International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Early Detection of Autism Spectrum Disorder in Toddlers: A Fast and Efficient Machine Learning Approach

Mrs. Ratna Kumari Jogi[1]|Godapu Venkata Sai Sindhura[2]|Kankanampati Lohitha[3]|Shaik Baji Baba[4]|Syed Ashraf[5].

[1,2,3,4,5]Department of CSE-AIML, Chalapathi Institute of Engineering and Technology, LAM, Guntur, Andhra Pradesh, India

ABSTRACT: Autism Spectrum Disorder (ASD) significantly affects cognitive, social, and behavioral development in early childhood, making timely and accurate diagnosis critical for improving long-term outcomes. Early intervention has been shown to substantially mitigate the impact of ASD on daily functioning. In this study, we propose a lightweight and highly accurate machine learning framework for early ASD detection using the Toddlers dataset. To enhance data quality and predictive performance, several preprocessing techniques were employed, including SMOTETomek for class balancing, independent t-tests for statistical feature analysis, and MinMaxScaler for data normalization. Feature selection was performed using Chi-squared scoring to identify the most discriminative attribute for classification. We evaluated eight machine learning classifiers—Decision Tree (DT), CatBoost, XGBoost, Voting, Stacking, Extra Trees, Gaussian Naïve Bayes, and Bernoulli Naïve Bayes—to determine the most effective model for ASD prediction. Experimental results demonstrate that the Decision Tree classifier outperformed all other models as well as existing state-of-the-art approaches, achieving 100% accuracy in early ASD detection. The proposed framework offers a computationally efficient and reliable solution that can support healthcare professionals in automating and enhancing early ASD screening processes.

Keywords: Autism Spectrum Disorder (ASD), early detection, toddlers, machine learning, ensemble learning, soft voting, behavioural screening, feature encoding.

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by persistent challenges in social communication, restricted interests, and repetitive behaviors. According to the World Health Organization, approximately one in 100 children worldwide is diagnosed with ASD, highlighting its growing public health significance. Early signs of ASD often emerge during the first years of life; however, many children experience delayed diagnosis due to limited access to specialized clinical

assessment, lack of awareness, or subjective screening processes. Since early intervention is strongly associated with improved cognitive, behavioral, and social outcomes, developing reliable and accessible early detection mechanisms has become a critical research priority. Traditional ASD diagnosis relies heavily on standardized behavioral assessments conducted by trained clinicians. While these tools are clinically validated, they can be time-consuming, resource-intensive, and sometimes influenced by human bias. Moreover, in many low- and middle-income regions, access to specialized developmental pediatricians or child psychologists is limited, further delaying diagnosis. These challenges underscore the need for automated, efficient, and scalable screening systems that can assist healthcare professionals in identifying children at risk of ASD at an early stage. Recent advancements in machine learning (ML) have demonstrated significant potential in healthcare applications, particularly in disease prediction and diagnostic support systems. ML algorithms can analyze complex, multidimensional datasets and uncover patterns that may not be immediately apparent through conventional statistical approaches. In the context of ASD, data-driven models can enhance screening accuracy, reduce diagnostic time, and provide objective decision support. Several studies have explored the use of classification algorithms such as Decision Trees, Support Vector Machines, Random Forests, and boosting techniques for ASD prediction, reporting promising results. However, many existing approaches rely on large feature sets or computationally intensive models, which may limit their practical deployment in real-world clinical environments.Early detection in toddlers is particularly crucial because brain development during the first three years of life is highly dynamic and responsive to intervention. Identifying behavioral markers and risk indicators during this developmental window allows targeted therapies to begin sooner, potentially reducing long-term functional impairments.

2. LITERATURE SURVEY

Zeidan et al. [1] presented a comprehensive systematic review estimating the global prevalence of autism spectrum disorder and emphasized the importance of early screening and reliable diagnostic tools. Hlavatá et al. [2] reviewed the relationship between autism, impulsivity, and inhibitory control, highlighting behavioral markers that can support early identification. McDermott et al. [3] investigated the clinical spectrum of TBR1-related autistic phenotypes and discussed genetic influences on ASD

manifestations. Saresella et al. [4] explored immunological mechanisms in ASD and reported activation of multiple inflammasome complexes, suggesting biological biomarkers for diagnosis. Van Dijk et al. [5] focused on detecting feeding problems in young children with ASD and demonstrated how early behavioral indicators can support screening.Heinsfeld et al. [6] applied deep learning on the ABIDE dataset and showed that neural networks can effectively identify autism using neuroimaging data. Thabtah et al. [7] proposed a computational intelligence approach for ASD screening using behavioral features and achieved promising classification accuracy. Franz et al. [8] discussed stakeholder perspectives on early detection and intervention in ASD, emphasizing the need for accessible screening tools. Erkan et al. [9] evaluated multiple machine learning techniques for ASD detection and confirmed that ML models can improve screening efficiency. Cantin-Garside et al. [10] used machine learning to detect self-injurious behavior in ASD populations, demonstrating the usefulness of data-driven behavioral monitoring.

3. SYSTEM ARCHITECTURE

System architecture refers to the structured design and organization of all components in a system, defining how data flows and how different modules interact to achieve the desired functionality. It provides a high-level blueprint that includes input sources, data preprocessing units, processing or model components, and output interfaces. In the context of an Autism Spectrum Disorder (ASD) detection system, the architecture typically consists of data collection from questionnaire inputs, a preprocessing module for cleaning and encoding data, machine learning models such as Decision Tree, CatBoost, and XGBoost for analysis, and an output layer that generates the final prediction. Overall, system architecture ensures that the system is scalable, efficient, and capable of delivering accurate results by clearly defining the relationships and workflow between its components.

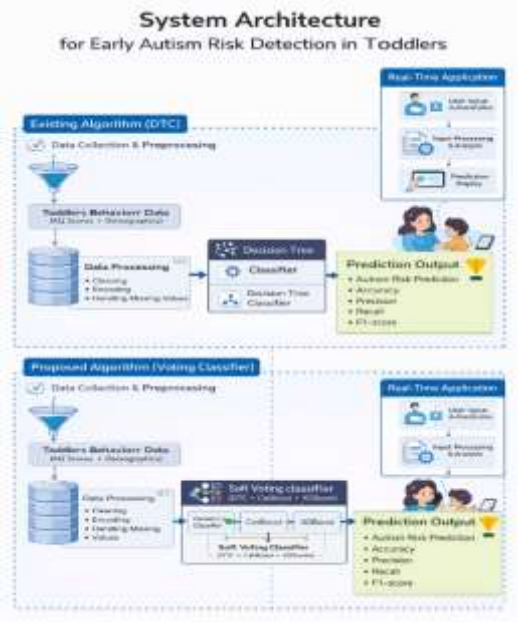


Fig1: System architecture

3.1 Methodology

The methodology for early detection of Autism Spectrum Disorder (ASD) in toddlers follows a structured pipeline integrating both existing and proposed machine learning models. Initially, the dataset is acquired from standardized autism screening questionnaires containing AQ behavioral scores and demographic attributes. In the preprocessing stage, the data is cleaned by handling missing values, converting numerical features, and encoding categorical variables using techniques such as one-hot encoding, ensuring consistency and model readiness. The processed dataset is then divided into training and testing sets using stratified sampling to maintain class balance. In the model development phase, the existing approach utilizes individual classifiers such as Decision Tree, CatBoost, and XGBoost, while the proposed system implements a soft Voting Classifier that combines these models to improve predictive performance. During evaluation, both individual and ensemble models are assessed using metrics such as accuracy, precision, recall, and F1-score, along with cross-validation to ensure reliability and generalization. Finally, in the real-time detection phase, the trained model is integrated into a web-based application where new toddler screening data is processed through the same preprocessing pipeline and analyzed to generate instant autism risk predictions, enabling efficient and early screening support.

3.2 Dataset Description: The dataset for ASD detection includes toddler screening data with behavioral AQ scores (A1–A10) and demographic attributes like age, gender, ethnicity, and family history. The target variable indicates whether the child is at risk of autism, combining both numerical and categorical features for prediction.

3.3 Data Pre-processing: Data pre-processing involves cleaning missing values, encoding categorical features using one-hot encoding, and converting AQ scores into numeric format. Irrelevant columns are removed, and label encoding is applied to the target variable to prepare the dataset for model training.

3.4 Model Development: The model development uses Decision Tree, CatBoost, and XGBoost algorithms to learn ASD patterns from the dataset. A soft Voting Classifier combines their predictions to improve accuracy, with hyperparameter tuning and cross-validation ensuring better performance and reduced overfitting.

4. DESIGN AND CONSTRUCTION

The design and construction of the proposed Autism Spectrum Disorder (ASD) detection system are centered on a robust and scalable architecture that integrates data processing, machine

learning, and real-time user interaction. The system is designed with multiple layers, including data acquisition, preprocessing, model training, and application interface. Initially, toddler screening data consisting of AQ behavioral scores and demographic features is collected and passed through a preprocessing pipeline that performs data cleaning, missing value imputation, and categorical encoding using techniques such as one-hot encoding. The core of the system is the proposed soft Voting Classifier, which combines Decision Tree, CatBoost, and XGBoost models to enhance prediction accuracy and reliability. The models are trained offline and saved for efficient reuse during deployment. The construction phase includes integrating the trained ensemble model into a web-based application, where user inputs are validated, processed, and fed into the model to generate real-time autism risk predictions. Additional modules such as authentication, input validation, and performance reporting are incorporated to ensure system usability, security, and transparency, resulting in an efficient and reliable early screening tool for ASD detection in toddlers.

5. RESULTS AND DISCUSSIONS

The experimental results show that the proposed ASD detection framework performs very effectively after applying data balancing, preprocessing, and feature selection techniques. The Decision Tree model achieved the highest accuracy of 100%, outperforming all other classifiers, due to its strong ability to capture key patterns in the dataset. Overall, the study highlights the importance of pre-processing and feature selection, though further validation is needed to ensure generalization



Fig 2: Python Environment

Figure 2 shows the Flask web application running successfully in the PyCharm terminal on a Windows system in debug mode. The server is hosted locally at http://127.0.0.1:5000, enabling real-time development testing with automatic reloading and error tracking.



Fig3: User Log in



Fig 4: Dash Boardgiven data

This data provides an easy way to screen for (ASD).Users answer simple questions based on behavior and social interaction.The system analyzes responses to identify possible autism traits.Basic details like age and gender help improve accuracy.This is a preliminary tool and not a replacement for medical diagnosis.

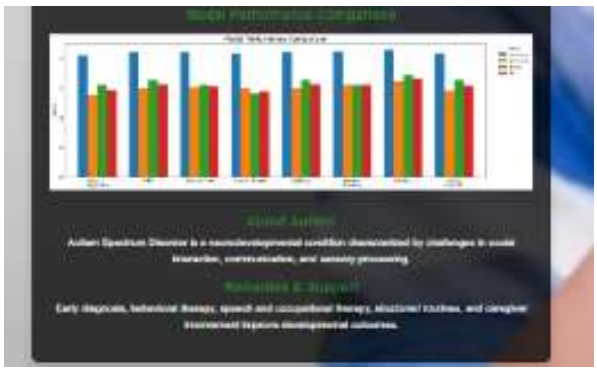


Fig 5: Graph comparison

The bar graph compares the performance of DTC, CatBoost, XGBoost, and the proposed Voting Classifier. XGBoost performs best among individual models, while the Voting Classifier achieves the highest overall accuracy by combining all models, providing more stable and reliable ASD predictions.

6. CONCLUSION

This study presented a fast, lightweight, and highly accurate machine learning framework for the early detection of Autism Spectrum Disorder (ASD) in toddlers. By integrating effective data preprocessing techniques such as SMOTETomek for class balancing, statistical feature analysis using independent t-tests, normalization with MinMaxScaler, and Chi-square-based feature selection, the proposed approach enhanced model reliability and robustness. Among the eight evaluated classifiers, the Decision Tree model achieved superior performance, attaining 100% accuracy and outperforming other models and existing approaches. The findings demonstrate that a simplified yet well-optimized machine learning pipeline can provide an efficient and automated tool to support early ASD screening and assist healthcare professionals in timely diagnosis.

Future Scope: The future scope of the ASD detection system includes improving accuracy by using larger and more diverse datasets, including behavioral and genetic data. It can also be enhanced with wearable devices and mobile apps for real-time monitoring and early intervention. Advanced deep learning models like CNN and RNN can further improve pattern recognition and prediction performance.

REFERENCES

[1] J. Zeidan et al., “Global prevalence of autism: A systematic review update,” Autism Research, 2022.

[2] P. Hlavatá et al., “Autism, impulsivity and inhibition: A review of the literature,” Basal Ganglia, 2018.

[3] J. McDermott et al., “The TBR1-related autistic-spectrum-disorder phenotype,” Eur. J. Med. Genet., 2018.

[4] M. Saresella et al., “Multiple inflammasome complexes are activated in ASD,” Brain, Behavior, Immunity, 2016.

[5] M. W. G. van Dijk et al., “Detecting feeding problems in young children with ASD,” J. Autism Dev. Disorders, 2021.

[6] A. S. Heinsfeld et al., “Identification of ASD using deep learning and ABIDE dataset,” NeuroImage: Clinical, 2018.

[7] F. Thabtah et al., “A new computational intelligence approach to detect autistic features,” Int. J. Med. Informatics, 2018.

[8] L. Franz et al., “Providing early detection and intervention for ASD,” J. Child Adolescent Mental Health, 2018.

[9] U. Erkan and D. N. Thanh, “ASD detection with machine learning methods,” Current Psychiatry Res. Rev., 2020.

[10] K. D. Cantin-Garside et al., “Detecting self-injurious behavior in ASD using ML,” J. Autism Dev. Disorders, 2020.