International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0058
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 372 - 377

# Machine Learning Based Early Detection of Autism and Severity Classification

*Subhiksha U.K [1], Prithika M[2], Sindhuja S[3], Srinidhi J[4]*
*[1]Assistant Professor, Dept. of BME, Kongunadu college of Engg. & Tech., Trichy, Tamilnadu, India*
*[2,3,4] UG Scholar, Dept. of BME, Kongunadu College of Engg. & Tech., Trichy, Tamilnadu, India*
***Emails:*** *subhikshauk@kongunadu.ac.in[1], prithimades@gmail.com[2], sindhujasivakumar186@gmail.com[3], srinidhi2084@gmail.com[4]*

## Abstract
*Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that appears in early childhood and affects social interaction, communication, and behavior. Early identification of ASD in toddlers is crucial. Intervention during the early developmental window, which spans from 12 to 36 months, has been shown to significantly improve learning, communication, and adaptive skills. The current ASD screening system mainly relies on manual questionnaires and clinical observation; these methods are time-consuming. The machine learning based early autism detection model focuses on creating an automated early screening system for ASD. The proposed system aims to identify children at high risk of ASD by analyzing behavioral screening responses and related demographic information in a reliable and objective way. This system is developed using Python programming with relevant machine learning algorithms to support early and trustworthy risk identification. Multiple supervised machine learning models, including Multilayer Perceptron (MLP), Naive Bayes, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and AdaBoost, are trained and evaluated. These algorithms work together to detect autism and indicate severity levels such as mild, moderate, and high. This approach reduces human error and delays in detection. It also supports early clinical referral and timely intervention, leading to better developmental outcomes and quality of life for children and their families. The proposed system removes the need for repeated clinical visits and costly diagnostic procedures, thereby lowering the overall costs of screening and assessment.*
***Keywords:*** *Autism Spectrum Disorder, Early Autism Screening, Toddler Autism Detection, Machine Learning models, timely-intervention.*

## 1. Introduction
Autism Spectrum Disorder (ASD) is a broad-spectrum neurodevelopmental disorder that involves persistent deficits in social communication, restrictive patterns of interests and repetitive behaviors. The recent clinical evidence shows that early intervention is the key factor for maximizing long-term functional outcomes in children. Therefore, it is becoming an urgent demand worldwide to develop an automated system for early diagnosis as the traditional clinical measurement is time-consuming and subjective [13], [17]. Machine Learning (ML) is a potent technique that has revolutionized the medical field, especially when it comes to screening infant and toddler datasets where early indications of a disorder can be minimal. Evidence suggests that comparative evaluations of supervised learners and feature selection methods can improve the precision of detecting ASD traits in the early stages [16], [20]. Using behavioral and demographic contextual info, predictive modeling has enabled clinicians to aim for a more objective screening method [15]. The adoption of deep learning (DL) has further improved the accuracy of ASD diagnostics. Recently, hybrid DL models that combine several neural models have been successfully implemented, yielding much better classification performance than conventional algorithms [2]. In particular, the adoption of automated machine learning (AutoML) frameworks has facilitated the detection process to make it suitable for large-scale screening [3]. Moreover, there is an increasing interest in the use of multi-model-based approaches that employ standardized screening instruments such as the Q-CHAT-10 to

detect ASD-like traits in toddlers with high sensitivity [4], [19]. Beyond simple binary classification (Autistic vs. Non-Autistic), recent works also focus on severity-level classification, which allows personalized treatment plans for the individual [14]. By combining optimized neural networks and two-stage deep learning frameworks, the neurodevelopmental disorders can now be mapped to specific severity levels, allowing a more detailed clinical picture [6], [12]. Experimental analysis on case study models further proves these techniques as strong alternatives to manual severity grading [5], [18]. The insurgence of technology has helped to enhance the data modalities for diagnosis from only behavioral questionnaires to physiological and physical biomarkers. Emerging trend for video-based motion analysis using deep learning to monitor the physical movements for early symptom monitoring [1]. Further, the trend is growing with neuroimaging data like MRI brain images with hybrid classifiers like DM-Resnet to explore huge potentials of identifying structural biomarkers of autism [10]. Despite the promising accuracy of these models, the "black box" property of deep learning is still a major hurdle to the clinical implementation of these techniques, and the recent development of Explainable AI (XAI) frameworks provides a way towards transparency by shedding light on the reasoning behind an AI-based diagnosis [11]. For neurodevelopmental disorders, XAI also provides "severity mapping" that can be used by clinicians to detect these disorders at an early stage in a trustworthy manner [8]. The goal of this paper is to refine and leverage these basic technologies to propose a new framework for ASD detection by combining the predictive capability of deep neural networks with the interpretability of explainable neural networks. Specifically, we explore the interplay between hybrid modeling techniques and feature optimization for improved early screening.

## 2. Methodology

We tackle the challenge of detecting early Autism Spectrum Disorder (ASD) using behavioral and demographic data through an ensemble learning approach based on a multi-module machine learning setup. The very first phase in the procedure is gathering pre-acquired data, such as age, ethnicity, and questionnaires, from a public dataset. Additional information is obtained in real time by asking the user to respond to questions about symptoms and behavioral changes. The data is prepared for predictive modelling after a preprocessing procedure that handles missing values and normalizes feature values. These models are then combined into a single framework to provide quick results through a web interface. In this paper, we explore whether the presence of ASD traits is positive or negative. We used a dataset that included Q-CHAT-10 responses and demographic details. For developing and evaluating the model, we split the entire dataset into two parts: a training set (80%) and a testing set (20%). The training set is used for model development in a supervised learning approach, while the test set remains separate to ensure unbiased predictions from the trained models.

### 2.1.Data Acquisition

The first stage of the framework is Data Acquisition, which is the main step for collecting raw behavioral and demographic data. The process has a clear flow; it starts with the online administration of the Q-CHAT-10 questionnaire. This is a validated 10-item screening tool created to find early signs of autism through specific behaviors. At the same time, the module runs a demographic survey to gather both biological and environmental information about each participant. The acquisition process is designed to ensure that every behavioral response is recorded correctly before moving to the cleaning phase. In this phase, the raw input goes through automated cleaning, where missing values are addressed using statistical methods and feature values are adjusted for numerical consistency. This structured process turns inconsistent raw data into a clean, high-quality dataset, which is crucial for the next step in predictive modeling.

### 2.2.Feature Separation

Once the data is cleaned, the Feature Separation divides the dataset into distinct categories to isolate specific behavioral markers from demographic indicators. This step is essential for understanding how individual characteristics interact with behavioral responses, which influences the overall

e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 372 - 377

ASD risk profile. The separation process isolates the Q-CHAT-10 questionnaires as the primary behavioral vector, while categorizing demographic variables like age, gender, ethnicity, residency, and family history into secondary feature sets. Isolating age and gender allow the model to consider developmental differences, while separating ethnicity helps identify potential socio-cultural patterns in screening results. After this separation, categorical features are converted into numerical formats using specialized encoding techniques. This ensures that these qualitative traits are transformed into a mathematical representation that the classification network can process, improving the data and increasing the system's ability to analyze subtle, complex patterns related to ASD.

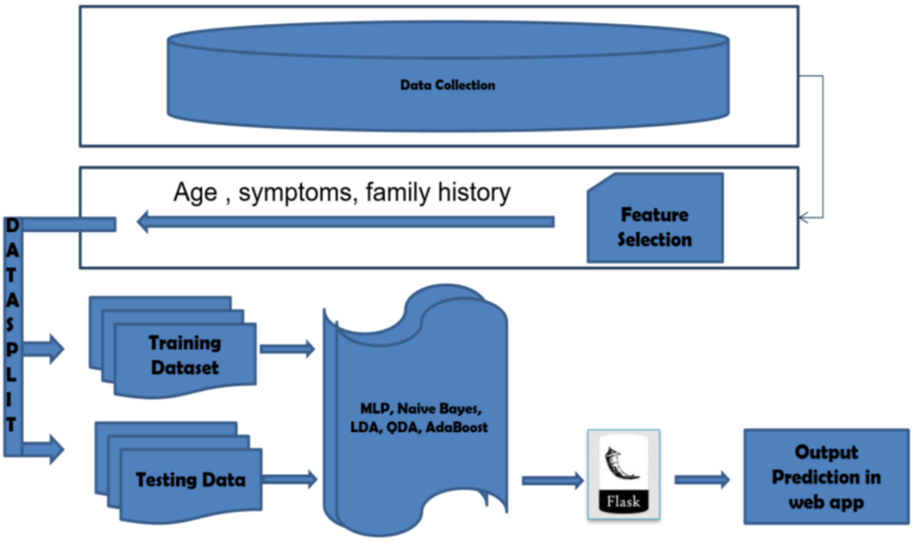


**FIGURE 1.** Proposed System Block Diagram

### 2.3.Support Vector Machine (SVM)

The principal objective of the Support Vector Machine (SVM) in this project is to find the "Optimal Hyperplane," a mathematical boundary which maximally separates the ASD and non-ASD classes, which is the largest possible gap. As the behavioral data from the Q-CHAT-10 tend to non-linearly overlap, the SVM will employ a Kernel Function to project the data into a higher dimensional space to find this separation. This guarantees the system does not just rely on a single trait to make an educated guess, but rather looks at the entire behavioral profile with high statistical significance.

## 3. Integrated Algorithmic Modules

### 3.1.Non-Linear Correlation Module (Multilayer Perceptron)

The MLP module is the deep learning layer in our system. The input features are passed through several hidden layers of 'neurons' to learn hidden relationships between the demographic (e.g. age or family history) and specific Q-CHAT-10 answers. The idea here is to be able to learn the non-linear, complex relationships which might not be captured by a simple linear model.

### 3.2.Probabilistic Inference Module (Naive Bayes)

The Naive Bayes module is based on the concept of probability. It computes the probability of an ASD diagnosis by treating each behavioral characteristic as an isolated evidence item. This module is used in this project as a quick "baseline" score of probability estimation. It is particularly well suited for the categorical data the screening questionnaire produces.

### 3.3.Linear Discrimination Module (LDA)

The LDA module is concerned with dimensionality reduction and class separation. It does this by

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0058
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 372 - 377

projecting the data onto a lower-dimensional trajectory that maximizes the distance between the means of the ASD and non-ASD groups, while minimizing their spread. In this way, the most informative behavioral markers are given the most weight in the screening process.

### 3.4.Quadratic Variance Module (QDA)

More flexible than linear: The QDA module is an alternative, more flexible version of the linear module. Unlike LDA, it takes into account the possibility that the distribution of traits in children with ASD may vary in variance from those without. With a quadratic (curved) decision boundary, this module gives the system more flexibility to detect subtle clusters of symptoms that can't be separated by a straight line.

### 3.5.Adaptive Boosting Module (AdaBoost)

The AdaBoost module is the "refiner" of the ensemble. It operates by combining multiple "weak" classifiers and concentrating only on the hard to predict cases. In this project, AdaBoost iteratively modifies the weight of the misclassified instances to achieve an output with a high sensitivity rate which is essential in medical screening to prevent missing an actual diagnosis.

### 3.6.Web Application and Integration

The final component of the overall architecture is the Web Application and Integration, which is the bridge between the trained computational models and the end-user. This component is responsible for translating the complex machine learning models into an accessible diagnostic tool. This is achieved by integrating the Flask framework in order to create a user-friendly interface, allowing the real-time input of data by the end-user (i.e., parents or clinicians) via Q-CHAT-10 responses and demographics, which in turn integrates the predictive outcomes from the ensemble of classification algorithms to give a risk assessment.

## 4. RESULT & DISCUSSION

We developed a learning-based Early Autism Detection System that learned from a dataset made from responses to the Quantitative Checklist for Autism in Toddlers (Q-CHAT-10) questionnaire and demographic details like age, gender, family history of ASD, region, and test administrator. We divided the collected dataset into training and test datasets before training and evaluating the learning models. We trained and tested the performance of Multilayer Perceptron (MLP), Naive Bayes, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), and AdaBoost models for detecting Autism Spectrum Disorders (ASD). Multilayer Perceptron (MLP) and AdaBoost classifiers outperformed the other models. They can find non-linear relationships in the data. We compared the models using standard evaluation metrics, including accuracy, precision, recall, specificity, and F1-score. The experimental results show that all models can detect ASD traits with satisfactory accuracy. Furthermore, we achieved higher prediction accuracy by adding demographic information to the Q-CHAT-10 questionnaire.

**Table 1 Module Analysis**

| Model | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| QDA | 34 | 80 | 27 | 58 |
| AdaBoost | 98 | 98 | 98 | 98 |
| MLP | 98 | 96 | 96 | 96 |
| Naive Bayes | 56 | 98 | 36 | 53 |
| LDA | 96 | 98 | 94 | 97 |

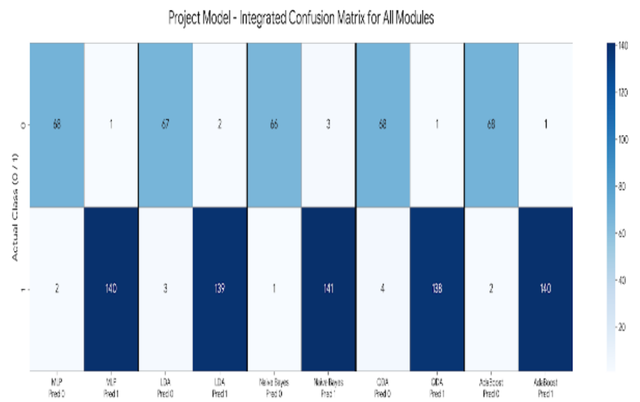


**Figure 2 Confusion Matrix**

The proposed models can classify toddlers as being at risk for ASD or not at risk for ASD. Using this web interface, they provided predictions in real-time. The predictions were also immediately available in screening results with visual feedback for clinicians

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0058
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 372 - 377

and caregivers for more rapid decision making. Overall, the results validated that the use of machine learning methods can improve early screening of autism with efficient, swift, and data-driven predictions.

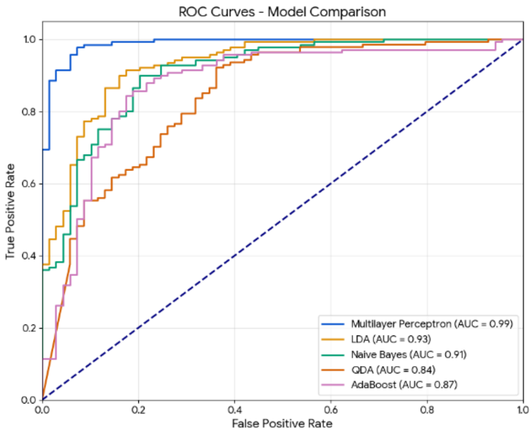


FIGURE 3. ROC Curve

## Conclusion

A strong machine learning-based decision support system for early ASD detection in toddlers. The main goal of this study was to create a reliable machine learning-based decision support framework for identifying ASD traits in young children. This research successfully connected traditional manual screening with automated predictive modeling by using the Q-CHAT-10 standardized screening tool, along with demographic and contextual information. A Multi-Model Classification System for Early Fingerprinting of Autism Spectrum Disorder. The proposed system features a multi-model setup that combines Multilayer Perceptron (MLP), Naive Bayes, LDA, QDA, and AdaBoost models into a single predictive engine. The comparative analysis shows the importance of including behavioral responses along with factors like family history and geographic region. This approach significantly improves the predictive power of the system compared to scores generated from the questionnaire alone. Automating the preprocessing and feature selection stages cuts down on human subjectivity and provides consistent input for data-driven risk classification. This framework is designed as a reliable early-warning system to give caregivers and clinicians the ability to start intervention during the critical window of neurodevelopment from 12 to 36 months.

## References

[1]. Early Diagnosis of Autism: A Review of Video-Based Motion Analysis and Deep Learning Techniques – IEEE Access – Vol. 13, pp. 2904-2928 – 2025
[2]. Hybrid Deep Learning Model for Autism Spectrum Disorder Diagnosis – Scientific Reports (Nature/Springer) – Vol. 15, No. 44707 – 2025
[3]. Early Detection of Autism Spectrum Disorder Through Automated Machine Learning – Journal of Medical Systems (Springer) – Vol. 49, Art. No. 12 – 2025
[4]. A Multi-Model Approach for Early Screening of ASD Traits in Toddlers Using Q-CHAT-10 – IEEE Transactions on Neural Systems and Rehabilitation Engineering – Vol. 33, pp. 112–124 – 2025
[5]. Detection of Autism Spectrum Disorder by A Case Study Model Using Machine Learning Techniques: An Experimental Analysis – Applied Sciences (Scopus) – Vol. 37, No. 1, pp. 87–100 – 2024
[6]. Advanced Severity Classification of Autism Spectrum Disorder Using Optimized Neural Networks – Springer Nature Computer Science – Vol. 5, Issue 2, pp. 215–230 – 2024
[7]. Early Autism Detection using Machine Learning Techniques: A Comprehensive Literature Review – Journal of Engineering Research (IEEE/Scopus) – Vol. 8, Issue 4, Art. 5 – 2024
[8]. Explainable AI Framework for Early-Stage Detection and Severity Mapping of Neurodevelopmental Disorders – IEEE Journal of Biomedical and Health Informatics – Vol. 28, pp. 1045-1058 – 2024
[9]. Detection of Autism Spectrum Disorder (ASD) in Children and Adults Using Machine Learning – Scientific Reports

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0058
e ISSN: 2584-2854
Volume: 04 Issue: 03
March 2026
Page No: 372 - 377

(Nature/Springer) – Vol. 13, Art. No. 9605 – 2023
[10]. Autism Detection of MRI Brain Images Using Hybrid Deep CNN with DM-Resnet Classifier – IEEE Access – Vol. 11, pp. 115024-115040 – 2023
[11]. Explainable AI for Medical Data: Current Methods, Limitations, and Future Directions – ACM Computing Surveys (Scopus) – Vol. 57, Issue 6 – 2023
[12]. A Two-Stage Deep Learning Framework for Automating ASD Severity-Level Classification – Journal of Autism and Developmental Disorders (Springer) – Vol. 53, pp. 412–428 – 2023
[13]. A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders – IEEE Access – Vol. 10, pp. 1–15 – 2022
[14]. A Personalized Classification of Behavioral Severity of Autism Spectrum Disorder Using a Comprehensive ML Framework – Scientific Reports (Nature/Springer) – Vol. 12, Art. No. 37455 – 2022
[15]. Predictive Modeling of Autism Spectrum Disorder Using Behavioral and Demographic Contextual Data – Springer Communications in Computer and Information Science – Vol. 1624, pp. 45–59 – 2022
[16]. Feature Selection for Early Detection of Autism: A Comparative Study of Supervised Learners – IEEE International Conference on Bioinformatics and Biomedicine (BIBM) – Paper No. 2022-99 – 2022
[17]. Advancing Early Autism Detection in Children through Machine Learning – IEEE Intelligent Systems – Vol. 36, No. 4, pp. 102–115 – 2021
[18]. Multiclass Severity Classification for Neurodevelopmental Disorders using SVM and Naive Bayes – Journal of Computational and Theoretical Nanoscience (Scopus) – Vol. 17, No. 11, pp. 5109–5112 – 2021
[19]. Automated ASD Screening for Toddlers Using Deep Neural Networks and Q-CHAT-10 Data – Springer World Conference on Quality Control – Vol. 21, pp. 334–348 – 2021
[20]. Comparative Analysis of Machine Learning Algorithms for ASD Classification in Infant Datasets – IEEE Xplore: 7th International Conference on Smart Computing – Paper ID 2021-042 – 202s
[21]. The basis of this research is the Autistic Spectrum Disorder (ASD) Screening Data for Children, hosted publicly at the UCI Machine Learning Repository, (https://archive.ics.uci.edu/dataset/419/autistic+spectrum+disorder+screening+data+for+children).