IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, April 2026
Impact Factor: 8.2

Hybrid Machine Learning Models for Children with Autism Spectrum Disorder

[1]Mrunmayee Aundhekar and [2]Ms. Madhubala Chaudhari
Department of Computer Science and Engineering[1,2]
Deogiri Institute of Engineering & Management Studies, Chh. Sambhajinagar, Maharashtra, India

Abstract: *Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in communication, social interaction, and behavioral patterns. Due to its highly heterogeneous nature, early diagnosis and the design of personalized educational interventions remain significant challenges. It is well established that each child with ASD exhibits unique characteristics, making it unlikely for a single teaching approach to be effective for all learners. Despite this, many educational systems continue to rely on generalized instructional methods, which often fail to address the individualized learning needs of children with ASD.*

To address this limitation, this study proposes a machine learning–driven framework for both ASD detection and personalized teaching strategy recommendation. The proposed system adopts a two-phase architecture inspired by recent advancements in data-driven ASD analysis. In the first phase, a hybrid ensemble model integrating Decision Tree and XGBoost algorithms is employed to identify the presence of ASD based on behavioral and screening- related features, In the second phase, feature selection techniques are utilized to determine the most significant behavioral, verbal, and physical attributes influencing learning patterns in children with ASD. The system leverages multiple attributes derived from screening questionnaires and behavioral indicators to enhance prediction accuracy and support personalized educational recommendations. By combining ensemble learning techniques with effective feature engineering, the proposed approach improves both diagnostic reliability and the personalization of teaching methods. The findings demonstrate that machine learning can serve as a valuable tool for educators and healthcare professionals in identifying ASD characteristics and recommending suitable instructional strategies, ultimately enabling children with ASD to receive education aligned with their unique abilities and learning preferences..

Keywords: Autism Spectrum Disorder, Machine Learning, Decision Tree , XGBoost algorithms

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects an individual's ability to communicate, interact socially, and exhibit typical behavioral patterns. The prevalence of ASD has increased significantly over the past decade, highlighting the need for effective diagnostic and intervention strategies. One of the major challenges associated with ASD lies in its heterogeneous nature, as individuals on the spectrum display a wide range of symptoms, abilities, and learning preferences. This diversity makes early diagnosis complex and necessitates highly personalized approaches to education and intervention.

Traditional educational practices often rely on general teaching methods, which may not adequately address the specific needs of children with ASD. Since each child demonstrates unique behavioral, cognitive, and communication characteristics, a one-size-fits-all teaching strategy can lead to ineffective learning outcomes. Consequently, there is a growing demand for intelligent systems that can assist educators in designing customized learning experiences tailored to individual requirements [1].

Recent advancements in machine learning (ML) have shown great potential in transforming healthcare and education by enabling data-driven decision-making. In the context of ASD, ML techniques have been widely explored for early

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, April 2026
Impact Factor: 8.2

detection using behavioral and screening data. However, limited research has focused on integrating diagnostic systems with personalized teaching strategy recommendations, which are equally critical for improving educational outcomes. This gap highlights the need for a unified framework that not only detects ASD accurately but also supports individualized learning interventions [2].

In this study, a machine learning–based framework is proposed to address both ASD detection and personalized teaching method recommendation. The system is designed using two-phase architecture. In the first phase, a hybrid ensemble model combining Decision Tree and XGBoost algorithms is employed to classify individuals based on behavioral and screening features, achieving high diagnostic accuracy. In the second phase, feature selection techniques are applied to identify the most influential characteristics affecting learning behavior, and the same hybrid model is used to recommend suitable teaching strategies tailored to each child's needs.

The proposed approach leverages ensemble learning and feature engineering to enhance prediction performance and ensure more reliable outcomes. By analyzing data derived from screening questionnaires and behavioral indicators, the system aims to bridge the gap between diagnosis and intervention. The integration of detection and recommendation within a single framework provides a comprehensive solution for supporting children with ASD [3].

The primary objective of this research is to assist educators and healthcare professionals in making informed decisions by providing accurate diagnostic insights and personalized teaching recommendations. Ultimately, this work seeks to improve the quality of education for children with ASD by aligning instructional methods with their individual abilities, thereby fostering better learning outcomes and overall development.

II. LITERATURE SURVEY

Autism Spectrum Disorder (ASD) has gained significant attention in recent years due to its increasing prevalence and the challenges associated with early diagnosis. Traditional diagnostic approaches rely heavily on clinical observations and standardized assessment tools, which can be time-consuming, subjective, and dependent on expert availability. To address these limitations, researchers have increasingly explored the use of machine learning (ML) techniques for accurate and early detection of ASD[4][5].

Several studies have demonstrated the effectiveness of supervised machine learning algorithms in ASD classification tasks. Algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbours (KNN) have been widely applied to behavioural and screening datasets. Among these, Random Forest has shown strong performance due to its ability to handle high-dimensional data and reduce overfitting through ensemble learning. Similarly, SVM has been effective in capturing complex decision boundaries, particularly in smaller datasets [6][7].

In recent years, ensemble and boosting techniques have gained prominence for improving classification accuracy. XGBoost (Extreme Gradient Boosting) has emerged as a powerful algorithm due to its ability to handle missing data, incorporate regularization, and model complex nonlinear relationships. Studies have reported that XGBoost outperforms many traditional algorithms when applied to structured datasets such as ASD screening questionnaires and demographic data.

Some researchers have proposed hybrid models that combine multiple algorithms to leverage their individual strengths. For instance, integrating Decision Tree with boosting techniques or combining Random Forest with gradient boosting methods has resulted in improved prediction accuracy and robustness. These hybrid approaches are particularly effective in handling imbalanced datasets and capturing diverse feature interactions[8][9].

Despite the progress in ASD detection, most existing studies focus primarily on classification and early diagnosis, with limited attention given to post-diagnosis interventions such as personalized teaching strategies[10][11]. This highlights a research gap in integrating diagnostic systems with recommendation frameworks for individualized education

III. SYSTEM DEVELOPMENT

The system development process is organized into two distinct phases. Phase I focuses on the diagnosis of autism spectrum disorder (ASD) by utilizing a combination of statistical techniques and machine learning algorithms applied

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 4, April 2026 Impact Factor: 8.2

to an ASD dataset. Phase II is dedicated to determining the most suitable teaching methods for children with ASD. This phase leverages the outcomes and insights derived from the diagnostic stage to recommend appropriate educational strategies tailored to individual learning needs. Figure 1 shows the block diagram for work to be done.

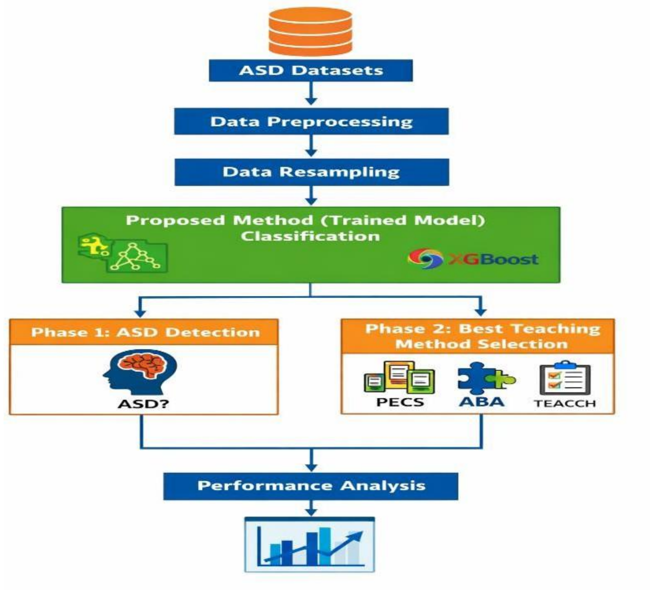

Figure 1. Block diagram for work to be done.

3.1 Dataset: The proposed framework utilizes two publicly available datasets focused on early ASD screening in toddlers. These datasets provide behavioral and screening-related attributes that are essential for both ASD detection and personalized teaching method recommendation.

a) Autism Spectrum Disorder Screening Data for Toddlers

This dataset contains screening information collected from toddlers to identify early signs of ASD. It includes a variety of features such as behavioral responses, social interaction indicators, communication abilities, and demographic details. The dataset is primarily based on questionnaire responses designed to capture early developmental traits associated with ASD. These attributes are used as input features in the machine learning model to detect the presence of ASD.

b) ASD Screening Data for Toddlers in Saudi Arabia

This dataset is a region-specific extension of ASD screening data, collected from toddlers in Saudi Arabia. It includes similar behavioral and screening features, along with additional contextual and demographic information relevant to the population. The dataset enhances the diversity and robustness of the model by incorporating data from different backgrounds, thereby improving generalization and prediction accuracy.

3.2 Teaching Methodology:

To provide personalized educational recommendations for children with autism spectrum disorder (ASD), this study considers three widely recognized and evidence-based teaching methodologies: Picture Exchange Communication System (PECS), Applied Behavior

Analysis (ABA), and TEACCH (Treatment and Education of Autistic and Communication- Handicapped Children). These approaches are selected due to their proven effectiveness in addressing communication, behavioral, and learning challenges in children with ASD.

3.3. Hybrid Method Based on Random Forest and XGBoost

Random Forest (RF) is an ensemble learning technique that generates multiple decision trees by using randomly selected subsets of data samples and features [12][13]. Each individual tree makes an independent prediction regarding the presence of ASD in a child, and the overall output is obtained through a majority voting mechanism. RF is known for its robustness against overfitting and its ability to effectively manage noisy and high-dimensional data, which are common characteristics of ASD screening datasets.

XGBoost is an advanced gradient boosting algorithm that constructs decision trees in a sequential manner, where each subsequent tree aims to correct the errors made by the preceding ones. It incorporates regularization methods to control overfitting and performs exceptionally well on structured tabular data, such as behavioural assessments and demographic information. Additionally, XGBoost can capture complex nonlinear patterns and feature interactions within ASD datasets. During the data preprocessing stage, ASD- related data is collected, including screening questionnaire responses, behavioural attributes, demographic information, and family history. The dataset then undergoes cleaning procedures to eliminate missing or inconsistent values, manage outliers, and convert categorical variables into numerical formats. Subsequently, feature scaling is applied where necessary, particularly to enhance the performance of models such as XGBoost. These teaching methods are incorporated into the proposed system as target classes for recommendation.

3.4. Hybrid Method Based on Decision Tree and XGBoost

To improve the accuracy and robustness of autism spectrum disorder (ASD) detection and personalized teaching recommendation [14][15], a hybrid machine learning approach combining Decision Tree (DT) and XGBoost is proposed. This hybrid model leverages the strengths of both algorithms to achieve better predictive performance on ASD screening datasets. The hybrid model first utilizes the interpretability of Decision Trees to identify important features and initial patterns in the dataset. Subsequently, XGBoost enhances prediction accuracy by learning complex relationships and reducing errors through boosting. This combination results in a more reliable and efficient model for ASD detection. In Phase I, the hybrid model is used to classify whether a child has ASD based on behavioural, social, and demographic features. In Phase II, the same model, along with feature selection techniques, is applied to recommend the most appropriate teaching method (such as PECS, ABA, or TEACCH) tailored to the child’s individual characteristics.

IV. RESULT ANALYSIS

The performance evaluation of the proposed system is conducted using two hybrid machine learning approaches: Random Forest combined with XGBoost and Decision Tree combined with XGBoost. The analysis is carried out on two benchmark datasets, namely the Autism Spectrum Disorder Screening Data for Toddlers and the ASD Screening Data for Toddlers in Saudi Arabia, to ensure robustness and generalizability of the results. The evaluation is performed across two distinct phases of the system. In Phase I, the focus is on ASD detection, where both hybrid models are assessed based on their ability to accurately classify whether a child exhibits ASD traits. In Phase II, the performance is measured in terms of the system’s capability to recommend the most appropriate teaching method—such as PECS, ABA, or TEACCH—based on the identified behavioural and developmental characteristics.

Various performance metrics, including accuracy, precision, recall, and F1-score, are used to compare the effectiveness of the two hybrid approaches. The comparative analysis demonstrates the strengths and limitations of each model in both diagnostic and recommendation tasks. Overall, this dual-phase and multi-model evaluation ensures a comprehensive assessment of the system’s efficiency in both ASD identification and personalized teaching method selection. The most used evaluation metrics in ASD classification problems include Accuracy, Precision, Recall (Sensitivity), and F1 Score.

The evaluation of classification models using performance metrics such as accuracy, precision, recall, and F1-score enables the development of more reliable and clinically effective ASD diagnostic systems. In this study, performance

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 6, Issue 4, April 2026 Impact Factor: 8.2

analysis is conducted using hybrid approaches, namely Decision Tree combined with XGBoost (DT + XGBoost) and Random Forest combined with XGBoost (RF + XGBoost), along with several individual machine learning algorithms including K-Nearest Neighbours (KNN), Random Forest, Decision Tree, XGBoost, and Support Vector Machine (SVM).

After a comprehensive comparison of all models, the results indicate that the hybrid RF + XGBoost approach achieves the highest performance, with an accuracy of 98.29%, outperforming both the standalone algorithms and the other hybrid method. This demonstrates the effectiveness of combining ensemble techniques for improving ASD detection accuracy. Table 1 shows the comparative results.

Table 1. Comparative results with other state of art methods

	model	cv_accuracy_mean	cv_accuracy_std	test_accuracy	test_precision_w	test_recall_w	test_f1_w
0	LogReg	1.000000	0.000000	1.000000	1.000000	1.000000	1.000000
1	SVM(RBF)	0.998169	0.002242	0.995726	0.995726	0.995726	0.995726
2	GBM	0.976189	0.007324	0.985043	0.985025	0.985043	0.985029
3	XGBoost	0.985342	0.008388	0.985043	0.985025	0.985043	0.985029
4	Hybrid(RF+XGB)_Vote	0.976184	0.010978	0.982906	0.982906	0.982906	0.982906
5	RandomForest	0.961527	0.009904	0.974359	0.974439	0.974359	0.974211
6	Hybrid(DT+XGB)_Vote	0.956035	0.021400	0.972222	0.972720	0.972222	0.972345
7	DecisionTree	0.931310	0.019900	0.959402	0.959336	0.959402	0.959364
8	KNN	0.927636	0.017317	0.955128	0.959098	0.955128	0.955744

In Phase II, the system focuses on identifying the most suitable teaching method among three approaches: PECS (Picture Exchange Communication System), ABA (Applied Behavior Analysis), and TEACCH (Treatment and Education of Autistic and Related Communication- Handicapped Children). During this phase, feature selection techniques are employed to extract the most relevant behavioral, verbal, and physical characteristics of children with ASD.

Feature	Type	Description
A1: Question 1 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A2: Question 2 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A3: Question 3 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A4: Question 4 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A5: Question 5 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A6: A6: Question 6 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A7: Question 7 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A8: Question 8 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A9: Question 9 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
A10: Question 10 Answer	Binary (0, 1)	The answer code of the question based on the screening method used
Region	String	List of Saudi regions
Age	Number	Toddlers (months)
Gender	String	Male or Female
Screening Score	Number	1-10 (the final score obtained based on the scoring algorithm of the screening method used. Less than or equal 3 no ASD traits; > 3 ASD traits)
Family member with ASD history	Boolean (yes or no)	Whether any immediate family member has a ASD
Who is completing the test	String	Parent, other
Class	Number	0 or 1 (0 : no ASD traits, 1: ASD traits)

Table 2: Screening Features Q-CHAT-10 Question

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 4, April 2026
Impact Factor: 8.2

Based on these selected features, the same hybrid model is utilized to recommend personalized teaching strategies tailored to the individual needs of each child, achieving an accuracy of 97%. The model relies on a set of screening features derived from 10 key questionnaire-based attributes, which play a significant role in influencing the outcome. These features and their impact on the results are presented in the corresponding table.2

	model	cv_accuracy_mean	cv_accuracy_std	test_accuracy	test_precision_w	test_recall_w	test_f1_w
0	Hybrid(DT+XGB)_Vote	0.984605	0.006301	0.977778	0.977814	0.977778	0.977695

Table 3: Phase 2 results

The analysis reveals that, after applying feature selection techniques and the hybrid algorithm, several features achieved higher importance rankings. As presented in Table 4.2, the XGBoost classifier identified features A6, A9, A5, A7, and A4 as the most significant. In contrast, the Decision Tree classifier ranked A9, A6, A2, and A7 as the top contributing features.

Furthermore, the results obtained through K-fold cross-validation indicate that the hybrid approach combining Decision Tree and XGBoost achieves superior accuracy compared to individual models, as illustrated in the corresponding table 3.

V. CONCLUSION:

This study presents a two-phase hybrid machine learning framework designed for the early detection of autism spectrum disorder (ASD) and the recommendation of personalized teaching methods for toddlers. In Phase I, two hybrid models—Random Forest combined with XGBoost, and Decision Tree combined with XGBoost—were evaluated. The RF + XGBoost model achieved a slightly higher accuracy of 98.29% compared to 97.22% obtained by the Decision Tree + XGBoost model. However, the difference in performance is marginal, indicating that the proposed model remains highly reliable. In Phase II, the Decision Tree + XGBoost model showed strong capability in recommending appropriate teaching strategies, achieving an accuracy of 97.77%

REFERENCES

[1] Hajjej, S. Ayouni, M. A. Alohali and M. Maddeh, "Novel Framework for Autism Spectrum Disorder Identification and Tailored Education With Effective Data Mining and Ensemble Learning Techniques," in IEEE Access, vol. 12, pp. 35448-35461, 2024, doi: 10.1109/ACCESS.2024.3349988

[2] Zoana, Zarin & Shafeen, Mahmudul & Akter, Nasrin & Rahman, Tanvir. (2023). Application of Machine Learning in Identification of Best Teaching Method for Children with Autism Spectrum Disorder. 10.48550/arXiv.2302.05035.

[3] Raj, Suman & Masood, Sarfaraz. (2020). Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques. Procedia Computer Science. 167. 994- 1004. 10.1016/j.procs.2020.03.399.

[4] Anita Vikram Shinde, Dipti Durgesh Patil,A Multi-Classifier-Based Recommender System for Early Autism Spectrum Disorder Detection using Machine Learning, Healthcare Analytics, Volume 4, 2023, 100211, ISSN 2772-4425

[5] Baadel, S., Thabtah, F., & Lu, J. (2020). A clustering approach for autistic trait classification. Informatics for Health and Social Care, 45(3), 309–326. https://doi.org/10.1080/17538157.2019.1687482

[6] Jabbar, Usama, Iqbal, Muhammad Waseem, Alourani, Abdullah, Shinan, Khlood, Alanazi, Fatmah, Sarwar, Nadeem, Ashraf, M. Usman, Machine Learning–Based Approach for Early Screening of Autism Spectrum Disorders, Applied Computational Intelligence and Soft Computing, 2025, 9975499, 27 pages, 2025.

[7] Toteja, S., & Bajpai, P. (2024). Leveraging Machine Learning to Identify Effective Teaching Strategies for Children With Autism Spectrum Disorder. 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), 2038-2043.

[8] Hasan, S M & Uddin, Md Palash & Mamun, Md. Al & Sharif, Muhammad & Ulhaq, Anwaar & Krishnamoorthy, Govind. (2022). A Machine Learning Framework for Early- Stage Detection of Autism Spectrum Disorders. IEEE Access. PP. 1-1. 10.1109/ACCESS.2022.3232490.

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 6, Issue 4, April 2026 Impact Factor: 8.2

[9] Themistocleous, C. K., Andreou, M., & Peristeri, E. (2024). Autism Detection in Children: Integrating Machine Learning and Natural Language Processing in Narrative Analysis. Behavioral Sciences, 14(6), 459. https://doi.org/10.3390/bs14060459

[10] Balaji, V., Kanaga Suba Raja, S. (2022). Recommendation Learning System Model for Children with Autism. Intelligent Automation & Soft Computing, 31(2), 1301– 1315. https://doi.org/10.32604/iasc.2022.020287

[11] Joudar, Shahad & Albahri, A.s & Amjed, Rula. (2022). Intelligent triage method for early diagnosis autism spectrum disorder (ASD) based on integrated fuzzy multi-criteria decision-making methods. Informatics in Medicine Unlocked. 36. 101131. 10.1016/j.imu.2022.101131.

[12] Pamungkas, Yuri & Rangkuti, Rahmah & Karim, Abdul & Sangsawang, Thosporn. (2025). Recent Advances in Artificial Intelligence for Dyslexia Detection: A Systematic Review. International Journal of Robotics and Control Systems. 5. 2016-2033. 10.31763/ijrcs.v5i3.2057.

[13] Aldrees A, Ojo S, Wanliss J, Umer M, Khan MA, Alabdullah B, Alsubai S and Innab N (2024) Data-centric automated approach to predict autism spectrum disorder based on selective features and explainable artificial intelligence. Front. Comput. Neurosci. 18:1489463. doi: 10.3389/fncom.2024.1489463

[14] M, T. ., Madhuri, D. K. ., Sundar, R. ., Parasa, G. ., Duraivelu, V. ., & Krishnaveni, N. . (2025). Recommendation learning management system for autism using deep convolutional neural networks and gene expression programming. Edelweiss Applied Science and Technology, 9(2), 910–935. https://doi.org/10.55214/25768484.v9i2.4625

[15] Jeon, I., Kim, M., So, D., Kim, E. Y., Nam, Y., Kim, S., Shim, S., Kim, J., & Moon, J. (2024). Reliable Autism Spectrum Disorder Diagnosis for Pediatrics Using Machine Learning and Explainable AI. Diagnostics, 14(22), 2504. https://doi.org/10.3390/diagnostics14222504