Autism Spectrum Disorder (ASD) Detection And Diagnosis In Young Children And Adults

Harshit Rathi *E19CSE211*, Khushi Singh *E19CSE170*, Priya Gandhi *E19CSE291*, Mayank Aggarwal *E19CSE275*, *Team Exclusives*

Introduction

Autism is a development disability which is cause by problems in the brain, it can be genetic or otherwise. Around 1 in 100 children in India is affected from autism and it usually begins at the age of 3 and can last till a person lives. These children in the spectrum face a lot of issues throughout their life. The symptoms include the following – they have delayed language and movement skills, they are slow learners, and they have slow cognitive skills, they are hyperactive impulsive and have /Unresponsive/mixed behavior, they suffer from epilepsy or seizure disorder, they have difficult mood, emotional reactions and complex sleeping habits and suffer from stress and anxiety.

These are just a short list of symptoms other than that social communication and interaction is very challenging from them. They have restricted and repetitive behavior or interest. Sometimes, parents/people are unable to identify these signs and symptoms of autism, they feel it is very normal for kids to behave like this which leads to long term problems because if autism is detected in an early age it can be controlled and the child on the spectrum can live a very better life. He can improve his communication skills, have control on himself and live with little to no support. But if not treated people with ASD need a full time help in their through-day life. So, we are doing this project so that the people on the spectrum can live like other people and rise to their full potential. Also, it is an interesting area of research, and it is a challenging project for us. There will a lot of learning opportunities for us in this project. We hope this project will help the people on the spectrum and to make their lives better along with their caretaker/parents who will be able to detect the issue soon and can work on it in an healthy way.

I. Problem Statement

Autism Spectrum disorder is a complex and socio-developmental disorder that hampers basic human body procedures. Simple communication, the ability to perceive, the ability to express. They lack socio-responsive behavior (they don't perceive if someone is calling their name or addressing them), the facial expression not matching what they are saying, unusual tone of talking and speaking that may sound singing a song or flat robot-like voice. ASD gives rise to poorly integrated verbal and nonverbal communication. We have realized the severeness of the poor detection and diagnosis of ASD. It has been proven and many pieces of research claim that the early section of ASD can help save and sustain the poor social behavioral and interaction skills of the person and save the person entering into a complex and distressful state a psychiatric dilemma.

Without a diagnosis and therapy and treatment can make the life of the patient difficult, causing them distress with everyday worlds. Getting them confused and put them in a state of chaos. This can turn into casting difficult added behaviors, forwarding to social isolation.

II. Background Research

Autism Spectrum disorder cases have been linked to being found back in 1747, the first case of autism that was officially documented was High Blair of Borgue's. “The modern sense of autism was made in 1938 and the name was coined, by Hans Asperger a resident of Vienna, at University Hospital. He went ahead and adopted Bleuler's terminology, autistic psychopaths, in a lecture in German about child psychology”. The propotionof autistic people is turned to be 1.5 percent of the entire world's entire population is classified with an autism spectrum disorder. Well, autism shows its roots in the human community for a long time. Modern techniques and other advancements to diagnose and detect mental disorders were first commenced in the late 19th century. Over the years with technological advancements and improving ML and deep learning models a lot of research has been made in this domain.

The Autism diagnostic has several features that are again classified in several classes. These classes include Deficit in Socio-interaction Behaviors and communication, Repetitive behavior, and restrictive interests.

Talking about the Poor Socio-interactions, the autistic person has shown patterns of abnormal social approach and they can't form a back-forth communication. They can't initiate or respond to social interaction. Difficulty in making constant eye contact and body language complexities or deficits in understanding and use of gestures or words or understanding facial expressions and moods. They have a poor sense of relationship with other humans/family members. They show repetitive patterns of behavior. They get extremely distressed at slight changes. They have persistent adherence to routines and having fixated ritualized patterns or verbal nonverbal behavior. Hyperactivity (ADHD).ASD if detected and diagnosed earlier can prevent so much distress. As a child develops with adequate medical assistance and we can curb further "psychiatric/mental" illnesses developing. Some Autistic people do benefit from medication, therapy, and lifestyle monitoring. Such routine treatment help with anxiety and clumsy mood or to help manage some behaviors, but these would need to be prescribed and monitored by a certified child and adolescent psychiatrist.

Earlier detection can control the worsening stages of impairment of socio-communication skills.11-39% of individuals with ASD also have epilepsy. Comorbid sleep disturbance feature is predominant in 50-80% of children with ASD and is correlated with daytime problem behaviors like uneasiness and fussy mood. The condition, symptoms, class of behavior, and the diagnosis approach used are all very nested and complex. This is where we need a highly proactive Deep learning model that helps us appropriately incubating every single detail and show us the pattern of the condition and its severity.

III. Previous Works

As over almost form the past 20 yspark etc. ears there is quite large rise in storage of data which in simple terms known as big data which is marked by the 2 factors the increasing rate in which data is produced and the other one is the different variety of the data coming this because due to affordability and assort ability of the infrastructure and the instruments. Also now a days there many ML algorithms like R, Spark etc. all the machine learning algorithms generally combines with the good hardware and because of these type of technologies the researchers got the extreme opportunities to work on these ML algorithms Now a days autism spectrum disorder has become a serious issue because as per the world health organist ion the cases of ASD is increasing day by day so the machine learning researchers are willing to find the solution of ASD. For detecting the presence of autism spectrum disorder the techniques which are manual can be very expensive and exhausting. For automatic classification of ASD it is important to find the correct classifier, which has more computational speed and its accuracy is satisfactory. Here we are going to compare going to have a comparison of all the techniques which use machine learning and deep learning to predict ASD.

To diagnose ASD there are various machine learning and pattern recognition techniques in which research is going on in the field of medical science. To name few of them we have SVM, decision trees, Random Forest scan, logistic regression CNN, K-Nearest Neighbor, ANN.

In researcher done by Suman Raj, Sarfaraz Masood, in the domain of

"Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques, Procedia Computer Science, ISSN 1877-0509"

(https://www.sciencedirect.com/science/article/pii/S1877050920308656). Their work showcased the use of several machine learning models such as Gaussian Naïve Bayes, Logistic regression , K-Nearest Neighbour, SVM Classifier, for predicting and analysis.

The UCI non-clinical dataset related to ASD screening in adult has 704 instances and 21 attributes and the dataset for children has 292 instances. We have proposed to conduct the experiment on the children and adult dataset

Adult subjects contain. Their experiment was carried for a wider domain it was inclusive of everyone (including a child, adolescent, or even an adult) and then predict the ASD. But with our project, we aim towards a more precise model to detect ASD in young children and an early phase. Which is quite a challenging task in itself. The majority of predominant features are not visible and diagnosed at the early phase. Considering the result of the fore mentioned research paper it is conclusive that can expect around 98.30% of accuracy. The model that gave the high accuracy was found out to be of CNN, rather than any other considered model building techniques.

Here are few of the citations and References that we bought together and mentioned that provided us we the idea and motivation to make this project where it deals with both machine learning as well as Artificial Intelligence to predict the detection of autism at a young age.

IV. Proposed Methodology

Our research mainly focuses on the fact that the ASD (Autism Spectrum Disorder) in children should be diagnosed at an early stage. According to the CDC, the cause of the autism is still unknown except one is the genetic condition [1]. The development of the autism people is different from most of the people but their abilities vary significantly. Dataset of 4-11 years old children has been collected considering the CDC research that clearly states the symptoms for autism begins at the age of 3 years. The symptoms may improve over time but can last till dead. The detailed explanation of each phase of the flowchart of the proposed methodology is explained below:

A. Dataset Collection

The ASD in children dataset for this research has been collected from a publicly accessible UCI Repository that contains 996 instances and 21 attributes. This dataset is called the AQ-10 a set of 10-adaptive behavioural features for children that takes a 10 questions screening test. Each question focusing on different domain like communication, attention switching, attention to detail, social interaction, responsiveness, expression and imagination. Scoring method of these questions is binary (either 0 or 1). The other attributes of the AQ-10 give personal and background details of the participant taking the autism assessment such as age, gender, ethnicity, born with jaundice, family member with PDD, country of residence, used the screening app before, result of the screening test, who is completing the test (relation), Class/ASD. The brief summary about the dataset has been provided in Table 1.

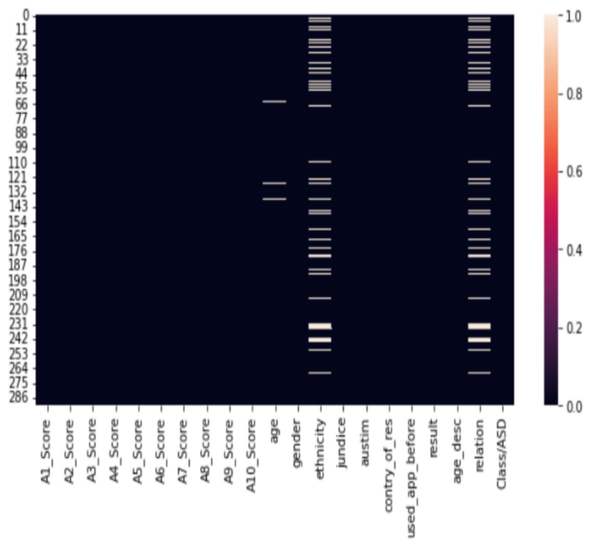


Figure 1: Visualization of the missing values using the seaborn heatmap.

B. Dataset Preprocessing

This is the most crucial step of this research to develop an effective prediction model. It is done to clean raw or noisy data before supplying it to the model.

The above Figure 1 clearly shows the missing values in the age, ethnicity and relation columns. Missing values van contribute to wrong predictions and these missing values are handled using the measure of central tendency method (imputation method), since the data is less and very sensitive. Imputation Method is fine-tier and low computation method to deal with Missing values problem. We convert the missing values to co-related non-missing values trend. For example we can use the Imputation(Mean/Median) method. Where we replace the missing values with the median of the non missing values of the same column.

We have used K-NN Imputation method to resolve and handle the missing values incorporated in our dataset.

i. K Nearest Neighbour (KNN)

K-NN functions on the lines of finding the K nearest neighbour of the missing value/element then finding analysing the ‘features similarity’ and predicting the missing value . We do the prediction of the missing value by looking and finding the k-closest neighbour to it. It is much more accurate and gives reliable results.

Our data has a number of missing values depicted by ‘?’ We have replaced them with NaN. Then this NaN values are imputed via K-NN Imputation technique. A majority of data is categorical so we drop the rows that has too many missing values. Because on trial we realised that computing all of these missing values and building the model gives us an overfitted model which is less generalised.

ii. One-Hot Encoding

After handling the missing values in the dataset, other pre-processing techniques like One-Hot encoding (for multi-class encoding) is used. We have made a vivid observation that the type of data is non-numeric as well. We have parameters as Country of residence , ethnicity and relation of the case. These all are of string data type but they hold a strong perseverance in the prediction. We convert these categorical values to non-categorical values via One hot Encoding. We have used Label Class encoding for converting the Class ASD which states whether the case has ASD or not. We have define to have numeric values of 0 or 1 instead of the string values “Yes” or “No”.

iii. Min Max Scaler

This is a classic normalisation technique where the input values are casted in range 0-1, its coherent range of floating point values that inhibits the most precision. Normalization for numerical values (using Min Max Scaler) is applied to the dataset to get the desired dataset for the model training. We scale the parameter of age and result of our dataset. We scale the value of age and result in a range of (0-1). Finally we end up highly clean and cohesively apt pre-processed data.

C. Features Engineering

In this phase, the correlation matrix will be used to select and extract the important features for the model. Some of the unnecessary and irrelevant features like used app before, age, desc are directly dropped from the dataset irrespective of their correlation value.

D. Training and Testing Model

As shown in the Figure 2 , the UCI ASD dataset is split into two parts – Training set (80%) and Testing Set (20%) and then for cross-validation purposes, again a splitting is done on the training set giving – Training set (80%) and validation set (20%). Thus, the whole dataset is divided into three sub parts namely Training set, Validation set and Testing set in following the ratio of 64:16:20 respectively.

1) Developing the Prediction Model

In this research, the model used for the prediction of ASD in children is a supervised classification learning model. We have used the many classifier Model to test them on our problem statement and draw comparison on the evaluation metrics.

Age-Group	Total Instances	Missing values	Attribute types	Percentage of male %	Percentage of female %
Children (0-9 years) Adults (over 18 years)	996	YES	Categorical, continuous and binary	71.23 %	28.77 %

Table 1 : A brief summary of the dataset

ttribute	Type	Description
Age	Number	Age in years
Gender	String	Male or Female
Ethcicity	String	List of common ethnicities in text
Born with jaundice	Boolean (yes or no)	The case was born with jaundice or not.
Family Member with PDD	Boolean (yes or no)	The case has any immediate family affected by PDD
Who is completing the test	String	Parent of the case, self, caregiver, clinician, medical staff
Country of residence	String	List of countries in text
Used Screening app before	Boolean (yes or no)	Whether the user has used a screening app before
Screening Method Type	Integer (0,1,2,3)	The type of screening method chosen based on age category (0 = infant baby, 1=child(2-3 years) , 2 =adolescent, 3 = adult)
Test Question 1 Answer	Binary (0 or 1)	The answer code associated with the screening method applicable to the case

Test Question 2 Answer	Binary (0 or 1)	The answer code associated with the screening method applicable to the case
Test Question 3 Answer	Binary (0 or 1)	The answer code associated with the screening method applicable to the case
Test Question 4Answer	Binary (0 or 1)	The answer code associated with the screening method applicable to the case
Test Question 5 Answer	Binary (0 or 1)	The answer code associated with the screening method applicable to the case
Test Question 6 Answer	Binary (0 or 1)	The answer code associated with the screening method applicable to the case
Test Question 7 Answer	Binary (0 or 1)	The answer code associated with the screening method applicable to the case
Test Question 8 Answer	Binary (0 or 1)	The answer code associated with the screening method applicable to the case
Test Question 9 Answer	Binary (0 or 1)	The answer code associated with the screening method applicable to the case
Test Question 10 Answer	Binary (0 or 1)	The answer code associated with the screening method applicable to the case
Screening Score	Integer	The Final score obtained based on the screening test associated with the case. Computed in an automated manner.

Table 2 : The description of the attributes. All the 21 instances used in the Model.

The 6 prominent models used for our research are Decision Tree, Random Forest, Logistic Regression, Gaussian Naive Bayes, SVM, KNN.

i. Support Vector Machine (SVM)

SVM is very simple and easy it creates a line which just separates the data into the two different classes. SVM is a method which first sees the data and then it sort the given data into the different categories. So to increase the accuracy we use multiple SVM to classify Autism spectrum disorder (ASD)In the following project we have tried SVM we used gamma value to be 2 and and use sklearn package to use SVM the c value for the same is 1.0 for the evaluation matrices we use cross validation score where the mean score was giving us accuracy of 1.0 moreover if we fit our model on data then fbeta_score is also coming out to be 1 where the average value is taken out to be binary where the beta value is 0.5

ii. Logistic Regression

In our project we have tried logistic regression we have used sklearn.linear_model where the cv value is 10 to calculate the cross validation AUC we use cross validation score with cv value to be 10. Moreover, if we fit our data the f-beta score will be coming out as 0.93073593073593064 where the beta value is 0.5 and the average is equal to binary. The accuracy is judged by using the ROC score here we have the roc score of 0.97043715846994538, 0.99740986834010081 respectively

iii. Decision Tree

Classifier and is used mostly in classification and sometimes in regression also. To predict whether the person had autistic traits or not, we chose decision tree classifier for prediction model. The whole dataset is passed to the tree root node and then the data is split using features. This process is going to continue recursively till node gets unique label. We can split the algorithm in two phases in which the first phase is building the tree and second would be classifying the test data.

We will first import the DT classifier from sklearn library in python and then we will select the best features for constructing the tree and then the training data is called from the function. After this we iterate over each data of feature by visualizing the graph using and check its max IG. If its value is 0 that means it is pure and will return leaf node if not, then it will be split into two parts (false and true). The function is going to run recursively and the and the decision tree will be formed from the branches. After the construction of tree, we classify the test data according to the leaf node prediction. We iterate over the tree, and it reaches the leaf node and gets classified. After this we finally import the image using Ipython.display.

iv. Random Forest

We will import the random forest classifier from sklearn.ensemble library in python and fit the training data in it. We can find the cross-validation scores using cross_val_score and generate its mean using cv_scores_mean. So, the process is we first create an array to store the decision trees. It is first initialized with NULL. The we store some number of decision trees suppose 'x' in tree array. We call the function to build tree x times and it gets stored in the array. For some number of random attributes suppose 'I' the decision tree is generated. For the test data to get classified we take votes from each decision tree. If a greater number of votes gets to yes, we classify the data as having autistic trait and or more number of no votes we classify it as non- autistic traits.

v. K-Nearest Neighbour

In k-nearest neighbour, training data points corresponding to all instances is stored in n dimensional space. It is a supervised machine learning algorithm. For predicting the class, it analyses the closest k instances and the value predicted will be the most common class and in case of real valued data the mean of k nearest neighbour will be returned. We can classify the training dataset directly using KNN. Here, K correspond to the nearest neighbour. We label a group of points and use them to label other points. For labelling a new point it looks at the other points and find out the k closest points and each point votes, so any point is labelled according to the vote. So, to predict the class of a new instance we need to search its k closest instances. The closest neighbour is determined using the Euclidean distance formula. Its formula is the square root of squared difference of new instance and the existing instance.

vi. Naive Bayes

Naïve bayes is based on joint probability distribution. It has a very simple structure in which has parent nodes and its class nodes. In this model we assume the attributes that are conditionally independent and find out the class conditional probability.

This is a variant of Bayesian network. Using the prior probability and likelihood this model calculates the posterior probability and executes in lesser training time in comparison of ME and SVM

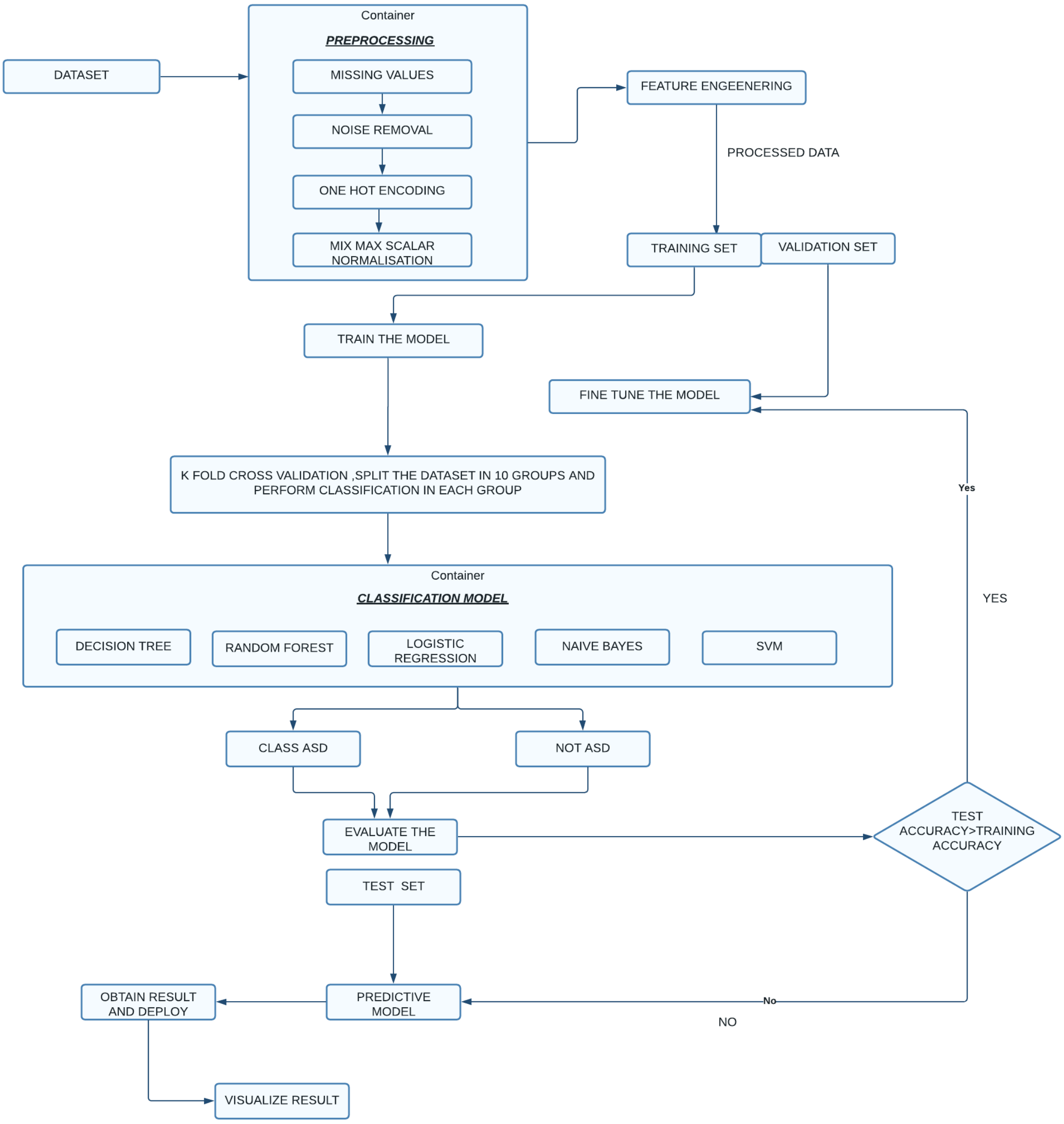


Figure 2: The flowchart of the proposed ASD Detection System

E. Evaluate the Model

The final predictive model which is built over all the aforementioned model is evaluate on all of these metrics to draw comparison and cast conclusion. We will be given the testing set and the performance of the model will be evaluated using the various classification metrics such as (Figure 4), precision score, recall score or sensitivity or true positive rate, specificity or true negative rate, F1 score and the accuracy. The results of the two classification algorithms will be visualized for better comparison. The following formulas are used to find the performance of the prediction model.

i. Accuracy

Accuracy is calculated as the number of the data points that are correctly predicted out of all the data points that exists. In other words, accuracy is defined as the percentage of correct positions of the test data.

Also, the need of the accuracy is like the AI learn continuously on its own so it is not possible to tell that the data that is given is correct or incorrect.

.

ii. ROC

For a binary classifier ROC (Receiver operating characteristic curve) is a graph which shows us a graph of true positive and false positive or shows the performance of the model at different classification thresholds

iii. AUC

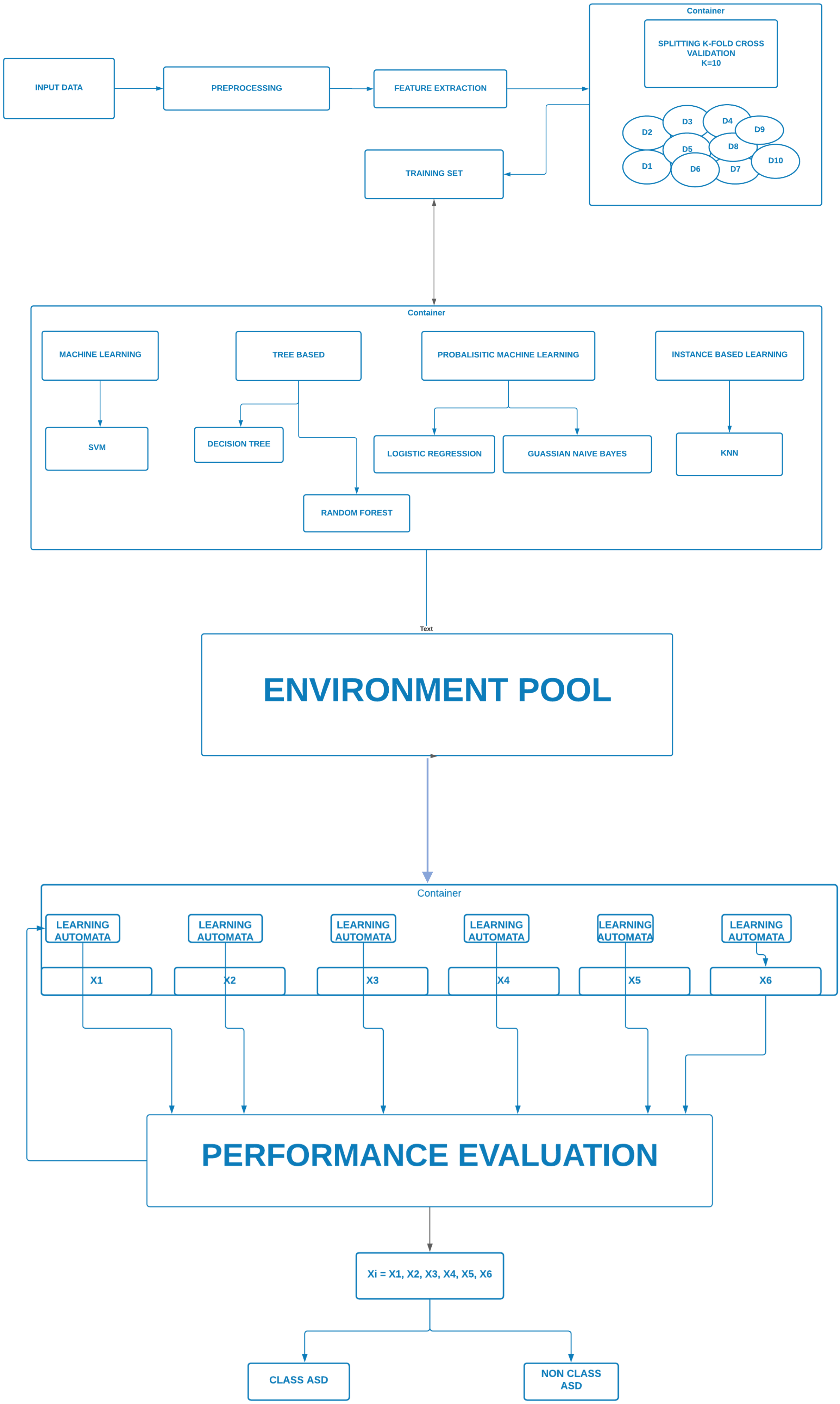
AUC acts as the summary of ROC curve. Also the result is the high if the area under curve(AUC) is more. More efficient is the model to make difference between the positive and negative classes AUC ranges in the value from 0 to 1 basically a model whose prediction is 100% wrong then AUC value is 0.0 and a model whose prediction value is 100 percent right the AUC value is 1.0 in general AUC of 0.7 to 0.8 are considered as acceptable and AUC of 0.8 to 0.9 are considered as excellent and the model having AUC of more than 0.9 is considered as outstanding in order to improve the AUC we have to improve the performance of the classifier if we compare AU with accuracy then AUC is known to be better measure then accuracy

iv. F-beta

The weighted harmonic mean of precision and recall is called as the f-beta score. The most optimal value of f-beta score is 1 and its worst value is 0. A good f1 score means that the false positives and the false negatives have a lesser value and were correctly identifying the threats. In some problem we are interested in F-measure with more importance to precision and in some problems with more importance to recall. So, f-beta score is the balance of both.

2) Fine Tune the Model

After a model is trained successfully, a simple criterion is used for adjusting the hyper parameters, in order to tune up the trained model, that is, if the testing accuracy of the model on the validation set is greater than training accuracy then the model is overfitting and if both the training and testing accuracy are high, then the model is under fitting. Thus, as shown in the Figure 3, a loop for fine tuning the trained model is used.



INPUT DATA
PREPROCESSING
FEATURE EXTRACTION
Container
SPLITTING K-FOLD CROSS VALIDATION K=10
D1
D2
D3
D4
D5
D6
D7
D8
D9
D10
TRAINING SET
Container
MACHINE LEARNING
TREE BASED
PROBALISITIC MACHINE LEARNING
INSTANCE BASED LEARNING
SVM
DECISION TREE
RANDOM FOREST
LOGISTIC REGRESSION
GUASSIAN NAIVE BAYES
KNN
Text
ENVIRONMENT POOL
Container
LEARNING AUTOMATA
LEARNING AUTOMATA
LEARNING AUTOMATA
LEARNING AUTOMATA
LEARNING AUTOMATA
LEARNING AUTOMATA
X1
X2
X3
X4
X5
X6
PERFORMANCE EVALUATION
Xi = X1, X2, X3, X4, X5, X6
CLASS ASD
NON CLASS ASD

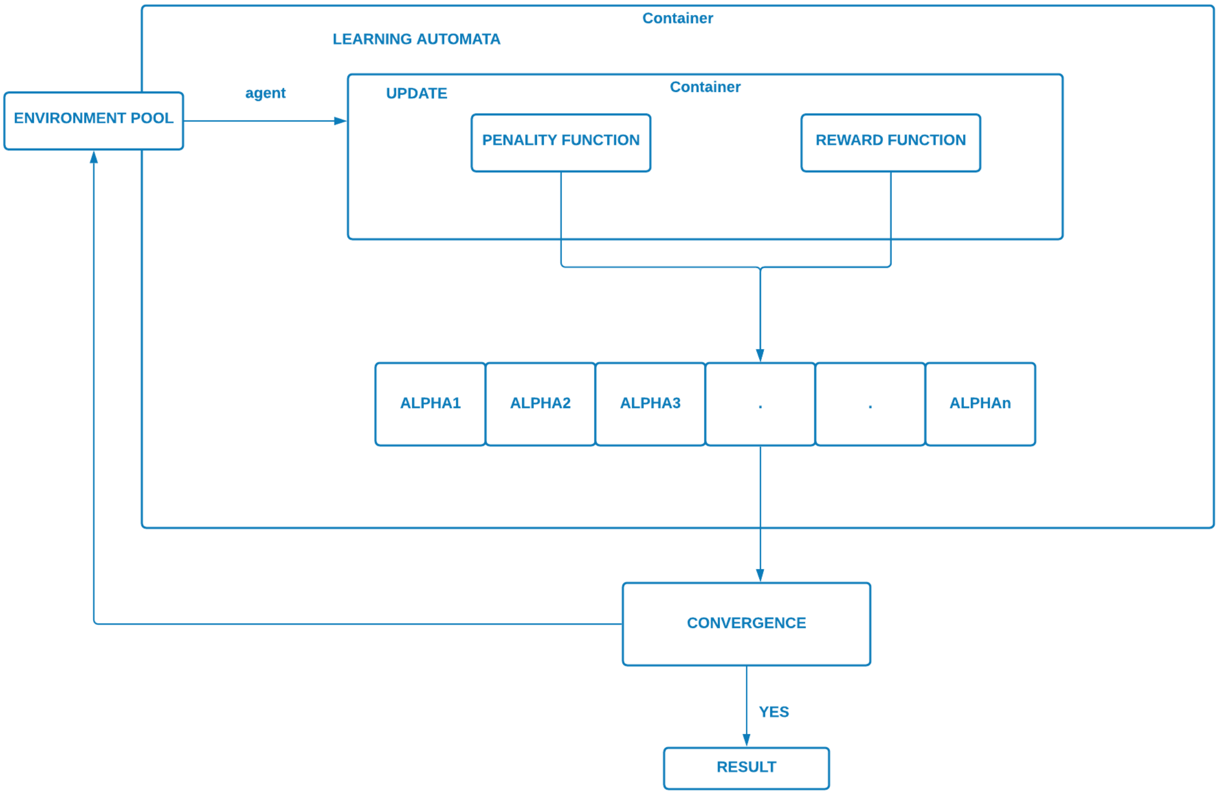


Figure 3 : The flowchart of showing the Fine Tuning of the Model using learning automaton model for hyper parameter tuning based on reinforcement learning.

i. Model Description

In our proposed model we take our input data set, pre-process and clean it then perform feature extraction technique and cast it in 10 Group of training data set(K-Fold Cross Validation). We then pass these training dataset to these different category based models such as (Machine Learning Models, Probabilistic Models, Tree-Based Models , Instance-Based Learning Models) now in the next step after training our each group dataset on each of these model. For example training dataset group-1 is trained on all of these models. We receive different result and different accuracy score for each group. We perform model evaluation and hyper parameter tuning in each case. If we look at *Fig.3.1* We can notice how each Machine learning model is attached to the corresponding learning automaton. This learning automaton performs the mechanism of hyper-parameter tuning of each of these Machine learning techniques and updates the hyper parameter to procure better results.

ii. Reinforcement Learning and Hype parameter tuning

Reinforcement Learning is what we describe as calculative approach to learn from the activity within a given system/environment, we can observe an *Environment Pool in the Fig. 3.1* to attain a particular desired goal. The goal of the agent is basically formalized in terms of special signal which is known as reward (Reward Function in *Fig. 3.1*) and it is transferred from environment to agent each and every time step. Like the motive/aim of the agent function is to increase to maximum the total amount of "scalar rewards" it generally receives. Basically it generally means to maximize not the immediate reward but the submissive total cumulative reward in the overall run, referred as return. Some of the parameters are known as the hyper parameters for eg learning rates, size of the neural network, exploration etc also the important point is that they are not automatically tuned during training.

iii. Learning Automaton

The crux of our model evaluation is the Learning Automaton. The Learning Automaton technique for hyper parameter tuning and we use the concepts of reinforcement learning to analyse the effect and the change of the values of hyper parameter on the result delivered. Basically, the tuning of the hyper parameters is to choose a particular set of the optical parameter to learn algorithms. it is a model whose value is set before the process of learning takes place. The main and the key feature of the ML algorithms is hyper parameter tuning .Learning automaton understands the most ideal action that is needed by interactions with its environment. Also, it is an adaptive decision-making unit. According to a specific probability the actions are chosen which can be updated according to the environment response and automaton which are obtained when we perform any specific activity.

iv. Learning Automaton Model Analysis

In our proposed system we pass the dataset given to each of these models to the specific learning automaton attached to it. Our Learning automaton is a simple machine that inputs the hyper parameters and the dataset and tunes the hyper parameters to analyse the result which if converges becomes the new result or if diverges is sent back to the Models and re-evaluated. It updates the hyper-parameters received and observers the changes which are sent as feedback to the result.The changes and effects are selected as per a desired probability distribution. Hence forth updates are carried out based on the environment response received by the automaton after performing a particular stimulus.

The changes or more appropriately called the feedback are split into two types in one is the ***reward signa***l and other one is ***penalty signals***. Our automaton has two essential functions that casts the update to the hyper parameter. First we have a penalty function that reduces the hyper parameter value, reducing its effect and weightage on the overall result. Next we have a Reward function that makes an addition to the value of the hyper-parameter adding more weightage and significance to it. The probability of the selected action is updated by each reinforcement signal which is received by the learning automata. There are four ways in which the amount is of reward or penalty is distributed. First is LRP in which have same number of reward and penalty, LRεP in which reward is several times more than penalty; LRI where amount of penalty is 0 and in LIP reward is 0.

For tuning parameters of reward and penalty three different modes have been defined. We will examine all the three modes of parameters for reward and penalty. We determine which setting and the mode is best for the parameters of reward and penalty. Based on the subject literature learning automata we can determine the numerical value of these parameters. We can think about a broad variety of these two parameters. To prove the effectiveness of proposed method in comparison to previous methods we tried to tune the parameters so that all the modes are considered. We receive the all the tuned hyper parameters as denoted by alpha values and we view the result if the result converges we claim the update , else we re-evaluate the model. This way we receive 6 result values and we hence evaluate the performance of each model on a particular dataset.

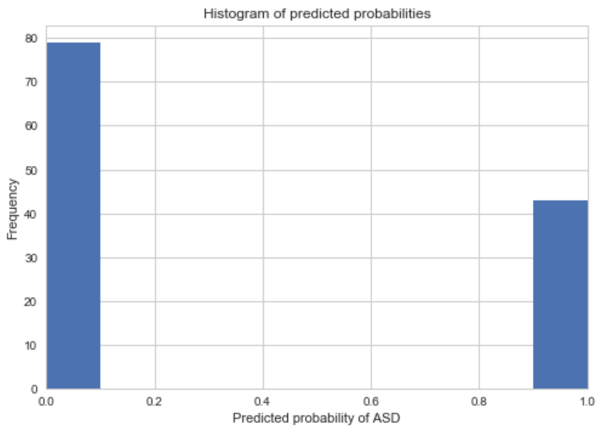
F. Result and Analysis

After using five discreet models and using specific evaluation metrics for deriving analysis and casting comparisons. We look into the result aspect of the model and their tendency towards different data and the way they give different optimised results and certain altered conditions.

i. Analysis after Decision Tree Classifier

After training our dataset on the decision tree model by using the sklearn library. We evaluate our model using specific metrics.

Figure 4: Histogram graph showing the predicted



probability of ASD

Decision Tree	
Evaluation Metrics	**Score**
Accuracy	1.0
Sensitivity	1.0
Specificity	1.0
Precision	1.0
FPR	0.0

Table 3.1: Decision tree evaluation metrics result

The decision tree model seems to a highly precise with the data. It nearly classifies all the data correctly. We use K-Fold cross-validation (with K=10) technique to gain a better insight of models accuracy.

Evaluation Metrics	Score
Model Accuracy K-Fold cross Validation	1.0
AUC Score	1.0
F-beta Score	1.0

Table 3.2: Decision tree evaluation metrics result

The score obtained after K-Fold Validation are impeccable and promise the good functioning of the model. But we know Table 3.1 shows a significant chance of Overfitting. The decision tree is not very apt towards working with data that has outliers or the data that is inclusive of small changes at certain instances. This could lead to miss-classification and cause wrong prediction which is not feasible. It is and unstable model which is relatively inaccurate most of the times. Even though we have achieved perfect scores in each evaluating metrics that des not assures us of the future classification of the data. A single outlier and some minor changes with the data could give wrong results. Hence to work on the line of high accuracy and model flexibility we cannot term Decision tree to give us the best optimised predictions.

ii. Analysis after Random Forest Classifier

Random Forest classifier is deployed by us to combat the discrepancy faced by Decision tree classifier. After Training the Model on the Random Forest Classifier we use the K-Fold Cross Validation (with K=10) again to ensure the better predictability of the model.

The results of the different evaluation Metrics are.

Random Forest	
Evaluation Metrics	**Score**
Model Accuracy K-Fold cross Validation	0.9933333333333334
AUC Score	0.9988095238095237
F-beta Score	1.0

Table 4: Random Forest evaluation metrics result

A far more reliability over the results is established for Random Forest. It reduces overfitting problem that we were facing in decision trees and also reduces the variance and therefore enhances the overall accuracy.

Random Forest is still computationally expensive and takes high amount of time since there are a number of decisions to be calculated to determine the class. We have some highly optimised Algorithms that are capable to give better results.

iii. Analysis after Logistic Regression

After analysing the two classifier model. Wes should realise the results that generated by a regression model. The logistic regression model does not enhances/optimise the overall result but gives us insights about the data dependability and estimates probabilities that helps to predict likelihood of ASD Positive/Negative class.

Logistic Regression	
Evaluation Metrics	**Score**
Model Accuracy K-Fold cross Validation	0.97043715846994538
AUC Score	0.99740986834010081
F-beta Score	0.93073593073593064

Table 5: Logistic Regression evaluation metrics result

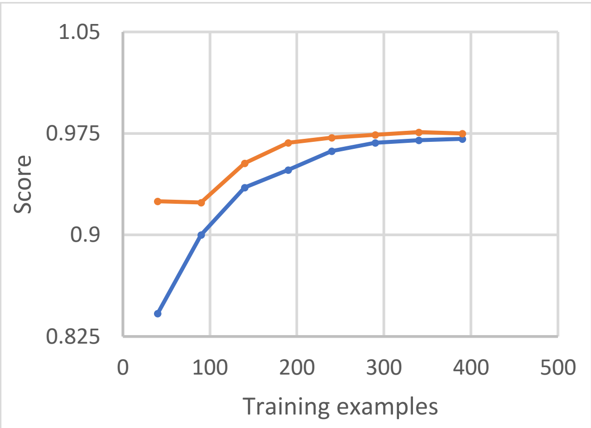


Figure 5: Learning rate curve of Logistic Regression

iv. Analysis by Gaussian Naive Bayes Model

Gaussian Naive Bayes	
Evaluation Metrics	**Score**
Model Accuracy K-Fold cross Validation	0.88500000000000001
AUC Score	0.94450904392764845
F-beta Score	0.83700440528634357

Table 6: Gaussian Naive Bayes Algorithm evaluation metrics result

We have used another probability oriented algorithm to analyse our problem. Naive Bayes Algorithm is quick and easy to implement. Naive Bayes works exceptionally with the training data that a number of categorical values. The Naive Bayes algorithm assumes all the features are independent. This is highly unlikely in real life. Since our data set is very cubical and has a large number of feature dependency we cant assure the productiveness and efficiency of Naive Bayes for ASD Classification.

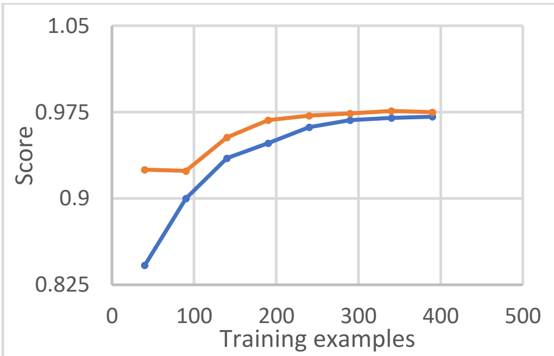


Figure 6: Learning rate curve of Gaussian Naive Bayes Algorithm

v. *Analysis of SVM Model*

We have then used an SVM Model with a kernel chooses as linear. By Far we observe the most feasible and accurate learning curve we realise. That the performance metrics is exceptionally good and the accuracy is feasible. The model does not encourages overfit. C is the hyper parameter (regularisation parameter) its value choose C=1.0. The range of C is (0.01 - 100). Another Hyper-parameter associated with SVM is Gamma range(0.0001 < gamma < 10)

Since we have a 'Linear' Kernel SVM we have to only Tune the C- Parameter and use optimisation approach, Gamma values shouldn't be very small.The basic issue with SVM is that it requires the dataset to have a perfect clear margin , line of separation between its data/features.

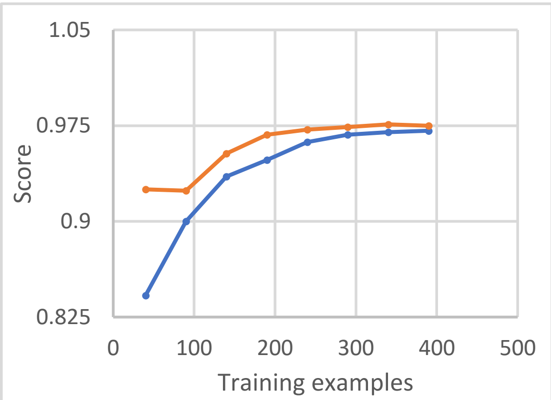


Figure 7: Learning rate curve of SVM Model

SVM separates the data into positive and negative classes only does not let any point to be misclassified this principal of SVM ruins the Generalisation theme of our model and causes a Overfit. Since real data is noisy and is high dependent we use these hyper parameters like C and Gamma associated with SVM to avoid Overfit. For our use case after several churns of trial w found the C=1.0 and Gamma Value = 2 to be best suitable for prediction. It learning Curve is cohesive and does not overfits.

Support Vector Machine	
Evaluation Metrics	**Score**
Model Accuracy K-Fold cross Validation	0.9965517241379309
AUC Score	1.0
F-beta Score	1.0

Table 6: Support Vector Machine evaluation metrics result

SVM by far is the best learning algorithm that helps us achieve the best results.

vi. *Analysis of KNN Model*

KNN Model with multi values of K is tested out on put dataset. The evaluation metrics score aren't very impressive as compared to SVM but somewhat good in comparison to Naive Bayes and Logistic regression. The choice of hyper parameter K is very crucial and hence we choose multi values of K and tested them to see if any optimisation can be done. We tested the value K in range of (10,50). We start with the K-value chosen as 10. After doing the traversal through all these values of K we find out that the K=48 gives the best score.

K-Nearest Neighbour K=10	
Evaluation Metrics	**Score**
Model Accuracy K-Fold cross Validation	0.94745901639344265
AUC Score	0.99300787498461918
F-beta Score	0.91928251121076232

Table 7: KNN Model evaluation metrics result

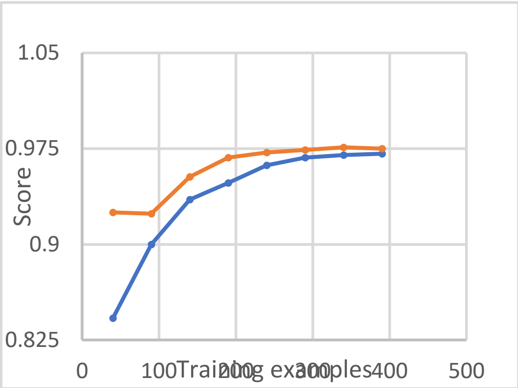


Figure 8: Learning rate curve of KNN Model

Conclusion

We have seen the trend, patterns and overall mechanism of Autism Detection by using several prominent classification models. A number of performance metrics were used to analyse the performance of the models implemented. We got through an entire journey of enchantment and optimisation of the algorithms used. As per our research and findings we can claim SVM to work the best for our problem statement. After Fine tuning the hyperparameter of our SVM Model is very precise and fluent for ASD Classification. The overall accuracy of ASD Classification by using SVM comes to be 98.30%-99.65%. If we carefully remove all the non-clinical Data and handle the missing values and then again fine tune the hyper parameters larger accuracy can be achieved. Also larger the dataset the more accuracy can be attained. With our case there is very limited dataset available due to with making all of these model work and making right analysis based on their metrics score is a tedious task. Nevertheless SVM has shown coherent and similar patterns resulting to similar range accuracy majority number of times. Playing with probabilities and different models has brought us to final conclusion that SVM can be used very effectively for ASD classification over any other Classifier or Model.

3) Best Model

We studied various algorithm to predict whether a person has autistic traits or not. We used logistic regression, decision tree, random forest tree, KNN and SVM. After exploring all these algorithms, we found out that each algorithm works fine but the model which worked best was SVM. Overall evaluation metrics performance of SVM was reliable and efficient the the AUC and F-beta score were 1.0 and the Accuracy score was around 0.998 onwards.

SVM is also a very popular classification model also it is a supervised technique. In this model people might get disappointing results because beginners can miss some of the important and significant steps, but SVM is better in term of accuracy, and it can perform well in n-Dimensional space. Also, in general SVM does not undergo overfitting and if the separation between the classes is clear then the SVM model is going to perform well and, in our project, there were only two classes autistic and non-autistic. In SVM we have two important parameters namely C and Gamma. The parameters which can we be adjusted are called hyper parameters and before training the model we need to set these hyper parameter. In SVM, for distinction between classes a decision boundary is created.

In standard SVM the black line in below figure will separate the classes by making the decision boundary. But this model is overfitting and too specific. This model will perform well in training set but will not perform on unknown value.

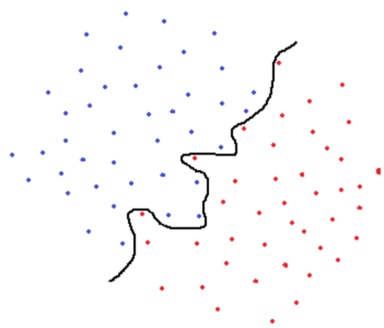


Figure 9: Showing an Overfitting of an decision boundary of SVM

To overcome this problem, we have a concept of soft margin which allows some of the data set to be misclassified or to be classified as wrong and this model is more generalized. In soft margin the decision boundary is a straight black line.

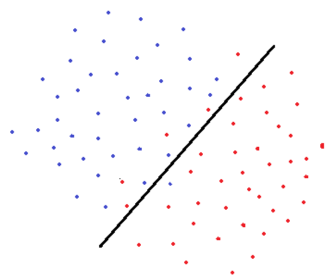


Figure 10:Soft Margin concept of SVM to provide generalisation

This model is more optimized because the points which are correctly classified are maximized and the distance of classes to the decision boundary is also maximized.

References

1. Suman Raj and Sarfaraz Masood. Analysis and detection of autism spectrum disorder using machine learning techniques. Procedia Computer Science,167:994–1004, 2020.1

Citations:

1. M. F. Rabbi, S. M. M. Hasan, A. I. Champa and M. A. Zaman, "A Convolutional Neural Network Model for Early-Stage Detection of Autism Spectrum Disorder," 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD), 2021, pp. 110-114, doi: 10.1109/ICICT4SD50815.2021.9397020.

2. A. SARANYA and R. ANANDAN, "Autism Spectrum Prognosis using Worm Optimized Extreme Learning Machine (WOEM) Technique," *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2021, pp. 636-641, doi: 10.1109/ICACITE51222.2021.9404729.

3. A. Sharma and P. Tanwar, "Deep Analysis of Autism Spectrum Disorder Detection Techniques.," *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, 2020, pp. 455-459, doi: 10.1109/ICIEM48762.2020.9160123.

References

1. Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., and Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the abide dataset. *Neuroimage Clin*. 17, 16-23.doi: 10.1016/j.nicl.2017.08.017.

2. Hyde, K. K., Novack, M. N., LaHaye, N., Parlett-Pelleriti, C., Anden, R., Dixon, D. R., et al. (2019). Applications of supervised machine learning in autism spectrum disorder research: a review. *Rev. J. Autism Dev. Disord*. 6, 128–146. doi: 10.1007/s40489-019-00158-x.

3. Jiao, Y., Chen, R., Ke, X., Chu, K., Lu, Z., and Herskovits, E. H. (2010). Predictive models of autism spectrum disorder based on brain regionalcortical thickness. *Neuroimage* 50, 589–599. doi: 10.1016/j.neuroimage.2009.12.047.

4. Aghdam, M. A., Sharifi, A., and Pedram, M. M. (2018). Combination of rs-fMRI and sMRI data to discriminate autism spectrum disorders in young children using deep belief network. *J. Digit. Imaging* 31, 895–903. doi: 10.1007/s10278-018-0093-8

5. Bi, X.-a., Liu, Y., Jiang, Q., Shu, Q., Sun, Q., and Dai, J. (2018). The diagnosis of autism spectrum disorder based on the random neural network cluster. *Front. Hum. Neurosci*. 12:257. doi: 10.3389/fnhum.2018.00257