AUTISM SPECTRUM DISORDER DETECTION

CONTENTS

1. INTRODUCTION

1.1 Motivation

The popularity of digital television has increased the consumption of web-series on various online streaming platforms such as Netflix. One such series which shows the struggles of a woman who faces autism is " Extraordinary Attorney Woo ". Autism is an extremely complex condition that affects social behavior in both children and adults. It is important that the signs of autism are recognised and the condition is diagnosed as early as possible. Once diagnosed, the person can understand themselves better and realize they are not alone in the way they feel.

1.2 Problem Definition

Autism spectrum disorder (ASD) is a neurological and developmental disorder that affects how people interact with others, communicate, learn, and behave. Although autism can be diagnosed at any age, it is described as a "developmental disorder" because symptoms generally appear in the first two years of life.

One of the hallmarks of autism is a lack of interest in or connection with other people. For example : An infant seems uninterested in making eye contact, never smiles, or can't bear to be held.In summary, those with ASDs may be lost in their own thoughts, or run right up to strangers and start conversations on obscure topics. Either way, people all across the autism spectrum suffer from an inability to understand the complexities of the social world.

1.3 Objectives of the Project

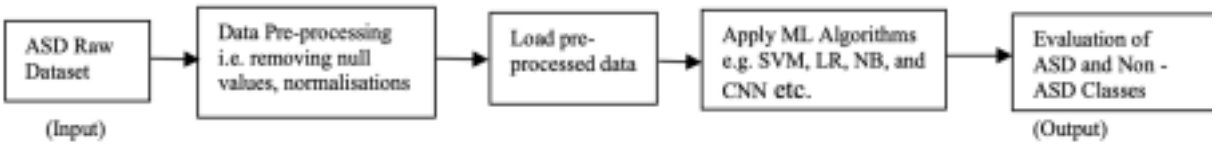
Propelled with the rise in use of machine learning techniques in the research dimensions of medical diagnosis, we attempt to explore the possibility to use Naïve Bayes, Support Vector Machine, Logistic Regression, KNN, Neural Network and Multilayer Perceptron for predicting and analysis of ASD problems in a child, adolescents, and adults.

1.4 Contributions

- **Balassubramanian Srinivasan(#G01360498):** Acquiring Dataset, proposal, midterm presentation, preparing test data, MLP Model, poster.
- **Murali Sai Lakith(#G01381718):** Acquiring Dataset, proposal, midterm presentation, data visualization, MLP model, poster.
- **Anjali Pudiyadatha (#G01395882):** Proposal, midterm presentation, midterm video, data visualization, SVM model, preparing test data, poster.
- **Ashlesha Ram Deokar (#G01387760):** Proposal, midterm presentation, midterm video, SVM model, preparing test data, poster.

2. METHODOLOGY

2.1 Preparing the data



The ASD Diognosis dataset extracted from the UCI machine learning repository being used consists of 21 features which include the patient's age, nationality, nation they currently reside in along with a few screening questions. The screening questions can be divided into three categories based on the patients' age: Child, Adolescent and Adult.

- ➢ The three datasets have 20 common features that are used for prediction.
- ➢ The dataset for Child has 292 samples, Adolescent has 104 samples and Adult has 704 samples.

The below image best represents the dataset being used for prediction in this project :

Table 1: List of ASD datasets

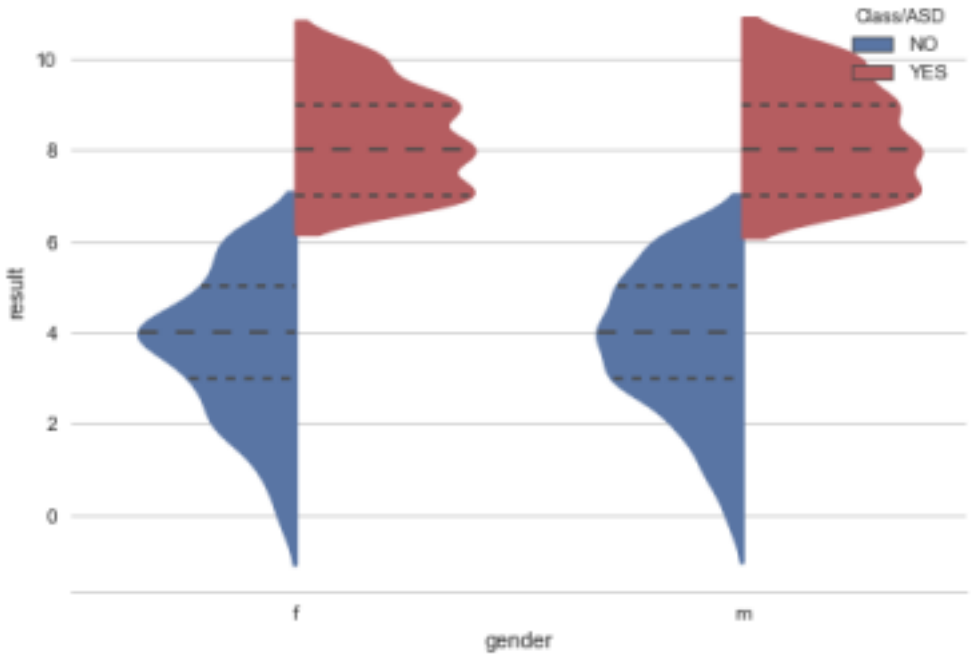
Sr. No.	Dataset Name	Sources	Attribute Type	Number of Attributes	Number of Instances
1	ASD Screening Data for Adult	UCI Machine Learning Repository [12]	Categorical, continuous and binary	21	704
2	ASD Screening Data for Children	UCI Machine Learning Repository [15]	Categorical, continuous and binary	21	292
3	ASD Screening Data for Adolescent	UCI Machine Learning Repository [16]	Categorical, continuous and binary	21	104

Table 2: List of Attributes in the dataset

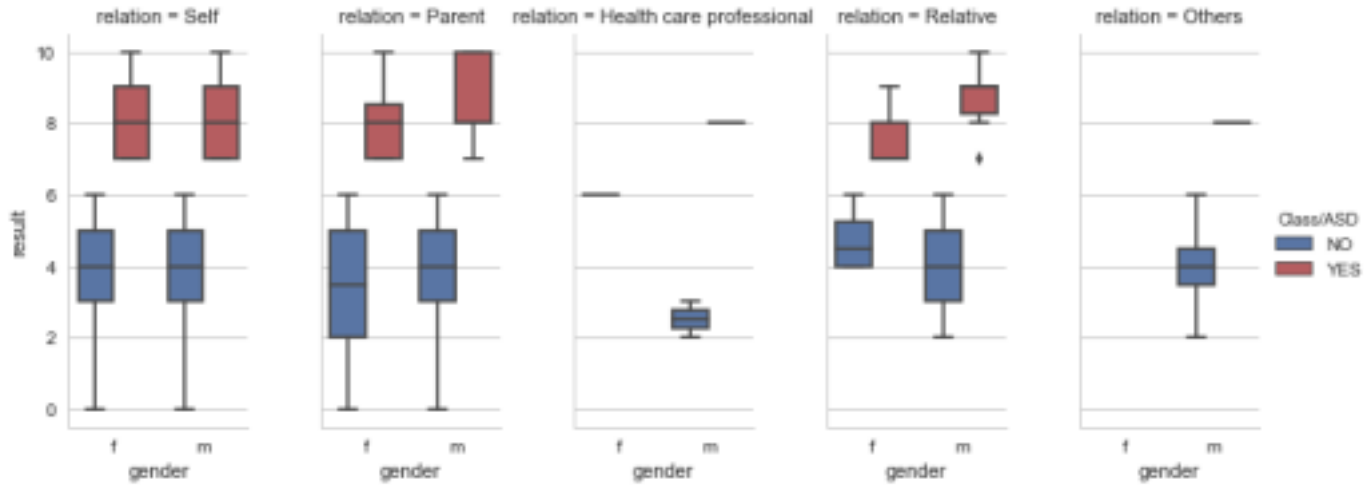
Attribute Id	Attributes Description
1	Patient age
2	Sex
3	Nationality
4	The patient suffered from Jaundice problem by birth
5	Any family member suffered from pervasive development disorders
6	Who is fulfilment the experiment
7	The country in which the user lives
8	Screening Application used by the user before or not?
9	Screening test type
10-19	Based on the screening method answers of 10 questions
20	Screening Score

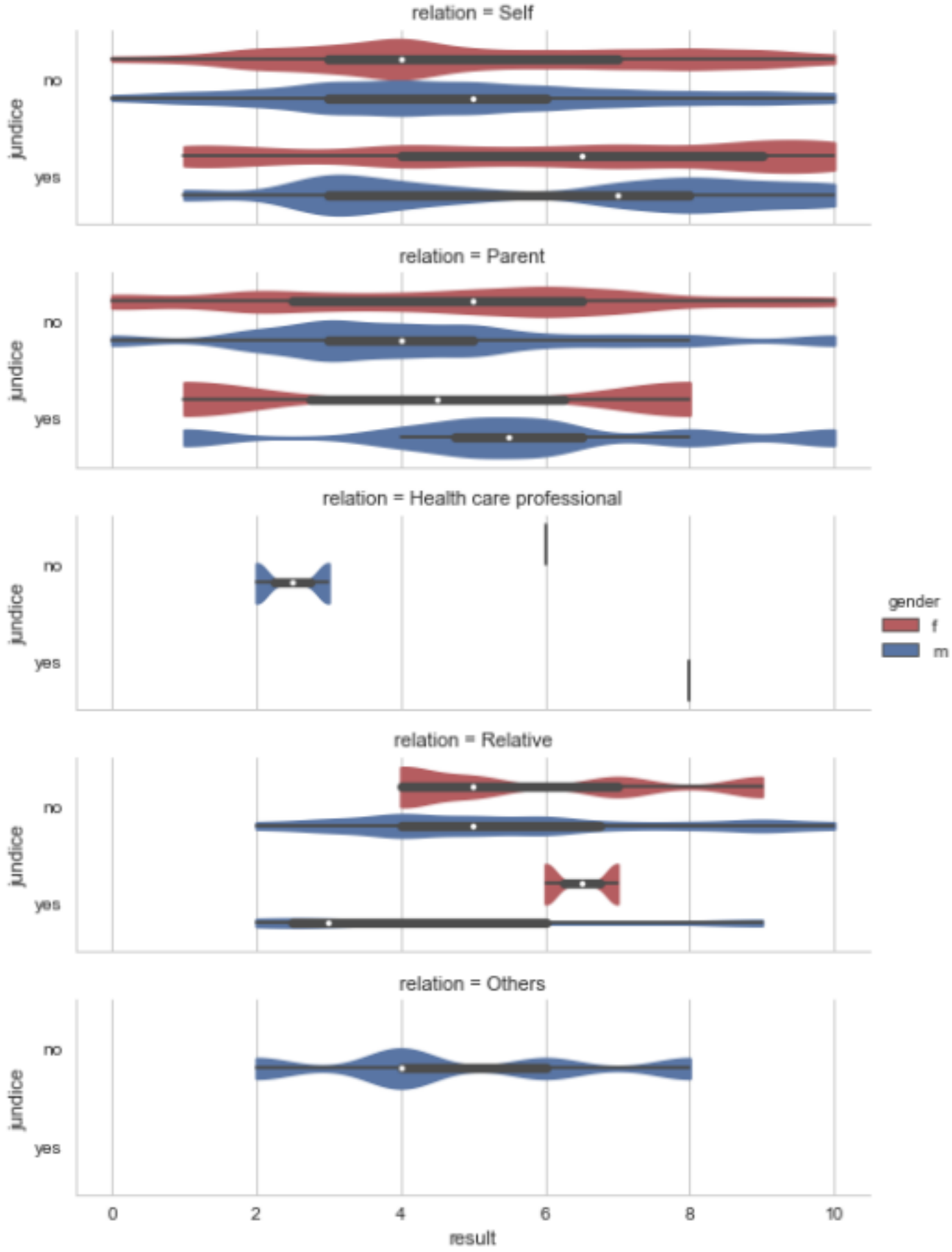
2.2 Data Visualization

Violin plot based on Gender



Factor Plot

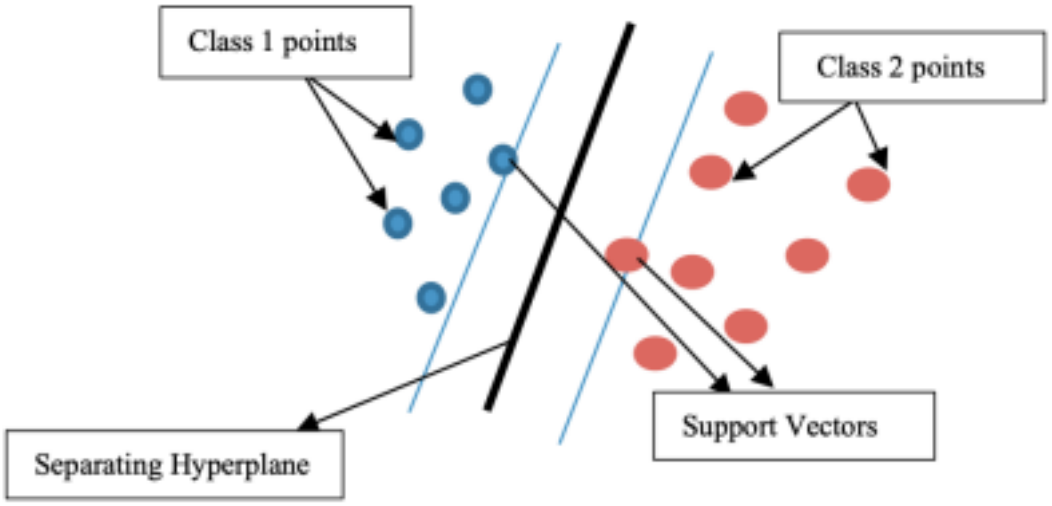



relation = Self
relation = Parent
relation = Health care professional
relation = Relative
relation = Others
jundice
no
yes
gender
f
m
0
2
4
6
8
10
result

2.3 Model

2.3.1 SVM

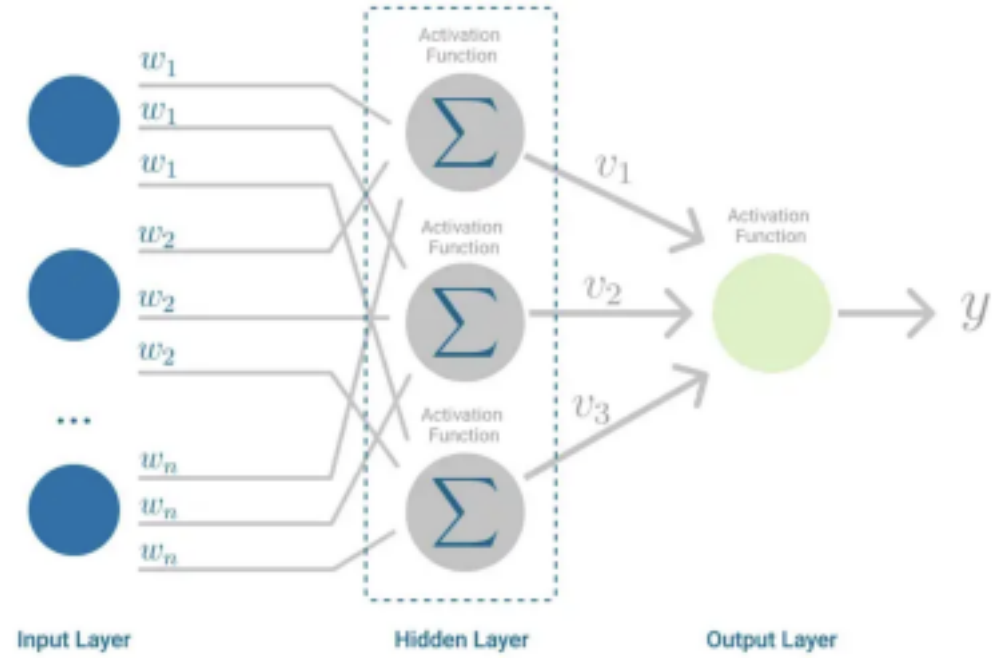
SVM is a linear supervised machine learning approach that is used for classification and regression. It is a pattern recognition problem solver. It does not cause the problem of overfitting. SVM separates the classes by defining a decision boundary. We used the kernel = 'linear' hyperparameter for our SVM model. Kernel function is responsible for using a linear classifier to solve nonlinear problems. We further went on the calculate the AUC and F1 beta score for the SVM model

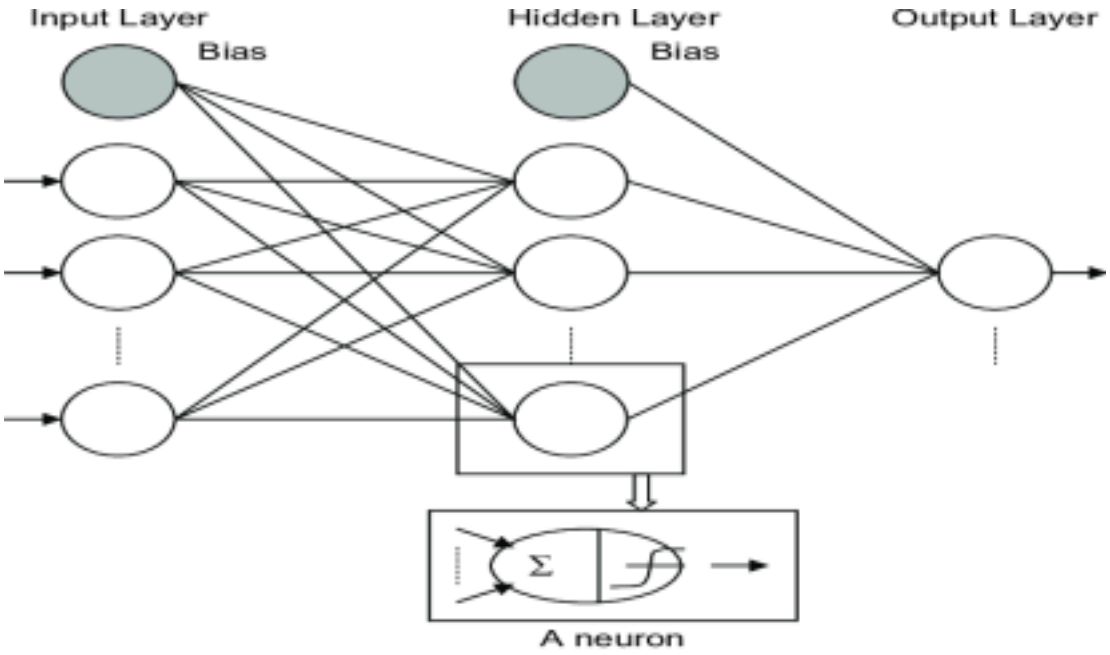


2.3.2 MLP

A Multilayer Perceptron has input and output layers, and one or more hidden layers with many neurons stacked together. And while in the Perceptron the neuron must have an activation function that imposes a threshold, like ReLU or sigmoid, neurons in a Multilayer Perceptron can use any arbitrary activation function.

We built a sequential model architecture of two layers having ReLU and Sigmoid as activation functions and a dropout layer as it helps to reduce overfitting.





3. EXPERIMENT

2.2 Evaluation Metrics

➢ **F Beta Score :**

$$F_\beta = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}$$

➢ **AUC :**

AUC represents the probability that a random positive example is positioned to the right of a random negative example. AUC ranges in value from 0 to 1.

➢ **MLP Evaluation ;**

To evaluate the MLP model on a training and testing set we use the Keras function model.evaluate().

MODEL	AUC	F1 BETA SCORE
SVM	1.0	1.0

MODEL	Accuracy
MLP	0.9508196711540222

GITHUB LINK TO THE CODE REPOSITORY:

We have uploaded our project in the following GITHUB profile :

https://github.com/apudiyad/Autism-Disorder-Detection

And the Repository Title is : Autism-Disorder-Detection

5. CONCLUSION AND FUTURE WORK

In this work, detection of Autism Spectrum Disorder was attempted using various machine learning and deep learning techniques. Various performance evaluation metrics were used to analyze the performance of the models implemented for ASD detection on non-clinical dataset from three sets of age groups viz. Child, Adolescents and the Adult. When comparing the result with another recent study on this problem, we observed better results for MLP classifier and SVM with including all its features attributes after handling missing values. In this work after handling missing values, both the SVM and MLP based models show the same accuracy of prediction for ASD Child dataset. Ultimately the prediction is the set of binary values {0,1}. '0' being showing negative results of autism and 1 being positive results of the test subject predicted as suffering from autism.

Apart from testing out different models and finding an optimal approach to deal with ASD detection, as this project progresses when are paralelly developing a GUI using Tkinter, so that the results ASD prediction using our model can also be accessed by various groups of target users keeping in mind the objective of helping out people to understand early deatection of Autism Spectrum Disorder.