International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

# Screening For Childhood Autism Spectrum Disorder Using Machine Learning

**T. Vedhakshari1, A. Bhargav Kumar2, R. Sneha Latha Reddy3, J. Narendra Shiva Sai4, V.L. Sowjanya5**

*[1-4] B.Tech Student, [5] Assistant Professor, LIET*

*[1,2,3,4,5] Computer Science and Systems Engineering, Lendi Institute of Engineering and Technology, Vizianagaram.*

---------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract** – Autism Spectrum Disorder (ASD) affects communication, behavior, and social interactions, and early detection is crucial for intervention. Current diagnostic method sare subjective and prone to delays or misdiagnosis. This project explores machine learning (ML) to automate ASD screening using behavioral data, diagnostic questionnaires, and developmental milestones. Supervised learning algorithms classify children into ASD or non-ASD categories with high accuracy. The approach aims to assist clinicians in early identification and intervention. It offers a scalable, efficient, and objective screening tool for ASD in clinical and educational settings.

**Key Words:** Autism Spectrum Disorder (ASD), Machine Learning, Early Detection, Screening, Behavioral Data, Supervised Learning, Diagnostic Tools, Developmental Milestones, Classification, Intervention, Artificial Intelligence, Predictive Modeling, Child Development, Healthcare Technology, Autism Diagnosis.

## 1. INTRODUCTION:

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects communication, behavior, and social interaction, with symptoms often becoming apparent in early childhood. Early detection of ASD is critical for implementing timely interventions that can significantly improve a child's developmental outcomes. However, traditional diagnostic methods primarily rely on subjective assessments, which can lead to delays in diagnosis or misdiagnosis.

In recent years, machine learning (ML) has emerged as a powerful tool for automating diagnostic processes, offering the potential for more objective, accurate, and efficient screening. This project aims to explore the use of ML techniques for early detection and classification of ASD in children. By leveraging various input features such as behavioral data, developmental milestones, and diagnostic questionnaires, the system seeks to predict the likelihood of a child being on the autism spectrum. The ultimate goal is to develop an automated, scalable, and objective screening tool that can support clinicians in early identification, leading to timely interventions that improve long-term outcomes for children with ASD.

The traditional approach to diagnosing ASD often involves subjective evaluation by clinicians, using observation, interviews, and standardized questionnaires. While these methods have been foundational in identifying ASD, they can be time-consuming, resource intensive, and prone to inconsistencies. As a result, there is a growing need for more efficient, objective, and reliable diagnostic tools to aid clinicians in early identification of the disorder.

## 2. METHODOLOGIES:

Random Forest for Classification: Random Forest is used to classify children as either ASD or non-ASD by building multiple decision trees. It enhances accuracy and reduces overfitting by averaging predictions, and also provides feature importance to identify key factors contributing to ASD.
Regression Analysis for Severity Prediction: Regression analysis predicts this verity or probability of ASD, providing a continuous out-come that quantifies the degree of autism. This helps clinicians' asses not just the presence of ASD but its intensity.

Model Evaluation:
Random Forest's performance is evaluated using accuracy, precision, and recall, while regression analysis is assessed using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared to measure prediction accuracy and model fit.

## 3. SYSTEMARCHITECTUREAND DESIGN:

The system architecture for ASD screening using machine learning consists of several layers to ensure efficiency and scalability. Data is collected from various sources, including behavioral data, developmental milestones, and diagnostic questionnaires. The data is preprocessed by cleaning, normalizing, and encoding features before being used for model training. Machine learning models, such as Decision Trees are trained on the prepared data, with cross-validation applied for optimization. The model saree valuated using metrics like accuracy and sensitivity, and hyperparameters are tuned for improved performance. The system features a user-friendly interface for clinicians to input data and receive real-time predictions.

**Overall Architecture:**

Our proposed prediction model for Autism Spectrum Disorder (ASD) utilizes data preprocessing and machine learning algorithms like Random Tree and Linear Regression, validated through robust data validation techniques. The system identifies key behavioral patterns to enable accurate ASD prediction and early intervention.

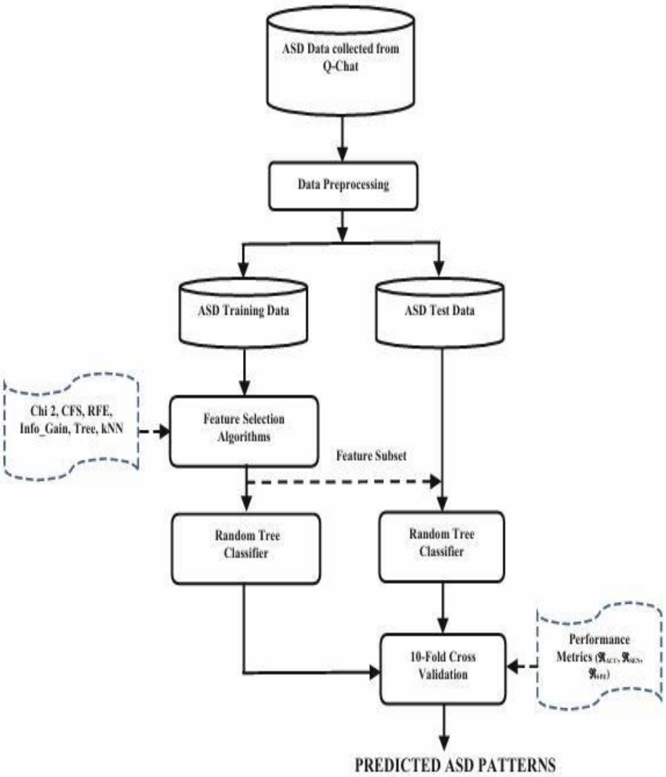
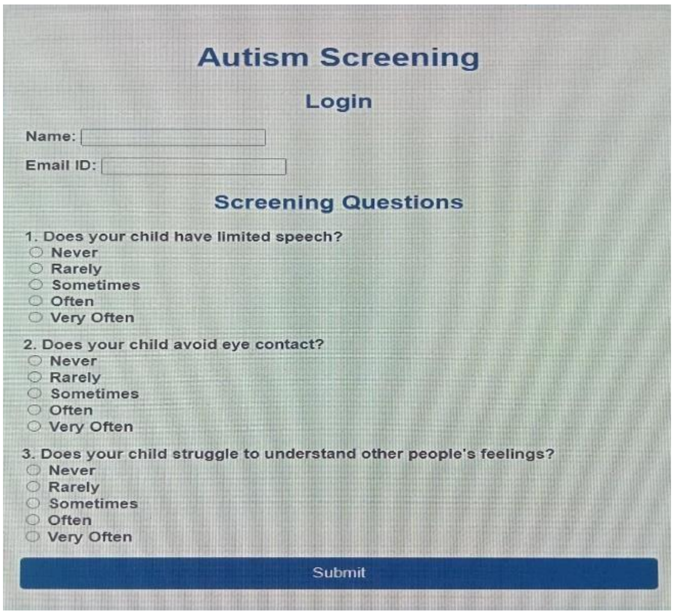



Fig. 1. Proposed prediction model for ASD data

## 4. MODULAR DESIGN:

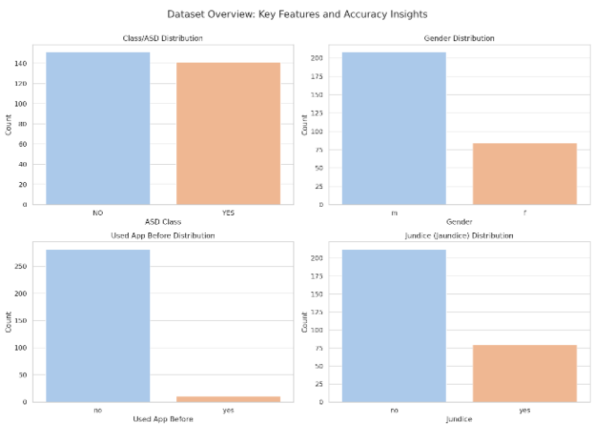
This project presents a modular design for early screening of Childhood Autism Spectrum Disorder (ASD) using a user-friendly interface. The system collects behavioral data through structured questionnaires, focusing on key indicators such as speech limitations, eye contact, and emotional understanding. Responses are categorized into frequency-based options to ensure detailed input. The collected data is processed using machine learning models to predict ASD likelihood. The modular approach ensures scalability, ease of use, and secure data handling.




**AUTISM MODEL:**

Our proposed autism detection model focuses on machine learning algorithms such as Random Forest and Linear Regression to analyze preprocessed data. By classifying inputs into ASD and non-ASD categories, the system provides an efficient and accurate tool for early autism screening.

## 5. DATA SET KEY OBSERVATION




The dataset consists of primarily integer fields, such as scores, and categorical variables, including gender and the Class/ASD column, which serves as the target variable. The Class/ASD column indicates whether a child is classified as having Autism Spectrum Disorder (ASD), with values of either "YES" or "NO." Additionally, the dataset contains demographic information, such as age, gender, ethnicity, relation, and country of residence. One of the key strengths of this dataset is the absence of missing values across all columns, ensuring a complete dataset for analysis. However, potential inconsistencies exist, particularly in the age column, which has only nine unique values. This suggests that age might have been encoded categorically, possibly representing predefined age ranges rather than exact ages.

The dataset exhibits a structured composition, primarily featuring integer fields such as scores and categorical variables like gender and Class/ASD, which serves as the target variable. The Class/ASD column identifies whether a child is classified as having Autism Spectrum Disorder (ASD), with values of either "YES" or "NO." Additionally, demographic information includes attributes such as age, gender, ethnicity, relation, and country of residence. Notably, the dataset contains no missing values across any columns, ensuring completeness for analysis.

However, some potential inconsistencies exist. The age column has nine unique values, suggesting categorical encoding, possibly representing age ranges. Ethnicity contains 11 unique values, which might include inconsistencies such as special characters or unexpected categories. The country of residence column has 52 unique values, potentially indicating typos or variations in country names. The age_desc column has only one unique value, which may lack meaningful variance for analysis. Several binary or categorical fields exist, including A1_Score to A10_Score, gender, jaundice, autism history, and Class/ASD. A visual summary of the dataset reveals key insights: the target variable, Class/ASD, appears balanced with a mix of "YES" and "NO" labels.

Gender distribution shows a binary split between male and female participants. The "Used App Before" column indicates whether participants have previously interacted with the ASD detection app. Additionally, the jaundice column highlights whether participants had jaundice during childhood. These factors provide valuable insights into the dataset's structure and potential areas for data cleaning and validation.

## 6. KEY FEATURES & VISUAL REPRESENTATION

The dataset comprises key inputs, including behavioral scores (A1 to A10) derived from diagnostic questionnaires, along with demographic details such as age, gender, and ethnicity. Additionally, it incorporates medical history indicators, such as whether an individual had jaundice or a family history of autism.

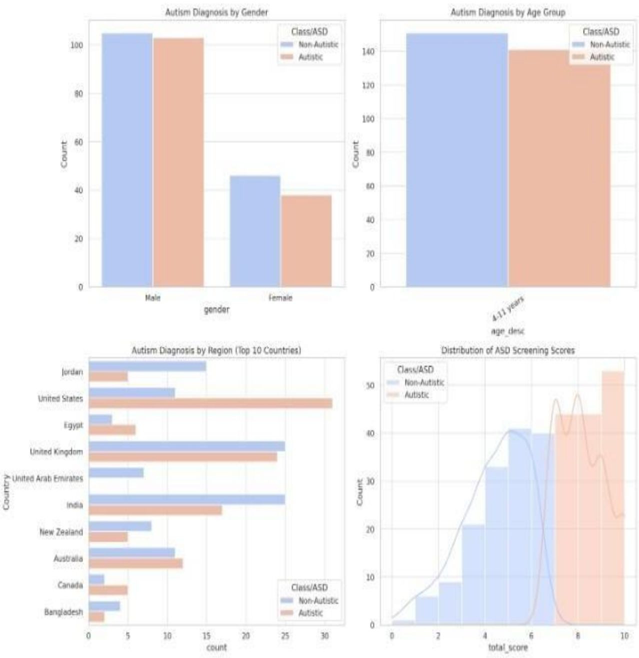
For machine learning analysis, a Random Forest Classification model is employed to categorize ASD vs. non-ASD cases, leveraging multiple decision trees to enhance accuracy and mitigate overfitting. Regression analysis is also utilized to predict severity levels, aiding in the assessment of ASD intensity.
The system architecture follows a structured workflow, beginning with data preprocessing, which involves normalization, encoding categorical variables, and handling missing values.

The model training phase uses cross-validation techniques to ensure robust predictions, while evaluation metrics, including accuracy, precision, recall, and Mean Absolute Error (MAE), assess the model's effectiveness. A user-friendly web-based interface allows individuals to enter behavioral scores and demographic details, providing real-time predictions on ASD likelihood and severity analysis for potential follow-up interventions.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 02 | Feb - 2025 SJIF Rating: 8.448 ISSN: 2582-3930


The results indicate that the model successfully differentiates between ASD and non-ASD cases. Visualizations further enhance insights by showcasing demographic trends, including ASD diagnosis rates based on gender, age, and geographical region, facilitating a deeper understanding of ASD patterns across populations.

## 7. ACCURACYANALYSIS




The visualizations provide an in-depth analysis of autism diagnosis across different demographic categories. The first chart shows that males are diagnosed with autism at a significantly higher rate compared to females. The second chart, focused on age group, indicates that the majority of diagnoses are in the younger age group (4-11 years). The regional breakdown of diagnoses highlights those certain countries, such as Egypt and India, have higher reported cases, while countries like Jordan and Bangladesh show relatively fewer. Finally, the distribution of ASD screening scores demonstrates a clear distinction between autistic and non-autistic individuals, with the score distribution indicating that autistic individuals tend to have higher screening scores. These visualizations can provide valuable insights for further research and targeted interventions.

Dataset:

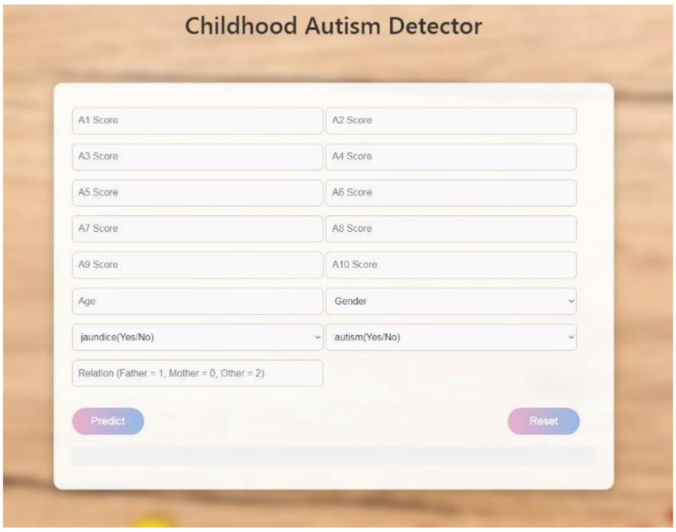
Screening Autism Data

| Name | Gender | Age | Country | Relation | ASD Result |
|---|---|---|---|---|---|
| John Doe | Male | 10 | USA | Father | Yes |
| Jane Smith | Female | 12 | Canada | Mother | No |

The table presents autism screening data for two individuals. John Doe, a10-year-oldmalefromthe USA,

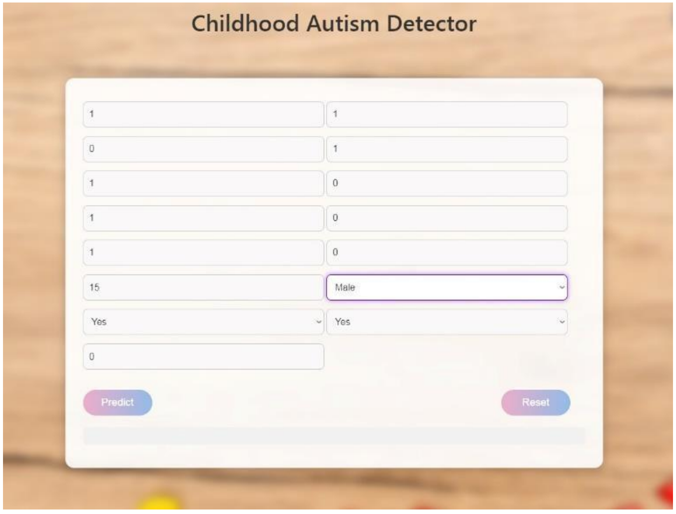
was screened with a positive result for ASD, reported by his father. Jane Smith, a 12-year-oldfemalefromCanada, received a negative result for ASD, as reported by her mother. This dataset captures essential demographic and screening outcomes for autism diagnosis.

## 8. RESULTS

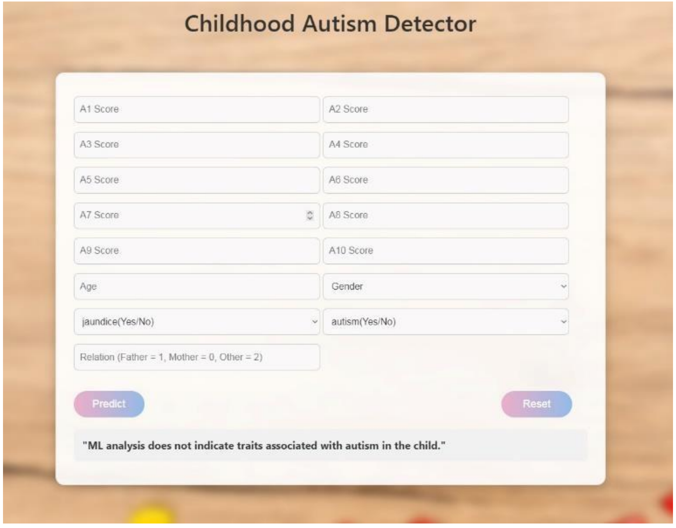


The image shows a form for a "Childhood Autism Detector" where users can input various scores and details to predict the likelihood of autism. The fields include ten score inputs (A1 to A10) that likely correspond to screening results, along with dropdowns forage, gender, jaundice (Yes/No), autism status (Yes/No), and the relation of the person filling out the form (Father, Mother, Other). Users can either predict the outcome by clicking "Predict" or reset the form using the "Reset" button. This interface allows easy data entry for autism screening purposes

The "Childhood Autism Detector" form is filled with the required data, including alternating scores of 1 and 0 for the A1 to A10 fields. The age entered is 15, and the selected gender is male. The formal so indicates a history of jaundice and autism, both marked as "Yes." The relation is set to 0, representing "Mother". With all fields populated, the user has the option to click the "Predict" button to process the data or "Reset" to clear the form and start over.




After the form submission in the "Childhood Autism Detector," the machine learning analysis result indicates that there are no traits associated with autism in the child. The message displayed at the bottom of the form states, "ML analysis does not indicate traits associated with autism in the child."




## 9. FUTUREWORK

The future scope of this project on ASD screening using machine learning holds several exciting opportunities for enhancement. First, integrating more advanced models like deep learning networks or ensemble methods could further improve prediction accuracy and handle more complex patterns in data. Expanding the dataset to include diverse populations from various regions would enhance the model's generalizability and make it more robust across different demographics. Additionally, incorporating real-time data collection tools such as mobile apps or wearable devices could enable continuous monitoring and dynamic feedback. The system could also be upgraded to provide personalized intervention recommendations based on the severity and specific needs of each child. Finally, expanding the system's application into broader clinical, educational, and community settings would allow for its integration into existing health and educational management systems, facilitating early intervention on a global scale. These developments would help evolve the project into a comprehensive, real-time tool for early diagnosis and intervention in autism spectrum disorder.

## 10.CONCLUSION

The "Childhood Autism Detector" project is designed to assist in the early detection of autism traits in children. By entering specific behavioral scores and demographic details, the system provides a quick assessment to help identify potential signs of autism. The user-friendly interface allows for easy input of data, such as age, gender, and family history, and the system outputs a prediction based on the given information. This tool can be valuable in guiding caregivers or healthcare professionals to consider further evaluations. In the final test case, the system indicated no traits associated with autism in the child. The project demonstrates the potential of technology in providing early insights for childhood development concerns.

## REFERENCES:

1. National Institute of Mental Health. https://www.nimh.nih.gov/health/topics/autism spectrum-disorders-asd/index.shtml

2. Penner, M., Anagnostou, E., Ungar, W.J.: Practice patterns and determinants of wait time for autism spectrum disorder diagnosis in Canada. Mol. Autism 9, 16 (2018)

3. Thabtah, F., Kamalov, F., Rajab, K.: A new computational intelligence approach to detect autistic features for autism screening. Int. J. Med. Inform. 117, 112–124 (2018)

4. Pratap, A., Kanimozhiselvi, C.S., Vijayakumar, R., Pramod, K.V.: Predictive assessment of autism using unsupervised machine learning models. Int. J. Adv. Intell. Para. 6(2), 113–121 (2014). https://doi.org/10.1504/IJAIP.2014.062174

5. Maenner, M.J., Yeargin-Allsopp, M., Van Naarden Braun, K., Christensen, D.L., Schieve, L. A.: Development of a machine learning algorithm for the surveillance of autism spectrum disorder. PLoS One 11(12), e0168224 (2016)

6. Thabtah, F.: ASDTests a mobile app for ASD, screening (2017). www.asdtests.com

7. Miotto, R., Wang, F., Wang, S., Jiang, X., Dudley, J.T.: Deep learning for healthcare: review, opportunities and challenges. Brief. Bioinf. 19(6), 1236–1246 (2018)

8. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Academic Press, Cambridge (2012)

9. Ramani, G., Selvaraj, S.: A pragmatic approach for refined feature selection for the prediction of road accident severity. Stud. Inf. Control23(1), 41–52 (2014)

10. Shanthi, S., Geetha Ramani, R.: Classification of vehicle collision patterns in road accidents using data mining algorithms. Int. J. Comput. Appl. 35(12), 30–37 (2011)