International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

Prediction of Autism Spectrum Disorder based on Machine Learning Approach

Dhanyatha Sriram[1], Greeshma A[2], Gouthami[3], Yeshwanth M[4], Prof. Y Shobha[5]

[1,2,3,4]Student, Dept. of Computer Science and Engineering, Bangalore Institute of Technology, Bengaluru, India
[5]Associate Professor, Dept. of Computer Science and Engineering, Bangalore Institute of Technology, Bengaluru, India

---

Abstract--In recent times Autism Spectrum Disorder (ASD) is gaining its momentum faster than ever before. As we know, ASD is a neurodevelopment syndrome which includes social communication and behavioural challenges. The effect of ASD and severity of symptoms are different in each person. Diagnosis of autism can be done at any age. According to studies and most observed behaviour patterns on autism affected patients are: Aggression, Self-Injury, Elopement, Tantrums, Obsession, On Compliance etc. Hence it becomes very important to detect any sign that leads to the severe ASD as early as possible. In this paper, Machine Learning (ML) approaches are applied, namely Random Forest, Naive Bayes (NB), Decision Tree, K-Nearest-Neighbour (KNN), Logistic Regression (LR) and Support Vector Machine (SVM). Performance measure helps to analyse the accuracy level of each data with all the applied algorithms to find out which gives accurate results in terms of time and accuracy, with this user will be able to find out if they are suffering from ASD or not.

Keywords-- Machine Learning, Hybrid Model, Random Forest, Naive Bayes (NB), Decision Tree, K-Nearest-Neighbour (KNN), Logistic Regression (LR) and Support Vector Machine (SVM).

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neuro development disorder that affects a person's communication, interaction, behaviour. The diagnosis of autism can be done at any age, its symptoms generally appear in the first two years of life and develops through time. Autism patients face different types of challenges such as difficulties with concentration learning disabilities, anxiety, obsession, depression and many more. Current explosion rate of autism around the world is numerous and is increasing at a very high rate. Early detection of autism can come to a great help because it will help the patient's condition from deteriorating and would help to reduce long term cost. ML algorithms are applied to diagnose ASD problem as a classification task in which prediction model are built based on chronological dataset, and then those patterns are used to predict if that person is suffering from ASD or not.

In this study the aim is to develop and implement a web application to detect ASD in a person of any age group with the help of ML algorithms in a small

period of time using datasets to get accurate results. To propose, new ML framework with dataset related to the autism screening of adults, adolescent and toddlers that have influential features and perform predictive analysis and Hybrid model using ML algorithms like Decision Tree (DT), Naive Bayes (NB), Random Forest, K-Nearest-Neighbor (KNN), Logistic Regression (LR) and Support Vector Machine (SVM).

2. LITERATURE SURVEY

Vast amount of work has been done in the past related to ASD.[1]David Peebles used, new ML framework related to autism screening of adults and adolescents that contain features and perform predictive analysis using Logistic Regression to reveal important information related to autism screening, but the datasets were anonymous, and it was only for adults and adolescents' children and toddlers could not take part in it, [2] developed and evaluated a ML approach to classify case studies in Autism and Disabilities monitoring. In ASD system and data children aged 8 years were taken for multiple clinical tests, but all symptoms weren't predicted, and it was only for children of 8 years, Fadi Thabtah [3], shed light on recent studies that employ ML in ASD classification to discuss the pros and cons. The development of intelligent diagnostic tools based on the ML, but the limited research had been conducted on identifying and evaluating ASD traits in the clinical environment under the DSM-5.

Armin Lawi [4], adult humans aged 17 years and over are used as sample datasets to diagnosis using the ML classification method to see the best estimation results and k-fold values in each classification method, but Logistic Regression had many classification errors in the testing process and K-Nearest Neighbours required number of iterations to produce best performance, [5] a comparative study to find whether children aged (4 to 11years) have ASD by using Linear Discriminant Analysis and K-Nearest Neighbour, Euclidean distance was used in the KNN algorithm, and the k-value was taken as 1, but Only children aged between 4 to 11 years could be tested and in paper [6], the use of ML techniques for grading autism in children. It is mainly used to predict the grade of childhood autism using five ML algorithms: Naïve Bayes, Support Vector, Decision Tree, K Nearest Neighbour, Support Vector in consistent with Childhood Autism Rating Scale (CARS) diagnostic criteria. the clinically diagnose the grades of autism in early stage, but the grading is only for children. This work included only a limited number of classification models.

Murat Gok [7], developed a ML model, trained using brain development gene expression data, but there are two early diagnostic tools for ASD in risk gene: imaging and biomarkers. Former is expensive, and latter does not meet the test requirements due to the complexity of the disorder and the increase in sequencing capacity, In paper [8], there was a partial improvement in a core

symptom of ASD by examine the prediction with ML methods with statistical estimation, information theories and mathematical learning, but their data were backdated and used short-term variables and clinician- rated improvement in autism,[9] the main purpose was to introduce deep learning-based framework, by using LeNet-5 architecture and implement functional MRI image.

KaziShahrukh Omar [10], an autism prediction model was developed by merging Random Forest-CART and Random Forest ID3 and a mobile application was developed based on the proposed prediction model but due to wide range of age their research was not accepted for people of all age group as screening approach. Rahul Mishra [11], predicted a model that was built based on chronological dataset, and then used those patterns to predict that the person is suffering from ASD or not, but this was least effective as its accuracy is just only 67.5564%. [12] detection of ASD was attempted using various ML and deep learning techniques. Various performance evaluation metrics were used to analyse the performance of the models implemented for ASD detection on non-clinical dataset from Adolescents and adults.

V. Geetha [13] evaluated the performance of ML Techniques Namely Multilayer Perceptron, J48 Classifier, Naïve Bayes, and Bayesian Network to predict ASD in adults at an early stage. This test would be done only to adults, In paper [14] a novel integrative approach to predict ASD phenotypes from biological processes defined by genetic alterations, but their limitations was analysing individual gene variants in favour of considering biological processes disrupted by a heterogeneous set of gene variants. Mary Hanleya [15], a mixed-method, multi-informant approach with teachers and parents were used to understand how sensory differences affect learning and school life for autistic pupils. Parents and teachers reported that a significant proportion of children with ASD were affected by unusual sensory reactions of sensitivities and the study aimed to understand these impacts.

3. PROPOSED SYSTEM

The aim is to develop and implement a web application to detect ASD in a person of any age group with the help of ML algorithms in a small period of time using datasets to get accurate results.

To propose, new ML framework with dataset related to Hybrid autism screening of adults, adolescent and toddlers that have influential features and perform predictive analysis using ML algorithms like Decision Tree (DT), Naive Bayes (NB), Random Forest, K-Nearest-Neighbor (KNN), Logistic Regression (LR).

A. Dataset Acquisition

Table 1

Sl. No	Attributes	Value Yes=1 N0=0
1	**Teenagers/Adults**: Do you hear small sounds when most	1/0

	of the people don't? **Toddlers**: Does your child tend give random answers to questio or make random comments?	
2	**Teenagers/Adults**: Do you concentrate on the whole pictur or do you give importance to small details in the picture? **Toddlers**: Does your child not respond to their name?	1/0
3	**Teenagers/Adults**: Can you do multiple work simultaneously? **Toddlers**: Does your child avoic eye contact?	1/0
4	**Teenagers/Adults**: In the midd of work if you get interrupted, can you go switch back to what you are doing quickly? **Toddlers**: Does your child not engage in pretend play with othe children?	1/0
5	**Teenagers/Adults**: Can you Concentrate on reading when someone is talking around you? **Toddlers**: Does your child struggle to understand other people's feelings?	1/0
7	**Teenagers/Adults**: While readi do you get the part, you are reading in just one read or do yo read it multiple times to understand it? **Toddlers**: Does your child have obsessive interests?	1/0
8	**Teenagers/Adults**: Do you like to collect information about categories of things? **Toddlers**: Is your child over or under-sensitive to smells, tastes or touch?	1/0
9	**Teenagers/Adults**: Is it hard fo you to make eye-contact? **Toddlers**: Does your child struggle to socialize with other children?	1/0
10	**Teenagers/Adults**: Do you find difficult to work out on people's	1/0
	intention? **Toddler:** Does your child avoid physical contact?	
11	**Age**	(4-11years)/ (12-17Years)/ (18+ years)
12	**Gender**	Male/ Female
13	**Ethnicity**	1-11
15	**Jaundice**	1/0
16	**Autism**	1/0
17	**Country_of_residence**	1-52
18	**Used app**	1/0
19	**Result**	1-10
20	**Age-description**	Toddler/ Adolescent/ Adult
21	**Relation**	Self/Parent
22	**Class1**	ASD/Not

Our dataset is taken from publisher Faizu Nabi to implement ML techniques. We have taken our dataset from there and also implement web application to detect ASD or not. Our dataset consists of 22 attributes and 1100 instances, out of 22 attributes the first ten (from 1-10 in the above table 1) consists of the questions. The output of the dataset consists of two values i.e., value 1 is YES, "person has ASD" (393 instances), value 2 is NO, "Not having ASD" (707 instances).

B. System Design

1) *System Architecture*:

Below Fig.1 shows the system architecture of our system. The Autism raw data is stored in a database, and it passed through Data Processing. Among all the features in the dataset, only few features are relevant for the prediction. This

phase selects the relevant features from the data to maximize the prediction accuracy. These features are then passed to a ML Model which is trained on large amounts of training data. Based on the measures of data by the user, the data point is classified as Autism or not.

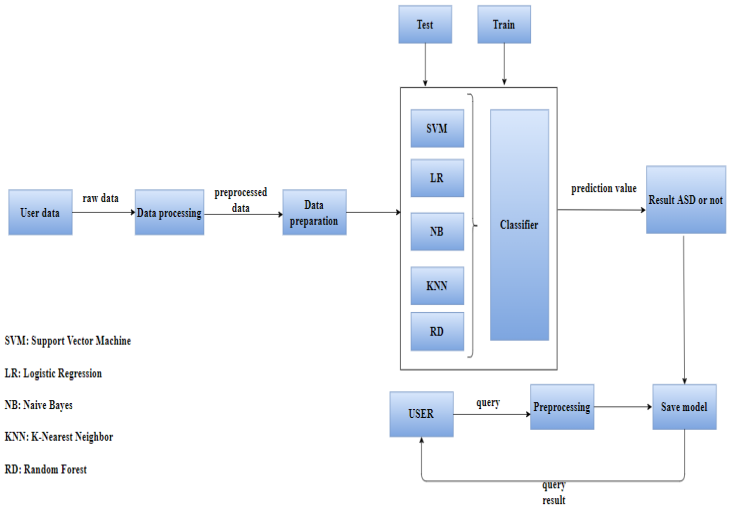

Fig 1. System Architecture

2) Interface Design:

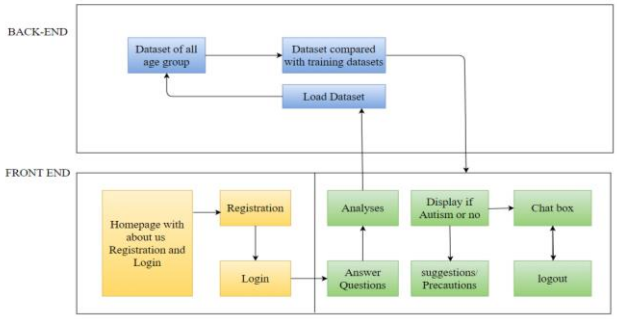

***Fig 2.* Interface Design**

The above Fig 2 is the Interface diagram for our system. The system starts with empty browser state, where the user enters the URL. If the server is down then user is directed back to empty browser for trying again. With server running successfully the user is directed to the home page. User can further select any module using the navigation bar. Upon selecting the Login/Register module, the user can register or login. After login, the user is provided with a questionnaire component which they must answer. Upon submitting, the data is loaded into the backend and then compared with the other training datasets and the result is sent to the user in the frontend. Selecting the Chat Bot Module, the user can ask questions is they have any. If the user selects any other module, the various visualizations for that respective module are displayed.

C. Methodology

1) ***Data pre-processing:*** This process of preparing the data and making it suitable for a ML model. It is the first and crucial step while creating ML model. Data Pre-processing is a process of cleaning the raw data i.e., the data is collected in the real world and is convert other words. Real time data collected from publisher Faizu Nabi.

2) ***Developing the hybrid model:*** To generate prediction of autism traits, using ML Algorithms like LDA (Linear Discriminant Analysis), Naive Bayes (NB), Classification and Regression Trees (CART), K-Nearest-Neighbor (KNN), Logistic Regression (LR), and Support Vector Machine (SVM) for classification and comparison of data for more accurate results.

3) ***Training and Testing Model****:* It refers to a predictive modelling problem where a class label is predicted for a given example of input data. The classifier using 'training data set', tune the parameters using 'validation set' and then test the performance of your classifier on unseen 'test data set'.

D. Implementation

This paper contains six algorithms and these are the main three:

1) ***Navies Bayes***: Is a classification algorithm based on bayes theorem and with an assumption of dependency among the predictors

$$P(H/E) = (P(E/H).\ P(H))/P(E)$$

Algorithm 1

Step 1: Start

Step 2: Load the dataset from the file and convert it into float from string.

Step 3: Bernoulli NB imported performs:

Splitting into test and train data (gives split ratio) Calculating mean of attributes of class values.

Calculate standard deviation (SD = root V). Then calculate Bernoulli probability density function.

$$\text{Sample Variance} = s^2 = \frac{\Sigma(X - \overline{X})^2}{n-1}$$

Step 4: Prediction is made using summarize and input vector.

Step 5: Get prediction method using summarize and test data set.

Step 6: Get accuracy method predicts accuracy using test set and prediction from get prediction.

Step 7: End

2) ***Logistic Regression***: Is used for classification i.e., binary classification and multilevel classification which has a dependent variable and an independent variable.

Sigmoid function

Y = 1 / 1+e^-x

x —> independent variable

e —> Euler's constant

Y —> converts independent variable into probability (0 to 1) with respect to dependent variable.

Algorithm 2

Step 1: Start

Step 2: Import required module.

Step 3: Generate dataset i.e., dependent and independent variable.

Step 4: Split the dataset into test and train samples.

Step 5: Perform logistic regression using train data.

Step 6: Perform logistic regression using train data.

Step 7: Make prediction using the model from the previous step.

Step 8: Display confusion matrix (where false negative and false positive must be minimized.).

Step 9: End

3) ***Support Vector Machine (SVM)*:** It is a supervised ML algorithm which can be used for classification and regression. Here we plot each data item as a point in n-dimensional space with the value of each feature being the value of particular coordinate then we perform classification by finding the hyper plane.

Algorithm 3

Step 1: Start

Step 2: Import sklearn library.

Step 3: Loading the data i.e., supervised learning data.

Step 4: Creating object to the user given data.

Step 5: Splitting the data into test and train.

Step 6: Generating the model.

Step 7: Model evaluation and prediction in the form of a graph.

Step 8: End

4. RESULTS

The system shows better performance comparing to the other existing approach of autism screening. The model can predict autism traits for different age groups. The user answers all the mentioned questions given in the form page and other required details thoroughly without missing any. Hybrid model which includes all the mentioned algorithm has been used for analyzing and detection of Autism. The final stage detects if a user has Autism or not. The web application also has chat box where the user can chat and ask frequently asked questions (FAQ) which referred to gain some insight on the importance of getting Autism test at early stage of occurrence to prevent from getting any worse. This system also provides a comparative view among different ML approach in terms of their performance.

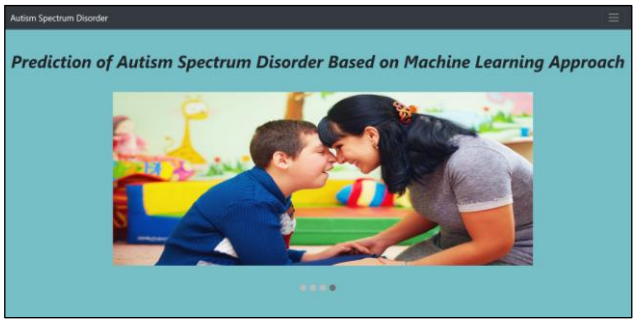

Fig 3. Home Page

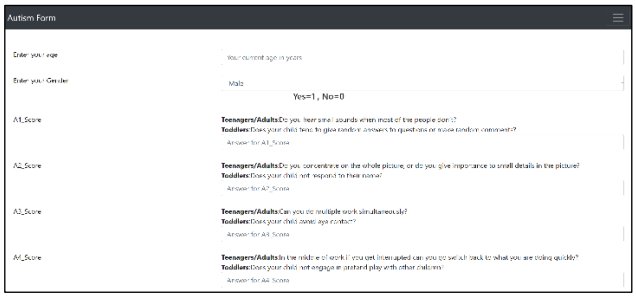

Fig 4. Form Page with Questions

Fig 5. Form Page with other details

Fig 6. Result Page with NO prediction of Autism

Fig 7. Result Page with Prediction of Autism

Fig 8. Chatbot and Logout page

A. *Performance Comparison*

The accuracy of ML Algorithms depends on certain factors like precision, A1 score, confusion matrix, recall, and many more. ML learns or approximates a function to best map inputs to outputs rom examples in the training dataset. The below Fig 9 shows the accuracy Result for Table 2 Accuracy Score.

Classifier	Test Accuracy
K Nearest Algorithm	89.5454%
Navies Bayes	95%
Logistic regression	91.12%
Support vector machine (SVM)	90.21%
Decision Tree	94%
Random Forest	85%
Hybrid Model	100%

Table 2. Accuracy Score

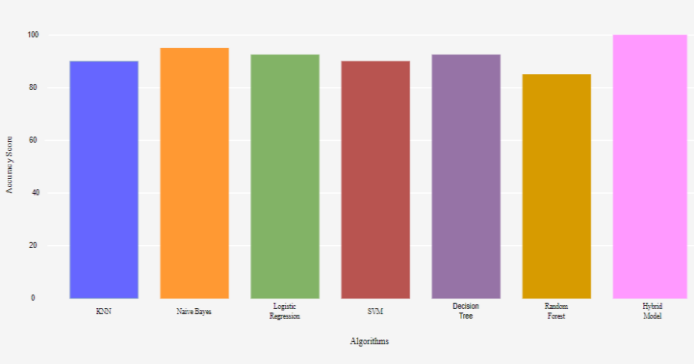

Fig 9. Accuracy Result

B. *Analysis*

Below Fig 10. Shows the analysis of positive to autism in toddlers and teenagers for different ethnicities. Same is done for adults.

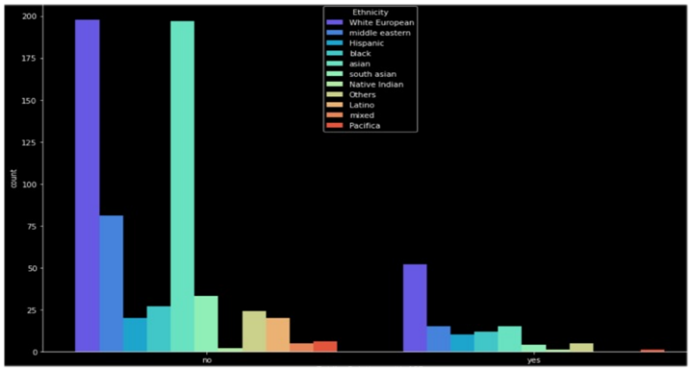

Fig 10. Analysis Result

5. CONCLUSION AND FUTURE WORK

A user-friendly web application helps different age groups predict autism traits easily and guides them at an early stage that will prevent the situation from getting any worse and reduce costs associated with delayed diagnosis. The framework not only improves the accuracy of autism screening but in additional, it helps in speeding up the process for a formal autism diagnosis procedure. Many inexperienced clinicians will not be well confident in certain autism case, hence there is a need of computer assist to predict the correct results. The proposed system attempts to show better performance comparing to the other existing approach of screening autism. The model can predict autism traits for different age groups, while many other existing approaches missed this feature. This system also provides a comparative view among different ML approach in terms of their performance.

In future, to predict the different level and to find the type of Autism. To be able to implement Mobile Application, to evaluate the usability and user experience of the model. To collect more data from various source to improve ML classifier.

ACKNOWLEDGMENT

We would like to express our heartfelt gratitude to our esteemed guide Y. Shobha, Associate Professor, Department of CS&E, BIT for her assistance, motivation and valuable suggestions during the entire period of our project development and research.

REFERENCES

[1] Fadi Thabtah, Neda Abdelhamid, David Peebles, ,A Machine Learning Autism Classification Method based on Regression Analyses2016.

[2] Matthew J. Maenner1,2*, Marshalyn Yeargin-Allsopp1, Kim Van Naarden Braun1, Deborah L. Christensen1, Laura A. Schieve11 National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention; Atlanta, GA United States of America, 2 Epidemic Intelligence Service, Centers for Disease Control and Prevention; Atlanta, GA, United States of America, Development of a Machine Learning Algorithm for the Surveillance of ASD, 2016.

[3] Fadi Thabtah, ASD Screening: Machine LearningAdaptation and DSM-5 Fulfillment. ICMHI '17, May 20-22, 2017, Taichung City, Taiwan© Association for Computing Machinery. ACM ISBN 978-1-4503-5224-6/17/05, 2017.

[4] Armin Lawi, Firman Aziz, Comparison of Classification Algorithms of the ASD Diagnosis.

The 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT) 2018 978-1-5386-8050-6/18/$31.00 ©2018 IEEE 218

[5] Osman Altay and Mustafa Ulas, Prediction of the Austim Spectrum Disorder Diagnosiswith Linear Discriminant Analysis Classifier and K-Nearest Neighbor. in 6th International Symposium on Digital Forensic and Security (ISDFS),2018.

[6]Dr.C .S. Kanimozhiselvi, Mr. D. Jayaprakash, Ms.K.S.Kalaivani, Grading Autism Children Using Machine Learning Techniques. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 5 (2019) pp. 1186-1188 ResearchIndiaPublicationshttp://www.ripublication.com, 2019

[7] Murat Go k1Received: 22 December Accepted: 20 April 2018 / Published online: 27 April 2018 - The Natural Computing Applications Forum 2018, A novel machine learning model to predict ASDs risk gene,2018.

[8] Mirac Baris Usta, Koray Karabekiroglu, Berkan Sahin, Muazzez Aydin,Abdullah Bozkurt, Tolga Karaosman, Armagan Aral, Cansu Cobanoglu,Aysegül Duman Kurt, Neriman Kesim, İrem Sahin & Emre Ürer ISSN: 2475-0573 (Print) 2475-0581 (Online) Journal homepage: https://www.tandfonline.com/loi/tbcp21, Use of machine learning methods in prediction ofshort-term outcome in autism spectrum disorders.

[9] Yang, Xin; Sarraf, Saman; and Zhang, Ning Deep Learning-based framework for Autism functional MRI Image Classification, Journal of the Arkansas Academy of Science: Vol. 72 , Article 11, 2018.

[10] Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, Md. Rezaul Karim Rizvi 2018 IEEE Paper,A Machine Learning Approach to Predict ASD.

[11] Bhawana Tyagi, Rahul Mishra, and Neha Bajpai, IEEE Paper, Machine Learning Techniques to Predict ASD,2018.

[12] Suman Raja ,Sarfaraz Masood A Machine Learning Autism Classification Method based on Regression Analyses,Fadi Thabtah , Neda Abdelhamid, David Peebles,International Conference on Computational Intelligence and Data Science (ICCIDS 2019) Analysis and Detection of ASD Using Machine Learning Techniques, 2019

[13] V. Jalaja Jayalakshmi, V. Geetha, R. Vivek, Classification of ASD Data using Machine Learning Techniques. International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6S, August 2019.

[14] Muhammad Asif, Hugo F.M.C, Martiniano IEEE Paper, Identification of biological mechanisms underlying a multidimensional ASD phenotype using machine learning. 2020.

[15] Elizabeth K. Jonesa,b, Mary Hanleya, Deborah M. Ribya. 1750-9467/ © 2020 The Author(s).