patterns in the human brain that are associated with ASD act as supplementary evidence for the diagnosis of ASD.

Based on our results, the Feedforward Neural network gave the highest overall accuracy at 68%, as it was easily able to identify and create patterns from various types of input. The SVM and LR models followed it at 67%, which may be because both models work well when there are understandable margins of dissociation between classes, and they are more productive in high dimensional spaces, attributes which are applicable for this dataset. On the other hand, the Naïve Bayes classifier was the least well-performing at an overall accuracy of 56%, likely due to its common trait of assuming that all features are independent or unrelated, so it cannot learn the relationship between features. In the case of the SVM classifier, our results exceed those of [2], [3], and [16]. Similarly, our GBM model had also performed better than [16], while our RF model achieved the same results the RF classifier in [16].

From the ROC curves (except for Decision Tree and SVM) and training loss and testing loss curves plotted of each ML and DL model, it can be theorized that the ML classifiers, with the exception of Decision Tree, are making negatively correlated predictions. The haphazard training loss and testing loss curves indicate that the DL models were unable to learn any pattern from the training data, and so, it will also behave abnormally during the testing portion. Both results imply all models are predicting randomly.

The aim of this paper was to prove whether ML and DL models could achieve higher levels of accuracy based on only fMRI data, as it was the case in [16], so as to not be biased by demographic data such as age, gender, etc. While papers applying datasets with demographic variables did perform exceedingly well, as seen in [1] and [4], our findings suggest that high-performance ML and DL models can be built with very few features from the fMRI data in a relatively small dataset, with the only cost being the probability of overfitting.

5.2. Future Work

For any future work regarding this study, we will be focusing on designing novel deep learning based models combining models that were able to most successfully diagnose the severity of ASD. We will also improve the performance of deep learning techniques by designing new data augmentation and simulation methods. It is our intention to test our models on different datasets from that of ABIDE. While the ABIDE dataset is one of the largest databases containing ASD patient brain imaging data from sites all over the world, the data is not generalized, which can lead to inaccuracies. Thus, in the future, it may help improve results by using the data of datasets containing more standardized data, which was previously not within the scope of this paper due to the lack of neuroimaging datasets being publicly available.

References

References

[1] R. Nur Syahindah Husna, A. R. Syafeeza, Norihan Abdul Hamid, Y. C. Wong, R. Atikah Raihan. 22nd April, 2021. Functional Magnetic Resonance Imaging for Autism Spectrum Disorder Detection using Deep Learning.

[2] Anibal Sólon Heinsfeld, Alexandre Rosa Franco, R. Cameron Craddock, Augusto Buchweitz, Felipe Meneguzzi. 30th August, 2017. Identification of autism spectrum disorder using deep learning and the ABIDE dataset.

[3] Taban Eslami and Fahad Saeed. 2019. Auto-ASD-Network: A Technique Based on Deep Learning and Support Vector Machines for Diagnosing Autism Spectrum Disorder using fMRI Data.

[4] Xiang Xiao, Hui Fang, Jiansheng Wu, ChaoYong Xiao, Ting Xiao, Lu Qian, FengJing Liang, Zhou Xiao, Kang Kang Chu, and Xiaoyan Ke. 2016. Diagnostic Model Generated by MRI-Derived Brain Features in Toddlers With Autism Spectrum Disorder

[5] Xin Yang, Paul T. Schrader, Ning Zhang.4th November 2020. A Deep Neural Network Study of the ABIDE Repository on Autism Spectrum Classification

[6] Yafeng Zhan, M.S., Jianze Wei, M.S., Jian Liang, Ph.D., Xiu Xu, M.D., Ran He, Ph.D., Trevor W. Robbins, Ph.D.. Zheng Wang, Ph.D. Am J Psychiatry 2021. Diagnostic Classification for Human Autism and Obsessive-Compulsive Disorder Based on Machine Learning From a Primate Genetic Model.

[7] Fahad Saeed. 2018. Towards quantifying psychiatric diagnosis using machine learning algorithms and big fMRI data. Big Data Analytics 3, 1 (2018), 7

[8] Decision Tree Classifier with Sklearn in Python. April 17, 2022. https://datagy.io/sklearn-decision-tree-classifier/

[9] Support Vector Machines. https://scikit-learn.org/stable/modules/svm.html

[10] Understanding Random Forest. June 12, 2019. https://towardsdatascience.com/understanding-random-forest-58381e0602d2

[11] An Introduction to Logistic Regression. July 11, 2021. https://www.analyticsvidhya.com/blog/2021/07/an-introduction-to-logistic-regression/

[12] Naive Bayes classifier. October 29,2022. https://en.wikipedia.org/wiki/Naive_Bayes_classifier

[13] Gradient Boosting Algorithm: A Complete Guide for Beginners. September 20, 2021. https://www.analyticsvidhya.com/blog/2021/09/gradient-boosting-algorithm-a-complete-guide-for-beginners/

[14] Gradient Boosting Classifier. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html

[15] Feedforward Neural Networks and Multilayer Perceptrons. November 24, 2021. https://boostedml.com/2020/04/feedforward-neural-networks-and-multilayer-perceptrons.html

[16] Gajendra J. Katuwal, Nathan D. Cahill, Stefi A. Baum, Andrew M. Michael. 2015. The Predictive Power of Structural MRI in Autism Diagnosis

[17] Wiggins, L.D.; Reynolds, A.; Rice, C.E.; Moody, E.J.; Bernal, P.; Blaskey, L.; Rosenberg, S.A.; Lee, L.; Levy, S.E. Using standardized diagnostic instruments to classify children with autism in the study to explore early development. J. Autism Dev. Disord. 2015, 45, 1271–1280.

[18] Christiansz, J.A.; Gray, K.M.; Taffe, J.; Tonge, B.J. Autism spectrum disorder in the DSM-5: Diagnostic sensitivity and specificity in early childhood. J. Autism Dev. Disord. 2016, 46, 2054–2063.

[19] Md. Mokhlesur Rahman, Ravie Chandren Muniyandi, Usman Opeyemi Lateef, Shahnorbanun Sahran. 2020. A Review of Machine Learning Methods of Feature Selection and Classification for Autism Spectrum Disorder

[20] Candida Claudia Luciano, Roberto Keller, Pierluigi Politi, Eugenio Aguglia, Francesca Magnano, Lorenzo Burti, Francesca Muraro, Angela Aresi, Stefano Damiani and Domenico Berardi. 2014. Misdiagnosis of High Function Autism Spectrum Disorders in Adults: An Italian Case Series

[21] Camilla Gesi, Giovanni Migliarse, Sara Torriero, Martina Capellazzi, Anna Caterina Omboni, Giancarlo Cerveri, Claudio Menacci. 2021
Gender Differences in Misdiagnosis and Delayed Diagnosis among Adults with Autism Spectrum Disorder with No Language or Intellectual Disability

[22] Lori M. Baudino. 2010. Autism Spectrum Disorder: A Case of Misdiagnosis