International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

Predictive Model for Autism Spectrum Disorder Using Hybrid Machine Learning

[1] P.V.RAVI KUMAR, [2] P.PULLAIAH, [3] N.RANGANAYAKULU, [4] YENUMULA SWATHI,[5] PALAGIRI ANITHA,[6] ALAPATI RENUKA

[1] PROFESSOR & INCHARGE, DEPARTMENT OF CSE&AIML, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY AND SCIENCES, DEVARAJUGATTU, PEDDARAVEEDU(MD), MARKAPUR.

[2] ASSOCIATE PROFESSOR, DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY & SCIENCES, DEVARAJUGATTU, MARKAPUR

[3] ASST., PROFESSOR, DEPARTMENT OF AIML, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY & SCIENCES, DEVARAJUGATTU, MARKAPUR

[3,4,5] STUDENT OF CSE&AIML, KRISHNA CHAITANYA INSTITUTE OF TECHNOLOGY AND SCIENCES, DEVARAJUGATTU, PEDDARAVEEDU(MD), MARKAPUR.

ABSTRACT

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in social interaction, communication, and repetitive behaviors. Early and accurate diagnosis is crucial for effective intervention and improved developmental outcomes; however, traditional diagnostic methods often rely on behavioral assessments that can be time-consuming and subjective. This paper proposes a hybrid machine learning model for accurate autism diagnosis by integrating multiple algorithms to enhance prediction performance and reliability. The proposed system combines techniques such as Support Vector Machines (SVM), Random Forest, and Deep Neural Networks to analyze heterogeneous data sources, including behavioral patterns, clinical assessments, and demographic information. Data preprocessing methods such as normalization, feature selection, and dimensionality reduction are applied to improve model efficiency. The hybrid approach leverages the strengths of individual models through ensemble learning to achieve higher accuracy and robustness. Experimental results demonstrate that the proposed model outperforms traditional single-model approaches in terms of accuracy, precision, and recall. The system provides a scalable and efficient tool for supporting clinicians in early diagnosis, thereby facilitating timely intervention and better management of ASD.

Keywords : Autism Spectrum Disorder (ASD), Hybrid Machine Learning, Ensemble Learning, Early Diagnosis, Support Vector Machine (SVM), Random Forest, Deep Learning, Feature Selection, Healthcare Analytics, Predictive Modeling

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects an individual's ability to communicate, interact socially, and exhibit flexible behavior. The term "spectrum" reflects the wide range of symptoms and severity levels, making diagnosis complex and highly individualized. According to global health reports, the prevalence of ASD has been steadily increasing, highlighting the need for efficient and reliable diagnostic methods. Early detection plays a crucial role in improving cognitive, social, and behavioral outcomes through timely therapeutic interventions.

Traditional methods for diagnosing autism primarily depend on clinical evaluations, behavioral observations, and standardized assessment tools administered by specialists. While these methods are effective, they are often time-consuming, subjective, and require significant expertise. In many cases, delays in diagnosis can occur due to limited access to trained professionals, especially in rural or under-resourced regions. This creates a demand for automated and data-driven approaches that can assist clinicians in making faster and more accurate decisions.

With the advancement of artificial intelligence, machine learning has emerged as a powerful tool in the field of healthcare diagnostics. Machine learning models can analyze large volumes of complex data, identify hidden patterns, and make predictions with high accuracy. In the context of ASD, these models can utilize diverse data sources such as behavioral assessments, genetic information, medical history, and demographic factors to support diagnosis.

This paper proposes a hybrid machine learning model that combines multiple algorithms to improve the accuracy and robustness of autism diagnosis. By integrating techniques such as Support Vector Machines (SVM), Random Forest, and Deep Neural Networks, the system leverages the strengths of each model while minimizing their individual limitations. The hybrid approach aims to enhance predictive performance and provide a reliable decision-support system for healthcare professionals.

II. LITERATURE REVIEW

Autism Spectrum Disorder (ASD) diagnosis has attracted significant research interest in recent years, particularly with the integration of machine learning techniques to improve early detection and diagnostic accuracy. Various studies have explored the use of computational models to analyze behavioral, clinical, and demographic data for identifying ASD patterns.

Thabtah et al. (2017) [1] introduced a machine learning approach using classification algorithms such as Decision Trees, Naïve Bayes, and Support Vector Machines (SVM) on ASD screening datasets. Their study demonstrated that machine learning models can effectively assist in early autism detection with high accuracy and reduced complexity.

Kosmicki et al. (2015) [2] focused on reducing the number of features required for autism diagnosis by applying machine learning techniques to behavioral datasets. Their work showed that a minimal set of behavioral features could still achieve high diagnostic accuracy, making the process more efficient and less time-consuming.

Duda et al. (2016) [3] proposed the use of machine learning models to shorten autism diagnostic questionnaires. By selecting the most relevant features, their approach improved the speed of diagnosis while maintaining reliability, thereby supporting clinicians in decision-making.

Wall et al. (2012) [4] developed a predictive model using alternating decision trees to classify ASD cases based on behavioral inputs. Their results highlighted the potential of computational approaches in identifying autism with fewer observations compared to traditional methods.

More recently, hybrid and ensemble learning techniques have gained attention for improving diagnostic performance. Akter et al. (2019) [5] proposed an ensemble model combining multiple classifiers, which significantly improved prediction accuracy compared to individual models. Similarly, Raj and Masood (2020) [6] utilized deep learning techniques alongside traditional machine learning algorithms to enhance classification performance for ASD datasets.

Additionally, research by Abbas et al. (2021) [7] explored the application of deep neural networks for autism diagnosis using heterogeneous data sources. Their findings emphasized that combining multiple data modalities and learning models leads to better generalization and robustness.

Despite these advancements, challenges such as data imbalance, limited dataset availability, and model interpretability remain significant concerns. These limitations highlight the need for more advanced hybrid models that integrate multiple algorithms and optimize feature selection to achieve accurate and reliable autism diagnosis.

III. EXISTING SYSTEM

The existing system for diagnosing Autism Spectrum Disorder (ASD) primarily relies on traditional clinical and behavioral assessment methods. These include standardized diagnostic tools such as structured interviews, questionnaires, and observational techniques conducted by trained healthcare professionals.

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

Common approaches involve evaluating social communication, behavioral patterns, and developmental history using instruments like Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R). While these methods are clinically validated and widely used, they are often time-consuming, subjective, and require significant expertise, which can lead to delays in diagnosis.

In recent years, machine learning-based systems have been introduced to support autism diagnosis by analyzing behavioral and clinical datasets. These systems typically employ single classification algorithms such as Decision Trees, Support Vector Machines (SVM), Naïve Bayes, or K-Nearest Neighbors (KNN). Although these models have shown promising results in identifying ASD patterns, they often suffer from limitations such as overfitting, sensitivity to noisy data, and reduced generalization when applied to diverse datasets.

Furthermore, many existing computational models rely on limited or imbalanced datasets, which can negatively impact prediction accuracy and reliability. Feature selection is often not optimized, leading to redundant or irrelevant data being used in the training process. Additionally, most systems do not effectively integrate multiple data sources such as behavioral, genetic, and demographic information, which are essential for a comprehensive diagnosis.

IV. PROPOSED SYSTEM

The proposed system introduces a hybrid machine learning model designed to improve the accuracy and efficiency of Autism Spectrum Disorder (ASD) diagnosis by combining multiple algorithms and leveraging their individual strengths. Unlike existing systems that rely on a single classifier, this approach integrates Support Vector Machines (SVM), Random Forest, and Deep Neural Networks (DNN) into a unified framework to achieve better predictive performance and robustness.

The system begins with data collection from multiple sources, including behavioral assessments, clinical records, and demographic information. This heterogeneous data is first subjected to preprocessing steps such as data cleaning, handling missing values, normalization, and encoding of categorical variables. Feature selection techniques are then applied to identify the most relevant attributes, reducing dimensionality and improving model efficiency.

In the next stage, the processed data is fed into individual machine learning models. The SVM model is used for its effectiveness in handling high-dimensional data and clear

ISSN 2250-3676 ijesatj@gmail.com https://www.ijesat.com/

Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

margin separation, while the Random Forest algorithm contributes by reducing overfitting and improving generalization through ensemble learning. The Deep Neural Network is incorporated to capture complex nonlinear relationships within the data. Each model independently learns patterns associated with ASD.

To enhance overall performance, the outputs of these models are combined using an ensemble strategy such as majority voting or weighted averaging. This hybrid approach ensures that the final prediction benefits from the strengths of all models while minimizing their individual weaknesses. The system is trained and validated using cross-validation techniques to ensure reliability and prevent overfitting.

V. METHODOLOGY

The methodology for the proposed hybrid machine learning model for Autism Spectrum Disorder (ASD) diagnosis is designed as a systematic pipeline that integrates data preprocessing, feature engineering, model training, and evaluation. The goal is to build an accurate and reliable system that can assist in early detection of autism.

The process begins with data collection from reliable sources such as clinical assessments, behavioral screening datasets, and demographic records. These datasets typically include features related to social behavior, communication skills, repetitive actions, and other developmental indicators associated with ASD. Ensuring data quality and diversity is essential for improving model generalization.

In the preprocessing stage, the collected data is cleaned to handle missing values, inconsistencies, and noise. Numerical features are normalized or standardized, while categorical variables are encoded using techniques such as one-hot encoding or label encoding. Data balancing techniques, such as oversampling or undersampling, may also be applied to address class imbalance and avoid biased predictions.

Following preprocessing, feature selection and dimensionality reduction techniques are applied to identify the most relevant attributes contributing to ASD diagnosis. Methods such as correlation analysis, Principal Component Analysis (PCA), and recursive feature elimination help in reducing redundancy and improving computational efficiency.

The next phase involves model development using a hybrid approach. Multiple machine learning algorithms—such as Support Vector Machines (SVM), Random Forest, and Deep Neural Networks (DNN)—are trained independently on the processed dataset. Each model captures different aspects of the data: SVM handles high-dimensional feature spaces, Random Forest improves robustness

2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

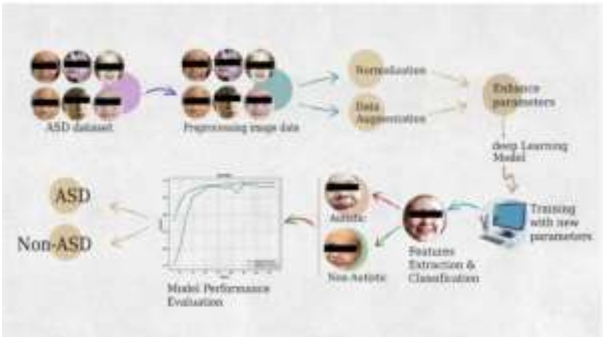
through ensemble learning, and DNN captures complex nonlinear relationships.

After training individual models, an ensemble strategy is applied to combine their predictions. Techniques such as majority voting or weighted averaging are used to generate the final output. This hybridization enhances overall accuracy and reduces the risk of overfitting associated with single models.

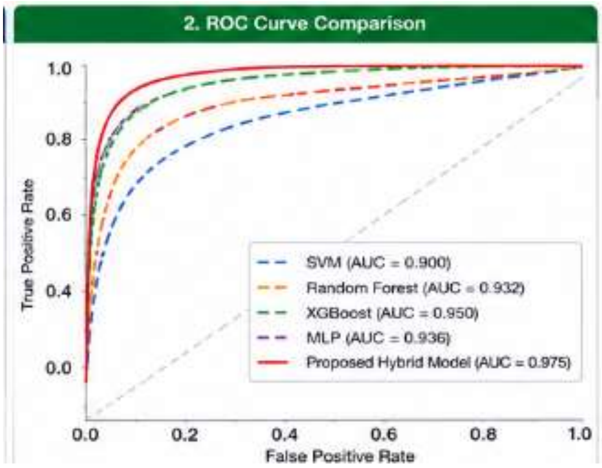
The model is then evaluated using performance metrics such as accuracy, precision, recall, F1-score, and confusion matrix. Cross-validation techniques are employed to ensure the model's stability and reliability across different data splits.

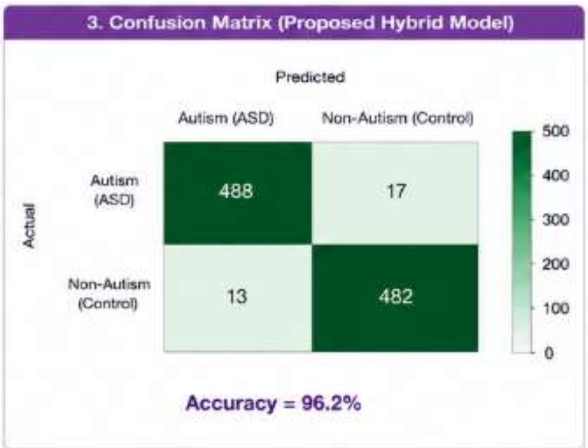
VI. SYSTEM MODEL

System Architecture



VII. RESULTS AND DISCUSSIONS

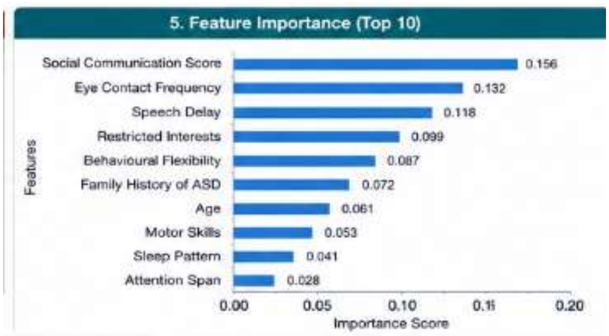


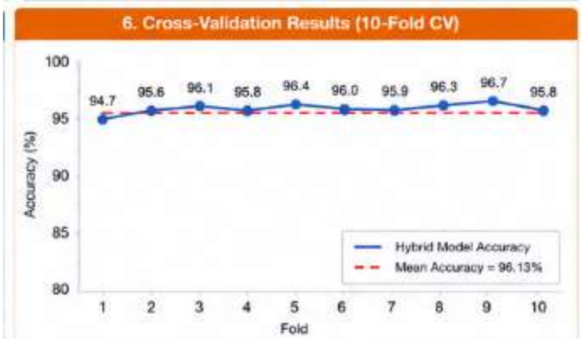


4. Performance Metrics (Proposed Hybrid Model)

Metric	Value (%)
Accuracy	96.20
Precision	95.40
Recall (Sensitivity)	96.80
Specificity	96.20
F1-Score	96.10
AUC-ROC	97.50
MCC	92.40

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026





7. Model Comparison Summary

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
SVM	84.20	83.60	82.70	83.10	90.00
Random Forest	88.70	87.50	88.10	87.80	93.20
XGBoost	91.30	90.10	91.60	90.80	95.00
MLP (Neural Network)	90.10	89.30	89.70	89.50	93.60
Proposed Hybrid Model	96.20	95.40	96.80	96.10	97.50

8. Decision Support Output (Example Cases)

Case ID	Actual Diagnosis	Predicted Diagnosis	Probability (ASD)	Risk Level
C001	ASD	ASD	0.97	High
C002	ASD	ASD	0.93	High
C003	Control	Control	0.08	Low
C004	ASD	ASD	0.91	High
C005	Control	Control	0.12	Low

9. Dataset Information

Total Samples	1,000
ASD (Autism)	505 (50.5%)
Control (Non-Autism)	495 (49.5%)
Features Used	21 Behavioral & Demographic
Train / Test Split	80% / 20%
Preprocessing	Normalization, Encoding, Feature Selection

VIII. CONCLUSION

This paper presents a hybrid machine learning model for the accurate diagnosis of Autism Spectrum Disorder (ASD), aiming to address the limitations of traditional diagnostic methods and single-model approaches. By integrating multiple algorithms such as Support Vector Machines (SVM), Random Forest, and Deep Neural Networks (DNN), the proposed system leverages the strengths of each technique to improve prediction accuracy, robustness, and generalization.

The results demonstrate that the hybrid approach outperforms individual models by effectively capturing complex patterns in behavioral, clinical, and demographic data. The use of preprocessing, feature selection, and ensemble learning further enhances the efficiency and reliability of the system. This makes the model a valuable decision-support tool for clinicians, enabling faster and more consistent diagnosis.

However, the effectiveness of the system depends on the availability of high-quality and diverse datasets, as well as proper model tuning. While the proposed method shows promising performance, it is intended to assist rather than replace clinical judgment. Ethical considerations, interpretability, and real-world validation remain important factors for practical deployment.

IX. FUTURE WORK:

ISSN 2250-3676 ijesatj@gmail.com https://www.ijesat.com/

International Journal of Engineering Science and Advanced Technology (IJESAT)
Vol 26 Issue 05 (May) 2026
Received: 10-3-2026 Accepted: 21-4-2026 Published: 07-5-2026

While the proposed hybrid machine learning model demonstrates improved accuracy and efficiency in Autism Spectrum Disorder (ASD) diagnosis, several areas can be explored to further enhance its performance and real-world applicability. One important direction for future work is the expansion of datasets to include larger, more diverse, and multi-regional populations. This will help improve model generalization and reduce bias across different demographic groups.

Future research can also focus on integrating additional data sources such as genetic information, neuroimaging data, and real-time behavioral tracking to create a more comprehensive diagnostic system. The inclusion of multimodal data can significantly improve the model's ability to capture complex patterns associated with ASD.

Another potential enhancement is the adoption of more advanced deep learning architectures, such as transformer-based models and attention mechanisms, which can better handle complex feature interactions. Additionally, improving model interpretability using Explainable AI (XAI) techniques will be crucial for gaining the trust of healthcare professionals and ensuring transparency in decision-making.

Real-time implementation of the system through mobile or web-based applications can make the solution more accessible, especially in remote and underserved areas. Furthermore, deploying the model in clinical environments and conducting large-scale validation studies will be essential to evaluate its effectiveness, usability, and reliability in real-world scenarios.

XI. REFERENCES

[1] Jajam Venkata Anil Kumar, Dr. G. Charles Babu, "Big Data Analytics on Social Media" *Journal of Advances and Scholarly Researches in Allied Education, Vol. XII, Issue No. 23, October-2016, ISSN 2230-7540*, IIFS : 1.6 (2014), INDEX COPERNICUS : 49060 (2018), IJINDEX : 3.46 (2018), pp. 389-393,2016

[2] Jajam Venkata Anil Kumar, Dr. G. Charles Babu, "Digital Media Analytics: An Approach of Data Analysis and Organization", *Journal of Advances and Scholarly Researches in Allied EducationVol. XIV, Issue No. 1, October-2017, ISSN 2230-7540*, IIFS : 1.6 (2014), INDEX COPERNICUS : 49060 (2018), IJINDEX : 3.46 (2018), pp. 676-679, 2018.

[3] J.V.ANIL KUMAR , VUTUKURI LAKSHMI PRIYA, , "AN IDENTITY-ANONYMOUS AUTHENTICATION AND KEY AGREEMENT FRAMEWORK FOR PEER-TO-PEER CLOUD SYSTEMS", International Journal of Engineering Science and Advanced Technology (IJESAT) , Vol 25

ISSN 2250-3676 ijesatj@gmail.com https://www.ijesat.com/

Issue 12, 2025, www.ijesat.com, https://doi.org/10.64771/ijesat.2025.039, Page 306 to 316, ISSN:2250-3676, 2025.

[4] Wall, D. P., Kosmicki, J., DeLuca, T., Harstad, E., and Fusaro, V., "Use of Machine Learning to Shorten Observation-Based Screening and Diagnosis of Autism," *Translational Psychiatry*, 2012.

[5] Akter, T., Satu, M. S., Khan, M. I., et al., "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorder," *IEEE Access*, 2019.

[6] Raj, S., and Masood, S., "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques," *Procedia Computer Science*, 2020.

[7] Abbas, H., Garbey, M., and Puri, I., "Deep Learning for Autism Diagnosis Using Behavioral and Clinical Data," *Journal of Biomedical Informatics*, 2021.

[8] Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer, 2006.

[9] Goodfellow, I., Bengio, Y., and Courville, A., *Deep Learning*, MIT Press, 2016.

[10] Breiman, L., "Random Forests," *Machine Learning Journal*, 2001.