2026 Volume: 02 Issue: 05 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-15

# ANALYSIS AND DETECTION OF AUTISM SPECTRUM DISORDER USING MACHINE LEARNING TECHNIQUES: A RANDOM FOREST-BASED CHATBOT SCREENING SYSTEM

***Gaurav Panjabrao Mandape**

India.

**Article Received: 29 March 2026**

**Article Revised: 19 April 2026**

**Published on: 09 May 2026**

***Corresponding Author: Gaurav Panjabrao Mandape**

India.

DOI: https://doi-doi.org/101555/ijrpa.2321

## ABSTRACT

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects social communi-cation, behavioral flexibility, and sensory processing. Delays in screening often postpone interven-tion, especially in settings where specialist access is limited. This paper presents a publication-style summary of an intelligent web-based ASD screening system that combines a conversational chat-bot interface with a supervised machine learning classifier. The proposed system uses a Random Forest model trained on validated ASD screening datasets from the UCI Machine Learning Repos-itory, integrating data from adult, child, and adolescent cohorts to support broader applicability. A 40-question assessment structure, inspired by the Autism-Spectrum Quotient (AQ), is delivered through a chatbot workflow to improve usability and reduce the perceived clinical burden of ini-tial screening. According to the supplied project documentation, the optimized model achieved a classification accuracy of 97.27% on held-out test data using 1,100 samples and a 300-tree Ran-dom Forest configuration. The full system combines a Flask backend, React frontend, and SQLite persistence layer to provide secure authentication, real-time inference, confidence-based feedback, and post-screening guidance. The paper argues that machine learning-enhanced conversational screening can improve accessibility and user engagement while remaining clearly positioned as a preliminary decision-support tool rather than a replacement for professional diagnosis.

**KEYWORDS:** Autism Spectrum Disorder, Random Forest, chatbot, machine learning, digital screening, conver-sational interface.

# 1 INTRODUCTION

Autism Spectrum Disorder is a heterogeneous neurodevelopmental condition characterized by per-sistent differences in social communication and restricted or repetitive patterns of behavior and interests. The World Health Organization estimates that approximately 1 in 100 children globally are autistic, reinforcing the need for timely and scalable screening pathways [1]. Early identification can improve access to behavioral support, family guidance, and specialist referral, yet conventional assessment pathways remain resource intensive and are frequently constrained by clinician avail-ability, waiting lists, and geographic inequities.

Digital health tools offer a practical opportunity to reduce these access barriers. In particular, conversational screening interfaces can make self-report and caregiver-assisted questionnaires easier to complete by presenting one question at a time in a familiar dialogue format. At the same time, supervised machine learning methods can learn nonlinear relationships among screening responses and demographic variables, supporting more robust risk classification than simple threshold-based rules in some settings.

This paper develops a concise research narrative around a web-based ASD screening platform described in the supplied project synopses. The system integrates a chatbot interface, a Random Forest classifier, and a lightweight web architecture for secure, responsive deployment. The main aim is to demonstrate how validated questionnaire data and machine learning can be combined into an accessible preliminary screening workflow that supports, rather than replaces, clinical judgment.

The key contributions of the work are as follows:

- A unified ASD screening pipeline trained on adult, child, and adolescent datasets from the UCI Machine Learning Repository;
- A 40-question chatbot-based assessment experience derived from AQ-style screening domains;
- A Random Forest classifier configured for high-accuracy binary prediction with confidence-oriented output; and
- A deployable full-stack implementation using Flask, React, and SQLite for authentication, result storage, and user guidance.

## 1.1 Related Work

Screening research in autism has long relied on structured instruments designed to capture behav-ioral traits associated with ASD. Baron-Cohen et al. introduced the Autism-Spectrum Quotient (AQ), a 50-item self-administered instrument spanning social skill, attention switching, attention to detail, communication, and imagination [2]. The AQ remains one of the most influential trait-based screening instruments in autism-related research and informed the behavioral framing of the present chatbot questionnaire.

Recent research has shown growing interest in machine learning for ASD classification and screening support. Supervised methods such as decision trees, support vector machines, neural networks, and ensemble classifiers have been used to identify salient behavioral features and improve classification accuracy across screening datasets [4, 5]. Among these methods, Random Forests are especially attractive because they are robust to noise, can model nonlinear feature interactions, provide feature importance estimates, and reduce overfitting through ensemble averaging [3].

Despite promising performance in offline experiments, many ASD-related machine learning studies stop at model evaluation and do not proceed to user-facing deployment. This leaves a gap between algorithmic performance and real-world usability. The present work addresses that gap by em-bedding the predictive model inside a conversational screening interface that can deliver real-time results and educational next steps.

## 2 MATERIALS AND METHODS

### 2.1 Research Design

The study follows an applied research design with an experimental machine learning workflow. The focus is not only on classifier accuracy, but also on end-to-end system usability, accessibility, and practical deployability. The overall process includes dataset preparation, feature preprocessing, model development, web application integration, and screening result presentation.

### 2.2 Datasets

The training corpus described in the uploaded project material combines three ASD screening datasets hosted by the UCI Machine Learning Repository: adult, child, and adolescent screening data [6, 7, 8]. Together, these datasets contain 1,100 instances. Each dataset includes responses to screening questions along with demographic and contextual attributes relevant to prediction.

**Table 1: Datasets used for model development.**

| Dataset | Instances | Attributes |
|---|---|---|
| Autism Screening Adult Data | 704 | 21 |
| Autism Screening Child Data | 292 | 21 |
| Autism Screening Adolescent Data | 104 | 21 |
| Combined dataset | 1,100 | 21 per source set |

The project synopsis indicates that the predictive feature space includes ten core screening variables and selected demographic descriptors such as age, gender, ethnicity, jaundice history, and family autism history. The target label is binary ASD classification.

To support publication-style clarity, it is useful to distinguish between the source datasets and the final conversational instrument. The datasets provide structured examples for supervised learning, while the chatbot interface acts as the user-facing delivery mechanism for collecting responses. In other words, the model learns from curated screening records, whereas the deployed applica-tion translates those signals into a more accessible interaction pattern suitable for preliminary community-facing use.

### 2.3 Questionnaire Design

Although the UCI datasets are based on shorter screening inputs, the proposed interface extends the user interaction into a 40-question flow inspired by AQ-style domains. The uploaded synop-sis organizes the questionnaire into five areas: social skills, attention patterns, communication, imagination, and routine preferences. Presenting the assessment through a chatbot enables pro-gressive disclosure, reducing fatigue by asking one question at a time while maintaining a coherent interaction sequence.

This domain-wise grouping is beneficial for both usability and analytics. From a user perspective,

**Table 2: Illustrative organization of the 40-question chatbot assessment.**

| Domain | Approximate questions | Purpose |
|---|---|---|
| Social skills | 8 | Capture interaction comfort, reciprocity, and social responsiveness |
| Attention patterns | 8 | Measure switching, focus, and preference for detail |
| Communication | 8 | Assess conversational ease and expressive understanding |
| Imagination | 8 | Probe flexible thinking and hypothetical engagement |
| Routine preferences | 8 | Examine insistence on sameness and behavioral rigidity |

The sequence feels structured rather than repetitive. From a modeling perspective, grouped ques-tions make it easier to inspect whether predictive strength is concentrated in specific behavioral domains or distributed across the entire response profile.

## 2.4 Data Preprocessing

The methodological workflow extracted from the project documentation includes the following preprocessing stages:

- handling missing values through median imputation for numerical variables and mode imputa-tion for categorical variables;
- removing irrelevant identifiers and non-predictive columns;
- encoding categorical variables for machine learning compatibility;
- standardizing selected numerical features; and
- using a stratified 80:20 training-test split to preserve class balance.

In addition, five-fold cross-validation is used during evaluation to reduce dependence on a single split and to assess generalization more reliably.

## 2.5 Problem Formulation

The learning task is framed as a supervised binary classification problem. Let $\mathbf{x}_i \in \mathbb{R}^p$ denote the encoded feature vector for the $i$th screening record and let $y_i \in \{0, 1\}$ denote the associated class label, where 1 indicates elevated ASD likelihood in the screening dataset and 0 indicates the alternative class. The objective is to learn a function $f : \mathbf{x}_i \mapsto \hat{y}_i$,
that minimizes classification error on unseen examples while preserving stable performance across age-group-specific source datasets.

In the deployed application, the classifier does not operate as a diagnostic oracle. Rather, it produces a risk-oriented screening output together with a confidence estimate,

$P(y = 1 \mid \mathbf{x})$,

which can be translated into user-facing guidance such as low-risk, moderate-risk, or high-risk follow-up recommendations. This distinction between predictive scoring and clinical diagnosis is essential for appropriate interpretation.

## 2.6 Model Development

The classifier used in this work is a Random Forest model [3]. Based on the uploaded project synopsis, the selected configuration is shown below:

| Hyperparameter | Value |
|---|---|
| Number of trees | 300 |
| Maximum depth | 25 |
| Minimum samples to split | 3 |
| Minimum samples per leaf | 1 |
| Maximum features | $\sqrt{p}$ |
| Bootstrap sampling | Enabled |
| Random state | 42 |

Random Forest was selected because behavioral screening data often contain mixed-type features, nonlinear decision boundaries, and complex interactions among self-report variables. Ensemble learning helps stabilize predictions while preserving interpretability through feature importance ranking.

At inference time, the ensemble prediction is formed by aggregating the votes of individual decision trees. If the forest contains $T$ trees, the predicted label may be expressed as

$$\hat{y} = \text{mode}\{h_t(\mathbf{x})\}_{t=1}^{T},$$

where $h_t(\mathbf{x})$ denotes the class predicted by the $t$th tree. The confidence score used in the web interface can similarly be estimated from the proportion of trees voting for the positive class. Such a mechanism is operationally attractive because it is straightforward to explain to non-technical stakeholders: the final result is derived from the consensus of many weak-to-moderate decision structures rather than a single brittle rule.

## 2.7 Evaluation Metrics

Although the supplied synopsis emphasizes overall accuracy, a screening system should be evaluated using multiple complementary measures. Let true positives, false positives, true negatives, and false negatives be denoted by $TP$, $FP$, $TN$, and $FN$, respectively. Then the standard metrics are given by

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN},$$

$$\text{Precision} = \frac{TP}{TP + FP}, \qquad \text{Recall} = \frac{TP}{TP + FN},$$

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

For early-stage health screening, recall is especially important because missed positive cases may delay referral and support. Precision also matters, however, because excessive false alarms can create anxiety and unnecessary follow-up burden. Accordingly, the most appropriate model is not simply the one with the highest raw accuracy, but the one that maintains a reasonable balance between sensitivity-oriented screening objectives and practical downstream triage capacity.

**Table 3: Recommended evaluation perspective for ASD screening models Metric Relevance to this study.**

| Metric | Relevance to this study |
|---|---|
| Accuracy | Summarizes overall correctness and is the principal value reported in the supplied project record |
| Precision | Indicates how often a positive screening output is likely to correspond to a truly elevated-risk case |
| Recall | Reflects the ability to detect potentially missed ASD-related cases during preliminary screening |
| $F_1$ score | Balances precision and recall when class distributions or screening priorities are uneven |
| Calibration | Evaluates whether probability outputs align with observed outcome frequencies in real use |

### 2.8 System Architecture

The screening system is implemented as a web application with the following layers:

- **Frontend:** a React-based chatbot interface for registration, progressive questioning, response capture, and results display;
- **Backend:** a Flask REST API for authentication, inference, and screening history management;
- **Model layer:** a serialized Random Forest model for prediction and confidence estimation; and
- **Database:** SQLite persistence for user accounts, sessions, and prior screening outcomes.

This architecture supports responsive interaction while keeping deployment lightweight enough for academic prototypes and pilot studies.

## 2.9 Screening Workflow and Interface Design

Beyond predictive performance, the proposed system is notable for how it operationalizes screening as an interactive digital journey. A publication-oriented description of the workflow is useful because many applied machine learning papers underdescribe the user experience layer that mediates real-world adoption.

The screening session can be viewed as a sequence of six stages: account creation, informed use ac-knowledgment, conversational question delivery, response aggregation, model inference, and follow-up guidance. The chatbot metaphor supports continuity between these stages because the user does not experience abrupt transitions from form filling to analytics; instead, the process appears as a single guided interaction.

**Table 4: End-to-end workflow of the chatbot screening platform.**

| Stage | System component | Main function |
|---|---|---|
| Registration | React + Flask | Create account and authenticate the user securely |
| Screening intake | Chatbot UI | Present questions sequentially and collect structured responses |
| Preprocessing | Backend service | Encode and normalize submitted values for model compatibility |
| Inference | Random Forest model | Generate class prediction and confidence-oriented score |
| Results | UI + database | Display recommendation and store outcome for session history |
| Guidance | Educational layer | Provide non-diagnostic explanation and professional referral advice |

This workflow offers several practical benefits. First, conversational pacing can reduce the cognitive burden associated with long questionnaires. Second, modular architecture makes it easier to revise questionnaire items or substitute alternative classifiers without redesigning the entire application. Third, storing prior results enables longitudinal monitoring in future versions, which could sup-port comparative tracking across repeated screening sessions while still preserving the distinction between screening and diagnosis.

An additional interface advantage is explainability at the communication layer. Even when the underlying model is mathematically complex, the application can still provide humanreadable summaries, such as indicating that the result reflects an observed pattern across social communi-cation, routine behavior, and attention-related responses. Such summaries do not fully solve model interpretability, but they improve transparency and user acceptance.

## 3 RESULTS AND DISCUSSION

The supplied project report states that the optimized Random Forest classifier achieved **97.27% classification accuracy** on held-out test data after training on the combined dataset of 1,100 instances. The system was also designed to provide probability-based outputs so that users re-ceive not only a binary screening result but also a confidence-oriented interpretation of the model response.

This reported performance suggests that the integrated feature set captures strong discriminative information for the benchmark screening records. Because the combined training corpus pools adult, child, and adolescent data, the result also indicates that a single ensemble model may be capable of learning stable cross-group patterns when preprocessing is handled carefully. In publication terms, this is important because it shifts the contribution from narrow dataset fitting toward broader screening-system feasibility.

At the same time, accuracy alone should be interpreted with caution. Screening datasets may contain structured regularities that are easier to classify than truly open-world clinical encoun-ters. Consequently, the reported result is best understood as evidence of technical promise under benchmark conditions rather than as proof of field-ready diagnostic performance. A strong paper therefore benefits from interpreting the numerical result in the context of deployment constraints, response variability, and the difference between curated data and live users.

From a usability standpoint, the platform adds value beyond the classifier itself. A conventional spreadsheet-like screening workflow may produce a prediction, but it does not necessarily encourage completion, comprehension, or appropriate follow-up. By coupling the model to a chatbot front end, the system enhances interaction quality in three ways: it improves response pacing, it frames outputs in plain language, and it links results to action-oriented guidance. These system-level improvements are especially relevant for preliminary screening in educational or community settings where specialist supervision may not be immediately available.

From a systems perspective, the experimental contribution is twofold. First, the model perfor-mance indicates that ensemble methods can provide highly competitive results for structured ASD screening data. Second, the conversational front end transforms the classification pipeline into a user-facing workflow that can be accessed through standard web browsers on both desktop and mobile devices.

The output of the platform is intentionally framed as a screening recommendation rather than a diagnosis. After prediction, the system provides explanatory text, educational resources, and guidance to seek professional assessment where appropriate. This framing is important for both ethical deployment and user trust.

### 3.1 Interpretation of Results

The published accuracy figure can be interpreted as supporting three claims. First, questionnaire-derived behavioral indicators remain informative inputs for machine learning-based ASD screening. Second, Random Forest remains a competitive baseline for medium-sized tabular health datasets because it accommodates feature heterogeneity with minimal manual feature engineering. Third, predictive performance becomes more practically meaningful when embedded in a responsive soft-ware system that users can actually complete.

Nevertheless, responsible interpretation requires acknowledging that benchmark success does not automatically imply generalizability. Data collected from public repositories may underrepresent linguistic diversity, socioeconomic variation, co-occurring conditions, and differences in who com-pletes the questionnaire (self-report versus caregiver report). For this reason, the strongest contri-bution of the present work may be its systems integration strategy rather than the isolated accuracy value alone.

### 3.2 DISCUSSION

The paper demonstrates the value of combining predictive analytics with conversational interaction design. Traditional ASD screening typically depends on static questionnaires or clinician-mediated forms. By contrast, the chatbot interface improves engagement through gradual question delivery, progress feedback, and a less intimidating user experience. This matters in preliminary screening contexts where user dropout and response fatigue can reduce practical utility.

The Random Forest model is a suitable choice for this application because it handles heterogeneous features well, requires relatively modest tuning, and remains resilient when feature relationships are nonlinear. Compared with single-model decision rules, ensemble prediction can improve robustness while retaining an interpretable structure for post hoc inspection.

Another important discussion point is the relationship between questionnaire expansion and model fidelity. The source datasets emphasize a limited set of validated screening features, whereas the chatbot interaction extends the experience to 40 questions for usability and domain coverage. This creates both an opportunity and a challenge. On the one hand, longer interaction can improve engagement, contextual richness, and perceived thoroughness. On the other hand, the deployed questionnaire should be mapped carefully back to the trained feature schema so that the model receives semantically consistent inputs. Any mismatch between user-facing items and training-time variables could weaken validity if not managed explicitly.

The project also highlights a broader design principle for applied clinical AI: performance and presentation are inseparable. A technically strong model that produces confusing outputs may have limited practical value, while an elegant interface built on an unreliable model may create misplaced trust. The strongest systems therefore align predictive rigor, interface clarity, and ethical messaging. In the present case, the chatbot layer, the backend inference service, and the post-screening educational guidance together form a single intervention pipeline rather than independent components.

At the same time, several limitations should be acknowledged. The study is based on secondary screening datasets and a project synopsis rather than a fully reproduced experimental notebook inside the present workspace. The 40-question chatbot interaction extends beyond the ten-question core signals represented in the source UCI datasets, so real-world validation with prospectively col-lected user data is still necessary. In addition, demographic fairness, calibration across populations, and longitudinal clinical utility require dedicated evaluation before the tool could be considered for broader deployment.

### 3.3 Threats to Validity and Reproducibility

Several threats to validity should be stated explicitly in a research publication. The first is *dataset validity*: repository-based screening datasets may reflect historical collection practices, limited ge-ographic diversity, and fixed questionnaire formats that differ from live deployment environments. The second is *construct validity*: a positive model output represents elevated screening likelihood, not confirmed clinical diagnosis. The third is *external validity*: performance on curated benchmark data may not transfer directly to community populations, multilingual contexts, or users with atypical comorbid presentations.

Reproducibility is another important consideration. The synopsis provides core hyperparameters, system components, and dataset sources, which supports partial transparency. However, full re-producibility would benefit from a versioned preprocessing pipeline, explicit feature mapping from chatbot answers to model inputs, dataset split documentation, and archived evaluation scripts. These additions would make it easier for future researchers to verify the reported result, com-pare against alternative models, and measure the impact of interface changes independently from classifier changes.

In a publication setting, stating these threats does not weaken the paper; rather, it improves credibility. A balanced discussion demonstrates that the work understands the difference between a promising prototype and a clinically validated decision-support platform.

### 3.4 Ethical and Practical Considerations

ASD screening tools interact with sensitive behavioral and health-related information. Accordingly, any deployable version of the system must protect privacy, minimize unnecessary data retention, and communicate the limits of automated inference clearly. The project synopsis includes secure authentication, session handling, and database-backed storage, which are appropriate foundations for a prototype. However, a publication-ready production system would also require explicit consent mechanisms, institutional review where applicable, and stronger external validation across diverse user groups.

The system should therefore be positioned as a decision-support aid for early screening, not as a substitute for licensed clinical diagnosis. This distinction is especially important in neurode-velopmental assessment, where contextual interpretation and multidisciplinary evaluation remain essential.

### 3.5 Future Deployment Opportunities

The proposed framework opens several practical directions for future extension. A multilingual chatbot could broaden access in linguistically diverse communities. Adaptive questioning could shorten the screening session by prioritizing the most informative items based on previous re-sponses. A clinician dashboard could summarize prior screening history, confidence trends, and question-domain patterns for follow-up discussion. In addition, explainability modules such as ranked feature contributions or domain-level summaries could help clinicians and families under-stand why a particular recommendation was produced.

There is also scope for methodological comparison. Although Random Forest offers a strong and interpretable baseline, future studies could benchmark gradient boosting, calibrated probabilistic classifiers, or hybrid rule-plus-ML pipelines. Such comparisons would be most useful if conducted not only on predictive accuracy, but also on calibration, fairness, latency, interpretability, and user comprehension. For a screening application, the best model is the one that balances predictive quality with transparent and ethically responsible deployment.

## 4 CONCLUSION

This paper presents a machine learning-enabled ASD screening framework that couples a Random Forest classifier with a web-based chatbot interface. Using 1,100 instances from validated UCI screening datasets and an optimized 300-tree ensemble, the documented system achieves strong predictive performance while addressing a practical usability problem in early screening access. The combination of conversational interaction, real-time prediction, and post-screening guidance makes the approach well suited for preliminary, accessible, and scalable deployment.

Future work should focus on prospective validation with real users, subgroup-wise error analy-sis, calibration studies, multilingual support, and comparison against alternative models such as gradient boosting and explainable neural approaches. With careful ethical oversight and clinical collaboration, conversational AI systems may become valuable front-end tools for widening access to early ASD screening pathways.

## ACKNOWLEDGEMENTS

The author acknowledges the publicly available UCI Machine Learning Repository datasets and the prior research literature that informed this study. Appreciation is also extended to the broader community of researchers, clinicians, educators, and caregivers whose work continues to improve early ASD screening, awareness, and support.

## REFERENCES

1. World Health Organization, “Autism,” 2025. Available: https://www.who.int/news-room/ fact-sheets/detail/autism.
2. S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley, “The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians,” *Journal of Autism and Developmental Disorders*, vol. 31, no. 1, pp. 5–17, 2001.
3. L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
4. F. Thabtah, “Autism spectrum disorder screening: Machine learning adaptation and DSM-5 fulfillment,” *Proceedings of the 1st International Conference on Medical and Health Informat-ics*, pp. 1–6, 2017.
5. F. Thabtah and S. Peebles, “A new machine learning model based on induction of rules

for autism detection," *Health Informatics Journal*, vol. 26, no. 1, pp. 264–286, 2020.

6. UCI Machine Learning Repository, "Autistic Spectrum Disorder Screening Data for Adult," accessed May 6, 2026. Available: https://archive.ics.uci.edu/dataset/426/autism+screening+adult.
7. UCI Machine Learning Repository, "Autistic Spectrum Disorder Screening Data for Children," accessed May 6, 2026. Available: https://archive.ics.uci.edu/dataset/419/autistic+spectrum+disorder+screening+data+for+children.
8. UCI Machine Learning Repository, "Autistic Spectrum Disorder Screening Data for Ado-lescent," accessed May 6, 2026. Available: https://archive.ics.uci.edu/dataset/420/autistic+spectrum+disorder+screening+data+for+adolescent.