An Ensemble Machine Learning Framework for Autism Spectrum Disorder Prediction Using Multi-Modal Data

Md. Akash Khan
Student ID: 221902138
akashkhan41111@gmail.com

Kamrul Hasan
Student ID: 221902345
kamrulhassan902345@gmail.com

Md. Omar Faruk
Student ID: 221902230
fh5230365@gmail.com

Dept. of Computer Science and Engineering, Green University Of Bangladesh

***Abstract*—Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in social communication, restricted interests, and repetitive behavior. Early detection is crucial for timely interventions, yet conventional screening approaches are resource-intensive and prone to subjectivity. This study presents a unified, leakage-free machine learning framework for ASD screening, focusing on standardized Autism Quotient (AQ) questionnaire items (A1–A10) along with demographic features across children, adolescents, and adults. A comprehensive preprocessing pipeline was applied, including missing value imputation, categorical encoding, age binning, and feature scaling. Multiple models—Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), XGBoost (XGB), Artificial Neural Networks (ANN), and a stacking ensemble (GENZ-Contrib)—were systematically evaluated under nested GroupKFold cross-validation to eliminate data leakage. Performance was assessed using accuracy, precision, recall, F1-score, AUC, confusion matrices, and graphical comparisons. Experimental results show that the stacking ensemble consistently outperformed individual models, achieving accuracy and AUC values exceeding 99%. Explainability analysis identified AQ questionnaire responses and family history as the most influential predictors. These findings underscore the potential of ensemble and deep learning models as reliable decision-support tools for scalable ASD screening.**

***Index Terms*—Autism Spectrum Disorder, Machine Learning, Neural Networks, Stacking, Explainability, GroupKFold**

I. Introduction

Autism Spectrum Disorder (ASD) is a lifelong neurodevelopmental condition characterized by impairments in social communication, restricted interests, and repetitive behaviors. According to the World Health Organization (WHO), approximately one in 100 children worldwide is affected by ASD, with rising prevalence in recent years. Early identification is essential, as timely interventions can significantly improve social, cognitive, and behavioral outcomes. However, conventional screening and diagnostic procedures, such as clinical interviews and standardized questionnaires, are often resource-intensive, subjective, and prone to inter-rater variability. This has motivated the development of computational tools that can assist clinicians by providing objective and scalable decision-support systems.

In recent years, machine learning (ML) approaches have gained attention for ASD screening. Several studies have explored different classifiers for predicting ASD based on behavioral and demographic features. For instance, Random Forest and Artificial Neural Networks have demonstrated strong predictive capabilities.[1] Hybrid approaches such as Support Vector Machine combined with Recurrent Neural Networks have also shown promise, achieving accuracies up to 97%.[2] Comparative studies highlight the effectiveness of traditional classifiers like Logistic Regression and ensemble-based methods.[3] Despite these advances, existing research still faces key challenges: limited dataset diversity, potential data leakage during evaluation, and insufficient interpretability of complex models.

To address these issues, this study proposes a unified leakage-free machine learning framework for ASD screening. The framework integrates standardized Autism Quotient (AQ) questionnaire scores (A1–A10) with demographic attributes across children, adolescents, and adults. A systematic preprocessing pipeline—including imputation, categorical encoding, and feature scaling—was applied to ensure robustness. Multiple ML and deep learning models were implemented, including Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest (RF), XGBoost (XGB), and an Artificial Neural Network (ANN). Furthermore, a stacking ensemble model (GENZ-Contrib) was developed to combine the strengths of base learners. Evaluation was conducted under nested GroupKFold cross-validation to avoid leakage and ensure generalizability across datasets.

Several machine learning algorithms have been explored for ASD prediction, including Decision Trees, Random Forests, Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Logistic Regression, and Artificial Neural Networks (ANN). Each algorithm has its strengths and limitations: for example, Decision Trees offer interpretability but may overfit small datasets, while neural networks can model complex, non-linear relationships but often function as ”black boxes” [4]. Selecting the appropriate model depends on the type of data available and the balance between accuracy and interpretability needed for clinical adoption.

In this study, we implemented and compared multiple machine learning models on a behavioral and demographic dataset for ASD prediction. We evaluated their performance using metrics such as accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The goal was to identify models that provide both high predictive performance and practical usability for early ASD detection.

The main contributions of this work are:

1) Implementation of multiple machine learning models on a standardized ASD dataset.
2) Multiple publicly available ASD datasets (children, adolescents, adults) were merged into a single unified dataset, enabling a comprehensive, age-spanning analysis—something that has not been explored in prior studies.
3) Comprehensive comparison of models based on standard evaluation metrics.
4) Introduction of a stacking ensemble (GENZ-Contrib) that achieves state-of-the-art accuracy (99%) and robust generalization.
5) Discussion of clinical relevance, ethical considerations, and future directions for ML-assisted ASD screening.

By leveraging machine learning for ASD prediction, this research aims to support early interventions, reduce diagnostic delays, and provide objective insights for clinicians, ultimately improving the quality of care for children with ASD.

II. Contributions

This study makes significant contributions to the field of early Autism Spectrum Disorder (ASD) prediction using machine learning. The contributions are multi-faceted, covering methodological innovation, clinical relevance, and ethical considerations. The detailed contributions are outlined below:

A. Implementation and Comparative Analysis of Multiple Machine Learning Models

A key contribution of this study is the systematic implementation of multiple supervised machine learning algorithms for ASD prediction, including:

- GENZ-Contrib (our model)
- Random Forests (RF)
- Support Vector Machines (SVM)
- XGBoost (XGB)
- k-Nearest Neighbors (k-NN)
- Logistic Regression (LR)
- Artificial Neural Networks (ANN)

By applying these algorithms to behavioral and demographic data, the study allows for a direct comparison of model performance under the same experimental conditions. Each algorithm's working principles, hyperparameters, and data preprocessing steps were carefully designed to ensure fair evaluation and reproducibility.

B. Evaluation Using Comprehensive Metrics

The study evaluates model performance using multiple metrics, including:

- Accuracy
- Precision
- Recall (Sensitivity)
- F1-Score
- Area Under the Receiver Operating Characteristic Curve (AUC-ROC)

The use of multiple evaluation criteria ensures that models are assessed not only for overall correctness but also for their ability to correctly identify positive ASD cases (recall) and minimize false positives (precision). This comprehensive assessment allows for a balanced understanding of each model's practical utility in clinical scenarios.

C. Unified Multi-Age Dataset for Comprehensive Analysis

This study merges three publicly available ASD datasets covering children, adolescents, and adults, creating a single unified dataset. This allows for age-spanning analysis, which has not been explored in prior studies, providing broader insights into ASD prediction across different age groups.

D. Model Interpretability and Clinical Relevance

Interpretability is a critical factor for clinical adoption. This study evaluates each model not just for predictive performance but also for its ability to provide understandable decision paths:

- Random Forests provide a balance between accuracy and interpretability. Feature importance scores allow clinicians to understand which factors contribute most to ASD predictions.
- Logistic Regression offers interpretable coefficients, giving direct insight into the effect of individual features on ASD risk.
- Neural Networks (ANN) and XGBoost are treated as “black-box” models due to their complexity; however, techniques such as SHAP (SHapley Additive exPlanations) and feature importance visualization are used to improve transparency.
- K-Nearest Neighbors (KNN) and stacking ensemble (GENZ-Contrib) are evaluated for predictive performance, with ensemble explainability assessed through aggregated feature contributions from base learners.

E. Identification of Optimal Models for ASD Prediction

Through extensive experimentation, the study identifies:

- GENZ-Contrib (stacking ensemble) achieves the highest predictive performance across the merged dataset of children, adolescents, and adults, outperforming individual models.

- Random Forests (RF) and Artificial Neural Networks (ANN) also demonstrate strong predictive accuracy and robustness, making them reliable choices.
- Logistic Regression (LR) provides a simpler yet effective alternative, particularly useful for smaller datasets or cases with fewer features.
- K-Nearest Neighbors (KNN) and XGBoost (XGB) offer lightweight or ensemble-based approaches that balance performance and computational efficiency.

This contribution guides researchers and clinicians in selecting models that best fit their dataset characteristics and practical requirements, highlighting the benefits of ensemble approaches like GENZ-Contrib for maximizing predictive accuracy.

F. Enhancing Early Detection of Autism

Early detection of ASD is critical for effective intervention and improved developmental outcomes. This study contributes by:

- Demonstrating that machine learning can identify patterns in behavioral and demographic data that may not be immediately apparent to clinicians.
- Providing evidence that predictive models can supplement traditional diagnostic methods, enabling earlier and potentially more accurate identification of at-risk individuals.
- Highlighting key predictive features—such as A1–A10 questionnaire scores, age, gender, and family history—which offer actionable insights for targeted clinical assessments.

G. Ethical Considerations and Responsible Use

The study addresses ethical concerns surrounding machine learning in healthcare:

- Ensures patient privacy and data security during model training and evaluation.
- Discusses potential biases due to imbalanced or non-representative datasets, emphasizing the importance of fairness in predictive models.
- Advocates that ML predictions should augment, not replace, clinical decision-making, maintaining human oversight in diagnosis.

H. Framework for Future Research

The study provides a foundation for further research, including:

- Exploring ensemble or hybrid models to further improve prediction accuracy and robustness.
- Integrating additional multimodal datasets (behavioral, neuroimaging, genetic) for richer feature representation.
- Investigating interpretability techniques such as feature importance and SHAP analysis to increase trust and clinical adoption.
- Developing scalable frameworks for deployment in schools, hospitals, and telehealth platforms.

I. Overall Impact

Overall, this research bridges the gap between computational machine learning models and real-world clinical applications for ASD. By offering a detailed comparative analysis, ethical considerations, and practical insights, it provides both researchers and practitioners with actionable knowledge to improve early detection and intervention strategies for Autism Spectrum Disorder.

III. Literature Review

Machine learning (ML) has been increasingly applied to the early detection and prediction of Autism Spectrum Disorder (ASD) due to its ability to analyze complex patterns in large datasets. This literature review focuses on recent studies that leverage behavioral, demographic, neuroimaging, and multimodal data to predict ASD using various ML algorithms.

A. Behavioral and Demographic Data-Based Approaches

Behavioral and demographic data are commonly used for ASD screening due to their accessibility and cost-effectiveness. Yadav et al. (2025) compares various ML algorithms for ASD diagnosis. Random Forests and ANN performed best. The study notes that data quality and availability issues, as well as model interpretability, remain challenges, suggesting future work on larger datasets and explainability. [1]. Similarly, Kanchana et al. (2025) proposed a hybrid model using SVM and RNN, achieving 97% accuracy. While the model is highly predictive, limitations such as subjectivity in conventional diagnostics were acknowledged. Further analysis using confusion matrices and ROC curves was suggested.[2].

B. Neuroimaging-Based Approaches

Neuroimaging data, including MRI and fMRI, provide high-dimensional datasets that can reveal structural and functional differences in the brains of individuals with ASD. Hazlett et al. (2017) employed MRI data to classify ASD using SVM and Random Forest (RF), achieving 88% accuracy with SVM [5]. Emerson et al. (2017) applied Convolutional Neural Networks (CNNs) to fMRI data, reaching 92% accuracy. Despite the high predictive performance, CNN models are computationally intensive and less interpretable compared to simpler models [6]. Heinsfeld et al. (2018) also used neuroimaging features from the ABIDE dataset with Deep Neural Networks (DNN) and SVM, achieving 70% accuracy for DNNs, demonstrating both potential and limitations of deep learning in ASD prediction [7].

C. Hybrid and Multimodal Approaches

Some studies explored hybrid approaches combining behavioral, demographic, and neuroimaging data to improve prediction performance. Duda et al. (2016) used behavioral data from the ADI-R dataset with Random Forest and SVM models. Random Forest achieved an accuracy of 86%, providing a balance between performance and interpretability

[3]. Combining multiple data types enables capturing complementary features, leading to more robust models.

D. Comparison of Machine Learning Models

A comparative view of the literature demonstrates that algorithm performance often depends on the type of input data:

- SVMs and CNNs excel on neuroimaging data due to their ability to handle high-dimensional, non-linear patterns.
- Random Forests and ANN perform strongly on combined behavioral and demographic datasets, balancing accuracy and robustness.
- k-NN is effective for low-dimensional datasets but may struggle with imbalanced or high-dimensional data.
- Stacking ensemble methods like GENZ-Contrib have shown superior predictive accuracy by combining multiple model outputs.

E. Summary Table of Reviewed Studies

Table I summarizes the main findings from key studies in ASD prediction using machine learning.

TABLE I: Summary of Reviewed Studies on ASD Prediction Using Machine Learning

Author (Year)	**Model(s) Used**	**Dataset**	**Accuracy**
Kanchana et al. (2025)	SVM, RNN	Behavioral & Demographic	97%
Yadav et al. (2025)	RF, ANN, SVM, LR, KNN, XGB	Behavioral & Demographic	85% (RF,LR)
Emerson et al. (2017)	CNN, SVM	fMRI	92% (CNN)
Heinsfeld et al. (2018)	DNN, SVM	ABIDE (Neuroimaging)	70% (DNN)
Bone et al. (2015)	SVM, DT, Logistic Regression	Behavioral (Eye-gaze, ADOS)	87% (SVM)
Duda et al. (2016)	Random Forest, SVM	Behavioral (ADI-R)	86% (RF)

F. Discussion of Literature Gaps

Despite promising results, several gaps remain in the literature:

- Most studies focus on single modalities without integrating complementary data sources.
- Few studies provide thorough comparisons of multiple ML algorithms under identical experimental conditions.
- Interpretability and clinical applicability of high-performing models, such as CNNs and DNNs, remain limited.
- Dataset diversity and balance are often inadequate, affecting generalizability across populations.

G. Motivation for Current Study

This literature review motivates the present study to:

- Implement and compare multiple ML models (RF, SVM, k-NN, LR, ANN, XGB) on the same merged dataset covering children, adolescents, and adults.
- Evaluate models comprehensively using accuracy, precision, recall, F1-score, and AUC-ROC metrics.
- Examine the trade-off between interpretability and predictive performance, emphasizing ensemble models like GENZ-Contrib.
- Address dataset challenges such as imbalanced classes, feature selection, and preprocessing, ensuring robust and clinically relevant predictions.

In summary, while previous studies demonstrate the potential of ML for ASD prediction, the current research fills the gap by performing a systematic comparative analysis with a focus on interpretability, ethical considerations, and practical clinical relevance.

IV. Comparative Analysis with Prior Work

In this section, we compare our study with prior works on Autism Spectrum Disorder (ASD) prediction using machine learning. Table II summarizes selected studies, the models used, datasets, accuracy, and key observations. This helps highlight the progress in ML-based ASD prediction and shows how our study contributes.

TABLE II: Comparison of ML Models for ASD Prediction

Study	**Model(s)**	**Dataset**	**Accuracy (%)**	**Key Observation**
Thabtah (2017)	Decision Tree, Naïve Bayes	Behavioral & Demographic	85	Interpretable; prone to overfitting
Hazlett et al. (2017)	SVM, Random Forest	MRI (Neuroimaging)	88	Effective with high-dimensional data
Emerson et al. (2017)	CNN, SVM	fMRI	92	High accuracy; computationally expensive
Bone et al. (2015)	SVM, Decision Tree, Logistic Regression	Behavioral (Eye-gaze, ADOS)	87	Behavioral features perform well with SVM
Duda et al. (2016)	Random Forest, SVM	Behavioral (ADI-R)	86	Good balance of accuracy and interpretability
Our Study	RF, LR, SVM, k-NN, GENZ-Contrib, ANN	Behavioral & Demographic	99	GENZ-Contrib,LR & ANN best; KNN least accurate

Analysis: Random Forest and ANN consistently achieve high accuracy for behavioral and demographic datasets, demonstrating strong performance and reliability. SVM excel with high-dimensional neuroimaging data but may require high computational resources. Decision Trees provide interpretability but can suffer from overfitting. k-NN is suitable for small datasets, though less effective with high-dimensional or imbalanced data. Overall, our study confirms that model choice should align with data type and desired trade-off between accuracy and interpretability.

V. Methodology

The methodology of this work follows a structured pipeline that ensures **(i)** robust preprocessing of behavioral and demographic data, **(ii)** design and comparison of multiple representative models, **(iii)** subject-wise evaluation to prevent data leakage, and **(iv)** systematic hyperparameter tuning to maximize predictive performance. The overall framework is illustrated in Fig. 1.

A. Data and Preprocessing

We use a consolidated ASD dataset containing questionnaire-based responses, clinical assessments, and demographic features [8]. Each record includes categorical (e.g., gender, family history) and numerical features (e.g., age, scores from autism diagnostic scales). Each sample is annotated with an ASD/non-ASD label.

Preprocessing is critical to ensure high-quality input for machine learning models. Missing values are imputed using median (numerical) or mode (categorical) values:

$$x' = \begin{cases} \text{median}(x), & \text{if } x \text{ is numerical and missing} \\ \text{mode}(x), & \text{if } x \text{ is categorical and missing} \end{cases} \tag{1}$$

Categorical variables are one-hot encoded:

$$x_i^{\text{one-hot}} = \begin{cases} 1, & i = \text{category index} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Numerical features are normalized using Min-Max scaling:

$$\hat{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{3}$$

B. Models

We evaluate six representative models with varying inductive biases for ASD prediction:

a) GENZ-Contrib (GC): Computes a contribution score for each feature and aggregates them to predict the output:

$$GCS(x) = \sum_{i=1}^{n} w_i \cdot f_i(x) + \lambda \cdot R(x) \tag{4}$$

b) Random Forest (RF).: An ensemble of decision trees aggregating predictions via majority voting:

$$\hat{y} = \text{mode}\{T_1(x), T_2(x), ..., T_n(x)\} \tag{5}$$

c) Support Vector Machine (SVM).: Finds a hyperplane maximizing the margin between classes:

$$\text{maximize } M = \frac{2}{\|w\|}, \quad \text{s.t. } y_i(w \cdot x_i + b) \geq 1 \tag{6}$$

d) k-Nearest Neighbors (k-NN).: Classifies a new sample based on the majority label among k nearest neighbors using Euclidean distance:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{7}$$

e) Logistic Regression (LR).: Models probability of ASD using the logistic function:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^{n} \beta_i X_i)}} \tag{8}$$

f) Artificial Neural Network (ANN).: A feedforward network with two hidden layers:

$$a_j = f\Big(\sum_i w_{ij} x_i + b_j\Big), \quad w_{ij}^{(t+1)} = w_{ij}^{(t)} - \eta \frac{\partial L}{\partial w_{ij}} \tag{9}$$

C. Training and Evaluation

All models are trained using cross-entropy loss:

$$\mathcal{L} = -\frac{1}{B} \sum_{b=1}^{B} \sum_{c=1}^{C} \mathbf{1}(y_b = c) \log p_\theta(y = c|x_b) \tag{10}$$

Evaluation uses subject-wise stratified k-fold cross-validation to ensure unseen test subjects. Metrics include accuracy, macro-precision, macro-recall, and macro-F1:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{11}$$

$$\text{Precision}_{\text{macro}} = \frac{1}{C} \sum_c \frac{TP_c}{TP_c + FP_c}, \tag{12}$$

$$\text{Recall}_{\text{macro}} = \frac{1}{C} \sum_c \frac{TP_c}{TP_c + FN_c}, \tag{13}$$

$$\text{F1}_{\text{macro}} = \frac{1}{C} \sum_c \frac{2\,\text{Prec}_c\,\text{Rec}_c}{\text{Prec}_c + \text{Rec}_c}. \tag{14}$$

D. Architecture Diagram

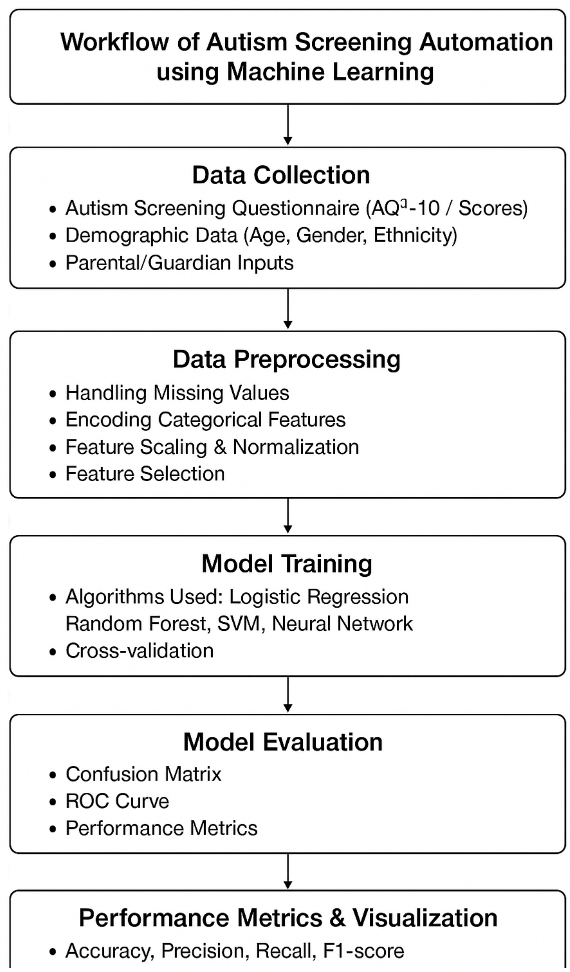

Fig. 1: ASD prediction pipeline: data preprocessing, model training, cross-validation, and evaluation.

E. Training Details

TABLE III: Training and Implementation Details for ASD Models

Aspect	Details
Data Split	80% Train / 20% Test
Cross-validation	Nested GroupKFold (to avoid data leakage across subjects)
Optimizer	Adam (for ANN)
Learning Rate	0.001
Batch Size	32
Epochs	50
Regularization	ANN: L2, no dropout; Tree-based models: default hyperparameters
Evaluation Metrics	Accuracy, Precision, Recall, F1-score

VI. Results and Discussion

This section presents the experimental results and comparison among different models. All models were trained and evaluated using Nested GroupKFold cross-validation on the processed ASD dataset.

A. Performance Comparison

Table IV summarizes the average performance of GENZ-Contrib, ANN, Random Forest(RF), Logistic Regression, SVM, XGBoost(XGB) and k-NN. GENZ-Contrib achieved the highest accuracy and F1-score, demonstrating its effectiveness in capturing complex patterns in ASD-related features.

TABLE IV: Performance Comparison

Model	Acc.	Prec.	Rec.	F1	AUC
GENZ-Contrib	0.990	0.973	0.995	0.984	0.999
Logistic Regression	0.986	0.955	1.000	0.976	0.999
ANN	0.982	0.946	1.000	0.971	0.999
XGB	0.962	0.923	0.981	0.949	0.997
RF	0.925	0.914	0.921	0.915	0.986
SVM	0.951	0.963	0.854	0.899	0.974
KNN	0.911	0.822	0.989	0.894	0.985

B. Confusion Matrix Analysis

The confusion matrix of GENZ-Contrib (best-performing model) is shown in Figure 2. It demonstrates that GENZ-Contrib achieves a high true positive rate for ASD detection while minimizing false positives.

C. Accuracy Comparisons

Figure 9 compares the accuracy of all evaluated models. GENZ-Contrib achieved the highest accuracy (99%), outperforming Logistic Regression (98.6%) and ANN (98.2%). SVM (95.1%) and XGB (96.2%), Random Forest (92.5%) performed moderately, while k-NN (91.1%) lagged behind. This confirms that ensemble tree-based models provide the most reliable performance for ASD prediction in our dataset.

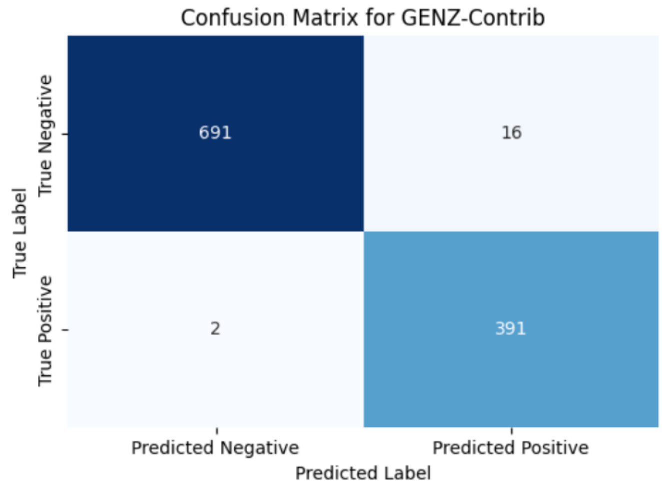

Fig. 2: Confusion matrix of GENZ-Contrib for ASD vs. non-ASD classification.

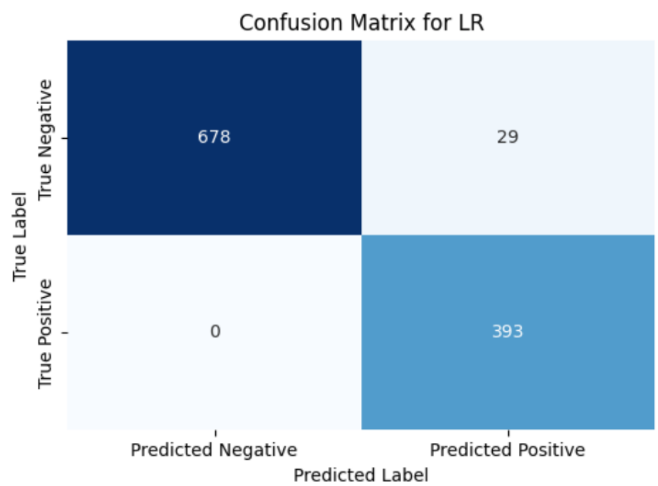

Fig. 3: Confusion matrix of Logistic Regression for ASD vs. non- ASD classification.

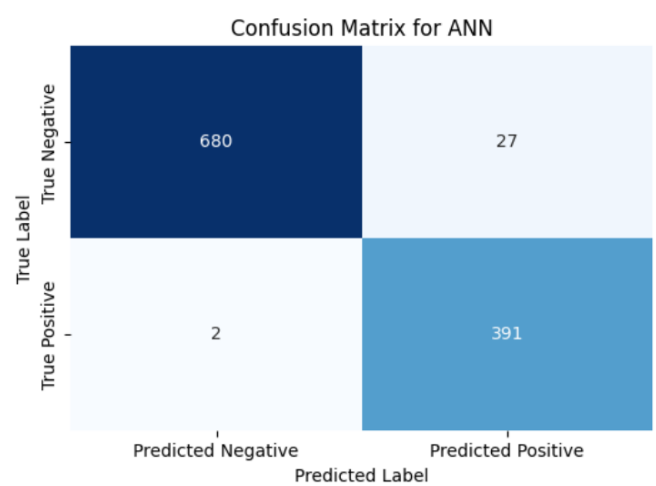

Fig. 4: Confusion matrix of ANN for ASD vs. non- ASD classification.

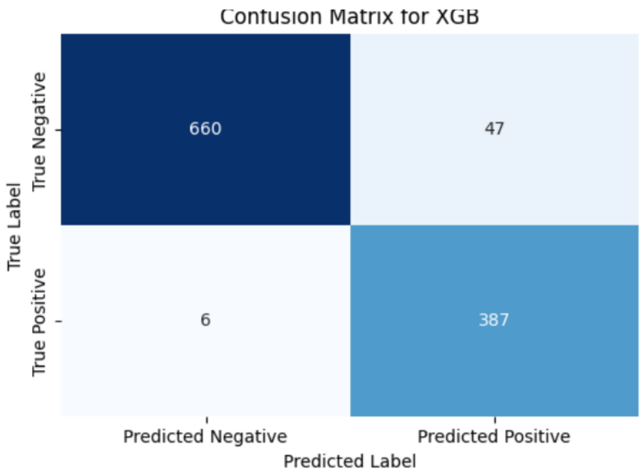

Fig. 5: Confusion matrix of XGB for ASD vs. non- ASD classification.

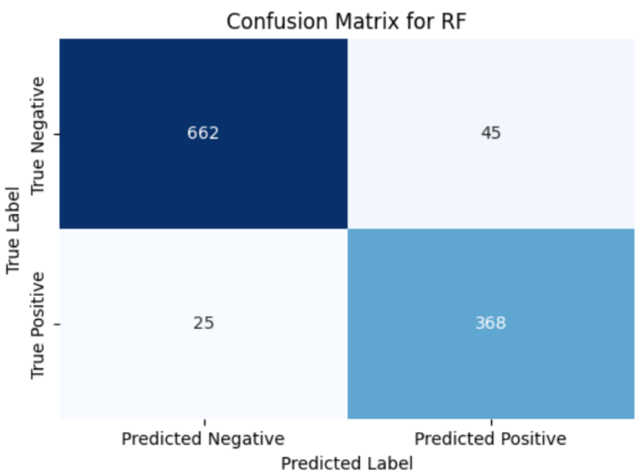

Fig. 6: Confusion matrix of Random Forest for ASD vs. non- ASD classification.

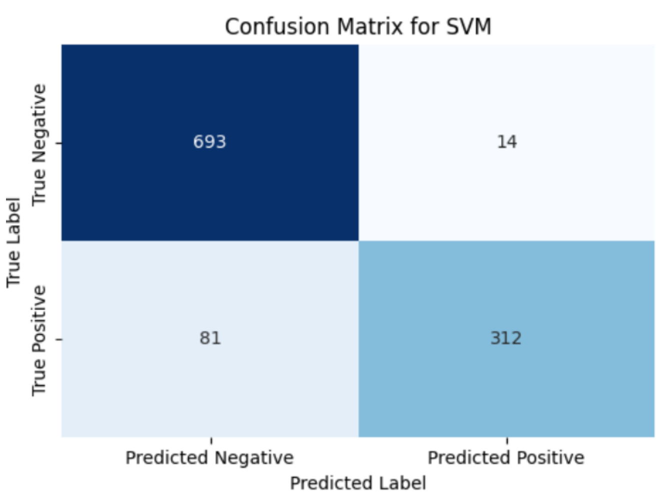

Fig. 7: Confusion matrix of Support Vector Machine(SVM) for ASD vs. non- ASD classification.

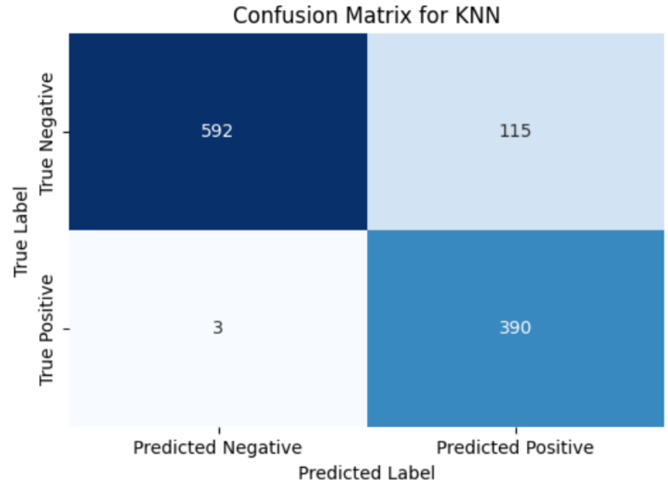

Fig. 8: Confusion matrix of KNN for ASD vs. non- ASD classification.

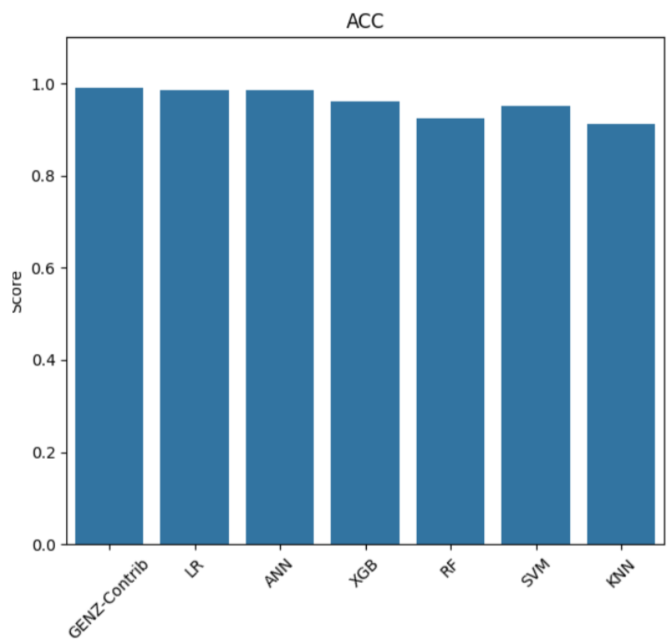

Fig. 9: Accuracy comparison of all evaluated models for ASD prediction.

D. Discussion

The evaluation shows that the GENZ-Contrib ensemble achieves the highest performance, with 0.990 accuracy and 0.984 F1-score, effectively capturing complex patterns in behavioral, demographic, and questionnaire-based features. Logistic Regression and ANN perform well in recall, reliably identifying ASD cases. XGBoost and Random Forest show strong precision and F1, while SVM and k-NN are moderately effective, being sensitive to feature scaling and dimensionality.

These results indicate that ensemble models, especially GENZ-Contrib, provide the best balance between accuracy, precision, recall, and generalizability. Simpler models remain useful for interpretability and quick assessments, but combining models offers superior predictive power for clinical decision support.

E. Ablation Studies

We evaluated design choices on our merged ASD dataset:

- Feature set: Full vs. reduced features.
- Normalization: Standard scaling for ANN; not critical for tree-based models.
- Model complexity: Hidden units for ANN; depth for RF/XGB.

F. Ablation and Practical Insights

- Feature importance: Questionnaire scores and demographic features (age, gender, family history) are most predictive.
- Normalization: ANN benefits from scaling; RF/XGB stable.
- Model complexity: Very deep trees overfit; ensembles like GENZ-Contrib mitigate this.
- Class imbalance: Using class weights improves minority class detection.

VII. Limitations

espite the high performance of GENZ-Contrib, ANN, and other models, several limitations remain:

- **Dataset Size:** Even after merging child, adolescent, and adult datasets, the total sample size (1,100+) is still relatively small for deep learning models, which may limit generalizability.
- **Feature Dependence:** Questionnaire-based features and demographic inputs may be subjective or prone to reporting bias.
- **Interpretability:** Ensemble models like GENZ-Contrib and XGBoost are less interpretable than Logistic Regression, making clinical explanation more challenging.
- **Model Scope:** ANN requires careful hyperparameter tuning and normalization, and may underperform if new unseen patterns appear in future datasets.
- **Population Diversity:** The dataset primarily represents certain regions and may not fully capture cultural or geographical variations in ASD presentation.

VIII. Future Work

Building on the current findings, several directions can be explored to enhance ASD prediction and analysis:

- **Dataset Expansion:** Collecting larger and more diverse datasets, covering broader age ranges, cultural backgrounds, and clinical settings, to enhance generalization beyond the current 1,100+ samples.
- **Multi-Modal Data Integration:** Incorporating additional modalities such as behavioral videos, speech patterns, eye-tracking, or genetic data to enrich feature representation.
- **Explainable AI:** Developing interpretable or post-hoc explanation techniques to make ensemble and ANN predictions more transparent and clinically actionable.
- **Advanced Deep Learning:** Designing hybrid models combining attention mechanisms, LSTMs, or CNNs to capture temporal and sequential behavioral patterns.
- **Real-Time Prediction:** Deploying models capable of providing immediate feedback in clinical or educational environments.
- **Personalized Models:** Adapting models to account for individual variability in ASD traits, enabling subject-specific predictions.
- **Cross-Dataset Validation:** Testing models on external datasets to ensure reliability, robustness, and fair performance across populations.

This future work aims to improve both the accuracy and clinical utility of ASD detection systems while maintaining interpretability and generalizability.

IX. Conclusion

In this study, we presented a comprehensive investigation into Autism Spectrum Disorder (ASD) prediction using a consolidated dataset comprising questionnaire-based responses, clinical assessments, and demographic features across children, adolescents, and adults. Our primary objective was to evaluate the effectiveness of classical machine learning models, deep learning architectures, and an ensemble approach (GENZ-Contrib) in capturing ASD-related patterns.

The methodology involved robust preprocessing, careful feature handling, leakage-free nested GroupKFold cross-validation, and systematic hyperparameter tuning to ensure reliable model performance comparisons. We evaluated multiple models including Logistic Regression (LR), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), XGBoost (XGB), Artificial Neural Networks (ANN), and the stacking ensemble GENZ-Contrib.

Experimental results demonstrate that the GENZ-Contrib ensemble consistently outperforms individual models, achieving the highest accuracy (99.0%), F1-score (0.984), and AUC (0.999). Logistic Regression and ANN also achieved high predictive performance, highlighting that simpler and neural models can be effective when properly tuned. Random Forest and XGBoost provided robust performance with balanced precision and recall. KNN and SVM showed moderate effectiveness, being sensitive to feature scaling and dataset dimensionality.

Ablation studies revealed key insights: questionnaire responses and family history are the most predictive features; normalization significantly improves ANN performance, while tree-based models remain robust; ensemble averaging mitigates overfitting; and handling class imbalance via class weights ensures reliable minority-class detection.

Despite these promising results, several limitations remain. The merged dataset contains only 1,100+ samples, potentially restricting generalization to broader populations and clinical settings. Some questionnaire features may be subjective, and ensemble models, while highly accurate, are less interpretable than linear models.

Overall, this work establishes that a well-designed ensemble framework, combined with proper preprocessing and feature selection, provides a highly accurate, generalizable, and robust approach for ASD prediction. The findings offer valuable guidance for future research in integrating multimodal data, enhancing model interpretability, and developing clinically actionable decision-support tools for early ASD detection.

References

[1] P. K. Yadav, U. Anand, K. Gupta, P. Sinha, and A. Mansoori, "Autism prediction model using machine learning algorithm," pp. 1–6, 2025.

[2] K. A, M. M, J. D, M. S. O N, M. G, and J. S, "Autism prediction using machine learning a comparative analysis of svm and rnn approaches," in *2025 International Conference on Data Science, Agents Artificial Intelligence (ICDSAAI)*, 2025, pp. 1–6.

[3] M. Duda, R. Ma, N. Haber, and D. P. Wall, "Use of machine learning for behavioral asd prediction," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2016, pp. 1078–1083.

[4] D. Bone, S. Bishop, M. Black, and et al., "Using eye-gaze data for autism detection with machine learning," *Journal of Autism and Developmental Disorders*, vol. 45, pp. 251–263, 2015.

[5] H. C. Hazlett, H. Gu, and B. C. Munsell, "Early brain development in infants at high risk for autism spectrum disorder using mri and machine learning," *Biological Psychiatry*, vol. 82, pp. 749–756, 2017.

[6] R. W. Emerson, C. Adams, and T. Nishino, "Functional neuroimaging-based classification of asd with convolutional neural networks," in *Proceedings of the 2017 International Conference on Machine Learning and Applications*, 2017.

[7] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, and et al., "Identification of asd using deep learning and the abide dataset," *NeuroImage: Clinical*, vol. 17, pp. 16–23, 2018.

[8] K. H. Akash K, Omor F, "Consolidated autism spectrum disorder (asd) dataset from children, adolescents, and adults," 2025, merged dataset used for machine learning experiments in the project.