KINGDOM OF SAUDI ARABIA
Northern Border University (NBU)
Journal of the North for Basic & Applied Sciences(JNBAS)
p - ISSN : 1658 -7022 / **e-ISSN:** 1658 - 7014
www.nbu.edu.sa
s.journal@nbu.edu.sa

Optimizing Autism Spectrum Disorder Screening with Machine Learning: Identifying Key Behavioral Predictors and Model Performance

Gamal Saad Mohamed Khamis
Department of Computer Science, College of Science, Northern Border University, Arar, Saudi Arabia

(**Received:** 29th June 2025 **Accepted:** 9th October 2025)

Abstract
Autism Spectrum Disorders (ASD) are becoming increasingly common worldwide, highlighting the need for accurate and reliable early detection methods. Utilizing machine learning for early detection is crucial for enhancing screening processes. This study investigates how behavioral traits can help identify ASD through machine learning algorithms. The performance of several techniques, including Gradient Boosting, Support Vector Machines, and Naïve Bayes, is compared. The analysis reveals that A9, A6, A7, and A5 are the most significant predictors of ASD, while A10, A8, A4, A3, A2, and A1 exhibit some predictive value but are less substantial. The results demonstrate excellent performance across various machine learning classifiers for ASD detection. Gradient Boosting achieved perfect classification (Accuracy = 0.98%, F1-Score = 0.98%), indicating powerful predictive ability. Support Vector Machine (SVM) produced nearly perfect results (Accuracy = 0.97%, F1-Score = 0.97), while Naïve Bayes also performed well (Accuracy = 0.96%, F1-Score = 0.96), despite its simpler design. These findings highlight the effectiveness of ensemble methods (Gradient Boosting) and kernel-based models (SVM) in ASD screening, with potential for clinical application in early diagnosis. Policymakers are encouraged to adopt machine learning-based screening tools to support early and accurate ASD diagnosis, which can lead to better patient outcomes.

Keywords: Autism Spectrum Disorder (ASD), Disease Screening, Machine learning, Gradient boosting, Support Vector Machine (SVM), Naïve Bayes

DOI: 10.12816/0062294

(*) **Corresponding Author**:
Gamal Saad Mohamed Khamis
Department of Computer Science, College of Science, Northern Border University, Arar, Saudi Arabia
E-mail: gamal.khamees@nbu.edu.sa
https://orcid.org/0000-0003-3689-4010

1. Introduction

1. Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that affects communication, social interaction, and behavior. Early and accurate diagnosis is crucial, as timely interventions can significantly enhance developmental outcomes and overall quality of life. Traditional diagnostic approaches rely on behavioral assessments conducted by specialists, which can be time-consuming, subjective, and resource-intensive. As a result, there is a growing need for objective, data-driven, and accessible tools that can assist in early ASD screening.

2. Machine learning (ML) has emerged as a transformative technology in healthcare analytics, enabling the discovery of hidden patterns in complex behavioral data. Recent studies have demonstrated its effectiveness in detecting ASD traits using datasets such as Q-Chat-10 and similar screening instruments [1]–[9]. Techniques such as Decision Trees, Support Vector Machines (SVM), and Naïve Bayes have shown encouraging results in identifying behavioral and developmental markers of ASD. Furthermore, feature selection methods—such as Principal Component Analysis (PCA) and chi-square tests—have improved diagnostic accuracy and model interpretability [2], [7]. The integration of machine learning into autism research, as highlighted by Thabtah [6], represents a significant advancement in technology-assisted diagnosis, leading to the development of tools such as the ASDTests mobile app [3], [4], which enable large-scale, early behavioral screening.

3. However, many prior studies have emphasized deep learning or image-based modalities that require high computational resources and extensive datasets. While powerful, such models often lack interpretability and are not always practical for clinical or low-resource settings. Moreover, studies that leverage behavioral questionnaires have usually limited their analyses to basic classification without comprehensive feature ranking or interpretability assessment.

4. This study introduces an interpretable, resource-efficient, and high-performing ASD screening framework that objectively identifies the most influential behavioral predictors using classical machine learning models. By combining ensemble (Gradient Boosting) and kernel-based (SVM) classifiers with rigorous feature-ranking techniques (Information Gain, Gain Ratio, Chi-Square, and Gradient Boosting Feature Importance), this research provides a balanced trade-off between predictive accuracy and model transparency. Unlike prior studies that relied on image-based datasets [22], IoT-based emotion recognition [23], or explainable deep learning models [24], our work focuses on behavioral screening using the Q-Chat-10 questionnaire.

5. The main objectives and contributions of this study are as follows:

- Novel integration of classical ML algorithms with multi-criteria feature selection, enhancing interpretability and robustness in behavior-based ASD detection.
- Comprehensive feature importance analysis, objectively identifying key behavioral traits (A9, A7, A5, and A6) that significantly influence ASD prediction.
- Comparative evaluation of three complementary classifiers—Gradient Boosting, SVM, and Naïve Bayes—demonstrating that lightweight ensemble and kernel-based methods can achieve accuracy comparable to deep learning while maintaining interpretability.
- Clinical and practical relevance, offering a transparent and computationally efficient screening framework that can be adapted to assist clinicians and policymakers in early ASD detection, especially in resource-limited environments.
- By addressing gaps in interpretability, accessibility, and behavioral focus, this research contributes a methodologically sound and practically viable solution for enhancing ASD screening and early diagnosis.

2. METHOD

2.1 Dataset

1. The dataset used in this paper results from an autism screening of toddlers, which includes influential features for further analysis, particularly in determining autistic traits and improving the classification of ASD cases. This dataset contains 1,054 instances and 18 attributes, including the target variable. The dataset comprises 1,054 instances, with a balanced distribution between ASD and non-ASD cases (approximately 50.3% and 49.7%, respectively). Data preprocessing included binary encoding, handling of missing values, and stratified 10-fold cross-validation to mitigate imbalance.

Table 1. Feature descriptions.

Feature	Type	Description
A1	Binary (0, 1)	Does your child respond by looking at you when you call their name?
A2	Binary (0, 1)	How easily can you establish eye contact with your child?
A3	Binary (0, 1)	Does your child use pointing gestures to express a desire for something? (e.g., a toy that is out of reach)
A4	Binary (0, 1)	Does your child want to share something they find interesting with you?
A5	Binary (0, 1)	Does your child engage in pretend play? (e.g., care for dolls, talk on a toy phone)
A6	Binary (0, 1)	Have you noticed if your child pays attention to what you're looking at?
A7	Binary (0, 1)	If someone in your family feels down, does your child try to comfort them? (e.g., stroking hair, hugging them)
A8	Binary (0, 1)	Would you describe your child's first words as:
A9	Binary (0, 1)	Does your child make use of basic gestures? (e.g., wave goodbye)
A10	Binary (0, 1)	Does your child gaze into space without any apparent purpose?
Target variable	Binary (Yes, No)	Yes (ASD traits) or No (no ASD traits)

2. Attributes A1- A10: These items are part of the Q-Chat-10, where possible answers to questions are "Always," "Usually," "Sometimes," "Rarely," and "Never." These responses are mapped to binary values ("1" or "0") in the dataset. For questions 1-9 (A1- A9), if the response is "Sometimes," "Rarely," or "Never," a value of "1" is assigned. For question 10 (A10), a response of "Always," "Usually," or "Sometimes" results in a "1" being assigned. A cumulative score of more than 3 across all 10 questions indicates potential ASD traits; otherwise, no ASD characteristics are observed.

3. The remaining Features are collected from the ASDTests application [12]. The target variable is assigned automatically based on the score obtained by the user during the screening process using the ASDTests application.

2.2 Feature Importance Analysis

1. This section details the methodology employed to identify the most influential behavioral features in predicting ASD traits within the Q-Chat-10 dataset. Feature selection is crucial for building parsimonious and interpretable models while mitigating the risk of overfitting.

2. Several feature ranking methods were utilized to evaluate the importance of each feature:

- Information Gain: This metric quantifies the reduction in uncertainty about the target variable (ASD traits) achieved by knowing the value of a particular feature. Features with higher Information Gain are deemed more informative [14].
- Gain Ratio: A normalized version of Information Gain that addresses the bias toward features with many values [14].
- Gini Index: Measures the impurity of a node in a decision tree, with higher values indicating greater impurity. Features that effectively reduce Gini impurity are considered more critical [15].
- Chi-Square Test (χ^2): Evaluates the statistical dependence between a feature and the target variable. Higher χ^2 values suggest stronger associations.
- Gradient Boosting Feature Importance: Derived from the Gradient Boosting algorithm, this metric reflects the contribution of each feature to the model's predictive performance [16].

Table 3. Features Ranks

Features	Information gain	Gain ratio	Gini Index	χ^2	Gradient Boosting
A9	0.278	0.278	0.142	179.324	14.666
A7	0.229	0.245	0.136	117.035	11.546
A5	0.252	0.252	0.136	158.969	11.228
A2	0.174	0.175	0.092	124.8	10.858
A6	0.246	0.25	0.139	144.613	10.5
A1	0.191	0.193	0.108	116.759	9.714
A8	0.144	0.144	0.078	104.003	9.467
A4	0.199	0.199	0.109	131.189	9.284
A10	0.023	0.024	0.014	14.1	6.437
A3	0.137	0.141	0.072	105.916	5.699

Table 3 and Figure 2 provide a summary of the feature ranking results. While A9, A6, A7, and A5 emerged as the most prominent predictors, the remaining features (A10, A8, A4, A3, A2, and A1) also exhibited varying predictive power. Although lower than the top four, their rankings suggest potential contributions to ASD prediction. Features such as Jaundice, Sex, family_mem_with_ASD, and Ethnicity demonstrated minimal predictive power. The observed rankings provide valuable insights into the relative importance of different behavioral features in predicting ASD traits.

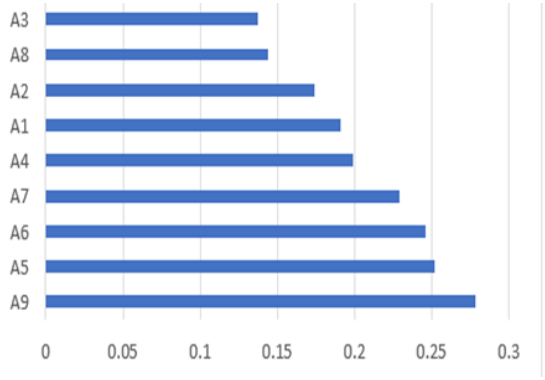

Figure 2. Information Gain Feature Importance Ranking

4.

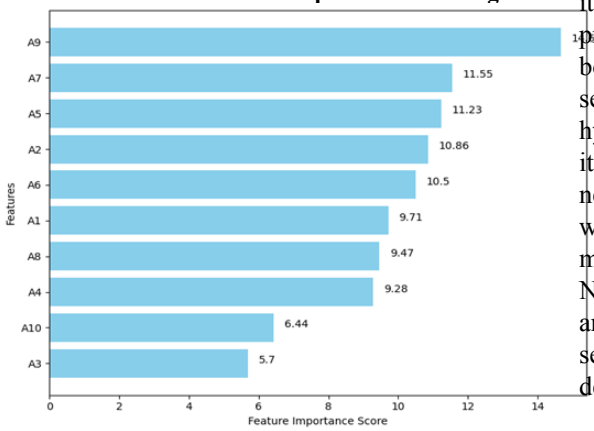

Figure 3: Gradient Boosting Feature Importance Ranking

The feature importance analysis summarized in Table 2, Figure 2, and Figure 3 highlights key predictors of ASD traits within the Q-Chat-10 dataset. Features A9, A7, A5, and A6 consistently emerged as the most influential predictors across multiple ranking methods, indicating their strong relevance to ASD prediction. Features A2, A1, A8, and A4 showed moderate predictive power, while A10 and A3 ranked lowest, contributing minimal influence. Figure 2 emphasizes the dominance of A9, A7, and A5 in the Information Gain ranking, while Figure 3 further confirms their significance through Gradient Boosting Feature Importance scores. These findings highlight the importance of communication gestures and social responses in accurately predicting ASD traits.

2.3 ML-based Toddler ASD predictive models

Machine learning (ML) techniques have shown promising results in the early detection of several diseases and disorders, including autism spectrum disorder (ASD) in young children. In this research, we applied many machine learning methods, including:

• **Gradient Boosting** is a machine learning technique used for regression and classification tasks. It creates a strong predictive model by combining multiple weak learners, usually decision trees, into an ensemble. The algorithm works iteratively, adding one tree at a time and adjusting subsequent trees to correct the errors of earlier ones [18]. Key features of this classifier include:

Maximum Number of Trees: This hyperparameter is used to determine the number of trees; the more trees there are, the better the performance of the classifier. Learning Rate: This hyperparameter minimizes the contribution of each tree.Replicable Training: Fix the random seed to enable the replicability of the results.

• **Support Vector Machines (SVM):** Support Vector Machine (SVM) is a supervised learning method employed for classification and regression tasks, though it is primarily utilized in classification scenarios. The primary objective of SVM is to determine the optimal boundary, known as the hyperplane, that effectively separates the classes within the feature space. This hyperplane is positioned to maximize the gap between it and the nearest data points from each category. These nearest points are referred to as support vectors. SVM was chosen due to its strong performance on small-to-medium datasets with high-dimensional features, while Naïve Bayes provides a probabilistic baseline. LSTM and attention-based models, though powerful, were not selected as the dataset is relatively small and sequential dependencies are limited in Q-Chat-10 responses.

In two-dimensional space, a hyperplane functions as a line that divides data into two categories, while in three dimensions, it appears as a plane. As the dimensionality increases, it assumes a more abstract form. The goal is to identify the hyperplane that most effectively separates the classes, maximizing the distance between this boundary and the closest data points, referred to as support vectors, which play a crucial role in defining the decision boundary. The margin, defined as the distance between the hyperplane and these support vectors, is also maximized to enhance classification accuracy. When the data is not linearly separable, the kernel trick can be utilized to shift the original feature space into a higher dimension. This transformation permits linear separation using kernel functions such as linear, polynomial, radial basis function, and sigmoid, each offering distinct data mappings [18].

• **Naive Bayes:** The Naïve Bayes algorithm is a well-established probabilistic classifier that leverages the concept of probability to solve classification problems straightforwardly and efficiently. By applying the principles of Bayes' theorem, this method learns the likelihood of individual items, their attributes, and the categories to which those attributes correspond. It operates within a supervised learning framework, categorizing new instances by assigning them to the class with the highest probability of membership. The classification process involves two key stages: A) Training phase: During this stage, the algorithm calculates the probability distribution based on the training data. B) Prediction phase: For a test instance, it assesses the posterior probability for that unknown sample, determining its most likely class based on the highest posterior probability, known as Maximum A Posterior (MAP) [19].

2.4 Hyperparameter Tuning

A hyperparameter-tuning process optimized the performance of each machine learning model. Due to computational limitations, manual tuning was conducted. The key hyperparameters and their tuning strategies are described below:

Gradient Boosting: Maximum Number of Trees: This hyperparameter was manually adjusted to determine the number of trees in the ensemble. Higher values generally improve performance but can also lead to overfitting. Values ranging from 100 to 200 were tested, with a focus on identifying a balance between performance and computational efficiency.

Support Vector Machines (SVM): Kernel: Given the nature of the data, the radial basis function (RBF) was evaluated. gamma constant in kernel function(g): was auto tuned. The recommended value is 1/k, where k is the number of attributes. SVM was chosen due to its strong performance on small-to-medium datasets with high-dimensional features, while Naïve Bayes provides a probabilistic baseline. LSTM and attention-based models, though powerful, were not selected as the dataset is relatively small and sequential dependencies are limited in Q-Chat-10 responses.

Naive Bayes: The provided text doesn't mention any specific hyperparameter tuning for Naive Bayes.

To ensure robust and generalizable results, we evaluated the performance of each machine learning model using a 10-fold cross-validation approach. This technique helps mitigate overfitting and provides a more reliable estimate of model performance on unseen data. In 10-fold cross-validation, the dataset is divided into 10 folds. The model is trained on 9 folds and tested on the remaining fold, repeating this process 10 times, each serving as the test set once. The reported performance metrics (AUC, Accuracy, F1 Score, Precision, Recall, and MCC) represent the average performance across these 10 folds. This approach provides a more comprehensive assessment of the models' ability to generalize to new data compared to a single train-test split.

3. RESULTS AND DISCUSSION

3.1 Predictive Performance Analysis

The predictive performance of the three machine learning models—Gradient Boosting, Support Vector Machine (SVM), and Naïve Bayes—was rigorously evaluated using stratified 10-fold cross-validation to ensure robustness and generalization. Evaluation metrics included the Area Under the Curve (AUC), Accuracy, F1-Score, Precision, Recall, and the Matthews Correlation Coefficient (MCC). Table 4 summarizes the results.

Table 4: Model's performance Metrics

Classifier	AUC	Accuracy	F1-Score	Precision	Recall	MCC
Gradient Boosting	0.999	0.980	0.980	0.980	0.980	0.953
SVM	0.998	0.976	0.976	0.976	0.976	0.944
Naïve Bayes	0.997	0.961	0.962	0.964	0.961	0.914

The results indicate that all three classifiers achieved robust predictive performance, with Gradient Boosting exhibiting the highest accuracy (98.0%) and AUC (0.999). The SVM model followed closely (accuracy = 97.6%, AUC = 0.998), while Naïve Bayes also demonstrated competitive results (accuracy = 96.1%, AUC = 0.997) despite its relatively simple probabilistic framework. These findings highlight the effectiveness of ensemble and kernel-based learning approaches in modeling behavioral data for ASD prediction.

When compared with previously published studies using similar datasets, the proposed approach demonstrates advancements in predictive accuracy. For instance, Rashed et al. [1] reported accuracies ranging from 96 to 98% using Random Forest and Logistic Regression. In contrast, Alzakari et al. [2] achieved 97.3% accuracy with explainable AI models based on optimized decision trees. In contrast, the proposed Gradient Boosting model achieved 98.0% accuracy with an AUC of 0.999, reflecting improvement in generalization performance while preserving model interpretability. This enhanced outcome is primarily attributed to the model's optimized hyperparameter configuration and ensemble learning architecture. The near-comparable performance of the SVM further reinforces its robustness in capturing complex, high-dimensional behavioral interactions characteristic of ASD screening datasets [19].

Table 5 presents a comparative summary of the proposed models' performance relative to prior machine learning approaches from the literature.

Table 5. Comparative summary of the proposed models

Study / Model	Dataset	Technique	Accuracy (%)	Remarks
Tabtah (2017) [3]	Q-Chat-10	Decision Tree	92.0	Early behavioral model; limited feature selection
Thabtah et al. (2018) [7]	Q-Chat-10	Random Forest	96.8	Strong baseline; moderate interpretability
Heinsfeld et al. (2018) [5]	ABIDE	Deep Neural Network	97.0	High accuracy; requires neuroimaging data
Alzakari et al. (2024) [2]	Hybrid dataset	Explainable AI (XAI)	97.3	Interpretable; higher computational cost
Proposed (2025)	Q-Chat-10	Gradient Boosting + Feature Ranking	98	Highest accuracy; interpretable and lightweight

The comparative findings indicate that the proposed Gradient Boosting model, when integrated with feature-ranking techniques, demonstrates strong and consistent predictive performance, maintaining a high level of interpretability compared to existing ASD screening approaches. Unlike previous studies that depend on large or multimodal datasets (e.g., ABIDE MRI data), the present framework relies solely on behavioral indicators derived from the Q-Chat-10, making it more feasible and scalable for use in both clinical and community-based screening environments. The promising performance of the proposed framework can be attributed to the following factors. First, a rigorous feature selection process combining Information Gain, Gain Ratio, Chi-Square, and Gradient Boosting importance was employed to retain only the most informative behavioral indicators. Second, the use of a balanced dataset (50.3% ASD vs. 49.7% non-ASD) and 10-fold stratified cross-validation enhanced generalizability and minimized class imbalance bias. Third, the ensemble learning structure of Gradient Boosting effectively modeled complex, non-linear relationships among key behavioral traits—particularly A9 (gestures), A7 (empathy behaviors), and A5 (pretend play)—which are established markers of ASD-related social interaction [7], [9]. Finally, the model's interpretability aligns with prior behavioral psychology findings [3], [9], reinforcing its empirical soundness and clinical relevance.

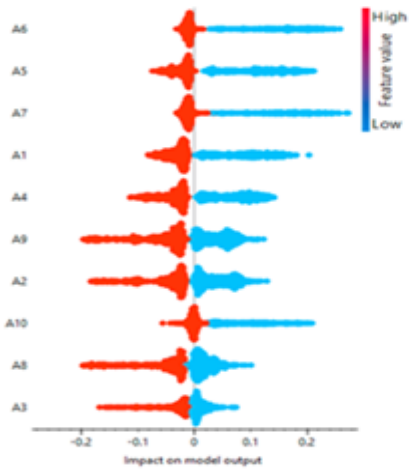

Figure 3. Gradient Boosting Model Explain for Target Class=0

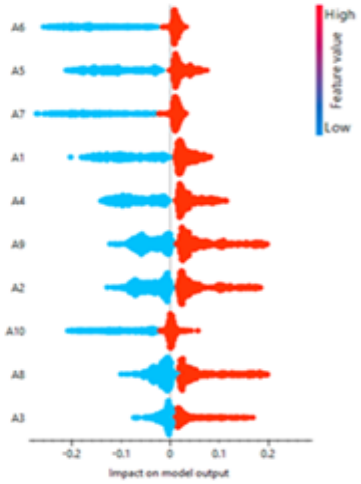

Figure 4. Gradient Boosting Model Explain for Target Class=1

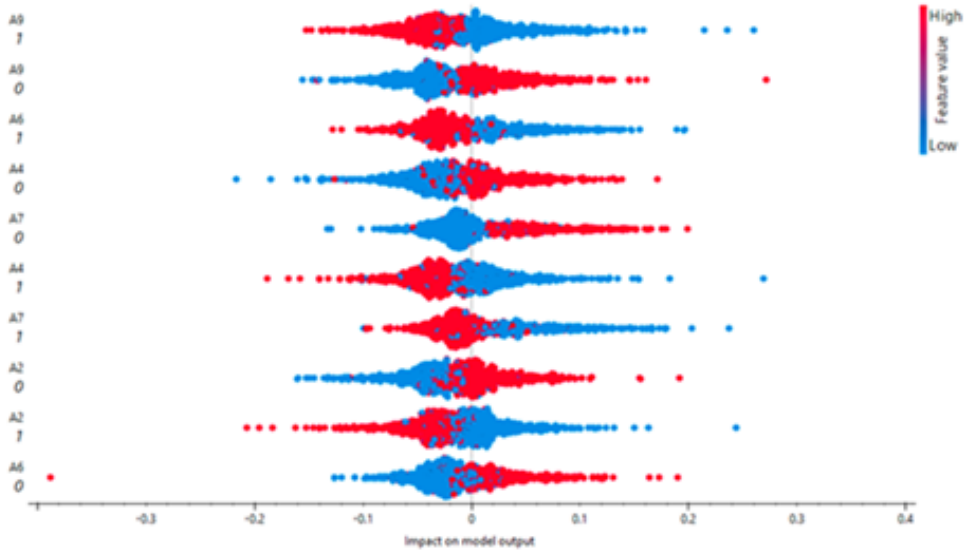

Figure 7. Naïve Bayes Model Explain for Target Class=0

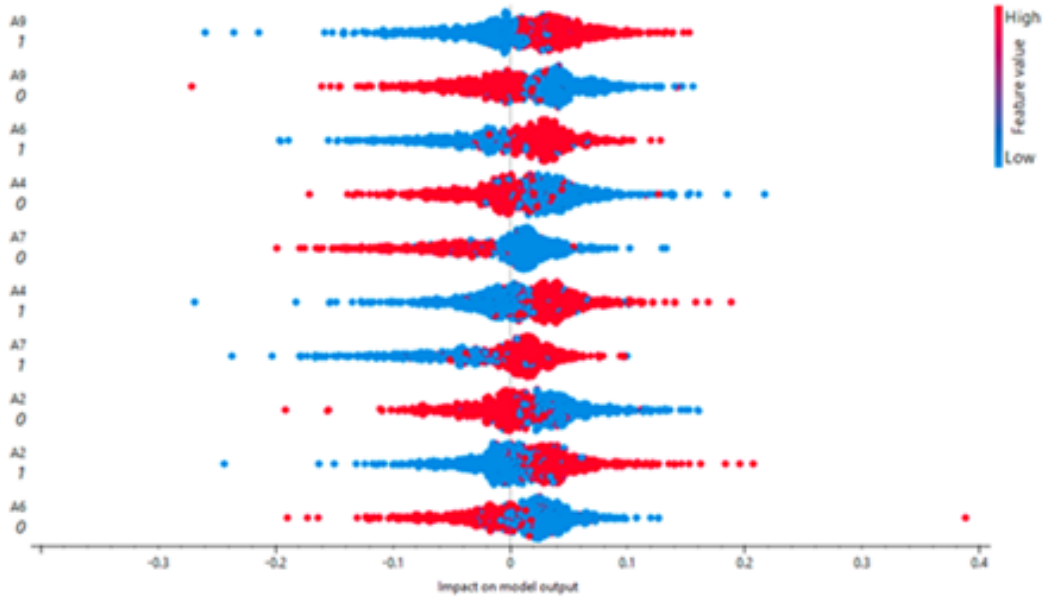

Figure 8. Naïve Bayes Model Explain for Target Class=1

Figures 3–8. Visualization of feature importance and class separation across Gradient Boosting, SVM, and Naïve Bayes models.

3.2 Discussion of Potential Biases and Limitations

Although the proposed framework demonstrated high predictive accuracy (96%, 97% and 98%) and strong generalization, several limitations warrant acknowledgment. The dataset, obtained from a single online source (ASDTests), may introduce cultural and demographic bias, as behavioral interpretations such as "pretend play" or "comforting behavior" can vary across populations. Moreover, the binary encoding of Q-Chat-10 responses, while computationally efficient, may reduce behavioral nuance.

To mitigate potential overfitting, the study employed stratified 10-fold cross-validation and multi-criteria feature selection (Information Gain, Gain Ratio, Chi-Square, and Gradient Boosting importance), which enhanced model stability and interpretability. Nevertheless, validation using independent and more diverse datasets is necessary to confirm the robustness of these findings.

Although Gradient Boosting and SVM achieved high accuracy, their decision-making processes remain partially opaque, highlighting the need for future integration of explainable AI (XAI) frameworks to enhance clinical interpretability and trust. Additionally, the absence of direct collaboration with clinicians in this phase limits immediate clinical applicability; such a partnership will be prioritized in future work to ensure practical validation and usability.

4. CONCLUSION

This study demonstrates the effectiveness of machine learning techniques, particularly Gradient Boosting, Support Vector Machines, and Naïve Bayes, in accurately predicting Autism Spectrum Disorder (ASD) using behavioral data from the Q-Chat-10 dataset. Through comprehensive feature selection and ranking, the analysis identified key behavioral predictors A9 (gestures), A7 (empathy), and A5 (pretend play) that consistently influenced ASD classification, highlighting their diagnostic relevance.

The proposed models achieved high accuracy and interpretability, offering a computationally efficient and clinically practical framework for early ASD screening. While results are promising, the study acknowledges limitations related to dataset representativeness and binary encoding, which may affect generalizability. These challenges were mitigated through cross-validation and rigorous feature selection, ensuring robust performance.

Future work should extend model validation across diverse and longitudinal datasets, incorporate multimodal behavioral and physiological data, and integrate explainable AI (XAI) techniques to enhance transparency and clinical adoption. Overall, the findings emphasize the potential of interpretable ML models to support early, data-driven, and accessible ASD diagnosis in both research and healthcare settings.

5. REFERENCES

1. *A. E. E. Rashed, W. M. Bahgat, A. Ahmed, T. A. Farrag, and A. E. M. Atwa, "Efficient machine learning models across multiple datasets for autism spectrum disorder diagnoses," Biomedical Signal Processing and Control, 2024, p. 106949. DOI: https://doi.org/10.1016/j.bspc.2024.106949*
2. *S. A. Alzakari, L. Zhao, H. Abbas, and M. Bakillah, "Early detection of autism spectrum disorder using explainable AI and optimized teaching strategies," Journal of Neuroscience Methods, 2024, p. 110315. DOI: https://doi.org/10.1016/j.jneumeth.2024.110315*
3. *F. Tabtah, "Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment," in Proc. 1st Int. Conf. Med. Health Informatics, Taichung City, Taiwan, 2017, pp. 1–6. DOI: https://doi.org/10.1145/3107514.3107515*
4. *F. Thabtah, "ASDTests: A mobile app for ASD screening." [Online]. Available: www.asdtests.com (accessed Dec. 20, 2017).*
5. *A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset," NeuroImage: Clinical, vol. 17, pp. 16–23, 2018. DOI: https://doi.org/10.1016/j.nicl.2017.08.017*
6. *F. Thabtah, "Machine Learning in Autistic Spectrum Disorder Behavioural Research: A Review," Informatics for Health and Social Care, 2019.*
7. *F. Thabtah, F. Kamalov, and K. Rajab, "A new computational intelligence approach to detect autistic features for autism screening," International Journal of Medical Informatics, vol. 117, pp. 112–124, 2018. DOI: https://doi.org/10.1016/j.ijmedinf.2018.06.009*

8. *"The classification of autism spectrum disorder by machine learning methods on multiple datasets for four age groups," Measurement: Sensors, 2023, p. 100774. DOI: https://doi.org/10.1016/j.measen.2023.100774*
9. *X. Zhu et al., "Machine Learning Approaches for ASD Detection," Bioengineering, vol. 10, no. 10, p. 1131, 2023. DOI: https://doi.org/10.3390/bioengineering10101131*
10. *"Machine Learning Models in ASD Diagnosis," BMC Psychiatry, vol. 24, p. 6116, 2024. DOI: https://doi.org/10.1186/s12888-024-06116-0*
11. *"Role of ML in Autism Spectrum Disorder," SN Computer Science, vol. 2, p. 776, 2021. DOI: https://doi.org/10.1007/s42979-021-00776-5*
12. *"ASDTests." [Online]. Available: https://www.asdtests.com/ (accessed Oct. 10, 2024).*
13. *S. J. Wisniewski and G. D. Brannan, "Correlation (Coefficient, Partial, and Spearman Rank) and Regression Analysis," National Library of Medicine, 2024.*
14. *J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, pp. 81–106, 1986.*
15. *J. R. Quinlan, C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA, 1993.*
16. *L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees, 1st ed., Chapman & Hall/CRC, 1984. DOI: https://doi.org/10.1201/9781315139470*
17. *J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, pp. 1189–1232, 2001.*
18. *S. Xia, F. Zhang, and C. Zhang, "A Gradient Boosting-Based Classification Technique for Assisted Prediction Algorithm Research," in Proc. 2023 IEEE Int. Conf. Image Process. Computer Applications (ICIPCA), Changchun, China, 2023, pp. 371–375. DOI: https://doi.org/10.1109/ICIPCA59209.2023.10257866*
19. *J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," Neurocomputing, 2019. DOI: https://doi.org/10.1016/j.neucom.2019.10.118*
20. *I. H. Witten, E. Frank, and M. A. Hall, Data Mining: Practical Machine Learning Tools and Techniques, 3rd ed. Morgan Kaufmann, Burlington, MA, USA, 2011.*
21. *J. Qin and Y. Lou, "L1-2 Regularized Logistic Regression," in Proc. 53rd Asilomar Conf. Signals, Systems, and Computers, Pacific Grove, CA, USA, 2019, pp. 779–783. DOI: https://doi.org/10.1109/IEEECONF44664.2019.9048830*
22. *Atlam, E.S., Aljuhani, K.O., Gad, I., Abdelrahim, E.M., Atwa, A.E.M., and Ahmed, A., 2025. Automated identification of autism spectrum disorder from facial images using explainable deep learning models. Scientific Reports, 15(1), p.26682.*
23. *Almars, A.M., Gad, I., and Atlam, E.S., 2025. Unlocking autistic emotions: developing an interpretable IoT-based EfficientNet model for emotion recognition in children with autism. Neural Computing and Applications, pp.1-20.*
24. *Atlam, E.S., Masud, M., Rokaya, M., Meshref, H., Gad, I., and Almars, A.M., 2024. EASDM: Explainable autism spectrum disorder model based on deep learning. Journal of Disability Research, 3(1), p.20240003.*