iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52417

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Prediction of Autism Spectrum Disorder based on Machine Learning Approach

Harshita Chandrappa[1], Swaathi BR[2], Prof. Swimpy Pahuja[3]
Dept of Computer Science and Information Technology, Reva University, Bangalore

Abstract: ***In recent years, awareness of autism spectrum disorder (ASD) has grown faster than before. As everyone is aware, ASD is a disorder of neurodevelopment that also encompasses problems with conduct and social interaction. The degree of symptom severity and each individual's experience with ASD vary. Any age can be used to diagnose autism. According to research, violence, self-harm, elopement. tantrums, preoccupation, and lack of obedience are behaviour patterns most frequently observed in people with autism. Therefore, it is imperative to spot any sign of severe ASD as soon as possible. ML methods like Random Forest, Naive Bayes, Decision Tree, K- Nearest Neighbour, Logistic Regression, and Support Vector Machine (SVM) are employed in this work. With the use of algorithms that have been used to decide which gives accurate results in terms of speed and accuracy, performance metrics assist in examining the degree of correctness of each piece of data across all users. ML methods like Random Forest, Naive Bayes, Decision Tree, K-Nearest Neighbour, Logistic Regression, and Support Vector Machine (SVM) are employed in this work. With the use of algorithms that have been used to decide which gives accurate results in terms of speed and accuracy, performance metrics assist in examining the degree of correctness of each piece of data across all users.***

Keywords: CART Model, Naive Bayes (NB), Random Forest, Support Vector Machine, Decision Tree are examples of machine learning techniques.

I. INTRODUCTION

Autism spectrum disorder (ASD) impacts a person's communication, behaviour as well as the ability to interact with the rest. This is a neuro-developmental illness and can be diagnosed at any age, although most of its symptoms appear at the initial two years of life and gets worse with time.

Autistic patients have to fight against a variety of issues, including attention problems, learning deficits, anxiety, obsessions, despair, and more. Autism cases are currently on the rise and are spreading rapidly over the world. An early diagnosis of autism prevents the condition of patient to deteriorate as well as lowers the cost in long-term. Machine learning algorithms are utilized as a classification task in order to discover ASD concerns. Prediction models are created based on past information, and then those patterns are used to forecast whether that person has ASD. The objective of this project is to design and construct a web application that makes use of machine learning (ML) techniques to rapidly and precisely detect ASD in people of any age group. To propose a novel ML framework, undertake predictive analysis, and create a hybrid model employing ML algorithms for the autism screening of adults, adolescents, and children.

II. LITERATURE SURVEY

Thabtah Fadi [1] provides information on the current research that uses ML for ASD classification and discusses the benefits and drawbacks. ASD features have received little attention in the minimal research that had been done on recognising and assessing them in the clinical setting under the

DSM-5, despite the introduction of sophisticated diagnostic tools based on ML. Armin Lawi [2] Using sample datasets of adult adults aged 17 and older, the best estimate results and k-fold values for each classification approach are investigated. However, Logistic Regression had a high rate of classification errors during testing, and K-Nearest Neighbours required more iterations to achieve the best performance.[3] Euclidean distance was utilized in the KNN method, and the k-value was set to 1, in a comparative study to determine whether children aged (4 to 11 years) have ASD. However, only those children of 4 to 11 years of age could be evaluated [4]. ASD symptoms are typically identified in youngsters as young as 18 months old. However, if the child has a little speech delay, ASD might not be identified until the start of the school year. When kids experience issues with their friends or interaction, the diagnosis is typically determined in these circumstances. Although an ML model was constructed and trained using brain development gene expression data, there are two early diagnostic methods for ASD in risk genes: imaging and biomarkers.

The former is costly, while the latter does not meet the test standards due to the disorder's intricacy and increased sequencing capacity [5]. The Alternating Decision Tree (ADTree) was used by Wall et al. [6] to speed up the identification of ASD traits and shorten screening times. They used the Autism Diagnostic Interview, Revised (ADI-R) strategy with 891 individuals and achieved a high degree of accuracy. The test, which was restricted to kids aged 5 to 17, was unable to detect ASD. In [7], the 'Red Flags' method was employed to screen for ASD with the Autism Spectrum Quotient in both children and adults., and they were then shortlisted to AQ-10 with greater than 90% accuracy.

III. PROPOSED SYSTEM

The goal is to create and build a web application that uses machine learning (ML) algorithms quickly and accurately identify ASD in individuals of any age range.

In order to do predictive analysis a novel ML framework with dataset relevant to Hybrid Autism Screening of Adults, Adolescents, and Toddlers (LR) is provided using ML methods such as Decision Tree (DT), Naive Bayes (NB), Random Forest, Support Vector Machine, Cart Model, and Logistic Regression.

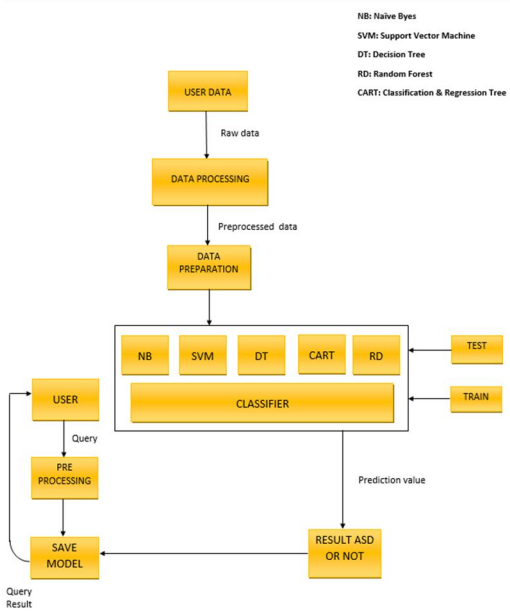
A. Acquisition of Dataset

Sl. No	Attributes	Value Yes=1 No=0
1	A1-10 screening questions for toddlers, adults, and teenagers focus on a variety of domains, including communication, imagination, and social interaction as well as attention to detail and attention switching.	1/0
2	Age	(4-11years)/ (12-17Years)/ (18+ years)
3	Gender	Male/Female
4	Ethnicity	1-11
5	Jaundice	1/0
6	Autism	1/0
7	Country_of_residence	1-52
8	Used app	1/0
9	Result	1-10
10	Age-description	Toddler/ Adolescent/ Adult
11	Relation	Self/Parent
12	Class1	ASD/Not

B. System Design

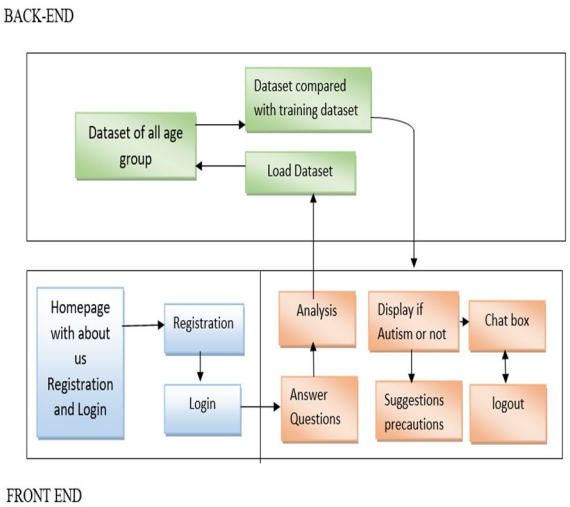
1) System Architecture

Figure 1 shows our system's system architecture. Database that has the stored raw data is being processed by data pre-processing. Only a few of the features in the dataset are relevant for prediction. This phase selects the relevant data features to improve prediction accuracy. These features are then fed into a machine learning model. Large amounts of training data were used to train the model. The data point is classified as Autism or not based on the user's data measures.



2) *Interface Design:*

The interface diagram for our system is shown in figure. The system starts with an empty browser state where the user enters the URL. If the server is down, the user is redirected to an empty browser to try again later. When the server is ready, the user is sent to the home page. The user may pick any module using the navigation bar. After selecting the Login/Register module, the user can register or login. After logging in, the user is given with a questionnaire that must be completed. The data is imported into the backend after submission, compared to the other training datasets, and the result is shown to the user on the frontend. The user may ask any inquiries they have by selecting the Chat Bot Module. The numerous visualisations for that module are presented when the user selects another module.



C. Methodology

1) *Pre-processing of Data:* It is the preparation of data for use with a machine learning model. It is the first and most crucial phase in the development of an ML model. The cleansing of raw materials, data is known as pre-processing. In the actual world, data is collected and turned into other words. The publisher, Faiza Nabi, offered real-time data.
2) *Creating the hybrid model:* To predict autism traits, use ML algorithms such as Linear Discriminant Analysis (LDA), Naive Bayes (NB), and Classification as well as Regression Trees (CART), K-Nearest Neighbour (KNN), and Logistic Regression (LR). For more accurate results, use Support Vector Machine (SVM) for data classification and comparison.
3) *Training and Testing Model:* This is a predictive modelling issue in which a class label is predicted for a given example of input data. The classifier is trained using a 'training data set,' the parameters are modified using a 'validation set,' and the performance of the classifier is assessed on an unknown 'test data set.

D. Implementation

The following are the main three algorithms in this paper:

1) *Navies Bayes:* A classification algorithm based on the Bayes theorem and the assumption of predictor dependency.

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

Algorithm 1
Step 1: Begin
Step 2: Load the dataset from the file and convert it from string to float.
Step 3: Imported Bernoulli NB performs:
Splitting data into test and training (provides split ratio) Calculating the mean of class attribute values.
Determine the standard deviation (SD = root V). The Bernoulli probability density function is then computed.

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

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) Support Vector Machine

It is a supervised machine learning approach that may be used for regression as well as classification. Each data point in n-dimensional space is represented as a point, and the value of each feature is the value of a certain coordinate. The hyperplane is located to perform classification.

Algorithm 2
Step 1: Start
Step 2: Import the necessary libraries, such as scikit-learn.
Step 3: Load the data that you want to use for training and testing the model.
Step 4: Create an object for the data that you loaded.
Step 5: Split the data into training and testing sets using the train_test_split() method from scikit-learn. Ensure that the model is not overfitting.
Step 6: Generate the SVM model by creating an instance of the SVC class in scikit-learn.
Step 7: Using the fit() function, train the SVM model using the training data.
Step 8: Evaluating the model's performance on the testing data requires making predictions with the predict() function and comparing the expected values to the actual values.
Step 9: Visualize the results using a graph or plot, if necessary.
Step 10: End

3) Random Forest

Classification and regression tasks are performed with it. It is an ensemble learning approach that integrates numerous decision trees in order to increase the model's accuracy and resilience.

Algorithm 3

Step 1: Select a random sample from a training or data set.

Step 2: For each training set, this algorithm will generate a decision tree.

Step 3: The choice tree will be averaged for voting

Step 4: As the final forecast outcome, select the prediction result that received the most votes.

4) CART Model

It is used to create decision trees for classification and regression. Used to tackle multi-class classification problems (it builds a binary tree for binary classification) and employs the Gini index is used as a statistic in the decision tree to examine the split of a feature node.

Algorithm 4

Step 1: Data collection.

Step 2: Split the data: Create training and test sets from the data. The testing set is used to assess the model's performance after it has been trained using the training set.

Step 3: Create a decision tree by recursively dividing the data depending on the values of the input features, create a decision tree using the training data. The goal is to construct a tree that best differentiates between the various classes or minimizes the variance of the target variable.

Step 4: Tree pruning (optional): The decision tree may need to be pruned by eliminating some of the branches or nodes if it is too complex or overfits the data. Techniques like cost-complexity pruning can be used to accomplish this.

Step 5: Make predictions based on the testing data using the decision tree.

Step 6: To measure the model's performance on the testing data, performance metrics like as accuracy, precision, recall, F1 score, mean squared error (MSE), and so on are used.

5) Decision Tree

Predictions are made using a decision tree. It follows a tree-like model of decisions and their possible outcomes. The technique operates by recursively separating the data into subsets depending on the most significant attribute at each node of the tree.

Algorithm 5

Step 1: Begin.

Step 2: Load the dataset you intend to utilise for model testing and training.

Step 3: To divide the data into training and testing sets, use the train_test_split() function from scikit-learn. This stage is critical for preventing the model from overfitting the training set of data.

Step 4: Utilise the DecisionTreeClassifier() or DecisionTreeRegressor() classes from scikit-learn to create a decision tree object with the proper hyperparameters.

Step 5: Using the fit() technique, train the decision tree model on the training set of data.

Step 6: Utilise the predict() method to predict the results for the testing data.

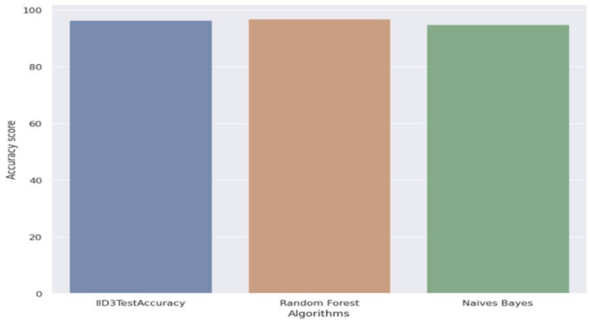
Step 7: Evaluate how well the model performed on the testing data by contrasting the projected values with the actual values and using a variety of performance indicators such as accuracy, precision, recall, and F1 score. Step 8: If necessary, visualise the decision tree using a graph or plot.

Step 9: Finish

IV. RESULTS

When compared to other known autism screening methods, this system performs better. The model can forecast autism traits of all age groups. The user thoroughly answers all of the questions on the form page as well as the other required details.

To analyse and identify autism, a hybrid model incorporating all of the previously stated techniques was applied. The final stage determines whether a user has autism or not. The web application also includes a chat box where the user can chat and ask frequently asked questions (FAQ) in order to gain some insight into the importance of getting an Autism test at an early stage of occurrence in order to prevent it from worsening. This system also provides a comparison of different ML approaches in terms of performance.



A. Performance Evaluation

Precision, A1 score, confusion matrix, recall, and other factors all influence the accuracy of ML algorithms. ML learns or approximates a function from examples in the training dataset to best map inputs to outputs. Accuracy Score is shown in Figure

B. Evaluation

Figure depicts the analysis of positive to autism in toddlers and teenagers of various ethnicities.

V. CONCLUSION AND FUTURE WORK

A user-friendly web tool aids and advises individuals of all ages in anticipating autism features early on, preventing the condition from deteriorating and lowering the expenses associated with delayed diagnosis. The framework not only improves the accuracy of autism screening, but it also aids in the speeding up of the formal autism diagnosis procedure. Many inexperienced clinicians will lack confidence in certain autism cases, necessitating the use of computer assistance to predict the correct results. The proposed system attempts to outperform the other existing approaches to screening for autism.

The model can forecast autism features for various age groups., which is something that many other existing approaches lack. This system also compares the performance of various machine learning algorithms. In future, to be able to implement a mobile application and evaluate the model's usability and user experience. To gather more data from various sources to improve the ML classifier. Also, to determine the stages i.e., initial or final stage of autism.

REFERENCES

[1] 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.

[2] 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

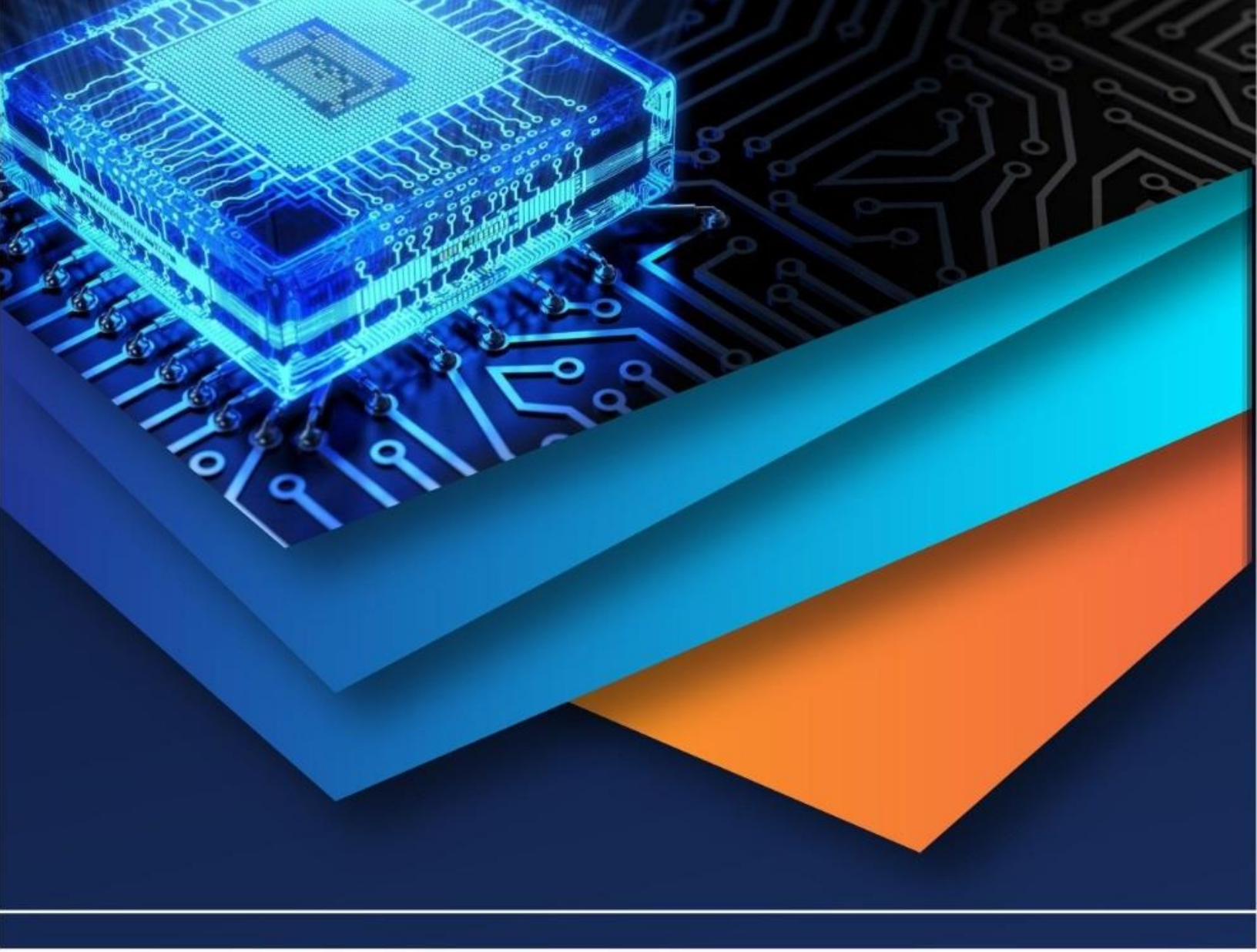
[3] Prediction of the Autism Spectrum Disorder Diagnosis Using Linear Discriminant Analysis Classifier and K-Nearest Neighbor, in 6th International Symposium on Digital Forensic and Security (ISDFS) 2018, by Osman Altay and Mustafa Ulas.

[4] S. J. Blumberg, M. D. Bramlett, M. D. Kogan, L. A. Schieve, J. R. Jones, & M. C. Lu (2013). Parent-reported autism spectrum disorder prevalence among US school-aged children from 2007 to 2011–2012 (No. 65). National Center for Health Statistics, Centers for Disease Control and Prevention, and the US Department of Health and Human Services.

[5] Go Murat k1Date received: December 22nd Accepted on April 20, 2018 / Published online on April 27, 2018 - The 2018 Natural Computing Applications Forum A novel machine learning algorithm for predicting ASD risk gene, published in 2018.

[6] D. P. Wall, R. Dally, R. Luyster, J.-Y. Jung, and T. F. DeLuca, “Use of artificial intelligence to shorten the behavioral diagnosis of autism,” PloS one, vol. 7, no. 8, p. e43855, 2012.

[7] C. Allison, B. Auyeung, and S. Baron-Cohen, “Toward brief “red flags” for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls,” Journal of the American Academy of Child & Adolescent Psychiatry, vol. 51, 2012



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY