International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders

A.T. Ligin

Student, Master of Computer Application, Dr.M.G.R. Educational and Research Institute, Chennai, Tamil Nadu, India.

ABSTRACT

This research paper discusses a machine learning framework for early detection of Autism Spectrum Disorder (ASD) across different age groups. Using four feature scaling methods and eight classifiers, AdaBoost and LDA achieved the highest accuracies on toddler, child, adolescent, and adult datasets. Feature selection techniques identified key ASD risk factors. Results show that well-tuned, simple ML models can effectively support ASD diagnosis.

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by a wide range of behavioral and cognitive challenges, making early diagnosis difficult yet crucial for effective intervention. Traditional diagnostic methods often rely on subjective assessments, which can delay timely identification and treatment. To address these challenges, this study proposes a machine learning-based framework that combines label encoding, Recursive Feature Elimination (RFE), and Logistic Regression to enhance the accuracy and efficiency of early ASD detection. By automating feature selection and classification, the framework aims to provide a reliable and objective tool to support clinicians in the early identification of ASD, ultimately improving patient outcomes.

2. EXISTING SYSTEM

This research aims to create a prediction model for autism detection across different ages using various machine learning methods. The datasets undergo preprocessing, including mean value imputation for missing data, One Hot Encoding for categorical features, and Random Over Sampling to address class imbalance. Feature scaling is then applied using four techniques: Quantile Transformer, Power Transformer, Normalizer, and Max Abs Scaler. Finally, eight classifiers AdaBoost, Random Forest, Decision Tree, K-Nearest Neighbors, Gaussian Naive Bayes, Logistic Regression, Support Vector Machine, and Linear Discriminant Analysis are used to evaluate model performance.

3. PROPOSED SYSTEM

This research aims to create an effective prediction model using different types of ML methods to detect autism in people of different ages. First of all, the datasets are collected, and then the preprocessing is accomplished the missing values imputation, Label encoding, and oversampling and Create an instance of RFE with the classifier and the desired number of features to select Logistic Regression classification of modeling, performance evaluation, and the results with improved accuracy.

4. MODULE DESCRIPTION

SYSTEM MODULES:

Module 1: Data Pre-Processing
Module 2: Algorithm Implementation
Module 3: Prediction

Module 1: Data Pre-Processing:
Four ASD datasets Toddlers, Children, Adolescents, and Adults are sourced from Kaggle, the UCI ML Repository, and the ASD Tests smartphone app. These datasets use QCHAT-10 and AQ-10 screening tools, where a score of 6 or above (out of 10) indicates a positive ASD case. Data from the app also supports open-source databases to facilitate ASD research. A brief description of each dataset is provided.

Module 2: Algorithm Implementation:
Three classification algorithms KNN, Logistic Regression, and Random Forest are used to predict ASD. Logistic Regression shows the highest efficiency in the analysis. Classifiers are applied to each clustered dataset, and the best models are identified based on low error rates.

Module 3: Prediction:
Several standard performance metrics such as accuracy, precision and error in classification have been considered for the computation of performance efficacy of this model. Preprocessed data are trained and input given by the user goes to the trained dataset

5. LITERATURE SURVEY

1. Title: Automated Autism Spectrum Disorder Detection Using Support Vector Machine

Author: Datcu, D., David, O., David, I. (2017)

Description: This study applied Support Vector Machines (SVM) to neuroimaging data for ASD detection. It demonstrated promising accuracy but faced challenges related to high dimensionality and limited model interpretability.

2. Title: Autism Spectrum Disorder Screening Using Behavioral Data: A Rule-Based Approach

Author: Thabtah, F., Peebles, D. (2019)

Description: The authors utilized decision trees and rule-based classifiers on behavioral datasets, emphasizing model explainability to assist clinical decisions. Their findings suggest that simpler models often offer better transparency, important for medical diagnostics.

3. **Title:** Recursive Feature Elimination and Random Forest-Based Autism Spectrum Disorder Classification

Author: Muneer, S., Ali, S., Imran, M. (2020)

Description: This work combined Recursive Feature Elimination (RFE) with Random Forest classifiers across multi-age datasets, resulting in improved accuracy and generalizability by selecting the most relevant features.

4. **Title:** Predicting Autism Spectrum Disorder with Logistic Regression: Insights and Interpretability

Author: Pouriyeh, S., Saha, S., Caragea, D. (2021)

Description: Focusing on Logistic Regression for ASD prediction, this study highlighted its effectiveness in binary classification and its ability to provide interpretable insights into feature importance

6. DESIGN AND IMPLEMENTATION

1. ARCHITECTURE DIAGRAM:

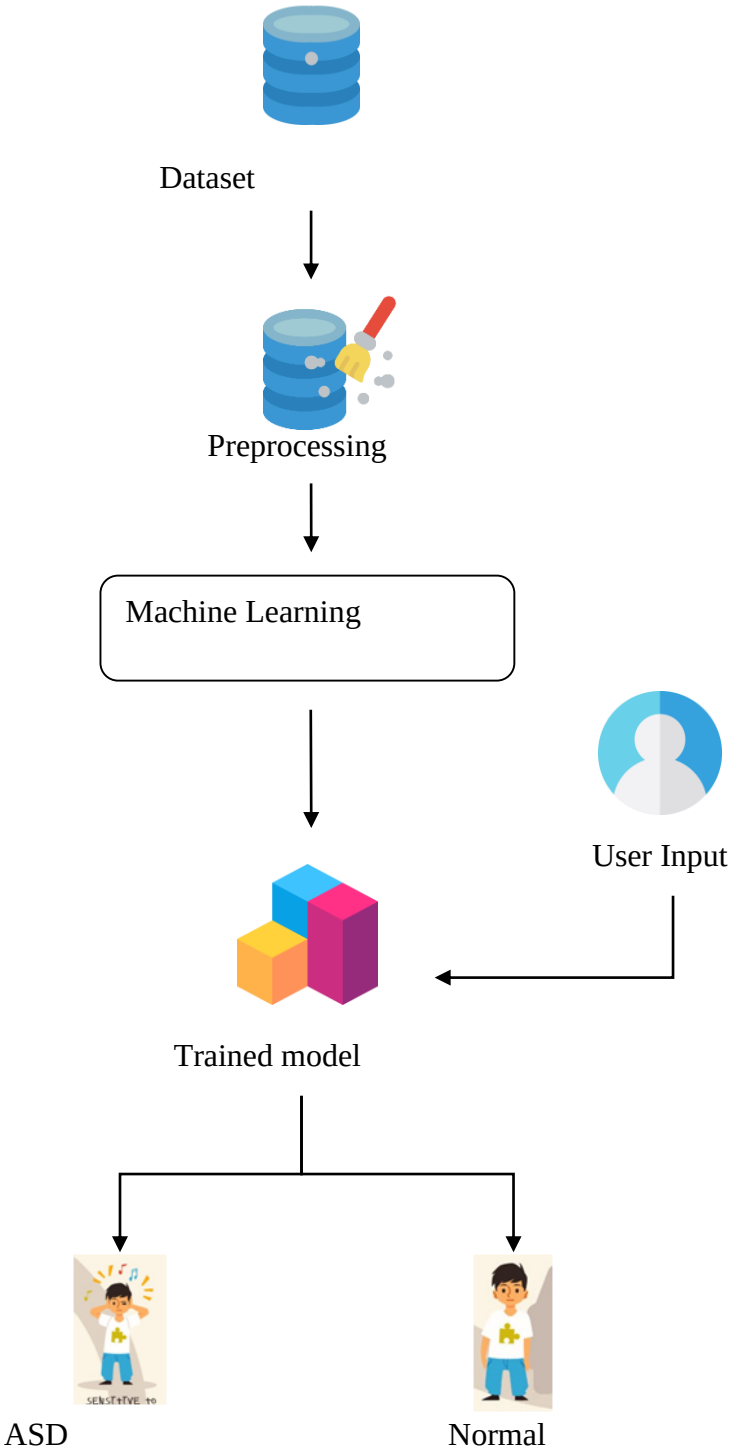


Figure1.Architecture Diagram

2. DATAFLOW DIAGRAM

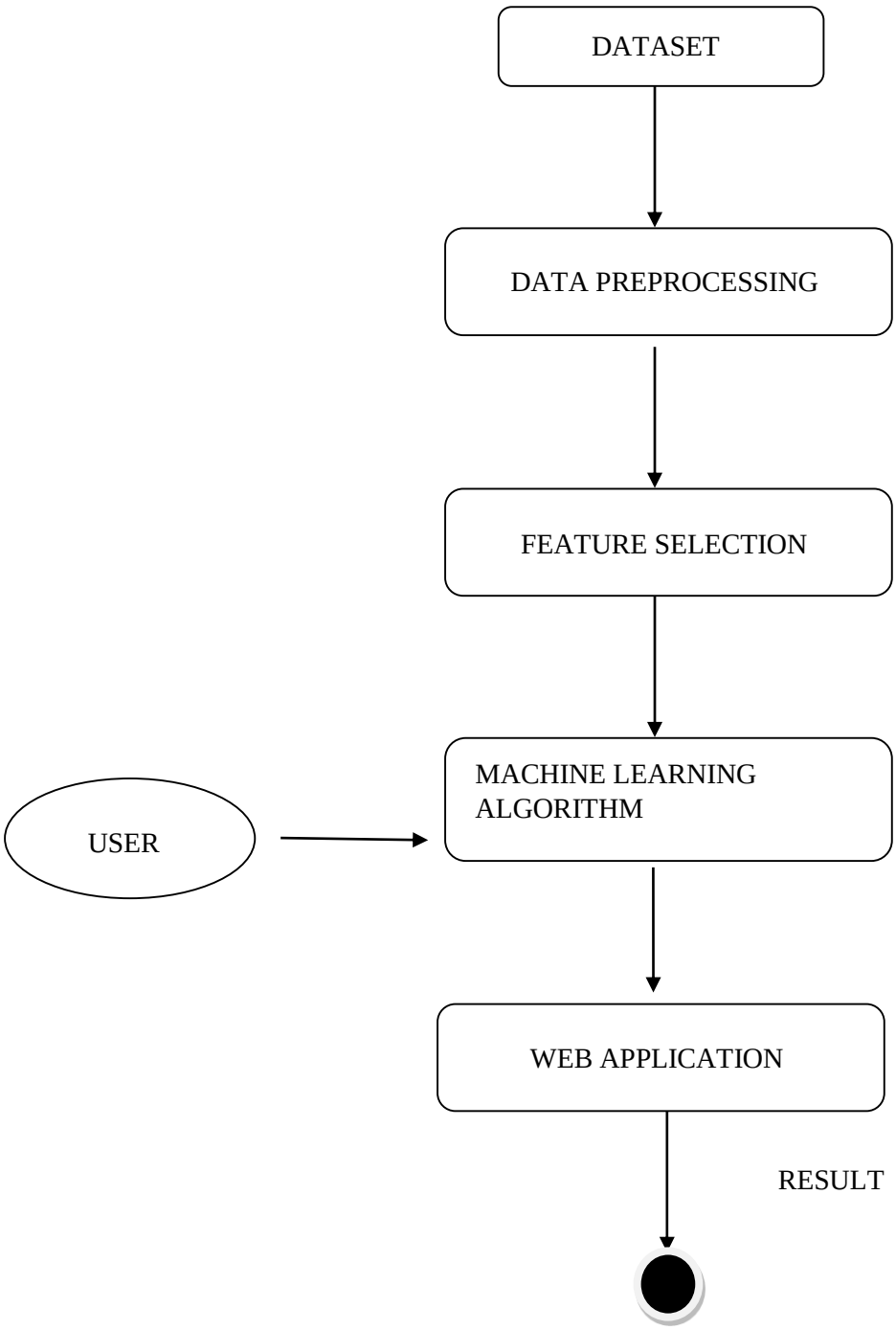


Figure2.Dataflow Diagram

3. SEQUENCE DIAGRAM:

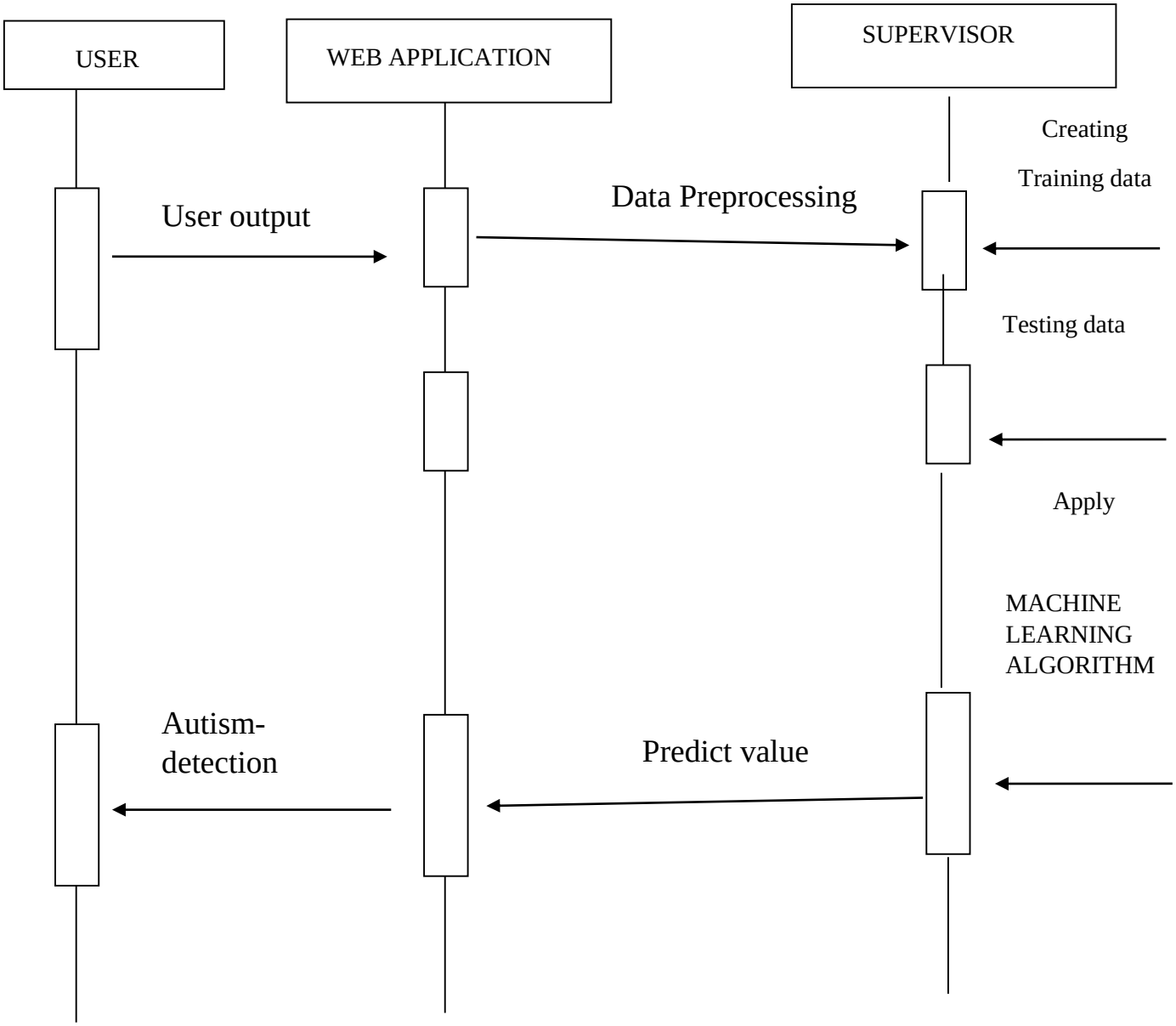
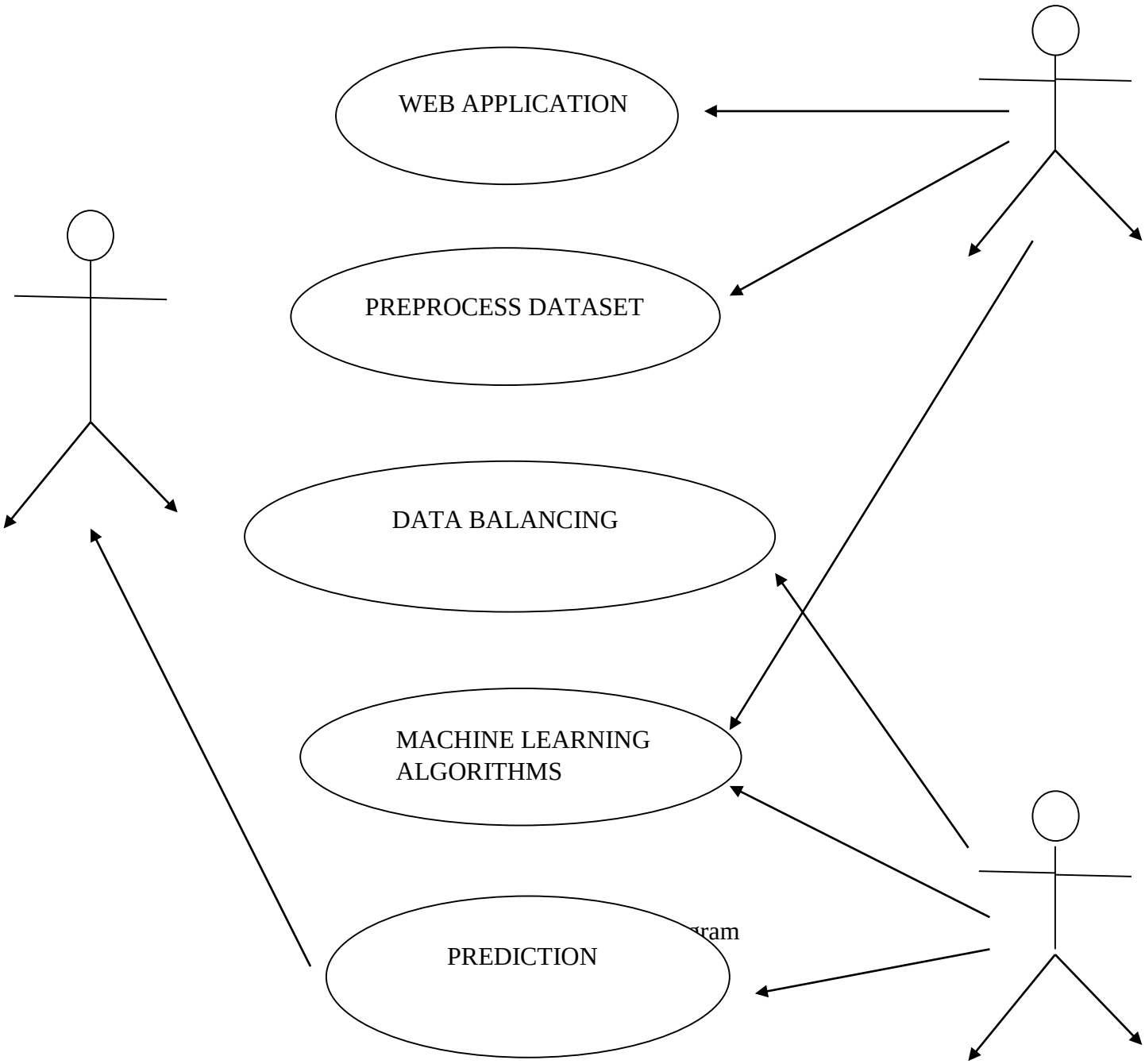


Figure3.Sequence Diagram

4. USE CASE DIAGRAM:



7. FUTURE ENHANCEMENT

Future enhancements of this study include integrating deep learning models like CNNs to capture complex patterns and incorporating multimodal data such as genetic, neuroimaging, and behavioral inputs for improved accuracy. Developing real-time diagnostic tools via mobile applications can enhance accessibility, especially in remote areas. The inclusion of explainable AI techniques would increase model transparency and clinical trust. Additionally, expanding datasets across age groups and cultures can improve generalizability and reduce prediction bias.

8. CONCLUSION

This research demonstrates the potential of machine learning in supporting early and accurate detection of Autism Spectrum Disorder (ASD). By combining label encoding, Recursive Feature Elimination (RFE), and Logistic Regression (LR), the proposed framework effectively enhances diagnostic performance while maintaining interpretability an essential aspect in clinical applications. The model addresses key challenges in traditional diagnosis by automating the identification of relevant features and reducing reliance on subjective assessments. Overall, this approach represents a promising step toward integrating AI-driven tools in healthcare, offering faster, more reliable screening and paving the way for future improvements through advanced algorithms and richer data sources.

REFERENCE

1. M. Bala, M. H. Ali, M. S. Satu, K. F. Hasan, and M. A. Moni, ``Ef_cient machine learning models for early stage detection of autism spectrum disorder," Algorithms, vol. 15, no. 5, p. 166, May 2022.
2. D. Pietrucci, A. Teofani, M. Milanesi, B. Fosso, L. Putignani, F. Messina, G. Pesole, A. Desideri, and G. Chillemi, ``Machine learning data analysis highlights the role of parasutterella and alloprevotella in autism spectrum disorders," Biomedicines, vol. 10, no. 8, p. 2028, Aug. 2022.
3. R. Sreedasyam, A. Rao, N. Sachidanandan, N. Sampath, and S. K. Vasudevan, ``Aarya_A kinesthetic companion for children with autism spectrum disorder," J. Intell. Fuzzy Syst., vol. 32, no. 4, pp. 2971_2976, Mar. 2017.
4. J. Amudha and H. Nandakumar, ``A fuzzy based eye gaze point estimation approach to study the task behavior in autism spectrum disorder," J. Intell. Fuzzy Syst., vol. 35, no. 2, pp. 1459_1469, Aug. 2018.
5. H. Chahkandi Nejad, O. Khayat, and J. Razjouyan, ``Software development of an intelligent spirography test system for neurological disorder detection and quanti_cation," J. Intell. Fuzzy Syst., vol. 28, no. 5, pp. 2149_2157, Jun. 2015.
6. F. Z. Subah, K. Deb, P. K. Dhar, and T. Koshiba, ``A deep learning approach to predict autism spectrum disorder using multisite resting-state fMRI," Appl. Sci., vol. 11, no. 8, p. 3636, Apr. 2021
7. K.-F. Kollias, C. K. Syriopoulou-Delli, P. Sarigiannidis, and G. F. Fragulis, ``The contribution of machine learning and eye-tracking technology in autism spectrum disorder research: A systematic review," Electronics, vol. 10, no. 23, p. 2982, Nov. 2021.

8. I. A. Ahmed, E. M. Senan, T. H. Rassem, M. A. H. Ali, H. S. A. Shatnawi, S. M. Alwazer, and M. Alshahrani, ``Eye tracking-based diagnosis and early detection of autism spectrum disorder using machine learning and deep learning techniques," Electronics, vol. 11, no. 4, p. 530, Feb. 2022.