**Machine Learning Approach To Predict Autism Spectrum Disorder**

**Mr.P. Purushotham, Assistant Professor in Department of Computer Science and Engineering**

**MLR Institute of Technology, Hyderabad, Telangana**

**Nagabandi Varshitha**

Computer Science Engineering

MLR Institute of Technology

Hyderabad, India

[nagabandivarshitha26@gmail.com](mailto:nagabandivarshitha26@gmail.com)

**Pottipally Sreehitha**

Computer Science Engineering

MLR Institute of Technology

Hyderabad, India

[sreehitha666@gmail.com](mailto:sreehitha666@gmail.com)

**Bondla Dikshitha**

Computer Science

Engineering

MLR Institute of Technology

Hyderabad, India

[bdikshitha26@gmail.com](mailto:bdikshitha26@gmail.com)

***ABSTRACT***

*Autism Spectrum Disorder (ASD) is becoming more and more popular in the modern era. Using screening tests to identify autism symptoms is highly costly and time taking. Development of AI and ML has made it possible to anticipate autism symptoms early on. Even though a lot of research uses various methodologies, these studies haven’t shown any conclusive results on the prediction of autistic features for various age groups. Therefore, the purpose of this work is to establish a website that allows people of any age to predict ASD and to offer an effective prediction model based on machine learning techniques. As a result of this study, Random Forest CART & ID3 the suggested prediction model is used to create a website . The AQ-10 dataset is used to access the model. According to the evaluations' findings, the prediction models excel the datasets in terms of accuracy , selectivity , responsiveness , precision , and false positive ratio .*

***KEYWORDS***

*AQ-10 datasets , Machine learning , CART, Random Forest , ASD ,ID3*

**I. INTRODUCTION**

A neurodevelopmental illness known as autism spectrum disorder impacts a person’s ability to interact , communicate , and learn . Autism can be diagnosed at any age, although symptoms usually start to show in the first two years of life and get worse as time goes by. Patients with autism spectrum disorders encounter a variety of obstacles , including difficulties focusing , learning disabilities , mental health issues like anxiety and depression , as well as motor and sensory issues.

Autism is currently spreading rapidly over the world , it is affecting many people . The WHO estimates that one child in every 160 has ASD . While some persons with the illness are able to live independently , others need care , assistance throughout their entire lives .

Autism diagnosis is a time- and money- consuming process. Early identification of autism can greatly benefit patients by enabling early prescription of appropriate medication . It prevents the patient's condition from worsening and assists cut down on the costs incurred over an extended period due to a delayed diagnosis. So, there is a great need for accurate , simple screening tests that can predict an individual’s qualities related to autism and decide if they need a thorough check for autism .

# **II. LITERATURE SURVEY**

The following section provides a quick overview of the work regarding ASD prediction approaches.

To shorten the screening duration and speed up the identification of Autism features, Wall et [1] used Adtrees or alternating decision trees. Using data from 891 individuals, they obtained a high degree of accuracy using AID revised approach. However, test’s age range was only five to seventeen, and it was not able to predict Autism disorder in different age divisions - child, adults, adolescents. For the same goal, Bone et al. [2] employed machine learning (ML) and a support vector machine (SVM) to achieve 89.2% sensitivity and 59% specificity. In their study, 462 people with non-ASD traits and 1264 people with ASD were included. Still, their findings were not approved for use as a screening tool for individuals of all ages due to the wide age range (4–55 years).Allison et al. [3] employed the “Red Flags” tool to screen adults and children for Autism using the ASQuotient, they were able to shortlist them for the AQ-10 with over 90% accuracy. In order to predict the traits of autism, Thabtah [4] compared earlier research on machine learning algorithms. In their search for comparatively more significant screening questions for the Autism diagnostic interview- revised and Autism diagnostic observation schedule screening techniques, Hauck & Kliewer [5] discovered that combining the two screening tests can improve results. Bekerom [6] identified traits associated with ASD in kids, like being obese, delayed growth and decreased regular exercise, using a variety of ML techniques, like Naive Bayes, SVM, random forest methods. The results were then compared. Wall et [7] work on using a brief screening test and validation to classify autism resulted in high sensitivity, specificity, and accuracy for both the ADTree and the functional tree. Using a significant brain scanning dataset Heinsfeld [8] used one DL algorithm and neural networks to find people with ASD. The results showed a mean classification accuracy of 70%, with a range of 66% to 71%. The Random Forest classifier produced a mean accuracy of 63%, compared to the SVM classifier’s 65% mean accuracy. Liu [9] used machine learning algorithms for examining a collection of eye movements in order to investigate whether patterns of face scanning might be helpful for identifying kids with Autism. This study showed an accuracy of 88.5%, a specificity of 86.2%, a sensitivity of 93.1%, and an AUC of 89.63%.In order to find problems with identification of conceptual problems, application of methods, and comprehension, Bone et [9] examined the earlier studies of Wall et [7] . They then replicated the findings using their machine-learning approach. The literature review makes it clear that, despite the fact that this field has seen a number of studies conducted, the researchers were unable to reach an agreement regarding the applicability of machine learning (ML) to the generalization of autism screening test results across age groups. Prior research has adapted various tools and techniques for autism screening tests; however, app-based solutions tailored to specific age groups have not been developed.

**III. BASIC CONCEPTS**

**A) PROPOSED METHODOLOGY**

The development of machine learning (ML) has made it possible to predict autism symptoms early on . The goal of this research is to provide a low-time , highly accurate , and economically feasible prediction model for ASD based on machine learning for individuals across various age groups.

**B) DATA COLLECTION**

The AQ-10 dataset , which is made up of 3 unique datasets derived from questions from the Aq 10 tool for screening, was utilised to generate an efficient predictive model. These three datasets include information for the age groups of children (four to eleven years) , adolescents (twelve to sixteen years) and adults (18 years or older). The AQ-10 , is a tool used to determine if someone needs to be referred for a thorough autism assessment. The aq 10 screening questions centre around a range of areas such as interactions, creative thinking, socialisation, paying attention to detail, and attention-grabbing switching. There is only one point available for each of the ten questions in the scoring system. Depending on how they respond to each question , users can receive 0 or 1 points. There are 292, 104 , and 704 instances in the datasets for children , adolescents , and adults , respectively. 21 traits , consisting of the combination of both category and numerical data, are present in each of the three datasets. These attributes include age, gender , ethnicity , if jaundice is inborn , a relative suffering from PDD, who is finishing the examination , previously used the screening app, Class , Result , Question 1-10 , Screening method Type.

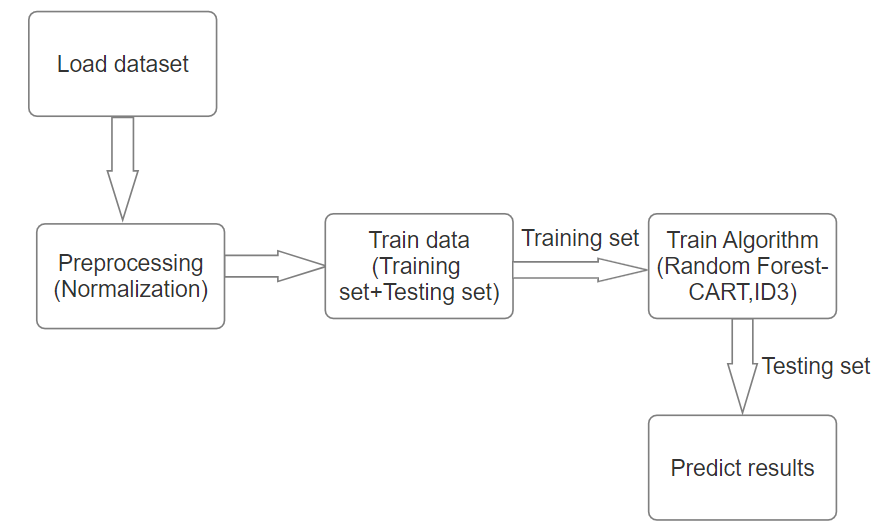
**C) DATA SYNTHESIZATION**

To eliminate features that weren’t relevant , the gathered data were combined .For example, the id field was removed since that seemed unnecessary for building a model for prediction. To deal with the null values, a list-wise deletion strategy was applied, whereby an observation was discarded if it included one or more missing values. Subsequently , a decision tree algorithm was employed to eliminate unnecessary features from the dataset . Based on the results , it was decided to remove the “age,” “age desc,” “used software before,” and “relation” columns in order to improve classification accuracy . Data set for the random forest classifier and regression technique , which are used to train and test the model.

**D) DEVELOPING PREDICTING MODEL**

Algorithms have been developed to predict autism traits, and their accuracy has been evaluated. Following the results of several supervised learning techniques, such as SVM , naive Bayes , and linear regression . Random forest was discovered to be more accurate and highly feasible than the other algorithms . Therefore , it was suggested to use Random Forest (CART) to implement the ASD predictive system . To achieve even better results, the algorithm underwent additional modifications.

**IV. SYSTEM ARCHITECTURE**



System architecture consists of several key steps.

**Load Dataset:** Import and gather the dataset containing information about individuals.

**Pre-processing:** Prepare the data by cleaning, normalizing, and handling missing values.

**Train Data:** Split the dataset into a training set for teaching the algorithm.

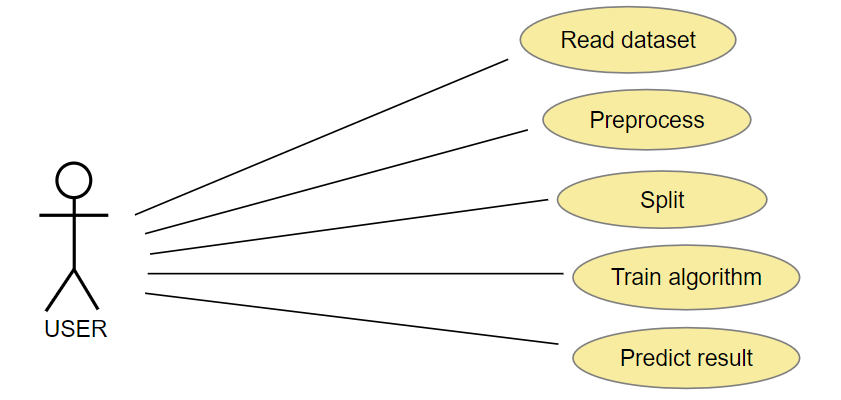
**Test Data:** Keep a separate testing set to assess the model's performance.

**Train Algorithm:** Select and train a machine learning algorithm on the training data.

**Prediction:** Use the trained algorithm to make predictions on new data for identifying ASD.

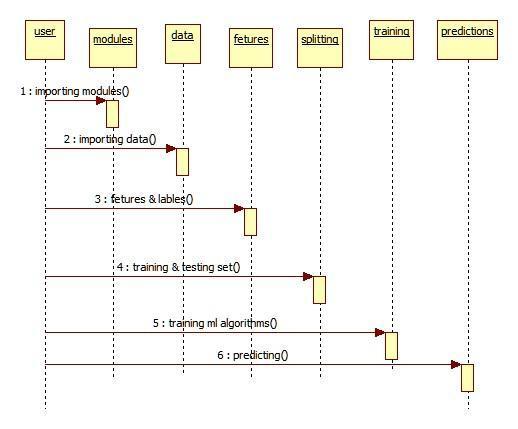
**V. UML DIAGRAMS**

**i) USE CASE DIAGRAM**

****

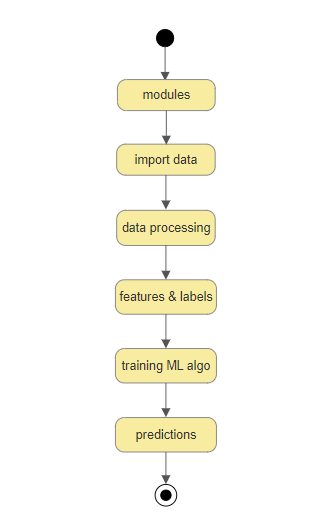
The Read dataset step in the use case diagram imports necessary personal data. Pre-processing entails data preparation and cleaning. Next, the data is divided into and training sets .The training data in the Train algorithm is used in ML model training. Finally, Predict result is the step where the trained model is used to make predictions about individuals potentially having autism spectrum disorder based on the provided data set.

**ii) SEQUENCE DIAGRAM**



The sequence diagram shows how the system receives raw data from the user that includes necessary features. The data is subsequently separated into test sets and training sets by the algorithm. The System’s algorithms are trained on the training set so that it may learn from the data .After being trained ,the system can use the testing set to provide predictions based on the patterns it has learnt. The flow of interactions and data inside the system architecture is shown by this sequential procedure.

**iii) ACTIVITY DIAGRAM**



Modules in the activity diagram list the parts and instruments that are utilized. The necessary data is then imported by the system. Data processing include operations such as transformation and cleaning. While training an ML algorithm (such as Random Forest, CART, or ID3)involves teaching the machine learning model, features and labels denote the preparation of the data, Lastly predictions indicates how the trained model is used to forecast autism spectrum disorder.

**CONCLUSION:**

This work offers a viable machine learning method for predicting autism spectrum disorder

(ASD), offering a cost-effective and accessible means to identify ASD traits in individuals of all ages. Random Forest-CART and Random Forest-ID3 algorithms were put together to form a model. When tested on various datasets, the constructed prediction model showed improved false-positive rates, sensitivity, accuracy, specificity, and precision when evaluated with different datasets. Data quality, and potential false positives remain key challenges. This research underscores the importance of responsible and collaborative integration of AI

Technology into healthcare, emphasizing the need for human expertise in the diagnostic process. The ultimate goal is to enhance early ASD detection and support while ensuring ethical and equitable access for all individuals.

**REFERENCES:**

1. "Use of artificial intelligence to shorten the behavioral diagnosis of autism,”Plos One,vol.7,

no.8,p.e43855,2012,D.P.Wall,R.Dally,R.Luyster,J-Y.Jung,and T.F.DeLuca.

1. Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion," Journal of Child Psychology and Psychiatry, vol. 57, 2016, D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S. Narayanan.
2. Journal of the American Academy of Child & Adolescent Psychiatry, vol. 51, 2012; C. Allison, B. Auyeung, and S. Baron-Cohen, "Towards brief "red flags" for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls."
3. In the Proceedings of the 1st International Conference on Medical and Health Informatics, F. Thabtah, "Adapting machine learning for autism spectrum disorder screening and fulfilling the DSM-5," was published in 2017. ACM, 2017.
4. "Machine learning for autism diagnostics: Applying support vector classification," by F. Hauck and N. Kliewer
5. "Using machine learning for autism spectrum disorder detection," B. van den Bekerom, 2017.
6. Use of machine learning to expedite observation-based screening and diagnosis of autism, D. Wall, J. Kosmicki, T. Deluca, E. Harstad, and V. Fusaro, Translational psychiatry, vol. 2, no. 4, p. e100, 2012.
7. "Identification of autism spectrum disorder using deep learning and the abide dataset," A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, NeuroImage: Clinical, vol. 17, 2018.
8. "Applying machine learning to facilitate autism diagnostics: pitfalls and promises," Journal of autism and developmental disorders, vol. 45, no. 5, pp. 1121–1136, 2015, D. Bone, M. S. Goodwin, M. P. Black, C.-C. Lee, K. Audhkhasi, and S. Narayanan.