IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14668

# Early-Stage Autism Spectrum Disorder Diagnosis Using Machine Learning

**Dr.R.Raja Kumar[1], Kuppala MadhuSudhan[2]**

Professor, Department of Computer Science and Engineering, Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, Andhra Pradesh, India[1]

M.Tech Student, Department of Computer Science and Engineering, Rajeev Gandhi Memorial College of Engineering and Technology, Nandyal, Andhra Pradesh, India[2]

**Abstract:** The project shows a way to use Machine Learning (ML) to find Autism Spectrum Disorder (ASD) early on, acknowledging the challenges of diagnosing the condition while striving to mitigate its severity through early interventions. The suggested system uses four typical ASD datasets, ranging from infants to adults, to test four Feature Scaling (FS) techniques: Quantile Transformer, Power Transformer, Normalizer, and Max Abs Scaler. Included scaled datasets are used for machine learning computations (like K-Nearest Neighbors, Gaussian Naïve Bayes, Logistic Regression, SVM, LDA, Ada Boost, and Random Forest). Factual estimations used to Find the best FS methods and classifiers for each age group. Babies, children, adolescents, and adults are the groups for which the voting classifier most accurately predicts ASD. The assignment includes an analysis of the relevance of a specific aspect. Employing four Component Determination Strategies to help medical care professionals with ASD screening and to emphasize the importance of calibrating machine learning approaches in predicting ASD across age groups. The suggested structure outperforms the existing early ASD finding methods. A group process that used a Voting Classifier with Random Forest (RF) and AdaBoost was able to get 100% accuracy, which made ASD recognition even stronger and more accurate.

**Keywords:** Machine Learning, Classification, Autism Spectrum Disorder, Feature Scaling, and Feature Selection Methods.

## I. INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental condition, includes a variety of behavioural and social communication issues that initially manifest in early childhood or infancy. [1], [2]. While ASD is represented by narrow and uninteresting norms of behaviour, the term "species" refers to a wide range of side effects and abilities [3, 4, 5]. Even in the absence of a long-term solution for ASD, early intervention and access to high-quality clinical care can significantly boost a child's development and help them develop their communication and behavior [6, 7, 8]. In fact, the identification and separating evidence of ASD is incredibly difficult and challenging, especially when using standard social exploration. Chemical imbalance is frequently examined at age two, although depending on how severe it is, it may occasionally be examined later [9], [10], and [11]. When feasible, a few therapy modalities are available to identify ASD. Until there is a significant chance that ASD will improve, these symptomatic approaches are typically not used in that mind-set.

A clear and concise agenda that applies to people at every stage of life—infancy, adolescence, prematurity, and adulthood—was presented by the authors of [12]. The ASD Tests portable apps framework was then built by the authors in [13] using a variety of poll summaries, AQ-10 methods, and Q-Visit, to identify ASD as quickly as possible. They also developed an open-source dataset using data from In response to additional developments in this field of study, they developed mobile applications and posted them to Kaggle and the University of California, Irvine (UCI) AI vault. A few studies using various Machine Learning (ML) techniques have been conducted recently with the aim of rapidly surveying and analysing ASD in addition to other conditions such as cardiovascular failure, diabetes, and stroke [14], [15] and [16].

Using Rule-based Machine Learning (RML) approaches, the authors in [17] investigated the traits of ASD and

  

DOI: 10.17148/IJARCCE.2025.14668

confirmed that RML improves arrangement exactness in models of order. In [18], the authors developed expectancy models for kids, teens, and adults using the Random Forest (RF) and Iterative Dichotomiser 3 (ID3) algorithms. A new assessment tool that combined ADI-R and ADOS ML approaches was given by the authors in [19], who also devised a few trait encoding algorithms in order to overcome the difficulties of non-linearity, irregularity, and incomplete information. Using mental registering, a second investigation by same authors [13] demonstrates a relationship between highlights and classes as well as between endlessly include themselves. The key characteristics from the data are chosen using support vector machines (SVM), decision trees (DT), and logistic regression (LR) in order to forecast the prevalence of pressure ulcers. Additionally, in [20], the authors examined cases of ASD (N = 11) and generally evolved (TD) (N = 19), where relationship-based property choice was used to understand the meaning of the attributes. The authors in [21] identified 15 preschool ASDs with only seven characteristics after looking at ASD and TD youth in 2015. They also explained how bunch analysis may be used to effectively examine multifaceted designs in order to speculate on the phenotype and variety of ASD. The authors of [22] investigated the classifiers' performance for predicting adult ASD using K-Nearest Neighbors (KNN), Logistic Regression, Linear Discriminant Analysis (LDA), Classification and Regression Tree (CART), Random Forest (RF), and Ensemble approaches.

## II.LITERATURE SURVEY

For this work, we collected ASD datasets from early children, children, adults, and adults [1] and applied different component choice methods. We evaluated the results using a variety of metrics, such as f1-measure, forecast precision, kappa insights, and AUROC, following the application of a few classifiers to these datasets. Additionally, we used a non-parametric factual importance test to analyze each classifier's performance. We found that SVM performed better than other classifiers for the newborn, child, juvenile, and adult datasets. In light of RIPPER, we achieved 97.82% accuracy for the baby subset; 99.61% accuracy for the kid subset using the Boruta CFS cross (BIC) technique and correlation-based feature selection (CFS); 95.87% accuracy for the juvenile subset using Boruta; and 96.82% accuracy for the CFS-based adult subset. After that, we used the Shapley Additive Explanations (SHAP) method, which produced the highest level of accuracy, to analyze the elements of several component subsets [1]. In order to investigate the function of the stomach microbiota in both health and disease, the 16S quality has recently been sequenced from waste samples. Mental imbalance range sickness (ASD) is a neurodevelopmental disorder that has gastrointestinal side effects. ASD has also been linked to dysbiotic stomach vegetation [2]. Despite a great deal of research, finding a typical dysbiotic profile in ASD individuals remains challenging [3, 4, 5]. Both external factors (such as dietary preferences) and specialist viewpoints (such as exploratory methods) cause differences between these investigations.

We compiled 959 samples from eight available projects (fourteen ASD and 49 Sound Controls, HC) in order to reduce the observed bias among investigations. Next, developed an indicator that could identify HC and ASD using an AI (ML) technique. Three calculations were evaluated and proceeded with: Slope Helping Machine, Backing Vector Machine, and Irregular Woodland. The definitions of five different genera, including Parasutterella and Alloprevotella, were consistent across the three methods. Additionally, our findings demonstrate that AI frameworks may detect shared ordered traits by comparing datasets obtained from different nations and concealed jumbling properties.

In many instances, mental imbalance manifests as a decline in social behaviors, correspondence skills, and interpersonal abilities [4]. The causes of this could be investigated by understanding how they interpret visual information. By observing where and when children look at picture boosts, the exploration work presented here examines how they behave. [3,4,5,9] By observing how the child looks and breaking down eye gaze borders, To test how a medically introverted child's visual discernment varies from that of a regular child, the fluffy based eye stare point assessment approach (FEGP) was created. It uses an exhibition level marker, perception, and deductions to assist discover visual conduct differences in mentally imbalanced youngsters. These findings could subsequently be used to modify educational plans so that children try to get to know their friends. Previous efforts have been made to identify and measure neurological conditions that have obvious adverse effects, such as hand tremors. Multiple sclerosis is one such illness that may be somewhat assessed by hand earthquake power. [5] This work contrives a method for capturing and analyzing the digitized sign of the standard Spirography test in order to achieve this

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025
DOI: 10.17148/IJARCCE.2025.14668

purpose. Equipment and programming improvements are being made to a device that can perform a standard Spirography test, record the sign, transfer it to a computer running relevant programming, and deconstruct it using highlight extraction and grouping calculations. Since Power Range Investigation demonstrates the significance of each recurring component for the overall development of the hand, it is advised as one of the product's core components. For Power Range Investigation, complex boundaries, factors—such as the signs' average Lyapunov range value and the greatest Lyapunov type—that are chosen as markers of the signs' level of chaos are also taken into consideration. A sign's complexity is demonstrated by its latency and inserting size, which work together to provide an imprecise list window when a sign is occasionally remade. The appropriate time-delay and the delay s-shaped models. Signals are seen as examples in highlighted space and are ordered by a prepared feed forward brain network. [16, 20] In light of the calculation of each subject's participation sign to the established classes of sound and sick gathering, the characterization work serves as a dynamic approach by which the physicist coordinates the appropriate therapies. A degenerative neuro-formative confusion, chemical imbalance range jumble (ASD) is puzzling. This work demonstrates that complex components, such as turbulent highlights, can be used to segregate signals from patients with and without hand tremors and can depict the dynamical method of behaving signs. The majority of current methods for identifying ASD use utilitarian attractive reverberation imaging (fMRI) with a rather small dataset.

This method has less conjecture and gives high precision [3, 4, and 5]. In order to address this need and improve the suitability of the computerized mental imbalance symptomatic model, we provide an ASD recognition model in this analysis that makes use of valuable availability components of fMRI data in the resting state. Two popular cerebrum map books, Craddock 200 (CC200) and Robotized Physical Marking (AAL), as well as two less often used chart books, Bootstrap Examination of Stable Bunches (BASC) and Power, are fully recalled by our proposed model. The characterization task is finished using a profound brain organization (DNN) classifier. Based on reenactment data, the proposed model achieves more accuracy than the state-of-the-art methods. The suggested model's mean exactness was 88%, but the mean accuracy of the best-in-class methods changed from 67% to 85%. The area beneath the working trademark bend (AUC) score of the collector, F1-score, and responsiveness of the suggested model were, respectively, 90%, 87%, and 96%. Relative analysis on a variety of scoring systems demonstrates the The BASC chart book has an edge over the other chart books that were previously addressed in terms of distinguishing between control and ASD.

## III. METHODOLOGY

### *A. Suggested Task*

Quantile Transformer, Power Transformer, MaxAbsScaler, and Normalizer are used in the suggested machine learning engineering to improve information and increase precision for early stage autism spectrum disorder (ASD) distinguishing proof. By breaking down many ASD datasets across age groups, The process makes a more accurate finding model, stresses improvement, and concentrates on significant risk factors. Improved preprocessing methods and automation boost ASD differentiating evidence and foster early mediation for better results.[3, 4], and 5]. To increase the strength and precision of ASD detection, a new method that combined a Voting Classifier with Random Forest (RF) and AdaBoost obtained 100% accuracy. To provide more precise expectations, this new methodology makes use of RF and Adaboost's distinct skills. For client testing, Carafe can provide a seamless, user-friendly front end.

### *B. System Architecture*

This study examines chemical imbalance at various ages using ML techniques to develop an expectation model. Preprocessing techniques used after dataset selection include oversampling, include encoding, and missing characteristics attribution. Missing dataset values are credited using Mean Value Imputation (MVI). One Hot Encoding (OHE) converts values to mathematical properties in a straightforward manner. RF, DT, KNN, GNB, LR, SVM, LDA, and stomach muscle are the eight machine learning techniques—are used to organize the component scaled datasets. The classifier grouping findings are used to determine the best characterization methods and FS tactics for every scaled ASD dataset element. The risk for ASD is still unknown after those reviews, and the four FSTs—IGAE, GRAE, RFAE, and CAE—are used to rank the most pertinent characteristics. The proposed research path would assess ASD datasets and identify the main risk factors for evidence that distinguishes ASD.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14668

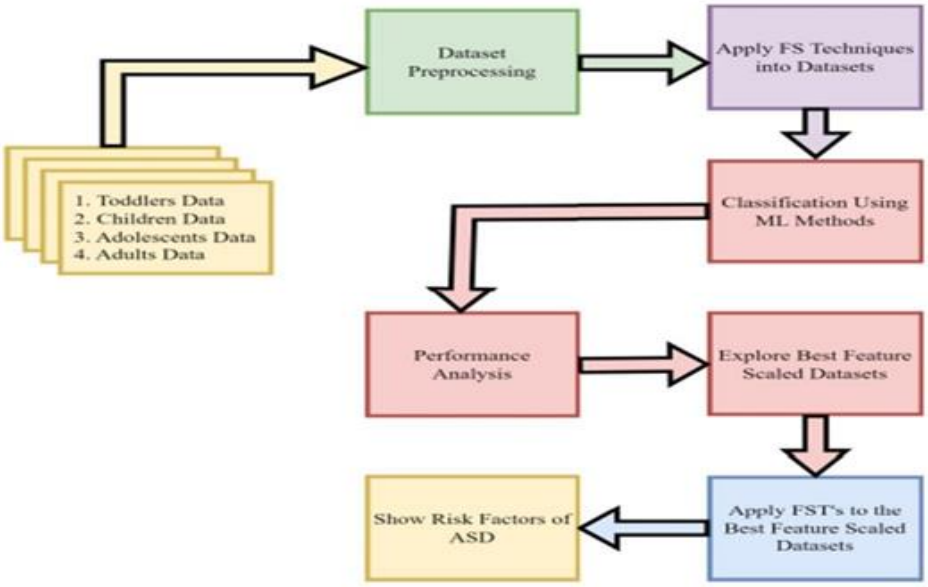


Fig. 1 Proposed Architecture

C. Gathering Datasets

Many datasets about ASD assessments for different age groups are stacked and examined in this example. Presumably, it will include elements like evaluating factors, verifying the information design, and exploring the dataset in further detail.

1) Decision Tree: This model resembles a tree, with an interior hub handling a quality test, a branch handling the test's result, and a leaf hub handling the class mark. Decision trees provide a powerful visual representation of dynamic techniques. They are interpretable and may help differentiate important traits, revealing importan perspectives that are used to predict ASD.
2) K-Nearest Adjacents: This non-parametric approach groups information points by using the greater part class of their k-nearest Adjacents in the component space. Finding designs in information without expecting a particular useful structure is made easier with KNN. It may identify close links in ASD datasets that are likely to go unnoticed by everyone. [12,13].
3) Naive Bayes: This probabilistic classifier relies on the Bayes theorem and the freedom of highlighting. Naive Bayes works well on high-layered datasets and is computationally efficient. It could potentially be used for the initial investigation of ASD data due to its speed and ease of use.
4) Logistic Regression: The computed capacity is used in this simple twofold characterization model to calculate the probability that a case will fall into a specific class. Interpretable, logistic relapse provides information on the connection between traits and autism spectrum disorder opportunities. In projects involving binary classification, it functions as a standard model.
5) Vector Machine Support: The method of supervised learning approach looks for the best hyperplane to divide classes over a high-layered space. SVM is good at handling confusing choice limits. It may detect nonlinear connections in ASD datasets, increasing the accuracy of classification [12,13].
6) Linear Discriminate Analysis: This technique determines direct component mixes that actually divide classes by reducing order and dimensionality. [23, 24] LDA is helpful for highlighting properties and reducing dimensionality. It could help further increase interpretability and identify important elements in the ID of ASD.
7) A Voting Classifier: Consolidation is a type of learning that occurs while different free classifiers are being prepared and a final prognosis is produced by combining their expectations. We have chosen Random Forest and AdaBoost as the core classifiers for our project.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14668

## IV.EXPERIMENTAL OUTCOMES

### *A. Precision*

Precision is used to measure the degree of correctly distinguished examples or tests among those grouped as certain. The precision might then be ascertained using the following formula:

$$\text{Precision}=\frac{True\ positive}{True\ Positive+False\ Positive}$$

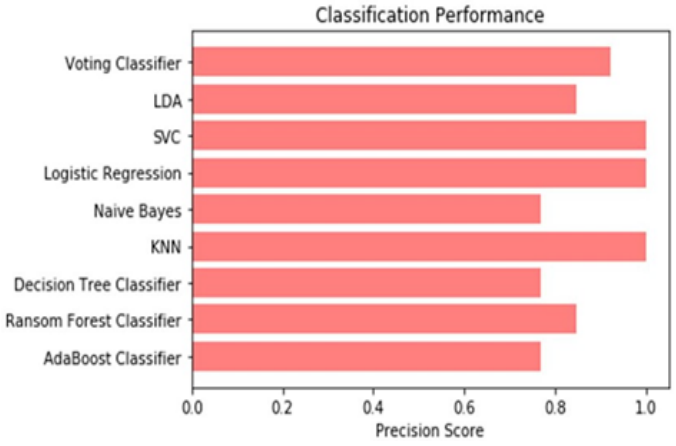


Fig. 2 Comparison Graph for Precision

### *B. Recall*

Recall is a machine learning statistic that assesses a model's capacity to recognize every instance of a particular class. It provides information on a model's final performance in collecting examples of a particular class. It is the proportion of all true class impressions that are positively predicted to those that are not.

$$Recall = \frac{True\ Positive}{True\ Positive+False\ Negative}$$

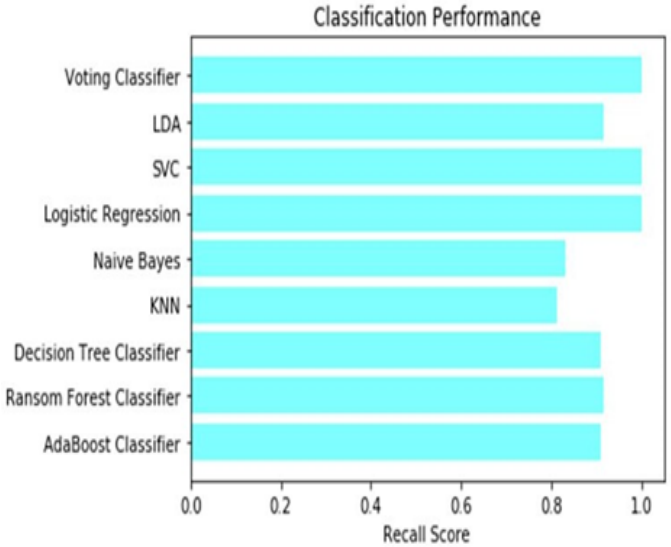


Fig.3 Comparison Graph for Recall

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14668

C. *Accuracy(Correctness)*

The capacity of a test to recognize good and patient examples is not a certainty. To assess a test's accuracy, we should note the percentage of true positives and true negatives in each broken-down case. This might be expressed quantitatively in the following way:

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative}$$

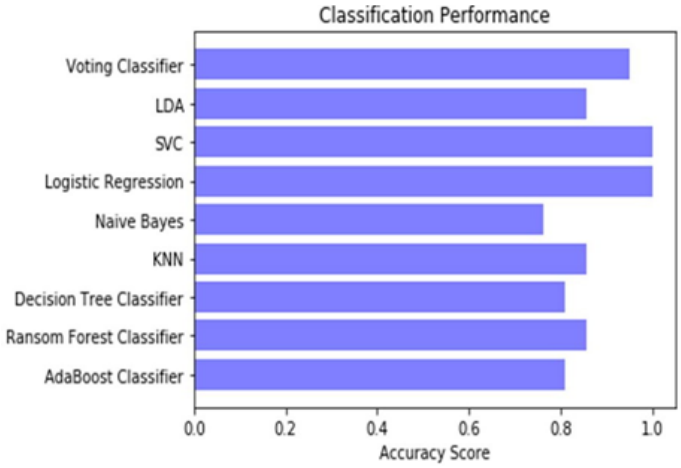


Fig.4 Accuracy Graph

D.F1 Score

The accuracy of a model is determined by the F1 score, a machine learning evaluation metric. It combines the accuracy and recall ratings of a model. The exactness metric calculates how many times a model made an accurate prediction over the full dataset.

$$F1\ Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \times 100$$

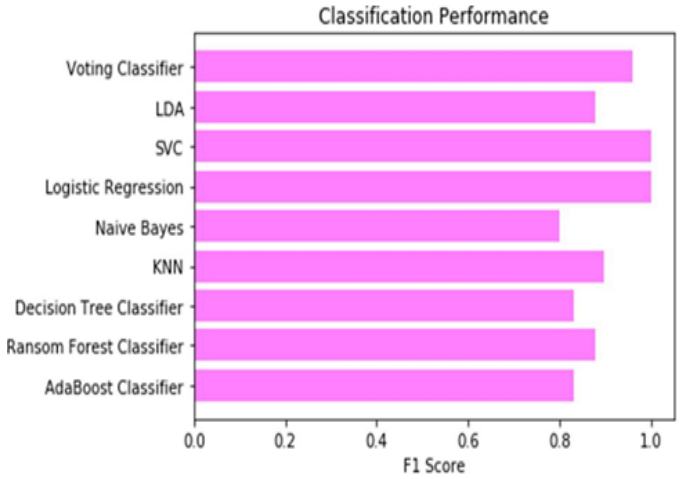


Fig.5 F1 Score

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14668

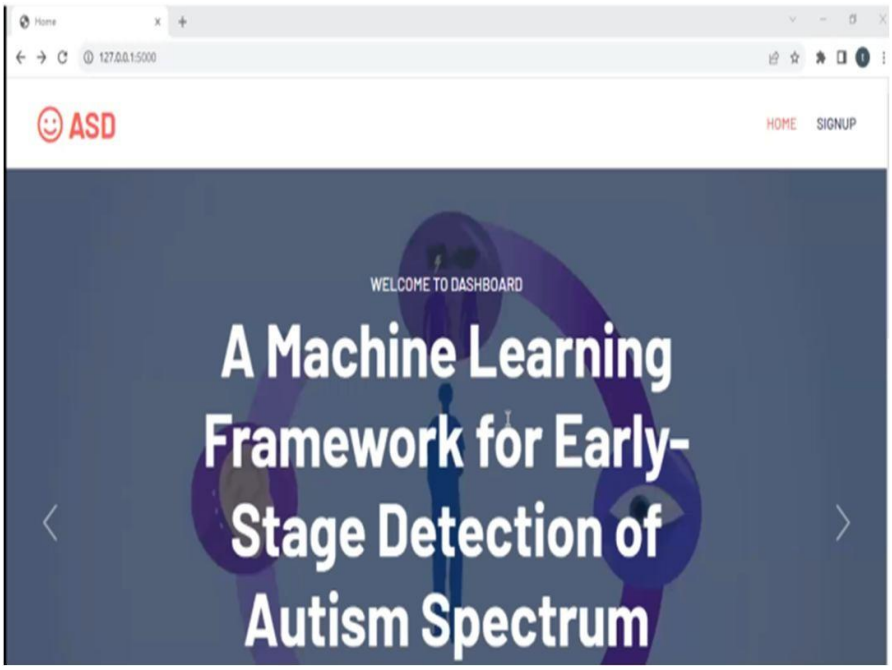


Fig. 6 Homepage

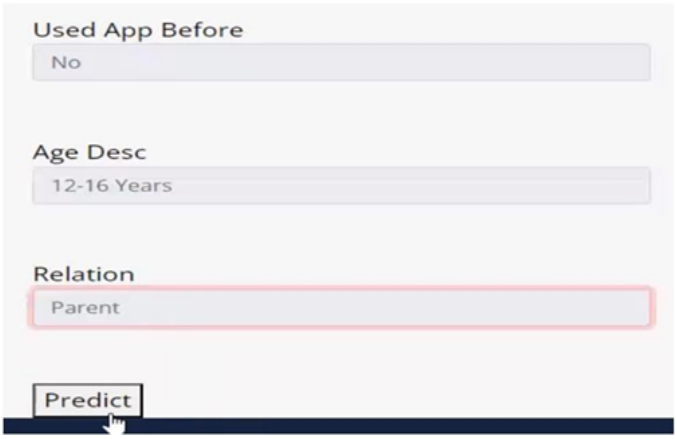


Fig. 7 Input from the Users


Fig. 8 Forecast Outcome

## V. CONCLUSION

The team has successfully created a novel machine learning framework for the early diagnosis of autism spectrum disease (ASD), or ID, using state-of-the-art computations and element scaling techniques. Using common ASD datasets for toddlers, adolescents, children, and adults, the system's robust presentation across age groups shows its versatility and remedial convenience [12,13]. This framework provides optimal grouping and element scaling computations for early ASD differentiation evidence, perhaps leading to practical medications. Through the use of AdaBoost and Arbitrary Random Forest, the group approach has improved the accuracy of ASD recognition. Its

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14668

frequent incorporation into an intuitive front end, where element values may be input and assessed, illustrates its usefulness and common sense in practical settings. In order to help accurately diagnose ASD, the evaluation makes use of persuasive qualities and a resolve to concentrate on important risk factors.

## VI. FUTURE SCOPE

The study intends to gather additional information on ASD and develop a more thorough expectation model for individuals of all ages in order to strengthen the evidence of ASD and other neuro-formative disorders [18]. The study could be broadened to encompass a more diverse and sizable sample of people with ASD. To increase the accuracy and reliability of ASD recognition, the project also suggests revamping the system or creating a more traditional expectation model using calculator learning approaches. The drive could investigate further neuro-formative issues and how the proposed framework could detect and forecast them. other data will be gathered, the model will be enhanced, and it may even be expanded to cover other neuro-formative difficulties.

## REFERENCES

[1] M. Bala, M. H. Ali, M. S. Satu, K. F. Hasan, and M. A. Moni, ‘‘Efficient machine learning models for early stage detection of autism spectrum disorder,’’ Algorithms, vol. 15, no. 5, p. 166, May 2022.
[2] D. Pietrucci, A. Teofani, M. Milanesi, B. Fosso, L. Putignani, F. Messina, G. Pesole, A. Desideri, and G. Chillemi, ‘‘Machine learning data analysis highlights the role of parasutterella and alloprevotella in autism spectrum disorders,’’ Biomedicines, vol. 10, no. 8, p. 2028, Aug. 2022.
[3] F. Z. Subah, K. Deb, P. K. Dhar, and T. Koshiba, ‘‘A deep learning approach to predict autism spectrum disorder using multisite resting-state fMRI,’’ Appl. Sci., vol. 11, no. 8, p. 3636, Apr. 2021.
[4] K.-F. Kollias, C. K. Syriopoulou-Delli, P. Sarigiannidis, and G. F. Fragulis, ‘‘The contribution of machine learning and eye-tracking technology in autism spectrum disorder research: A systematic review,’’ Electronics, vol. 10, no. 23, p. 2982, Nov. 2021.
[5] Mahammad, F. S., & Viswanatham, V. M. (2020). Performance analysis of data compression algorithms for heterogeneous architecture through parallel approach. The Journal of Supercomputing, 76(4), 2275-2288.
[6] Karukula, N. R., & Farooq, S. M. (2013). A route map for detecting Sybil attacks in urban vehicular networks. Journal of Information, Knowledge, and Research in Computer Engineering, 2(2), 540-544.
[7] Farook, S. M., & NageswaraReddy, K. (2015). Implementation of Intrusion Detection Systems for High Performance Computing Environment Applications. Inter national journal of Scientific Engineering and Technology Research, 4(0), 41.
[8] Sunar, M. F., & Viswanatham, V. M. (2018). A fast approach to encrypt and decrypt of video streams for secure channel transmission. World Review of Science, Technology and Sustainable Development, 14(1), 11-28.
[9] Mahammad, F. S., & Viswanatham, V. M. (2017). A study on h. 26x family of video streaming compression techniques. International Journal of Pure and Applied Mathematics, 117(10), 63-66.
[10] Devi,S M. S., Mahammad, F. S., Bhavana, D., Sukanya, D., Thanusha, T. S., Chandrakala, M., & Swathi, P. V. (2022).” Machine Learning Based Classification and Clustering Analysis of Efficiency of Exercise Against Covid-19 Infection.” Journal of Algebraic Statistics, 13(3), 112-117.
[11] Devi, M. M. S., & Gangadhar, M. Y. (2012).” A comparative Study of Classification Algorithm for Printed Telugu Character Recognition.” International Journal of Electronics Communication and Computer Engineering, 3(3), 633-641.
[12] Devi, M. S., Meghana, A. I., Susmitha, M., Mounika, G., Vineela, G., & Padmavathi, M. MISSING CHILD IDENTIFICATION SYSTEM USING DEEP LEARNING.
[13] V. Lakshmi chaitanya. "Machine Learning Based Predictive Model for Data Fusion Based Intruder Alert System." journal of algebraic statistics 13, no. 2 (2022): 2477-2483.
[14] Chaitanya, V. L., & Bhaskar, G. V. (2014). Apriori vs Genetic algorithms for Identifying Frequent Item Sets. International journal of Innovative Research &Development, 3(6), 249-254.
[15] Chaitanya, V. L., Sutraye, N., Praveeena, A. S., Niharika, U. N., Ulfath, P., & Rani, D. P. (2023). Experimental Investigation of Machine Learning Techniques for Predicting Software Quality.
[16] Lakshmi, B. S., Pranavi, S., Jayalakshmi, C., Gayatri, K., Sireesha, M., & Akhila, A. Detecting Android Malware with an Enhanced Genetic Algorithm for Feature Selection and Machine Learning.
[17] Lakshmi, B. S., & Kumar, A. S. (2018). Identity-Based Proxy-Oriented Data Uploading and Remote Data Integrity checking in Public Cloud. International Journal of Research, 5(22), 744-757.
[18] Lakshmi, B. S. (2021). Fire detection using Image processing. Asian Journal of Computer Science and Technology, 10(2), 14-19.
[19] Devi, M. S., Poojitha, M., Sucharitha, R., Keerthi, K., Manideepika, P., & Vasudha, C. Extracting and Analyzing

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14668

Features in Natural Language Processing for Deep Learning with English Language.

[20] Kumar JDS, Subramanyam MV, Kumar APS. Hybrid Chameleon Search and Remora Optimization Algorithm-based Dynamic Heterogeneous load balancing clustering protocol for extending the lifetime of wireless sensor networks. Int J Commun Syst. 2023; 36(17):e5609. doi:10.1002/dac.5609

[21] Prediction Of Covid-19 Infection Based on Lifestyle Habits Employing Random Forest Algorithm FS Mahammad, P Bhaskar, A Prudvi, NY Reddy, PJ Reddy journal of algebraic statistics 13 (3), 40-45

[22] Machine Learning Based Predictive Model for Closed Loop Air Filtering System P Bhaskar, FS Mahammad, AH Kumar, DR Kumar, SMA Khadar, ...Journal of Algebraic Statistics 13 (3), 609-616