A Machine Learning Framework for Age-Inclusive Autism Spectrum Disorder Detection Using U.S. Screening Data

Conference Paper · April 2025

DOI: 10.1109/ICICT64420.2025.11005015

6 authors, including:

Mohammad Anisur Rahman
East Texas A&M University

Intiser Islam
University of Bridgeport

Shariar Islam Saimon
University of Bridgeport

Md Rafiuddin Siddiky
Wilmington University

Proceedings of the International Conference on Inventive Computation Technologies (ICICT-2025)
DVD Part Number: CFP25F70-DVD; ISBN: 979-8-3315-1223-1

A Machine Learning Framework for Age-Inclusive Autism Spectrum Disorder Detection Using U.S. Screening Data

Mohammad Anisur Rahman
Department of Marketing & Business Analytics
East Texas A&M University, 2104 University Dr, Commerce, TX 75428, USA
mrahman25@leomail.tamuc.edu

Intiser Islam
School of Engineering,
University of Bridgeport, 126 Park Ave, Bridgeport, CT 06604, USA
iislam@my.bridgeport.edu

Shariar Islam Saimon
School of Engineering,
University of Bridgeport, 126 Park Ave, Bridgeport, CT 06604, USA
ssaimon@my.bridgeport.edu

Md Rafiuddin Siddiky
College of Technology
Wilmington University, New Castle, Delaware, USA
mdrafiuddinsiddiky@gmail.com

Md Mahadi Hasan
School of Information Technology,
Washington University of Science and Technology, Alexandria, Virginia, USA
mahadih384@gmail.com

Seaam Bin Masud
College of Technology
Wilmington University, New Castle, Delaware, USA
seaam.masud@gmail.com

Corresponding author: Mohammad Anisur Rahman (mrahman25@leomail.tamuc.edu)

Abstract*— The growing prevalence of Autism Spectrum Disorder (ASD) is thus prompting the development of efficient, accurate, and accessible diagnostic tools for all ages. Although these include traditional diagnostic methods, such as the Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R), these methods are resource-intensive and appear to have been designed primarily for early childhood, leaving adolescents and adults (40–55 years) at risk of missed diagnosis. We aim to develop a machine learning (ML) based framework to boost ASD detection in the U.S. Autism Screening Dataset across disparate age groups. The dataset comprises 704 records, including demographic, behavioral, and diagnostic attributes. Some data preprocessing techniques, normalized and encoded, improved inconsistency and the model performance. Then, principal component analysis (PCA) and correlation analysis were applied to determine the most relevant diagnostic features. Four supervised ML models over the Decision Tree, Random Forest (Gini and Entropy criteria), and Extra Trees Classifier are trained and evaluated using accuracy, precision, recall, f1 score, and ROC-AUC as performance metrics. Among these, the Extra Trees Classifier surpassed the other models, with 96.67% accuracy, 97.50% precision, 95.00% recall, and 96.17% F1 score, proving to be a good model to predict and robust. This highlights the potential of ML-driven approaches to complement traditional diagnostic frameworks by providing scalable, efficient, and age-independent detection of ASD. This framework addresses existing gaps by improving diagnostic accuracy and accessibility, which will lead to earlier intervention and better outcomes for people of all ages with ASD.

Keywords- Autism Spectrum Disorder (ASD), Autism Screening Dataset, Predictive Modeling, Machine Learning (ML), Age-Inclusive Diagnostics

I. INTRODUCTION

Autism is a lifelong disability that affects the ability to communicate and relate to others [1]. People with autism have difficulties in four main areas: Communication, social interaction, stereotyped activities, and sensory abnormalities, such as (hypoactivity, hyperactivity, and fluctuations between these states), which is claimed to be the most common complication [2, 3]. The Centers for Disease Control and Prevention (CDC) reports increasing ASD prevalence, now affecting 1 in 36 children in the U.S., with significant lifelong impacts. Early diagnosis and intervention have been shown to significantly improve outcomes by providing timely therapeutic and educational support. However, the diagnostic process is still a course of considerable difficulty even with improving diagnostic tools. The DSM-5 presents guidelines for identifying ASD, but these tools are time-consuming, need skilled personnel, and are expensive. Furthermore, conventional diagnostic frameworks place considerable emphasis on early childhood, to the exclusion of older children, adolescents, and adults, rendering many misses or delayed diagnoses.

Most existing methods used to diagnose ASD are based on clinical assessment of items like Autism Diagnostic Observation Schedule (ADOS) [4]. Although, these tools are good, but are not discriminatory and are primarily loaded in favor of certain age groups predominantly children [5]. It has this major gap in the way ASD symptoms are addressed in older populations, who can show subtle or masked behaviors that are different from early developmental signs. Challenges of early detection are exacerbated by the lack of scalable, automated, and age-inclusive diagnostic tools. This limitation poses a risk that many individuals across different age groups might go undiagnosed, potentially denying them valuable interventions to help them live a better life.

The primary objective of this study is to develop a machine-learning framework for detecting Autism Spectrum Disorder (ASD) across all age groups using a comprehensive U.S.-based Dataset. Specifically, this study focuses on finding patterns and correlations from the demographic, behavioral, and diagnostic data that can be beneficial in ASD detection. We employed various machine learning models: Decision Tree, Random Forest using Entropy criterion (RF-Ent), Random Forest using Gini criterion (RF-Gini), and Extra Trees Classifier (ET21) to determine which model best addresses screening for ASD.

The goal of this study is to make these models able to generalize across age groups, as limitations in current methods tend to be specific to certain age ranges. To further evaluate model performance, metrics such as accuracy, precision, recall, F1 score, and ROC-AUC are used to determine their reliability and applicability. In addition to that, this research intends to induce information on how to optimize these models for better detection abilities and to mold the design of scalable, efficient, and inclusive diagnostic tools for ASD detection. the capacity to allow informed decision-making.

The paper is organized mainly as follows: Section II summarizes relevant work; Section III uses Methodology to describe the technique, including the dataset description, recommended strategy, and model interpretability. Coverage of the experimental design, feature importance analysis, assessment metrics, and interpretability outcomes in Section IV provides the results and discussion. Section V concludes the report with ideas and comments for future endeavors.

II. RELATED WORK

The identification of ASD in the initial developmental stage is a significant area of research interest in neurodevelopmental disorders, and ML and DL are receiving considerable attention to improve diagnosis. In this section, recent achievements in Employing, ML, and DL frameworks for ASD detection are discussed concerning various datasets.

Several studies have attempted to detect and diagnose ASD using various machine learning (ML) techniques.
ASD has been analyzed across different age groups by M. Abdelwahab et al. [6] through the application of six machine learning techniques including Support Vector Machines (SVM), Random Forest (RF), Naïve Bayes (NB), Logistic Regression (LR), K-Nearest Neighbors (KNN) and Decision Tree (DT). The study used four different public ASD datasets available from Kaggle and UCI ML repositories. The models were tested in terms of effectiveness in enhancing the diagnostic experience, especially the early-stage diagnoses. The evaluated methods included in this study revealed that Logistic Regression (LR) showed better accuracy at 96.69% and performed better on the selected dataset. Thabtah et al. [7] proposed a new model combining Self-Organizing Maps (SOM) with classification algorithms to address biases in autism screening. The SOM was used to derive new class labels from features related to communication, repetitive traits, and social traits, which were then refined to improve accuracy. The model, evaluated on a real-life autism screening dataset with over 2000 instances, outperformed traditional methods, achieving higher accuracy, precision, and recall. In [8], S. Raj et al. analyzed and compared the possibilities of employing Naive Bayes, SVM, Logistic Regression, KNN, Neural Networks, and CNN for the ASD forecast at kids, adolescents, and nearly grown-up individuals. Evaluating three non-clinical adult, children, and adolescent ASD datasets, the authors found that CNN models were superior to other models with an accuracy of 99.53%, 98.30%, and 96.88% respectively establish the effectiveness of the CNN for ASD screening. Using Support Vector Machines (SVM) based Machine Learning (ML) algorithms Bone et al., [9] proposed methods that help improve the screening and diagnosing of ASD. Based on ADI-R and SRS for 1,264 participants with ASD and 462 without ASD, the algorithms displayed increased accuracy in comparison with the previous approaches. The screener has 89.2% and 86.7% sensitivity and 59.0% and 53.4% specificity for people under and over the age of 10 using only the five behavioral codes, the results suggest that ML is promising for better, faster, and tailored ASD diagnosis. Vakadkar et al. [10] used SVM, RFC, NB, and ML to predict infants with ASD by analyzing the datasets. The study also sought to establish the frailties that make an individual prone to developing ASD with a view of cutting short the procedures of diagnosing the condition. From the above, a relatively high accuracy was attained by the Logistic Regression model that makes it suitable for optimized search for ASD. In addition to being more accurate and efficient than conventional behavioral observation methods, B. Kamala et al. [11] proposed the use of machine learning (ML) to enhance early ASD detection. ML was shown to be able to analyze brain region relationships, discover genetic factors, and create good interventions for ASD diagnosis. In its application in evaluating an automated machine learning (ML) implementation via a feed-forward artificial neural network (fANN), Achenie et al. [12] Examined the use of M-CHAT-R data for ASD screening in 14,995 toddlers. The accuracy of the ML models was high with 99.92% correct classification for certain subgroups (into the White toddler subgroup). Results showed ML could effectively replace human scoring and follow-up, cut labor and human error by orders of magnitude, and use far fewer items than traditional methods. Akter et al. [13] used a variety of feature transformation methods and machine-learning techniques on early-detected datasets of ASD in toddlers, children, adolescents, and adults. For toddlers, SVM looked the best, for children and adolescents, respectively, Adaboost and Glmboost. For toddlers, the sine function rather than the other feature transformations performed best; for children

and adolescents, Z-score was most effective. The results showed that optimized machine learning models are capable of accurate ASD status prediction, and may be helpful for early detection of ASD.

III. METHODOLOGY

A. *Dataset Description*

There are 704 records related to Autism Spectrum Disorder (ASD) screening of 21 columns in this dataset. Those include 10 behavioral screening scores [19] (A1_Score thru A10_Score) and 10 binary indicators representing responses to various ASD screening questions. The demographic information also includes age, gender, ethnicity, and country of residence. Moreover, the dataset covers medical and family history factors such as whether the subject had jaundice at birth, whether the person had an earlier ASD diagnosis, and the respondent's relationship to the subject being screened. In addition to this, the dataset has screening information, such as whether the individual used an ASD screening app before (variable "used_app_before") and the result. The final classification essentially equals our target variable, which is "Class/ASD" (YES/NO), and tells us whether a person has ASD or not.

B. *Data Preprocessing*

Missing values were handled for data preprocessing at every step although for numerical data only median imputation was applied. To avoid numerical problems from models that rely on distance comparisons, numerical features were normalized to be on the same scale. For gender and ethnicity, categorical variables were encoded with one-hot / label encoding based on the required model. The purpose of that preprocessing was to clean the dataset, ensure its consistency, and ensure that it was ready for analysis and machine learning modeling.

C. *Proposed Methodology*

1. *ML Based Autism Detection Framework*

This study intended to increase the detection of ASD by utilizing machine learning algorithms in the improvement of accuracy and efficiency in the early diagnosis of Autism. The framework involves many machine learning models including supervised learning techniques to classify and predict the likelihood of ASD using clinical and diagnostic data. It includes data preprocessing, feature extraction, model training, and classification steps which have been attached to each to ensure the ruggedness and reliability of the detection system. This paper uses the dataset in the study and the preprocessing steps include data cleaning, normalization, and feature extraction. Then various techniques such as Principal Component Analysis (PCA), Correlation Heatmap, and Pearson Correlation are applied to extract relevant features from the raw data to transform it into something that can train a model. To cope with the curse of dimensionality, feature selection is performed to select the most influential variables, thus reducing the dimensionality of the dataset to improve the model's generalization from one case of ASD to another. The framework uses the Decision Tree, Random Forest (RF_Ent), Random Forest (RF_Gini), and Extra Trees (ET21) to perform training using machine learning algorithms. We train these algorithms on a dataset of ASD-related features and measure their performance by using performance metrics like accuracy, precision, recall, and F1 score. These metrics are designed to make sure these models can tell whether an individual is likely to have ASD or not. Once trained, the models are used to predict the ASD likelihood of new data to support early identification and intervention for people at risk of becoming an ASD individual. The process is reliable and an efficient step towards supporting timely clinical decision-making.

The overview of our workflow is illustrated in Fig. 1.

2. *Models*

We evaluated the framework with Decision Tree, Random Forest (RF_Ent), Random Forest (RF_Gini), and Extra Trees (ET21) techniques that are described as follows:

- **Decision Tree:** A supervised machine learning model takes the form of a Decision Tree, which is a tree using a tree-like structure to classify data or predict the outcome of some event [14] . A root node, internal nodes containing test conditions, and leaf nodes describing final class labels (or outcomes). Finally, it divides the dataset into smaller subsets by the feature. At each step, the tree classifies the dataset into smaller subsets until it reaches the classification of the leaves.

The Decision Tree analyzes diagnostic features to decide if an individual will likely have ASD within ASD detection. It is simple, interpretable, and a useful tool because of its simplicity and interpretation of the process of classification in a clear visual way, whether the data are categorical or numerical. For this study, the Decision Tree was used to identify patterns and classify individuals, demonstrating its utility in supporting early ASD detection through human-readable diagnostic rules. For classification, the output at a leaf node is the majority class c .The predicted class $\hat{y}$ is determined by:

$$\hat{y} = arg\max_{c} p(C)$$

where $p(c)$ is the proportion of data points in class c at the leaf node.

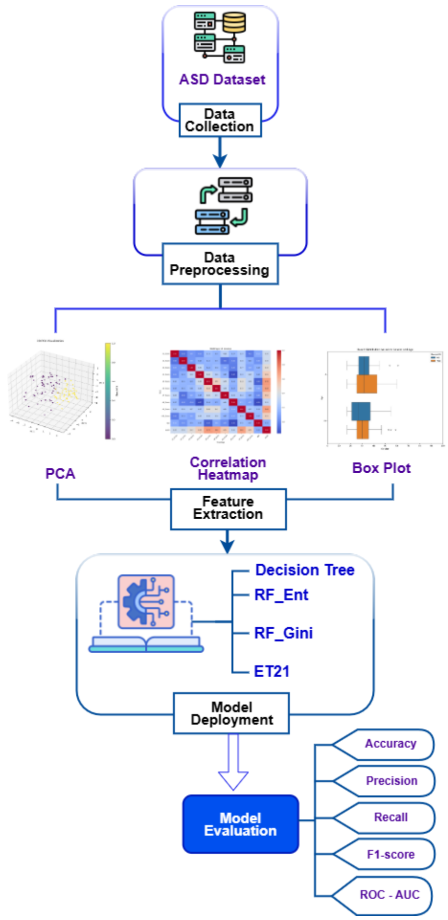

Fig. 1. Overview of ML Based Autism Detection Framework

- **Random Forest with Entropy (RF_Ent):** Robust Random Forest with Entropy (RF_Ent) is an ensemble machine learning algorithm that builds up multiple decision trees and outputs their results in an attempt to deliver better classification accuracy. Entropy is used by each tree to evaluate splits, using features that maximize Information Gain (IG) at each node [15]. Entropy, which measures uncertainty, is calculated as:

$$H(s) = -\sum_{i=1}^{c} p_i log_2(p_i)$$

where p_i is the proportion of instances in class i. Information Gain is then determined by:

$$IG = H(S) - \sum_{j=1}^{k} \frac{|S_j|}{|S|} H(S_j)$$

where S_j represents subsets formed by the split.
In ASD detection, RF_Ent analyzes diagnostic features across multiple trees, each trained on different data subsets. The final prediction aggregates individual tree outputs through majority voting:

$$\hat{y} = arg\ \max_{c} \frac{1}{n} \sum_{i=1}^{n} T_i\,(x)$$

Where $T_i(x)$ is the prediction of the i-th tree. RF_Ent's robustness and ability to handle complex data make it ideal for this study, effectively identifying patterns and supporting early ASD detection with high reliability.

- **Random Forest using Gini Index (RF_Gini):** The RF_Gini ensemble model is an ensemble model of many decision trees that fits multiple trees and gives a higher classification accuracy [16]. Gini impurity is calculated as follows, and each tree splits data by minimizing:

$$Gini = 1 - \sum_{i=1}^{C} {p_i}^2$$

where p_i is the proportion of data in class i. Lower values indicate purer splits.
The high dimensional nature of ASD diagnostic data poses a challenge to harvesting complex patterns, RF_Gini is particularly well suited for this kind of dataset and this is our motivation to use it. This allows us to gain insight into feature importance and helps us to determine which diagnostic factors are important. To carry out reliable classifications for early ASD detection, RF_Gini was used for this study for its ability to handle numerical and categorical data efficiently.

- **ET21:** In the same way that Random Forest is an ensemble where you create lots of trees for each observation, Extra Trees is essentially an ensemble of lots of trees for each observation, but with extra randomness built into making all these trees [17]. Extra Trees selects splits randomly, in contrast to Random Forest which has each split based on ascertaining Gini or entropy, an increase in diversification to the model [18]. Extra Trees constructs multiple decision trees out of randomly selected features and thresholds for ASD detection to decrease variance and avoid overfitting. The final prediction involved majority voting over all trees in the ensemble. The randomness in its architecture improves its ability to generalize well to unseen data, making it a good tool for finding patterns in ASD diagnostic datasets. The mathematical calculation for predictions combines the output of N trees as follows:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} T_i\,(x)$$

where $T_i(x)$represents the prediction of the i-th tree for input x.

IV. Result & Discussion

A. Experimental Setup:

We used Google Colab for our research, writing in Python. Using **scikit-learn**, model implementation was conducted. The workflow consisted in data preparation and training assessment to explain model predictions.

B. PCA, Correlation Heatmap and Box Plot outcomes:

Fig. 2 displayed the 3D PCA visualization, where the first principal components can explain the variation in the Early-Stage Autism Detection dataset, particularly those who have or do not have ASD. In class 0, the non-ASD cases are more scattered out among the data vectors, and the PCs to which those vectors are assigned are placed on the axes for PC1 PC2, and PC3, whereas class 1 has a greater population of density along a subset of axes. Components that explain the most variance as well as those whose contribution to distinguishing ASD from non-ASD cases is greatest are shown. Identifying these key components helps us select the most relevant features for further analysis so that the model can thus better classify individuals according to their likelihood to have ASD. By analyzing the PCA results, we can identify key features for ASD detection, enhancing the model's ability to accurately classify ASD and non-ASD cases.

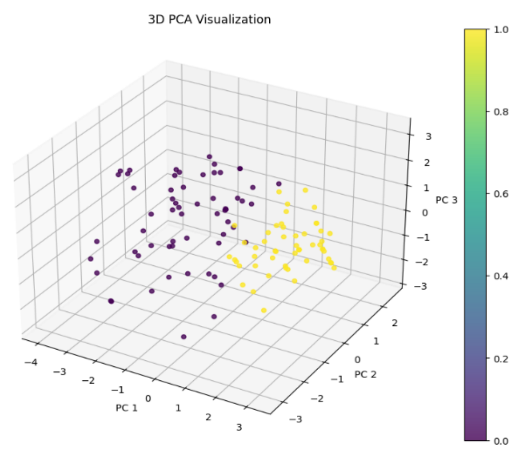

Fig. 2. PCA Plot Revealing Variance of Components for ASD and Non-ASD

Fig. 3 shows the first correlation heatmap, which helps to show the relationships of some key interrelationships among the diagnostic features for the refinement of ASD detection algorithms. For example, strong correlations associated with scores such as "A1" and "A2" indicate shared prediction patterns, and thus dimensionality reduction through feature combination is possible. Similarly, it finds much stronger relationships between attributes such as 'A3' and 'A4', which allows for the model to simplify its feature set without diminishing predictive accuracy. On the contrary, the weaker correlations between variables such as "Gender" and "Relation", signify its use as an independent predictor for the model to distinguish between ASD and non-ASD cases. Other variables also interact with features like "Age" and "Result" to reveal diagnostic patterns. The feature selection approach developed in this work optimizes the overall efficiency and accuracy of the model while also helping achieve the objectives of early ASD detection systems.

Fig. 4 shows the boxplot of the relationship between gender, age, and ASD classification and provides a clearer understanding of demographic trends using the data set. These patterns for age distribution in ASD-positive male and female subjects agree with established prevalence patterns in that males have a broader age range than females. Such a difference highlights the importance of gender-specific considerations in detection models. Evidence of atypical diagnostic scenarios, i.e. outliers, particularly among older participants within the ASD-positive group, is provided for adaptive algorithms. Through analysis of medians and interquartile ranges (IQRs), age is found to be an important variable in determining diagnostic features, supporting the development of ASD detection models tuned for more diverse populations.

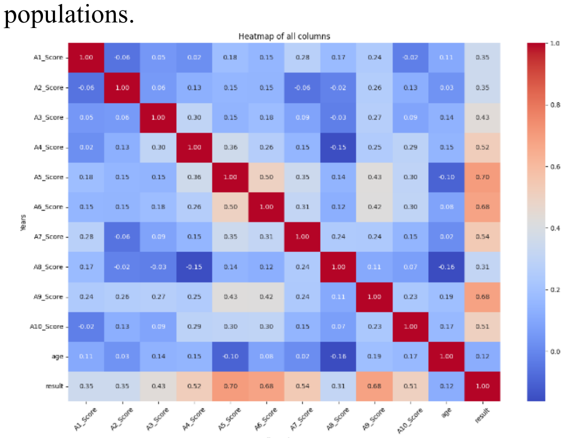

Fig. 3. Feature Correlation Heatmap

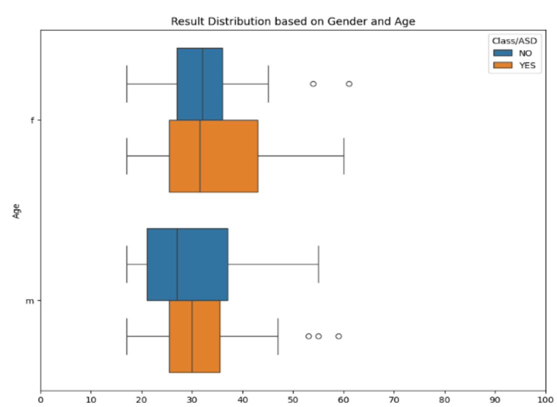

Fig. 4. Boxplot Showing Result Distribution Based on Gender and Age

C. Evaluation Metrics

In this study on early autism detection using diagnostic and demographic features, we evaluated four machine

Proceedings of the International Conference on Inventive Computation Technologies (ICICT-2025)
DVD Part Number: CFP25F70-DVD; ISBN: 979-8-3315-1223-1

learning models: It uses Decision Tree, Random Forest with Entropy (RF_Ent), Random Forest with Gini Index (RF_Gini), and Extra Trees (ET). Table 1 shows a summary of the results with metrics for our model such as the accuracy, precision, recall, F1-score, and ROC-AUC scores.

Overall, the Extra Trees (ET) model, achieved 96.67% accuracy, 97.50% precision, 95.00% recall, and 96.17% F1-score. With an ROC AUC score of 96.50%, and strong discriminatory power, ET is a good choice for early autism detection because of its good balance of precision and recall. RF_Ent, the Random Forest with Entropy model, got very close to the ET performance with 96.67% accuracy and a 96.85% high F1 score. With an ROC AUC score of 96.25%, the model clearly has the ability to perform the ASD vs. non-ASD case well and is another reliable model for the task. The Random Forest with Gini Index (RF_Gini) model had strong

TABLE I

PERFORMANCE METRICS OF DIFFERENT ML MODELS FOR ASD DETECTION ON U.S DATASET

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
Decision Tree	94.44	92.46	98.00	94.98	94.08
RF_Ent	96.67	96.67	97.50	96.85	96.25
RF_Gini	94.44	95.56	93.00	93.74	94.50
ET21	96.67	97.50	95.00	96.17	96.50

performance at 94.44% on accuracy and 95.56% on precision. Its slightly lower recall of 93.00% but an F1-score of 93.74% suggests it was occasionally less likely than ONDBias to underidentify cases of ASD. ROC-AUC score 94.50% further persualage its overall effectiveness but it lags behind RF_Ent and ET. The model developed by using the Decision Tree gave an accuracy of 94.44% and a recall of 98.00%, showing that it correctly classified most ASD cases. Its 92.46 % precision and 94.98 % F1 score indicate that occasionally there will be false positives, which might be a drawback in clinical scenarios. Its ROC-AUC of 94.08 states it's competitive but in the same breath a little limited scoring relative to the other models.

Among all the discussed models, the ET model is the most balanced and the most suitable for early autism detection, with its robustness to preserve high accuracy and high recall while managing a minimal number of false positives. The potential of RF_Ent is also shown to be strong for comparable level of performance.

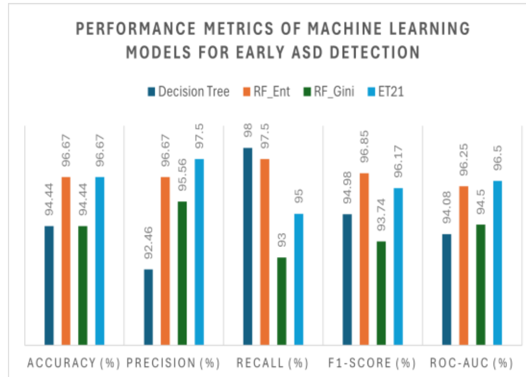

Fig. 5. Comparative Performance Metrics of ML Models for Early ASD Detection

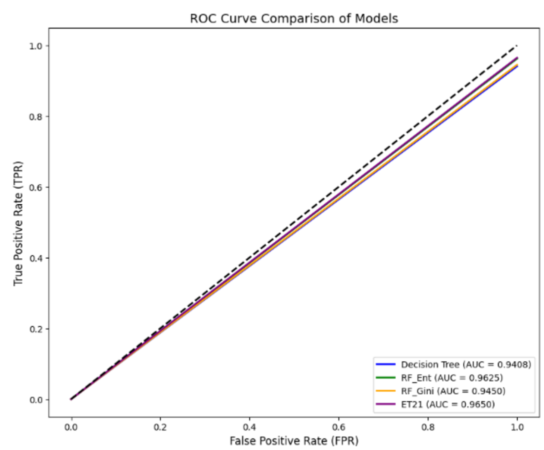

Fig. 6. Comparative Performance Metrics of ML Models for Early ASD Detection

With the Decision Tree and RF_Gini models as effective as they are, they lack some tradeoff in precision and recall, and ultimately are not as good as the ET and RF_Ent models on this dataset. It will be clear that sophisticated ensemble methods such as ET and RF_Ent are more appropriate to the more nuanced task of early ASD detection.

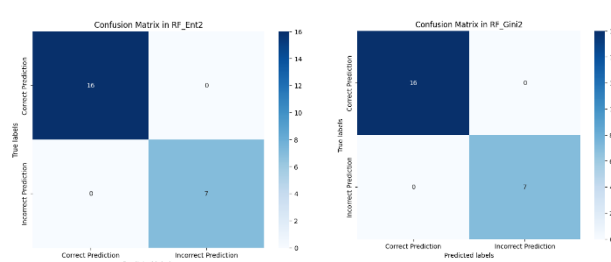

Fig. 7. Confusion Matrics for RF_Ent2(left side) and RF_Gini2(right side)

V. CONCLUSION

The current work presents a machine learning (ML)—based approach to Autism Spectrum Disorder (ASD) diagnostics, which will close a significant void in existing techniques. Traditionally, the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R) are among the popular diagnostic tools, but they are expensive, resource-dependent, and early age-dependent. This study uses ML techniques to increase the precision, efficiency, and inclusivity of ASD diagnostics in all age groups.

Here, we used a preprocessed dataset with demographic and diagnostic features. Key steps were taken to train the model on the dataset, such as Normalizing the data, encoding the features, and reducing dimensionality using PCA. Important variables affecting ASD prediction were also noticed in the correlation analysis, leading to the generation of streamlined models. Four were evaluated to examine the supervised ML models: decision tree, random forest using entropy and Gini criteria, and classification of extra trees. These models also showed promising results regarding comprehensive performance metrics: accuracy, precision, recall, F1 score, and ROC-AUC. With an accuracy of 96.67%, a balanced precision of 97.50%, and a Recall of 95.00%, the Extra Trees Classifier had the highest performance in terms of robust generalizability and avoiding overfitting. It is observed that the Random Forest (Entropy) model also produces decent results. In contrast, Decision Tree and Random Forest (Gini) models perform well with compromises in precision and recall, and ensemble methods are needed for fine-grained diagnostics.

The findings of this study show that the diagnosis can be performed in a shorter time and with less expertise using an ML-based framework. By expanding accessibility and inclusivity in identifying ASD for individuals that could otherwise easily fall through the cracks using more traditional means, these models can provide better detection of ASD, especially for older individuals or, in particular, presentations of the disorder. Future research can expand the dataset by a broader range of demographics and diagnostic criteria and increase prediction accuracy with additional ML techniques such as deep learning. These models will be eventually deployed in clinical test beds to validate their practicality and to refine them further. Yet, it is still essential to enhance the model's interpretability so that clinicians and stakeholders can trust the AI predictions. Key limitations include dependence on high-quality, diverse data, limited interpretability of some models, significant computational requirements, and the need for real-world clinical validation to ensure effectiveness and reliability.

The findings of this study show that machine learning can be a helpful tool in diagnosing ASD and could pave the way to timely interventions and better outcomes for people with ASD. With growing awareness and understanding of ASD, AI-driven approaches to diagnosis will be instrumental in tearing down the barricades in the way of early and inclusive detection.

REFERENCES

[1] S. Ion, "Social Assistance for Children and Young People with Autism Spectrum Disorders (ASD)," *Social Work Review/Revista de Asistenta Sociala,* no. 2, 2021.

[2] D. Simmons, L. McKay, P. McAleer, E. Toal, A. Robertson, and F. Pollick, "Neural noise and autism spectrum disorders," *Perception,* vol. 36, pp. 119-120, 2007.

[3] A. Sadaf, H. Waseem, S. Rose, U. Ahsan, K. Abbas, and W. A. Ahad, "Relationship Among Sensory Issues and Repetitive Behaviors In Children With Autism Spectrum Disorder," *Asian Journal of Allied Health Sciences (AJAHS),* pp. 59-64, 2020.

[4] J. A. Bastiaansen *et al.*, "Diagnosing autism spectrum disorders in adults: the use of Autism Diagnostic Observation Schedule (ADOS) module 4," (in eng), *J Autism Dev Disord,* vol. 41, no. 9, pp. 1256-66, Sep 2011, doi: 10.1007/s10803-010-1157-x.

[5] S. Nadel and J. E. Poss, "Early detection of autism spectrum disorders: screening between 12 and 24 months of age," (in eng), *J Am Acad Nurse Pract,* vol. 19, no. 8, pp. 408-17, Aug 2007, doi: 10.1111/j.1745-7599.2007.00244.x.

[6] M. M. Abdelwahab, K. A. Al-Karawi, E. Hasanin, and H. Semary, "Autism spectrum disorder prediction in children using machine learning," *Journal of Disability Research,* vol. 3, no. 1, p. 20230064, 2024.

[7] F. Thabtah *et al.*, "Autism screening: an unsupervised machine learning approach," *Health Information Science and Systems,* vol. 10, no. 1, p. 26, 2022/09/08 2022, doi: 10.1007/s13755-022-00191-x.

[8] S. Raj and S. Masood, "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques," *Procedia Computer Science,* vol. 167, pp. 994-1004, 2020/01/01/ 2020, doi: https://doi.org/10.1016/j.procs.2020.03.399.

[9] D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S. Narayanan, "Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion," (in eng), *J Child Psychol Psychiatry,* vol. 57, no. 8, pp. 927-37, Aug 2016, doi: 10.1111/jcpp.12559.

[10] K. Vakadkar, D. Purkayastha, and D. Krishnan, "Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques," *SN Computer Science,* vol. 2, no. 5, p. 386, 2021/07/22 2021, doi: 10.1007/s42979-021-00776-5.

[11] B. Kamala, K. S. M. Pooja, S. Varsha, and K. Sivapriya, "ML Based Approach to Detect Autism Spectrum Disorder (ASD)," in *2021 4th International Conference on Computing and Communications Technologies (ICCCT)*, 16-17 Dec. 2021 2021, pp. 313-318, doi: 10.1109/ICCCT53315.2021.9711826.

[12] L. E. K. Achenie, A. Scarpa, R. S. Factor, T. Wang, D. L. Robins, and D. S. McCrickard, "A Machine Learning Strategy for Autism Screening in Toddlers," (in eng), *J Dev Behav Pediatr,* vol. 40, no. 5, pp. 369-376, Jun 2019, doi: 10.1097/dbp.0000000000000668.

[13] T. Akter *et al.*, "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders," *IEEE Access,* vol. 7,

pp. 166509-166527, 2019, doi: 10.1109/ACCESS.2019.2952609.

[14] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Medical Informatics and Decision Making,* vol. 19, no. 1, p. 281, 2019/12/21 2019, doi: 10.1186/s12911-019-1004-8.

[15] Y. Wang, C. Song, and S. T. Xia, "Improving decision trees by Tsallis Entropy Information Metric method," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 24-29 July 2016 2016, pp. 4729-4734, doi: 10.1109/IJCNN.2016.7727821.

[16] Y. Qi, "Random Forest for Bioinformatics," 2012.

[17] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning,* vol. 63, no. 1, pp. 3-42, 2006/04/01 2006, doi: 10.1007/s10994-006-6226-1.

[18] A. Ziegler and I. R. König, "Mining data with random forests: current options for real-world applications," *WIREs Data Mining and Knowledge Discovery,* vol. 4, no. 1, pp. 55-63, 2014/01/01 2014, doi: https://doi.org/10.1002/widm.1114.

[19] Fink, N. (2025). A Deep Learning and Machine Learning Ensemble Approach for Autism Screening with Over 99% Accuracy.