



## Prediction of Autism Spectrum Disorder using Various Machine Learning Techniques

<sup>1</sup> Sparsh Lamba,<sup>2</sup> Abhishek Singh

<sup>1</sup> B.Tech,<sup>2</sup> B.Tech,

<sup>1,2</sup>Department of Computing Technologies,

<sup>1,2</sup>SRM Institute of Science and Technology, Tamilnadu, India.

**Abstract :** Autism spectrum disorder (ASD) is an unusual developmental and behavioral condition that directly impacts the way of life of a human in terms of his mental and physical wellbeing. It restricts the use of interactional, behavioral - cognitive abilities as well as qualities of the directly affected identity and although its signs might very well differ from person to person. Artificial intelligence's subset, machine learning is implemented to diagnose ASD using various classification techniques in which prediction methods were developed based on chronological data source, and also used those information to predict that the person who suffers ASD before or not. Also it can be used mostly for making decisions under ambiguity. On ASD dataset eight machine learning were build, among these Artificial Neural Network give very good accuracy 99% and the lease accuracy 42% was achieved by Quadratic Discriminant Analysis, other six classification models are Decision Tree Classifier, K Neighbors Classifier, Multinomial NB, Logistic Regression, Linear Discriminate Analysis and Random Forest Classifier.

**IndexTerms - Autism, Random Forest, Logistic Regression, Decision Tree Classifier.**

### I. INTRODUCTION

Autism spectrum disorders (ASD) brings on indefinite difficulties in the social, behavioral, and emotional sensitivity of an individual. People diagnosed with this difficulty, tend to have various kinds of symptoms and that's why this is commonly called as a 'spectrum' disease. Since ASD is actually a neurological mental disability, there is truly no particular medical test for it, thus trying to make the treatment of ASD an extremely difficult task. Now although these diseases can be developed in childhood, there have been some situations where the signs really aren't given a diagnosis until youth or adulthood. ASD, even in these modern technological times still has no exact treatment. An accurate detection and a head start in treatments can possibly lead to better outcomes. The data source used is the ASD monitoring data. The sets of data focus on providing observations into different factors that affect the predictive model of the disease. Learning algorithm techniques like Artificial Neural Network (ANN) were being used for checking out the optimal feature for every dataset. Several performance measures have been used in analyzing and make a comparison on every angle of the framework from each and every possible way.

### II. RELATIVE WORK:

This article utilizes eight most influential Machine Learning (ML) algorithms and provides an empirical analysis about their performances in trying to diagnose autism disorder on 4 different standard datasets, which seems to be slightly higher comparatively and also comes exclusively from the Qchat, aq-10-child, and aq-10-adult medical tests. In doing just that, we also incorporate accuracy, responsiveness, precision, and classification accuracy performance measures to carefully examine their performances. According to the experimental results, the most effective results are acquired with c-svc, a classifier based on a SVM classifier. More importantly, when it comes to c-svc results evaluated even lead to 100% in all sources of data. Multiple logistic regression has indeed been taken second place. Alternatively, the lowest results obtained with the c4.5 automated system, a choice tree-based algorithm [1].

In this employment, we present an interactive machine caregiverchild communication (c3i) system that helps youngsters with ASD practicing IJA skill sets. c3i is a innovation done by computers, representing a system that incorporates a care provider in trying to teach loop, while also maintaining some great benefits of both person and computer-administered interference. A feasibility analysis with 6 dyads (caregiver-child with ASD) was performed. A near substantial increase with small effect size on IJA efficiency was observed. In the meantime, physiology-based fatigue analysis has shown that c3i did not continue to increase due to the caretakers over the duration of the investigation. To the best of knowledge, this is really the first autonomous system

planned for trying to teach IJA skills to youngsters with ASD trying to incorporate caretakers inside the loop to boost the possibilities for generalized statement in real-world [2].

Our results indicate a number of speeches, audio, and page which includes instructions of ASD intensity in Hebrew talking youngsters. The above includes can also be used by a CNN architecture to produce quite accurate predictive model of ADOS, managed to score when trying to apply an architecture that uses various human speech sample that was taken from every youngster in tandem. We propose that this speech processing algorithm might well have great clinical utility in evaluating early ASD threat and as an innovative primary outcome for measuring ASD consequences that changes with time and following treatments [3].

In this document, we've taken the information of grown up adults from the age of 20 to 65 years of age as well as we did attempt to make the autism diagnosis by trying to apply data mining algorithms. There is sufficient range after all research question in the set of data that has been used in our study. The KNN, SVM, LR, Cart, Naïve Bayes and LDA algorithms were being used in classifying. Here we transform the data about some characteristics into a numeric value [4].

This paper mainly focuses on proposed framework that would assist in predictive model of ASD, in an ordinary person such that specific diagnostic can indeed be completed and then further treatment options may well be decided to follow. The set of data used is of the autistic spectrum disease monitoring data. The sets of data focus on providing information and insight into major factors influencing the predictive model of both the disease. Machine learning algorithms like Decision tree, Random forest, logistic regression, SVC and ANNs are used for finding out the optimal model for each set of data. Several performance indicators have been used in order to examine and make a comparison of each framework from each and every angle possible [5].

### III. DATASET DETAILS:

The economic effect of autism as well as the increasing number of ASD cases the world over exposes an urgent need to get the growth of easily accomplished and effective assessment techniques. Therefore, a time-efficient and widely available ASD monitoring is imminent to help healthcare providers and inform individual people whether they could perhaps pursue traditional clinical treatment. The rapid increase in the number of ASD deaths worldwide prescribes sets of data behavior and social traits. however, such sets of data are rare making it more difficult to undertake throughout analyzing to enhance the performance, sensitivity, specificity and predictive accuracy of the ASD testing process. Currently, very restricted autism sets of data related to clinical or testing can be found and most of them are multiple genes in nature. Hence, suggest a new set of data genetically based testing of individuals that supposed to contain 20 features includes to be used by for further evaluation helping to optimize influencing autistic traits and going to improve the categorization of ASD cases which are given in TABLE 1.

TABLE 1: Dataset Features Description

Feature	Type	Description
Age	Number	toddlers (month), children, adolescent, and adults(year)
Gender	String	Male or Female
Ethnicity	String	List of common ethnicities in text format
Born with jaundice	Boolean	Whether the case was born with jaundice
Family member with PDD	Boolean	Whether any immediate family member has a PDD
Who is completing the test	String	Parent, self, caregiver, medical staff, clinician, etc.
Country of residence	String	List of countries in text format
Used the screening app before	Boolean	Whether the user has used a screening app
Screening Method Type	Integer	The type of screening methods chosen based on age category
A1: Response of Q1	Binary	See Table 2 for details Q1
A2: Response of Q2	Binary	See Table 2 for details Q2
A3: Response of Q3	Binary	See Table 2 for details Q3
A4: Response of Q4	Binary	See Table 2 for details Q4
A5: Response of Q5	Binary	See Table 2 for details Q5
A6: Response of Q6	Binary	See Table 2 for details Q6
A7: Response of Q7	Binary	See Table 2 for details Q7
A8: Response of Q8	Binary	See Table 2 for details Q8
A9: Response of Q9	Binary	See Table 2 for details Q9
A10: Response of Q10	Binary	See Table 2 for details Q10
Scoring Result	Integer	See Table 2 for details
ASD	Boolean	adults diagnosed with ASD

In this dataset there are ten cognitive and behavioral includes ten questions which are detailed in TABEL 2 and ten individual's characteristic features total 20 features that do have demonstrated to be successful in trying to detect the ASD situations from control system in behavior patterns science.

TABLE 2: Details of Variables Mapping

Variable	AQ-10-Adult Features (16 and older)
Q1	I often notice small sounds when others do not
Q2	I usually concentrate more on the whole picture rather than the small details
Q3	I find it easy to do more than one thing at once
Q4	If there is an interruption, I can switch back to what I was doing very quickly
Q5	I find it easy to read between the lines when someone is talking to me
Q6	I know how to tell if someone listening to me is getting bored
Q7	When I'm reading a story I find it difficult to work out the character's intentions
Q8	I like to collect information about categories of things
Q9	I find it easy to work out what someone is thinking or feeling just by looking at their face
Q10	I find it difficult to work out people's intention

**IV. PROPOSED SYSTEM**

An autism data set is a collection of numbers or values that is related to a particular subject. This dataset is used to implement the autism system. Firstly, this dataset undergoes data preprocessing and in order to make this process easier, it is divided into four stages: data cleaning, data integration, data reduction, and data transformation. In this four stages reprocessed in order to check missing values, noisy data, and other inconsistencies before executing it to the algorithm.

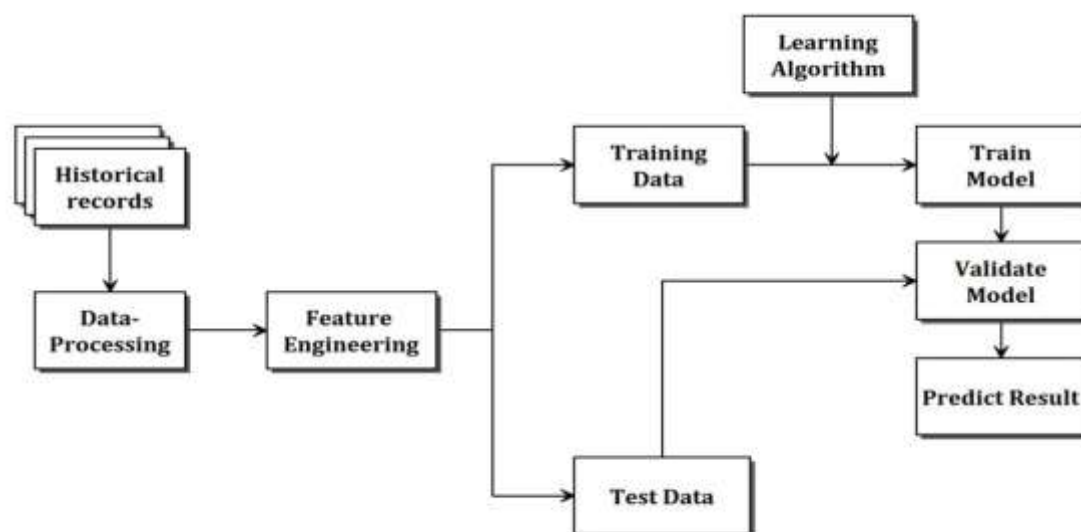


Fig1: Proposed System

Here a raw dataset will be converted into clean and valuable dataset (labeled dataset). This clean data split into two type of dataset like train and testing dataset, for this splitting dataset we use Scikit-learn (Sklearn) library. This trained dataset have classes, with help of this class our system can help in training our model with help of Deep Learning (DL) algorithm like Decision Tree Classifier, K Neighbors Classifier, Multinomial NB, Logistic Regression, Linear Discriminate Analysis, Quadratic Discriminate



Analysis, Artificial Neural Network (ANN), and Random Forest Classifier. This trained model is stored in a local system, if user provides a testing data, the trained model will give an output. Fig 1 shows this process clearly.

## V. RESULT ANALYSIS:

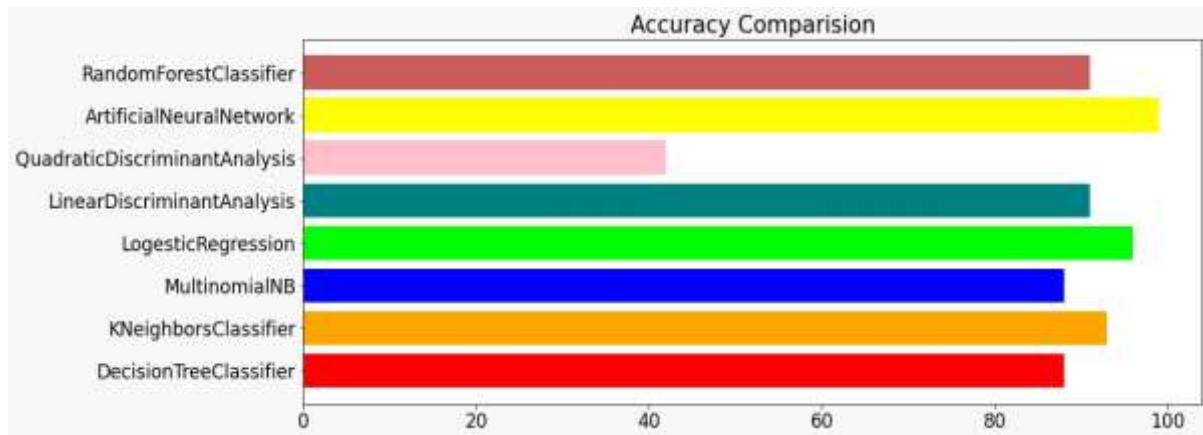


Fig 2: Accuracy Comparison Graph

After effective implementation the algorithms and the outcomes of their analyzed effectiveness were acquired. The outcomes show that the accuracy of Decision Tree Classifier - 88, K Neighbors Classifier – 93, Multinomial NB – 88, Logistic Regression – 96, Linear Discriminate Analysis – 91, Quadratic Discriminant Analysis – 42, Artificial Neural Network (ANN) – 99, Random Forest Classifier – 91. The best performance algorithm is ANN and least performance algorithm is Quadratic Discriminant analysis algorithm.

## VI. CONCLUSION:

The aim is to improve the classification accuracy of autism spectrum disorder using machine learning algorithms. For this objective, eight classification techniques are being used. The Artificial Neural Network (ANN) achieved 99% much better than all other algorithms. Based on this results, it is arrived at the conclusion that artificial neural network (ANN) algorithm performed much better than other algorithm.

## REFERENCES

- [1] Hidayet Takçı, Saliha Yeşilyurt, “Diagnosing Autism Spectrum Disorder Using Machine Learning Techniques” 2021 6th International Conference on Computer Science and Engineering (UBMK), IEEE, Sept. 2021, Ankara, Turkey.
- [2] Guangtao Nie, Akshith Ullal, Zhi Zheng, “An Immersive Computer-Mediated Caregiver-Child Interaction System for Young Children With Autism Spectrum Disorder”, IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society, May 2021.
- [3] Marina Eni, Ilan Dinstein, Michal Ilan, Idan Menashe, Gal Meiri, Yaniv Zigel, "Estimating Autism Severity in Young Children From Speech Signals Using a Deep Neural Network", IEEE Access ( Volume: 8), July 2020.
- [4] Bhawana Tyagi, Rahul Mishra, Neha Bajpai, “Machine Learning Techniques to Predict Autism Spectrum Disorder”, 2018 IEEE Punecon, Dec. 2018, Pune, India.
- [5] Astha Baranwal, M Vanitha, “Autistic Spectrum Disorder Screening: Prediction with Machine Learning Models”, 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Feb. 2020, Vellore, India.