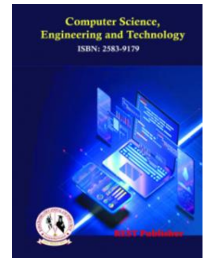
Computer Science, Engineering and Technology

Vol: 3(2), June 2025

REST Publisher; ISSN: 2583-9179

Website: https://restpublisher.com/journals/cset/

DOI: https://doi.org/10.46632/cset/3/2/7



Enhancing Autism Spectrum Disorder Detection Using Machine Learning Classifiers and Ensemble Models with Data Analytics and Interactive Dashboard Integration

*** Aparna Pandey, Urva Bhatnaggar, Udit Saran Bhatnagar, Mokshda Dubey**

Bhilai Institute of Technology, Raipur Chhattisgarh, India.

*Corresponding author Email: aparna95pandey@gmail.com

Abstract: *The neurological problem autism spectrum disorder, affecting many people globally, is multifarious in description; consequently, this makes quick detection of autism spectrum disorder (ASD) critical in the process of expert intervention. This work attempts to compare the use of multiple classifiers, logistic regression, k-nearest neighbor, and ensemble techniques such as random forest, CatBoost, and XGBoost. In the project, we have used Power BI visualizations to build a live interactive dashboard that will allow us to better understand the features present in the pre-acquired dataset and hence, help us with additional analysis and accuracy. Finally, the key findings of the project show that logistic regression did realize the test accuracy of 99.1 percent, while CatBoost, as well as random forest, got even better on a balanced dataset with the test accuracy of 99.5 percent and 99.6 percent, respectively. The results demonstrate that ensemble models are appropriate in labelling if someone has autism spectrum disorder (ASD). By conjoining analytical tools and machine learning approaches, one can presume smoother workability of project work for healthcare professionals, hence making the process comparatively less work intensive. Future work on the project may be to include advanced methodologies and to get somewhere close to developing an optimal model for clinical setups. This research work points out how amalgamating machine learning and data analytics could be beneficial for the advanced screening and diagnosis of ASD.*

Keywords: *Autism spectrum disorder (ASD) detection, ensemble learning Models, random forests, data analytics in healthcare, Power BI visualization.*

1. INTRODUCTION

Autism spectrum disorder (ASD) is a convoluted neurological condition demarcated by problems in social interaction and showcasing repetitive behaviors present uniquely across individuals [1]. Symptoms can emerge as early as 18 months, ranging from subtle behavioral indicators in toddlers to more pronounced cognitive and emotional patterns in older populations [2]. Initial signs often include late verbal expression, inadequate eye contact, reiterative actions, and difficulties adapting to change [3]. Given the spectrum nature of ASD, accurate detection across various age groups necessitates meticulous observation and thorough analysis. Understanding these varied manifestations is crucial for developing individualized interventions that assist individuals in navigating their distinct challenges and improving their overall quality of life. Early identification of ASD is instrumental in enhancing long-term outcomes by facilitating timely behavioral therapies, educational initiatives, and supportive care [3], [4]. However, conventional diagnostic methodologies often rely heavily on observational evaluations, clinical interviews, and developmental histories, which are labour-intensive and susceptible to biases and inconsistencies. Additionally, access to specialized practitioners is limited in many regions, leading to delays in crucial diagnoses [5]. These constraints highlight a significant gap in research. the need for scalable, data-centric diagnostic tools that can augment clinical discernment and optimize early identification processes [6]. In recent years, machine learning (ML) has emerged as a transformative advancement in healthcare diagnostics, providing the ability to detect patterns within extensive datasets that may escape human analysis [7], [8]. For instance, researchers have developed AI-based screening systems capable of identifying toddlers at elevated risk of autism. For children under two years old, these systems utilize machine learning to analyze data from large cohorts, focusing on measures such as age at first smile and eating habits. Initial tests have demonstrated the model's effectiveness in identifying ASD across various age groups, with peak accuracy observed in children aged two to four. Despite these promising results, experts caution against relying solely on these findings for early diagnoses, emphasizing that such algorithms should complement, not replace, traditional clinical methods when combined with advanced data visualization.

Platforms like Power BI ML can enhance interpretability and provide real-time insights for clinicians and caregivers. Despite this considerable potential, the integration of predictive models with user-friendly dashboards for ASD detection remains underexplored. This research aims to establish a machine learning-driven system for autism detection that integrates with an interactive Power BI dashboard, thereby enhancing diagnostic efficiency and accessibility. The goal is to develop a model that achieves optimal accuracy across selected algorithms and enables stakeholders to visually interpret patterns through an intuitive interface. The significance of this endeavour lies in its interdisciplinary approach, combining statistical modelling, advanced classification algorithms, and data visualization to bridge the gap between raw data and actionable insights by utilizing algorithms such as logistic regression, random forests, and CatBoost in conjunction with GridSearchCV for hyperparameter optimization. The study ensures optimal predictive efficacy. The incorporation of Power BI empowers end-users to dynamically engage with datasets, facilitating reasoned decision-making [9]. In conclusion, this study not only advances the field of ASD detection by providing a robust predictive framework but also demonstrates the practical applicability of machine learning in healthcare when integrated with user-centric analytical tools.

2. LITERATURE REVIEW

In current times, machine learning approaches are being explored to augment ASD detection, aiming to ameliorate efficacy and workability. This review examines the current landscape of ML, ensemble models, data analytics, and the integration of interactive dashboards in ASD detection. Multiple studies have investigated the application of ML classifiers for ASD diagnosis; for instance, researchers Parikh, Li et al. [10] employed ML models using personal characteristic data (PCD), such as age, sex, handedness, and IQ, for ASD classification. Their findings demonstrated that certain classifiers could efficaciously predict ASD, highlighting the potential of ML in capitalizing on readily available data for diagnostic purposes.

In another study, advanced ML techniques were utilized to enhance the accuracy and dependability of ASD diagnosis. The paper by Rony et al. [11] emphasized the importance of incorporating diverse data sources to improve predictive performance, suggesting that a multifaceted approach could yield more robust diagnostic models. Ensemble learning, which combines multiple classifiers to improve predictive performance, has shown promise in ASD detection.

A study by Alotaibi, Ali Alghamdi et al. [12] presented a two-tier metaheuristic-driven ensemble deep learning model for effective ASD diagnosis in disabled individuals. This approach integrated various deep learning techniques optimized through nature-inspired algorithms, resulting in enhanced diagnostic accuracy. Similarly, ensemble classification methods utilizing structural MRI data have been explored. Researchers Zhang-James, Buitelaar, Rooij et al. [13] developed an ensemble ML pipeline that included multiple base models followed by ensemble techniques to improve performance. This approach demonstrated the potential of conjoining structural neuroimaging data with ensemble learning for ASD classification.

Integrating data analytics with interactive dashboards can facilitate the visualization and recognition of complex datasets, aiding clinicians in the diagnostic process. The development of interactive dashboards for analyzing ASD data using ML techniques has been proposed to enhance user engagement and decision-making in the paper by G & Ramasubramanian [14]. Such dashboards can provide realtime insights, making the diagnostic process more efficient. Saha et al. [15] created a dashboard that uses machine learning to analyze data on autism spectrum disorder and improve accuracy. Along with a dataset of 1054 cases and 19 toddler-related factors, the researchers used a variety of algorithms. All things considered; the study makes a substantial contribution to the profession by providing a practical answer to the issues that arise when diagnosing ASD.

The paper by Talukdar, Gogoi & Singh [16] provides a comprehensive evaluation of various machine learning classifiers, including Naive Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF). This comparative approach helps identify the most effective methods for ASD detection. The study reports significant accuracy levels achieved by different classifiers, with Random Forest demonstrating the highest accuracy of 93.69% for toddlers and 93.33% for adolescents. The paper outlines a clear methodological framework for applying machine learning techniques to ASD classification. This includes data preprocessing steps, model training, and performance assessment metrics, providing a valuable reference for future research in this area.

The work by Akter et al. [17] improves the efficacy of classification algorithms in several significant ways, advancing the field of autism diagnosis using machine learning. The study employs a number of featureengineering methods, including log z-score and sine functions. The results show how these steps may proportionately improve the accuracy of ASD diagnostic models. Highlighting the value of the sine function for toddler support vector machines and the z-score for children, according to the results, SVM, AdaBoost, and GLMBoost are assessed in this study to see how well they forecast ASD. While SVM was the greatest choice for toddlers, AdaBoost fared better for children and teens. The paper by Omar, Mondal, Khan, Rizvi, & Islam [18] presents several significant contributions to the field of autism detection and prediction. The research proposes an effective prediction model for autism spectrum disorder (ASD). This model merges two algorithms: Random ForestCART (Classification and Regression Trees) and Random Forest-ID3 (Iterative dichotomiser 3), which enhances the accuracy of autism trait predictions.

3. METHODOLOGY

Fig. 1. provides a structured overview of the methodology used for the detection of Autism Spectrum Disorder (ASD) through a data-driven approach. The process begins with an introduction to ASD detection, supported by preliminary research and a feasibility study. Upon continuation, the methodology is initiated through several systematic stages.
The first stage involves the importation of essential libraries, including NumPy, Pandas, Seaborn, and Scikitlearn, which provide the foundational tools for data analysis and machine learning. Data acquisition follows, sourcing relevant datasets from platforms like Kaggle and integrating Power BI for dashboard visualization. The live dashboard makes it much easier to understand the different features of the dataset acquired for analysis. Exploratory Data Analysis (EDA) is then conducted to understand data structures, identify missing or duplicate values, and analyze categorical data patterns. This is succeeded by data preprocessing, which includes addressing class imbalance, managing missing data, label encoding, dropping duplicates, and standardizing features.

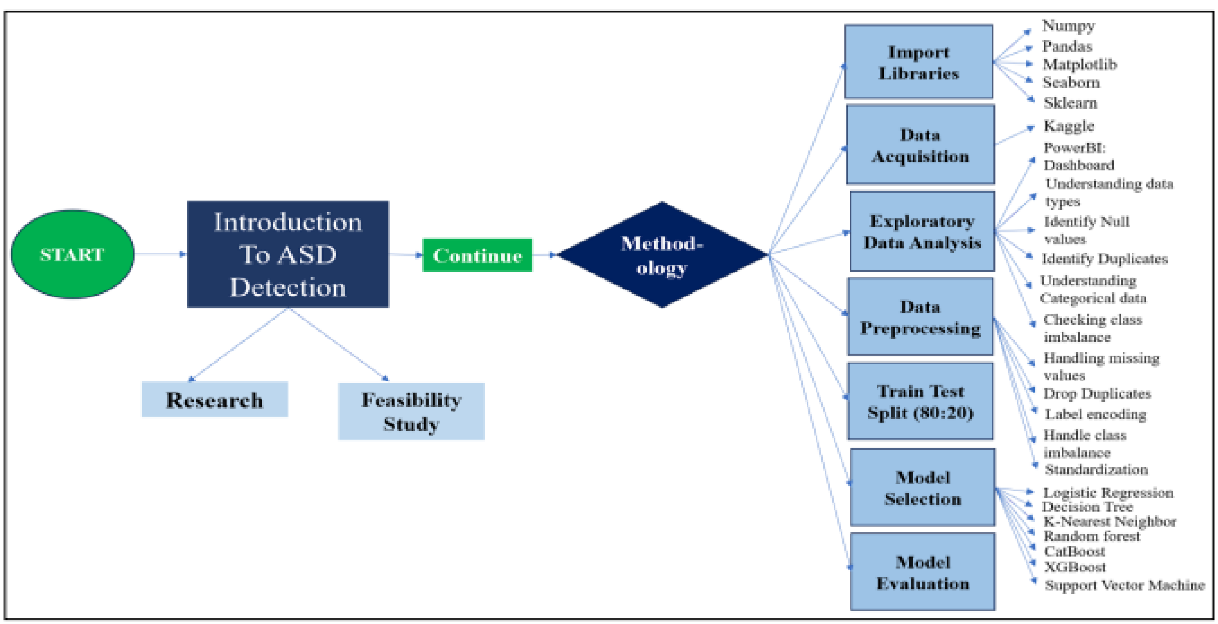


FIGURE 1. ASD workflow diagram

A. Dataset: Overview and Power BI Integration The study employs a publicly available dataset comprising 1,128 instances, each representing responses to an autism spectrum disorder (ASD) screening questionnaire for individuals aged 18 and above. The dataset includes 22 attributes, encompassing ten binary screening scores (A1– A10), demographic variables (e.g., age group, gender, ethnicity, country of residence), and relevant medical and familial indicators (e.g., history of jaundice, familial ASD, prior app usage, respondent relation).

The target variable, Class/ASD, is binary (“YES”/ “NO”), enabling supervised classification tasks. The dataset’s heterogeneity supports model generalization across diverse populations. The subsequent Power BI dashboard is shown in Fig.2.

FIGURE 2. ASD Dashboard

B. Model Selection and Training The dataset is subsequently partitioned into training and testing subsets using an 80:20 split. The model selection phase applies various machine learning classifiers such as logistic regression, decision tree, K-nearest neighbor, random forest, CatBoost, XGBoost, and support vector machine (SVM), ensuring a comprehensive comparison of algorithm performance. Finally, model evaluation is performed to assess the accuracy and reliability of the predictive models. Metrics such as precision, recall, F1- score, and overall accuracy are likely utilized to determine the most effective classifier, with the goal of deploying a robust and interpretable model for ASD detection supported by visual insights from the Power BI dashboard.

4. RESULT AND DISCUSSION

Case 1: Here, we have taken an imbalanced dataset having 1028 data points and 22 features, which has been cleaned and transformed using various methods like standardization, handling categorical variable using label encoding and feature selection. There is an imbalance in the target class of the dataset showcasing 721 instances of ‘yes’ in Class/ASD column and 399 instances of ‘no’ in Class/ASD target column. Upon applying various machine learning algorithms like support vector machine, logistic regression, K-Nearest Neighbors, decision tree, random forest, CatBoost and XGBoost the following results were obtained which are summarized in Table. I.

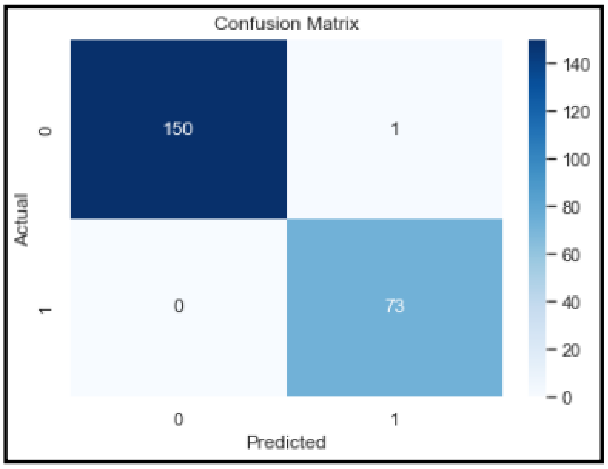


FIGURE 3. Confusion matrix logistic regression.

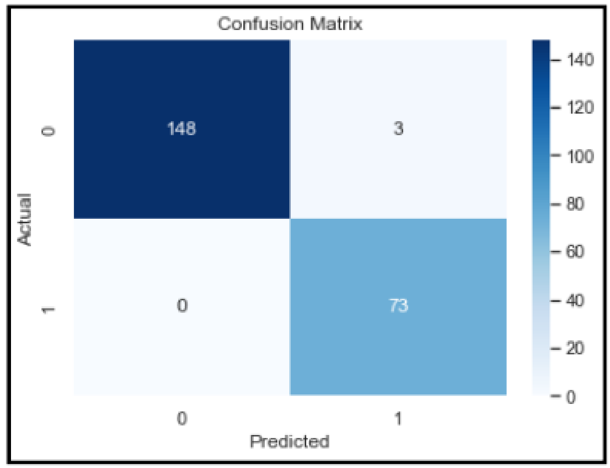
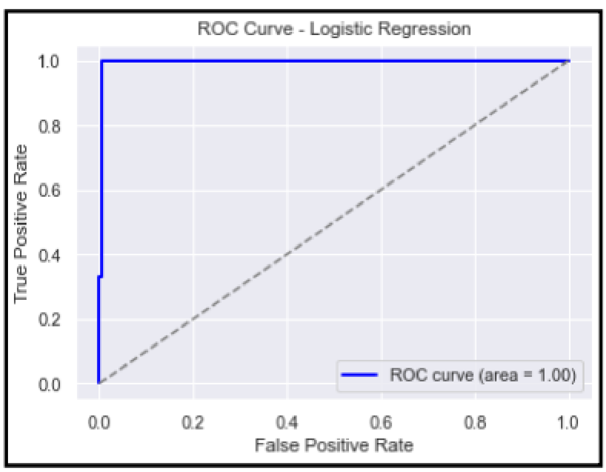


FIGURE 4. Confusion matrix random forest

Fig. 3. and Fig. 4. demonstrate the confusion matrix for the algorithms logistic regression and random forest, respectively.



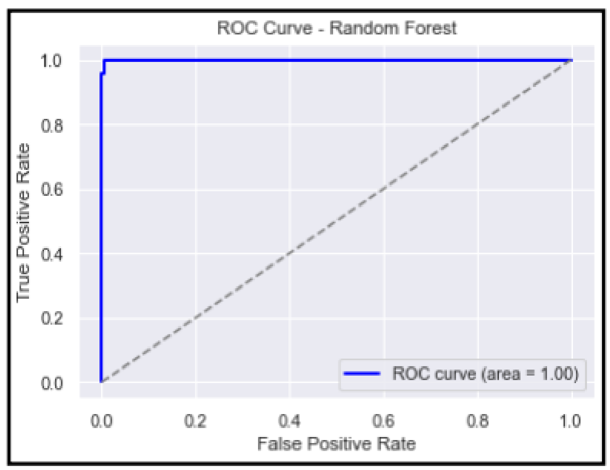


FIGURE 5. Receiver operating curve logistic regression **FIGURE 6.** Receiver operating curve random forest

Fig. 5. and Fig. 6. demonstrate the Receiver Operating Characteristic (ROC) Curve for the algorithms, logistic regression and random forest, respectively.

Case 2: On balancing the dataset using the Synthetic Minority Oversampling Technique combined with the Edited Nearest Neighbors (SMOTEENN) algorithm, we get an improved performance matrix for algorithms Random Forest, CatBoost, and XGBoost, realizing 99.6 percent, 99.5 percent, and 99.5 percent test accuracy, respectively. The results so obtained are tabulated in Table. II.

Fig. 7 and Fig. 8 demonstrate the confusion matrices for logistic regression and random forest, respectively. Fig. 9, Fig. 10, Fig. 11, and Fig. 12 illustrate the receiver operating characteristic (ROC) curves for logistic regression, random forest, CatBoost and XGBoost respectively.

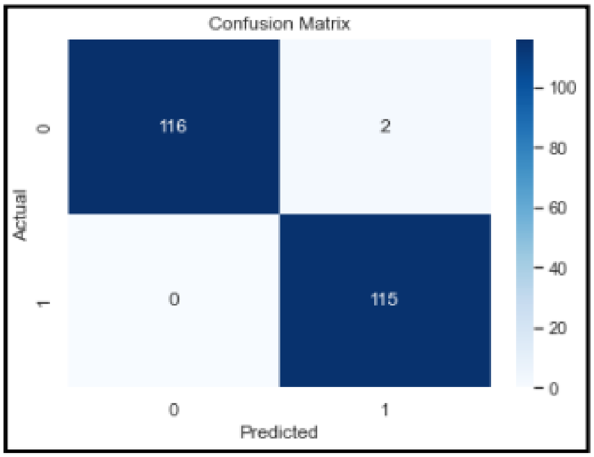


FIGURE 7. Confusion Matrix logistic regression balanced

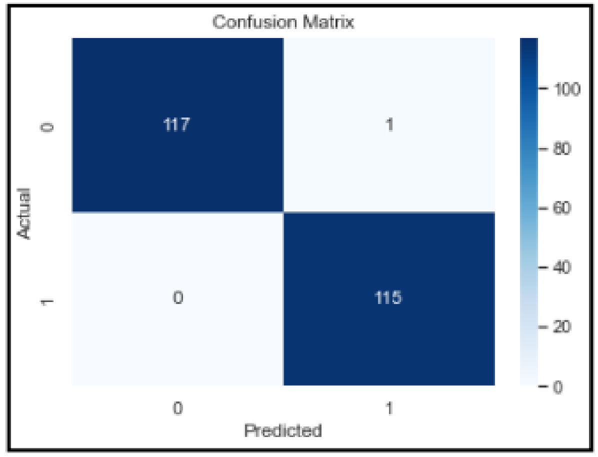


FIGURE 8. Confusion matrix random forest balanced

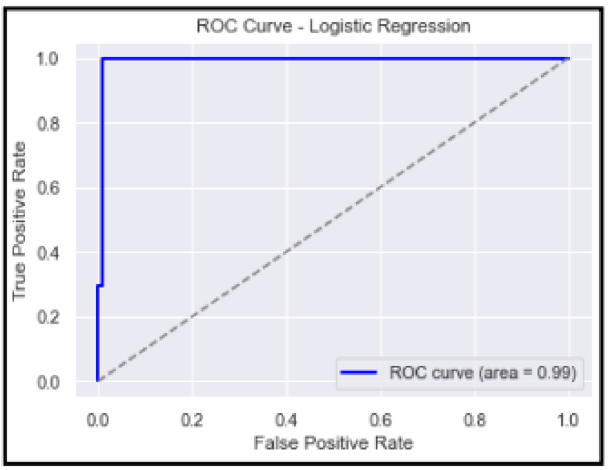


FIGURE 9. Receiver operating characteristic logistic regression balanced,

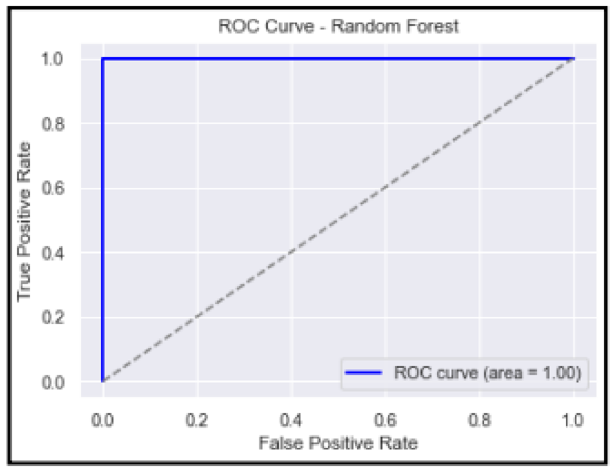
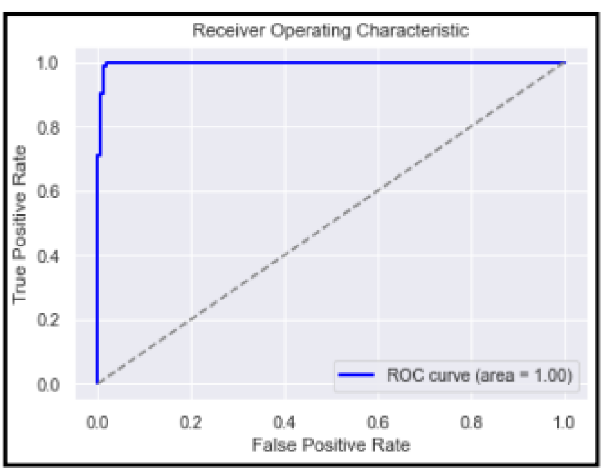


FIGURE 10. Receiver operating characteristic random forest balanced



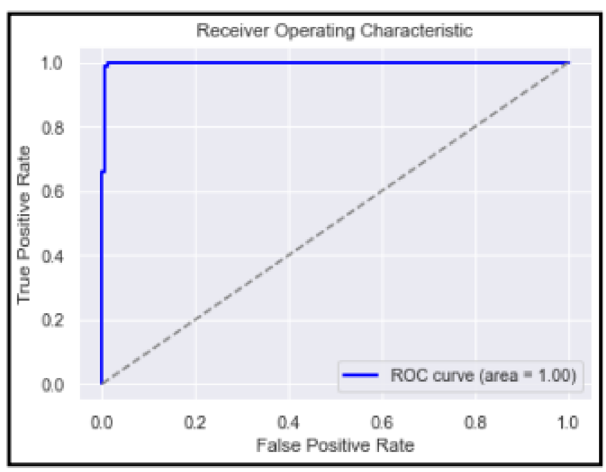


FIGURE 11. Receiver Operating Characteristic CatBoost Balanced

FIGURE 12. Receiver Operating Characteristic XGBoost Balanced

5. CONCLUSION

This study provides a detailed approach to autism spectrum disorder (ASD) detection by leveraging machine learning and ensemble learning techniques on a combined dataset through rigorous experimentation and evaluation. The random forest classifier demonstrated superior performance, achieving the highest accuracy of 99.6 percent on a balanced dataset. This underlines its robustness and reliability in effectively identifying ASD traits. The integration of Power BI for live interactive dashboard visualizations further enhances the interpretability and applicability of the model, offering an intuitive platform for clinicians. The convergence of high-performing algorithms with dynamic visualization tools not only facilitates accurate diagnostic support but also bridges the gap between complex machine learning models and user accessibility. The findings substantiate the potential of ensemble techniques in clinical decision support systems.

TABLE 1. Result for Imbalanced Dataset

Algorithms		Test Accuracy	Precision		Recall		F1 Score		ROC-AUC Score
			0	1	0	1	0	1	
Basic Classifiers	Support Vector Machine	99.1 %	1.00	0.97	0.99	1.00	0.99	0.99	0.9933774834437087
	Logistic Regression	99.6%	1.00	0.99	0.99	1.00	1.00	0.99	0.9966887417218543
	K Nearest Neighbors	96.9%	0.99	0.92	0.96	0.99	0.98	0.95	0.9732831352626327
	Decision Tree	98.2%	0.99	0.96	0.98	0.99	0.99	0.97	0.9832169100970698
Ensemble Classifiers	Random Forest	98.7 %	1.00	0.96	0.98	1.00	0.99	0.98	0.990066225165563
	CatBoost	98.7%	1.00	0.96	0.98	1.00	0.99	0.98	0.990066225165563
	XGBoost	98.6%	1.00	0.96	0.98	1.00	0.99	0.98	0.990066225165563

TABLE 2. Result for Balanced Dataset

Algorithms		Test Accuracy	Precision		Recall		F1 Score		ROC-AUC Score
			0	1	0	1	0	1	
Basic Classifiers	Support Vector Machine	99.1 %	1.00	0.98	0.98	1.00	0.99	0.99	0.9915254237288136
	Logistic Regression	99.1%	1.00	0.98	0.98	0.99	0.99	0.99	0.9915254237288136
	K Nearest Neighbors	96.9%	0.99	0.97	0.97	0.99	0.98	0.98	0.9829403095062638
	Decision Tree	99.1%	1.00	0.98	0.98	1.00	0.99	0.99	0.9915254237288136
Ensemble Classifiers	Random Forest	99.6 %	1.00	0.99	0.99	1.00	1.00	1.00	0.9957627118644068
	CatBoost	99.5%	1.00	0.99	0.99	1.00	1.00	1.00	0.9957627118644068
	XGBoost	99.5%	1.00	0.99	0.99	1.00	1.00	1.00	0.9957627118644068

Further research can explore the deployment of these models in real-world clinical settings and integration with electronic health records. (EHRs) could significantly enhance diagnostic workflows. Expanding the dataset diversity and including behavioral genetic and imaging data could also refine prediction granularity, paving the way for more personalized ASD intervention strategies.

Acknowledgment: We would like to sincerely thank Aparna Pandey, Assistant Professor at the Bhilai Institute of Technology Raipur, for her invaluable advice, inspiration, and unwavering support over the course of this study. Her input significantly influenced the focus and scope of this research work. We also like to express our gratitude to the Department of Computer Science & Engineering's faculties and staff for providing the required resources and academic setting. Lastly, we would like to express our gratitude to the academics and developers whose work on Machine Learning models and AI in autism detection served as the basis for our work's motivation and basic knowledge.

REFERENCES

[1]. T. Amy and D. Katerina, "The Emergence of Autism Symptoms Prior to 18 Months of Age: A Systematic Literature Review," Journal of Autism and Developmental disorders, vol. 51, pp. 973-993, 2021.
[2]. H. Doi, N. iijima, A. Furui, Z. Soh, R. Yonei, K. Shinohara, M. Iriguchi, K. Shimatani and T. Tsuji, "Prediction of autistic tendencies at 18 months of age via markerless video analysis of spontaneous body movements in 4-month-old infants," Nature, 2022.
[3]. T. Abualait, M. Alabbad, I. Kaleem, H. Imran, H. Khan, M. M. Kiyani and S. Bashir, "Autism Spectrum Disorder in Children: Early Signs and Therapeutic Interventions," MDPI, vol. 11, no. 11, 2024
[4]. G. Dason, "Early behavioral intervention, brain plasticity,and the prevention of autism spectrum disorder," Development and psychopathology, vol. 20, pp. 775-803, 2008.
[5]. M. Elbattah, O. A. S. Ibrahim and G. Dequen, "Editorial: Improving autism spectrum disorder diagnosis using machine learning techniques," Front.Neuroinform., 2024.
[6]. M. S. Durkin, M. Elsabbagh, J. Barbaro, M. Gladstone, F. Happe, R. A. Hoekstra, L.-C. Lee, A. Rattazzi, J. Stapel-ax, e. L. Stone, H. Tager-Flusberg, A. Thrum, M. Tomlinson and A. Shih, "Autism screening and diagnosis in low resource settings: Challenges and opportunities to enhance research and services worldwide," Autism Res., pp. 473-6, 2015.
[7]. R. Jeyarani and R. Senthilkumar, "Eye Tracking Biomarkers for Autism Spectrum Disorder Detection using Machine Learning and Deep Learning Techniques: Review," Research in Autism Spectrum Disorders, vol. 108, 2023.
[8]. R. A. Rasul, P. Saha, D. Bala, S. R. U. Karim, M. I. Abdullah and B. Saha, "An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder," Healthcare Analytics, vol. 5, 2024.
[9]. B. H. A. Halim and Maryam, "Enhancing Healthcare Transparency: Leveraging Machine Learning, GIS Mapping and Power BI for Private Hospital Insurance Claims Analysis," figshare, 2025.
[10]. M. N. Parikh, H. Li and L. He, "Enhancing Diagnosis of Autism With Optimized Machine Learning Models and Personal Characteristic Data," FrontComput Neurosci., 2019.
[11]. M. A. T. Rony, F. T. Johora, N. Thaliji, A. Raza, N. L. Fitriyani, M. Syafrudin and S. W. Lee, "Innovative Approach to Detecting Autism Spectrum Disorder Using Explainable Features and Smart Web Application," MDPI, vol. 12, no. 22, 2024
[12]. S. S. Alotaibi, T. AliAlghamdi and R. Alharthi, "Two-tier nature inspired optimization-driven ensemble of deep learning models for effective autism spectrum disorder diagnosis in disabled persons," nature, 2025.
[13]. Y. Zhang-James, J. K. Buitelaar, D. v. Rooij and S. V. Faraone, "Ensemble classification of autism spectrum disorder using structural magnetic resonance imaging features," JCPP Advances, 2021.
[14]. D. V. G and C. Ramasubramanian, "Optimized Machine Learning Classification Approaches for Prediction of Autism Spectrum Disorder," Annals of Autism & Developmental Disorders, vol. 1, no. 1, 2020.
[15]. A. Saha, D. Barua, M. Mishu and Z. Mohib, "Development of an Interactive Dashboard for Analyzing Autism Spectrum Disorder (ASD) Data using Machine Learning," International Journal of Information technology and Computer Science, vol. 4, pp. 14-24, 2022.
[16]. J. Talukdar, D. K. Gogoi and T. P. Singh, "A comparative assessment of most widely used machine learning classifiers or analysing and classifying autism spectrum disorder in toddlers and adolescents.," Healthcare Analytics, pp. 2772-4425, 2023.
[17]. T. Akter, M. S. Satu, M. I. Khan, M. H. Ali, S. Uddin, P. Lio, J. M. W. Quinn and M. A. Moni, "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders," IEEE Access, 2019.
[18]. K. S. Omar, P. Mondal, N. S. Khan, M. R. K. Rizvi and M. N. Islam, "A Machine Learning Approach to Predict Autism Spectrum Disorder," International Conference on Electrical, Computer and Communication Engineering (ECCE), 2019.