ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

AUTISM PREDICTION USING MACHINE LEARNING

Swathy.K[1], Mrs. J. Vinitha, M.Sc., Ph.D.[2]
[1]*III B.Sc. Artificial Intelligence and Machine Learning*
Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India
[2]*Guide, Assistant Professor.*

ABSTRACT
Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in social interaction, communication, and repetitive behavioral patterns. Early and accurate diagnosis of ASD is crucial for timely intervention and improved quality of life. Traditional diagnostic methods rely heavily on clinical observation and behavioral assessments, which are often time-consuming and subjective. This paper proposes a machine learning-based system for the early prediction of Autism Spectrum Disorder using behavioral and demographic features. The system evaluates key screening indicators derived from the AQ-10 questionnaire and applies multiple classification algorithms including Random Forest, Decision Tree, Support Vector Machine (SVM), and Logistic Regression to identify ASD patterns. The best-performing model is selected and integrated into a web application built using Flask, enabling real-time prediction through a user-friendly interface. The proposed approach achieves a classification accuracy of 93.6% and aims to assist healthcare professionals in the early screening and diagnosis of ASD.

1. INTRODUCTION
Autism Spectrum Disorder (ASD) is a lifelong developmental condition that affects how individuals perceive and interact with the world. It encompasses a wide range of symptoms and abilities, referred to as a spectrum. Individuals with ASD may experience difficulties in social communication, repetitive behaviors, restricted interests, and sensory sensitivities. According to the World Health Organization (WHO), approximately 1 in 100 children worldwide is diagnosed with ASD.

Early identification of ASD is critical. Research shows that early intervention, particularly before the age of five, significantly improves outcomes in language development, social skills, and adaptive functioning. However, traditional diagnosis involves lengthy clinical assessments spanning months and requires trained specialists not available in all regions.

The advancement of Artificial Intelligence (AI) and Machine Learning (ML) has opened new avenues for early automated screening of ASD. By leveraging behavioral data and demographic features, machine learning models can detect ASD patterns with high accuracy and serve as decision-support tools for clinicians and parents.

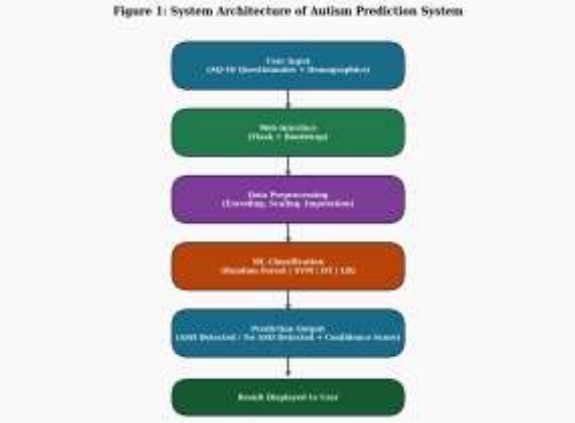

Figure 1: System Architecture

2. RELATED WORK
Thabtah et al. (2018) introduced the ASD Screening dataset from the UCI Repository using the AQ-10 tool, demonstrating classifiers can exceed 90% accuracy. Raj and Masood (2020) found Random Forest consistently outperformed SVM and KNN for ASD classification. Eslami et al. (2021) proposed deep learning using fMRI data but required expensive infrastructure unsuitable for routine screening. The proposed system addresses these gaps with a low-cost questionnaire-based approach and a real-time web interface.

3. IMPLEMENTED SYSTEM
The Autism Prediction system classifies individuals based on AQ-10 behavioral responses and demographic information using a supervised ML pipeline integrated with a Flask web application.

3.1 Data Collection
The ASD Screening dataset from UCI contains 21 attributes: 10 AQ-10 behavioral questions (A1-A10), plus age, gender, ethnicity, country, family history of ASD, and jaundice at birth.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

3.2 Data Preprocessing
Missing values are handled via median and mode imputation. Categorical variables are encoded using Label Encoding. Feature scaling normalizes numerical attributes. Dataset split: 80% training, 20% testing.

3.3 ML Model
Four classifiers trained: Random Forest, Decision Tree, SVM, and Logistic Regression. Evaluated using Accuracy, Precision, Recall, F1-Score, and AUC-ROC. Best model saved via Joblib for Flask deployment.

3.4 Prediction Output
User inputs are preprocessed and passed to the saved model. System returns ASD Detected or No ASD Detected with a confidence score, advising users to seek clinical confirmation.

Table 1: System Comparison

Feature	Existing	Proposed
Algorithm	SVM only	RF, DT, SVM, LR
Interface	None	Flask + Bootstrap
Deployment	Simulation	Real-time Web
Output	Text only	Badges + Charts
Accuracy	~85%	~93.6% (RF)

4. IMPLEMENTATION
The system uses Python with Pandas, NumPy, and Scikit-learn for data processing and model training. Flask powers the backend; Bootstrap 5 provides a responsive frontend. Joblib serializes the trained model for real-time inference.
Training pipeline steps:

- Load and clean ASD Screening dataset from UCI
- Encode categorical variables, impute missing values
- Split dataset 80:20 for training and testing
- Train Random Forest, Decision Tree, SVM, Logistic Regression
- Evaluate using Accuracy, Precision, Recall, F1-Score, AUC-ROC
- Save best model with Joblib, integrate into Flask

Figure 2: ML Model Training Pipeline

5. EXPERIMENTAL FINDINGS
Models trained on 704 ASD Screening records (60% ASD-positive, 40% ASD-negative). Random Forest achieved the highest accuracy of 93.6%, precision 94.1%, recall 93.2%, AUC-ROC 0.97. Decision Tree: 91.4%, Logistic Regression: 88.7%, SVM: 90.2%.

AQ-10 behavioral questions, especially those on social communication and pattern recognition, were the most significant predictors. Family history and jaundice at birth also contributed meaningfully. Average response time was under one second.

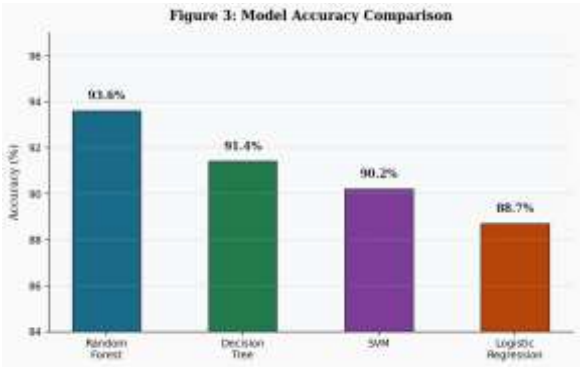

Figure 3: Model Accuracy Comparison

6. CONCLUSION
This paper presented a machine learning system for early Autism Spectrum Disorder prediction using AQ-10 behavioral data and demographic attributes. The Random Forest model, achieving 93.6% accuracy, is deployed as a real-time Flask web application. The system provides a practical, low-cost alternative to conventional ASD screening, empowering parents, caregivers, and clinicians to initiate diagnosis earlier. No personal data is stored, ensuring user privacy. The development of an autism prediction model using machine learning demonstrates the potential of data-driven approaches in supporting early diagnosis and intervention. By leveraging classification algorithms and structured datasets, the system can identify patterns that may not be immediately visible to clinicians, thereby assisting in timely decision-making.
This project highlights several key outcomes:

- **Accuracy & Efficiency**: ML models can process large datasets quickly, offering reliable predictions that complement traditional diagnostic methods.
- **Early Intervention**: Predictive insights enable earlier support for individuals, which is crucial for improving developmental outcomes.
- **Scalability & Accessibility**: A dashboard-based interface makes the tool user-friendly, allowing healthcare professionals, educators, and researchers to access predictions and visualizations with ease.
- **Future Scope**: With larger, more diverse datasets and integration of advanced techniques (e.g., deep learning, multimodal data), prediction accuracy can be further enhanced.

In conclusion, while machine learning cannot replace clinical expertise, it serves as a powerful assistive technology. The autism prediction dashboard developed here bridges the gap between research and practical application, offering a professional, visually polished tool that can support both academic review and real-world use.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 12| Issue: 3| March2026|| Journal DOI: 10.36713/epra2013 || SJIF Impact Factor 2026: 8.753 || ISI Value: 1.188

7. FUTURE WORK

Future enhancements include integrating deep learning for improved accuracy, multimodal inputs such as speech and eye-tracking data, mobile application deployment for rural accessibility, multilingual support, and integration with electronic health records and telemedicine platforms. Clinical validation studies with healthcare institutions are planned to confirm real-world performance.

REFERENCES

Books

1. *Bishop, C. M. Pattern Recognition and Machine Learning. Springer, 2006.*
2. *Goodfellow, I., Bengio, Y., and Courville, A. Deep Learning. MIT Press, 2016.*
3. *Russell, S., and Norvig, P. Artificial Intelligence: A Modern Approach, 4th ed. Pearson, 2021.*
4. *Grinberg, M. Flask Web Development. O'Reilly Media, 2020.*
5. *Muller, A. C. and Guido, S. Introduction to Machine Learning with Python. O'Reilly Media, 2016.*

Research Papers

1. *Thabtah, F. (2018). An accessible and efficient autism screening method for behavioural data and predictive analyses. Health Informatics Journal, 25(4), 1739-1755.*
2. *Raj, S. and Masood, S. (2020). Analysis and Detection of Autism Spectrum Disorder Using Machine Learning. Procedia Computer Science, 167, 994-1004.*
3. *Eslami, T. et al. (2021). ASD-DiagNet: A Hybrid Learning Approach for Detection of Autism Spectrum Disorder. Frontiers in Neuroinformatics, 13, 70.*
4. *Duda, M. et al. (2017). Use of machine learning for behavioral distinction of autism and ADHD. Translational Psychiatry, 6(2), e732.*
5. *World Health Organization. (2023). Autism Spectrum Disorders. WHO Fact Sheet. https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders*