ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar1921

Autism Spectrum Disorder Prediction

K. Sushma[1]; Balasankula Srihitha[2]; Bandari Bhavani[3]; Eslavath Pavan Kumar[4]

[1]Department of CSE(Data-Science), B V Raju Institute of Technology Narsapur, Medak District, Telangana, India
[2]Department of CSE(Data-Science), B V Raju Institute of Technology Narsapur, Medak District, Telangana, India
[3]Department of CSE(Data-Science), B V Raju Institute of Technology) Narsapur, Medak District, Telangana, India
[4]Department of CSE(Data-Science), B V Raju Institute of Technology Narsapur, Medak District, Telangana, India

Publication Date: 2026/04/09

Abstract: There is a significant Autism Spectrum Disorder (ASD). affect communication, socialization and reaction to. environmental stimuli. Timely diagnosis of ASD assists in giving. help and counsel to the injured persons on the right. time. Conven- tional diagnosis methods usually demand in-depth. consideration and assistance of professionals. This paper recommends the use of a machine learning model to assist in early screening of. ASD in terms of behavioral and demographic variables. Various random forest, decision tree, and other methods of classification. XGBoost, are trained and tested upon the customary preprocessing. of data, such as the processing of missing data and data conversion. categorical variables. The effectiveness of the different techniques is evaluated. The results indicate that ensemble methods are. more precise and valid than other methods.

Keywords: *Autism Spectrum Disorder (ASD), Machine Learning, Early Screening, Behavioral Data, Demographic Data, Decision Tree, Random Forest, XGBoost, Data Preprocessing, Classification Algorithms, Predictive Modeling*.

How to Cite: K. Sushma; Balasankula Srihitha; Bandari Bhavani; Eslavath Pavan Kumar (2026) Autism Spectrum Disorder Prediction. *International Journal of Innovative Science and Research Technology*, 11(3), 3690-3694. https://doi.org/10.38124/ijisrt/26mar1921

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a typical neurological condition characterized by challenges in social interaction, communication, and repetitive behaviors.In recent years,the occurrence of ASD has increased significantly across all age groups worldwide, making it a major public health concern. Early detection of ASD can significantly improve develop- mental, social, and behavioral outcomes, helping individuals achieve a better quality of life. Autism is affected by a mix of genetic and environmental factors.Risk factors may include genetic susceptibility, low birth weight, older parental age, and certain environmental influences. Common behavioral indicators of ASD include limited eye contact, lack of social responsiveness, difficulty in communication, no sensitivity of pain, repetitive behaviors,inappropriate laughing, giggling and attachment to objects, and resistance to changes in routine. The research will create a machine learning-based prediction model of the Autism Spectrum Disorder and assess its effectiveness with the help of typical classification metrics. The model suggested here is aimed at helping parents, teachers, and medical workers to identify the people who are at risk of ASD at a young age. Automation of screening data analysis. Lastly, the inclusion of machine learning in ASD screening can revolutionize the diagnostic process since it can be performed faster, more conveniently, more precisely, which means that it can intervene in time, and better long-term results of people with the autism spectrum.

II. RELATED WORK

Machine learning (ML) has received significant attention for this disorder prediction as a way to provide early, objec- tive, and scalable screening. Traditional evaluation depends on expert behavioral analysis, which can be time-consuming and subjective. To address these limitations, researchers have adopted supervised learning methods to behavioral data.

Recent research has highlighted the role of feature selection, data preprocessing, and explainability in improving model reliability and clinical utility.

Despite these advances, challenges persist, including small or imbalanced datasets, overfitting, limited generalization abil- ity across populations, and lack of explainability. Many earlier studies relied on a single

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar1921

classification algorithms, which limited comparative review and robustness. In contrast, the present study integrates three supervised learning algorithms— Random Forest, Decision Tree, and XGBoost—using feature engineering, robust preprocessing and cross-validated hyperpa- rameter tuning. Demonstrating the effectiveness of ensemble learning for ASD prediction. This integrated approach aims to develop a more generalizable and clinically useful ASD screening framework.

III. SYSTEM DESIGN

The system design describes the overall architecture, func- tional modules, data flow, and interactions among compo- nents in the Autism Spectrum Disorder (ASD) Prediction system.The design is based on a modular and layered approach to achieve better working model.The system analyzes various data to predict the risk of ASD using machine learning algorithms.

➢ *Data Input Module*

The Data Input Module is responsible for collecting and loading the dataset into the system. The dataset includes scores of behaviour, demographic information, and clinical indicators related to ASD.

➢ *Data Preprocessing Module*

This module prepares raw data for machine learning by cleaning and transforming it into a usable format.

- *Functions:*

✓ Handling missing values (mean/median/mode imputa- tion).
✓ Normalizing and scaling numerical features
✓ Removing duplicates and noisy data.

➢ *Feature Engineering and Selection Module*

This module identifies the most relevant features that con- tribute significantly to ASD identification.

- *Functions:*

✓ Feature importance ranking using Random Forest / XG- Boost.
✓ Removal of redundant and irrelevant features

➢ *Model Training Module*

This module builds predictive models using supervised ML models.

- *Algorithms Used:*

✓ Random Forest
✓ Decision Tree
✓ XGBoost

- *Functions:*

✓ Splitting data into train and test sets.
✓ Training multiple models.
✓ Hyperparameter tuning using cross-validation.

➢ *Model Evaluation Module*

Hyperparameter tuning using cross-validation.

- *Functions:*

✓ Calculates Accuracy, Precision, Recall, F1-score.
✓ Compares performance of different models.

➢ *Prediction Module*

This module uses the trained model to predict ASD risk for new input data.

- *Functions:*

✓ Accepts new patient/user data.
✓ Applies preprocessing and feature selection.
✓ Generates ASD risk prediction (Yes/No)

IV. ALGORITHM DESCRIPTION

➢ *Random Forest:*

This is a ML technique which creates many decision trees during the model development stage and integrates the predictions using majority voting for classifica- tion tasks. On training a number of decision trees are created based on randomly sampled subsets of the original data.

One of the major strengths this method is resistance to over- fitting,robustness to noise, and has ability to model complex relationships. It also yields an intrinsic mechanism for feature importance estimation, which is useful for determining the most influential behavioral and demographic factors in ASD prediction.

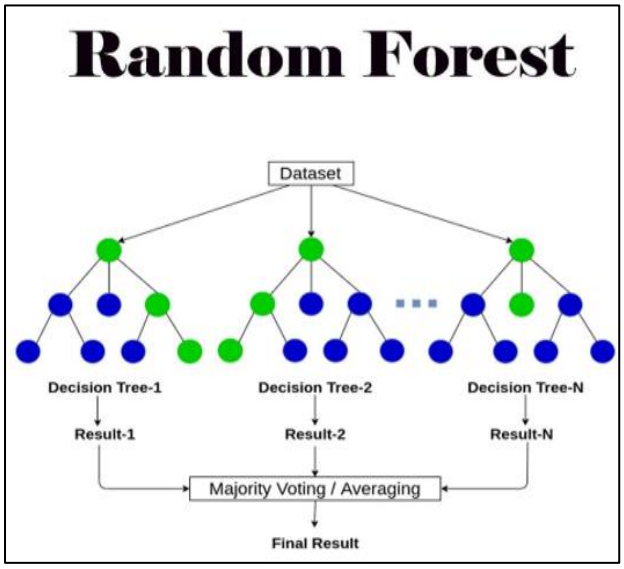

Fig 1 Random Forest

➢ *Decision Tree:*

This is a supervised learning method used for classification that organizes data into a tree-structured model through a sequence of feature-based splits. At each decision

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar1921

point, the feature that best separates the data is selected using required criteria. The last prediction in the classification is acquired through the way to the tree to a input node depending on the feature values of input.

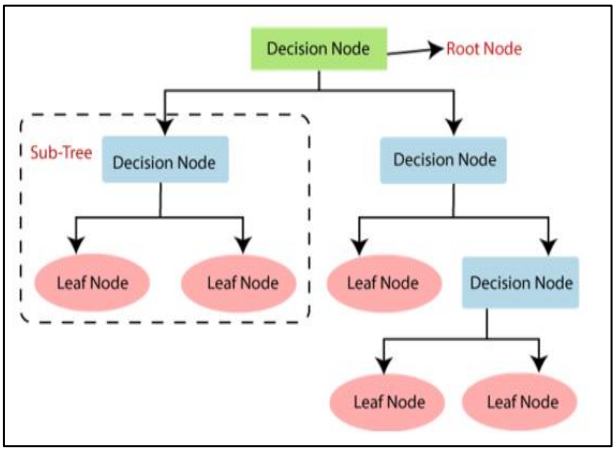

Fig 2 Decision Tree

➢ *XGBoost:*

Extreme Gradient Boosting is a sophisticated boosting which consists of one decision tree after another, where each subsequent tree tries to reduce the errors of the previous trees. This is an additive learning algorithm that forms part of the ensemble algorithm in contrast to the other ensemble algorithms which are used to optimize a regularized objective function. It is a technique which combines L1 regu- larization with L2 regularization to regulate the complexity of the model and avoid overtraining.

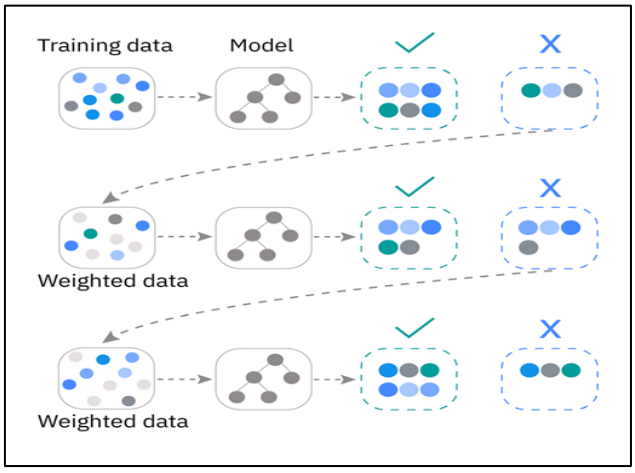

Fig 3 XGBoost

V. IMPLEMENTATION

➢ *Feature Selection and Model Training*

Improving the work of the model is done through feature se- lection whereby a correlation matrix is constructed to indicate the relations between the features and omitting features which are not very important. The method is adopted to enhance effectiveness of the model by decreasing its size. The dataset is divided into a training set and a testing set in 80:20 percent ratio and three machine learning models are employed to train the model. The Decision Tree model and the Random Forest Classifier is trained with the most appropriate number of trees and the best depth, whereas the XGBoost Classifier is trained with the help of the gradient boosting with the best learning rate and the optimum number of estimators.

➢ *Evaluation and Prediction*

The models are tested by evaluating the performance mea- sures that include accuracy, precision, recall, and F1-score of the models and a confusion matrix of the model is created after the training process and the final model is applied on the input data and the output would be either the presence or absence of ASD.

➢ *Introduction*

The implementation phase is mainly concerned with the de- velopment of the Autism Spectrum Disorder (ASD) prediction system.This phase is concerned with the development of the proposed methodology into a Python executable code. The implementation phase deals with the reading of the dataset,data preprocessing,feature selection,development of machine learn- ing model,evaluation of the model,and finally the prediction of the results.

➢ *Environment Setup*

The system is implemented using Python in Jupyter Note- book. The following libraries are used:

- Pandas – for data handling and manipulation
- XGBoost – for advanced classification
- Matplotlib & Seaborn – for visualization

➢ *Dataset Loading*

The dataset is loaded into the system using Pandas.

- *Steps:*

✓ The CSV file containing ASD data is imported.
✓ The dataset is inspected using .head(), .info(), and .de- scribe() functions.
✓ The number of rows, columns, and data types are verified.

➢ *Data Preprocessing Implementation*

During this phase, the raw data is processed. Later,those values in numerical variables are handled by imputing them with the mean or median, while the other missing values in categorical variables are replaced with the mode. Categorical variables such as gender, family history, and similar variables are then converted to numerical values using Encoding. Lastly, normalization of the data is achieved by scaling the numerical variables using StandardScaler.

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar1921

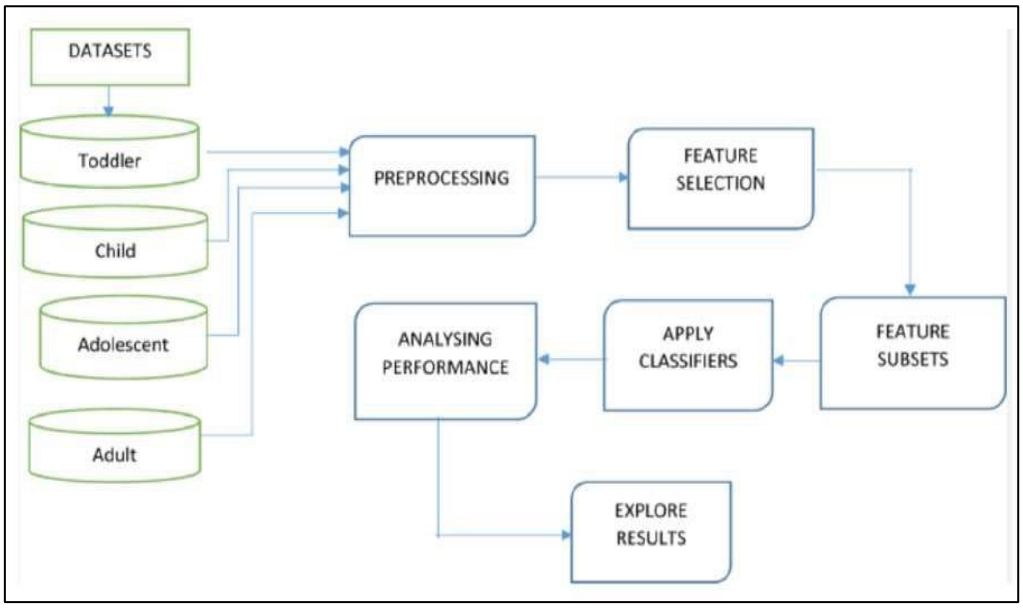

Fig 4 Architecture

VI. EXPERIMENTAL RESULTS

We have run a cross-validation test on Decision Tree classifier, Random Forest classifier, and XGBoost Classifier to add the three models and obtained the accuarcy of 85%. Nonetheless, the findings indicate that the models are better in categorizing Autism Spectrum Disorder.

```
Decision Tree Model Performance
Accuracy: 0.8281

Random Forest Model Performance
Accuracy: 0.8789

XGBoost Model Performance
Accuracy: 0.8867
```

Fig 5 Results

VII. DISCUSSION

The outcome of the experiment suggests that machine learn- ing models are strong at predicting Autism Spectrum Disorder (ASD) using behavioral and demographic variables and this process has an accuracy of 85%, which implies that the model has best generalization ability. The improved performance of the Random Forest classifier can be attributed to its ability to counter the problem of overfitting by aggregating the predic- tions of multiple decision trees and introducing randomness in feature selection. The performance of the XGBoost classifier was also outstanding, which validates the significance of gradient boosting in ASD prediction.

In addition, the preprocessing of data, including the han- dling of missing values,feature selection, and normalization, played an important role in improving the performance of the classifiers. The class balancing method that used SMOTE was effective in detecting additional ASD-positive instances, which is important in the medical sector because false negatives can cause a delay in treatment.

In spite of the fact that the proposed system has proved to yield better outcomes, it does not substitute ASD diagnosis. Rather, it can be implemented as a decision-support system that can assist the healthcare professionals during the initial stages of detection. In future research, it can be assumed that larger and more complicated datasets will be analyzed, deep learning models will be occasionally used, and more interpretable models will be developed.

VIII. CONCLUSION

By applying ML in the prediction of ASD is a significant achievement in the treatment. The usage of complex algo- rithms such as behavioral analysis makes machine learning algorithms more efficient and accurate in the diagnosis of ASD. This project has demonstrated that machine learning can be an effective tool in the early screening of ASD. The models developed are capable of early detection, which is critical in early intervention and also capable of a personalized approach to the ASD. Although the system cannot be used for diagnosis, the proposed system provides a quick, objective, and efficient way of early risk screening for ASD. Furthermore, the application of complex algorithms such as deep learning and ensemble learning makes the models more efficient and significantly better than the existing diagnosis methods.

ACKNOWLEDGMENT

We would like to add that we are extremely thankful to the faculty of the department of CSE (Data Science) members who participated actively in the development of this project, provided valuable advice and support, and cheered us up all the time. Their responses and effective

ISSN No: -2456-2165 https://doi.org/10.38124/ijisrt/26mar1921

recommendations were quite helpful in the creation and improvement of our work.

Without the help, support and technical advice of our mentor, Mrs. K. Sushma, we cannot have achieved this and we would like to extend our special thanks. Her experience and knowledge made us cross numerous difficulties and success- fully undertake this project. Finally, we would like to say a few words about the efforts of our team members who assisted us in fulfilling the completion of this project.

REFERENCES

[1]. N. Nigrou, "Predicting Autism Spectrum Disorder Using Ma- chine Learning Classifiers," Technical Report, National School of Applied Sciences, Al Hoceima, Morocco, Jan. 2024, doi: 10.13140/RG.2.2.35833.44646.

[2]. S. Raja and S. Masood, "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques," *Procedia Computer Science*, vol. 167, pp. 994–1004, 2020.

[3]. B. S. Bhagyalaxmi and O. K. Durrani, "Innovative Autism Spectrum Disorder Prediction Using Machine Learning," *International Journal of Scientific Development and Research (IJSDR)*, vol. 9, no. 8, pp. 607– 611, Aug. 2024.

[4]. S. Patra, A. Giri, R. Koley, R. Mandal, and R. Ray, "Autism Spec- trum Disorder Prediction," *International Journal of Creative Research Thoughts (IJCRT)*, vol. 13, no. 5, pp. c469–c471, May 2025.

[5]. M. Duda, R. Ma, N. Haber, and D. P. Wall, "Use of machine learning for behavioural distinction of autism and ADHD," *Translational Psychiatry*, vol. 6, no. 5, p. e732, 2016.

[6]. D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S. Narayanan, "Use of machine learning to improve autism screening and diagnostic instruments: A systematic review," *Autism Research*, vol. 9, no. 11, pp. 1274–1293, 2016.

[7]. A. Garg, A. Parashar, D. Barman, S. Jain, D. Singhal, M. Masud, and M. Abouhawwash, "Autism Spectrum Disorder Prediction by an Ex- plainable Deep Learning Approach," *Computers, Materials & Continua*, 2022, doi: 10.32604/cmc.2022.022170.

[8]. D. P. Wall, J. Kosmicki, T. F. DeLuca, E. Harstad, and V. A. Fusaro, "Use of machine learning to shorten observation-based screening and diagnosis of autism," *Translational Psychiatry*, vol. 2, no. 4, p. e100, 2012.

[9]. D. P. Wall, R. Dally, R. Luyster, J.-Y. Jung, and T. F. DeLuca, "Use of artificial intelligence to shorten the behavioral diagnosis of autism," *PLOS ONE*, vol. 7, no. 8, p. e43855, 2012.

[10]. J. A. Kosmicki, V. Sochat, M. Duda, and D. P. Wall, "Searching for a minimal set of behaviors for autism detection through feature selection- based machine learning," *Translational Psychiatry*, vol. 5, no. 2, p. e514, 2015.

[11]. F. Thabtah, "Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment," in *Proc. Int. Conf. Medical and Health Informatics*, 2017, pp. 1–6.

[12]. F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward," *Information, Health & Social Care*, vol. 44, no. 3, pp. 278–297, 2019.

[13]. F. F. Thabtah, "Autistic Spectrum Disorder Screening Data for Adult," UCI Machine Learning Repository, 2017. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult

[14]. F. F. Thabtah, "Autistic Spectrum Disorder Screening Data for Children," UCI Machine Learning Repository, 2017.

[15]. F. F. Thabtah, "Autistic Spectrum Disorder Screening Data for Adoles- cent," UCI Machine Learning Repository, 2017.

[16]. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[17]. C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY: Springer, 2006.

[18]. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[19]. A. Narayanan et al., "Artificial intelligence techniques for autism spec- trum disorder diagnosis: A review," *IEEE Access*, vol. 9, pp. 158081– 158112, 2021.

[20]. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.