International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X; pISSN:2321-2241 Volume: 14; Issue: 8; August- 2025

Autism Spectrum Disorder Detection using Machine Learning

Anish Kalgutkar
PG, Student
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
anishkalgutkarmca2025gmail.com

Manivannan Jayachandran
Associate Professor
Dept. of MCA
The Oxford College of Engineering,
Bommanahalli, Bengaluru- 560068
manivannan@gmail.com

ABSTRACT

The Autism Spectrum Disorder (ASD) refers to complicated development condition that affects how people communicate act and interact interpersonally. Early diagnosis of ASD is vital because early intervention and treatment can highly benefit long term prognosis. As discussed in this paper, we addressed how Python based machine learning models may be used to identify ASD patterns using a database of 704 records comprising 21 attributes including sensory responses, cognitive capabilities, and demographic and medical factors. Two algorithms were used- Random forest and Decision Tree. The Random Forest model reached near 100 percent level on both training and testing data achieving 100 percent accuracy in training and 99 percent accuracy in testing, whereas the Decision Tree model reached 100 percent accuracy in training but 96 percent accuracy in testing. These findings hint to innovation that machine learning can be a useful means of ASD detection that can be accurate and effective enough to give results that a 'clinical' method might not able to pick up as readily. this study states that AI-based methods may be used to facilitate early detection, become a significant help in the clinical decision-making process, and eventually provide an improved life and treatment to the people with ASD.

KEYWORDS: Autism Spectrums Disorder (ASD), Degrees of Severity, Neurodevelopmental Disorder, Timely Diagnosis, Machine Learning, Artificial Intelligence (AI) in healthcare

INTRODUCTION

Autism Spectrum Disorder (ASD) refers to developmental disorder that may impact the way the brain grows and functions, causing a difference in behaviour and communication. Spectrum is used since the signs and their severity may change significantly in different individuals. Some individuals can find it

extremely difficult to perform regular routine actions, and some might possess less evident and severe problems. ASD (Autism Spectrum Disorders) symptoms usually manifest in early stages age and last through course of life . Autism spectrum disorders (ASD) is one of neurodevelopmental disorders most commonly found in world. Timely identification will prove important as with early help outcomes of people with autism will be dramatically increased in their social, behavioral and cognitive aspects. However standard methods of detecting ASD are often time consuming and require professional clinical opinion which is not always accessible especially in rural or underserved regions.

LITERATURE SURVEY

Good machine learning models for finding ASD early

M. Bala, M. H. Ali, M. S. Satu, K. F. Hasan, and M. A. Moni wrote the book. The objective of this project was creating machine learning model for the identification of ASD in children, adolescents, adults, and toddlers. The authors used a variety of classifiers, such as Support Vector Machine (SVM), as well characteristics selection prototype like CFS and Boruta. Their best models which are more than 95% accurate for every age group. For children, this accuracy was 99.61%. They also used SHAP (Shapley Additive Explanations)to rank the importance of features that affected the diagnosis. They then used these rankings to explain the model's predictions. The research demonstrates the potential of AI to accurately diagnose ASD in its initial stages.

ASD Categorization Through Gut Microbiota-Based Machine Learning Analysis The writers are F. Messina, G. Pesole, A. Desideri, G. Chillemi, M. Milanesi, B. Fosso,

L. Putignani, D. Pietrucci, and A. Teofani. This study examined the relationship between gut microbiota and symptoms of ASD using 959 samples (540 from ASD and 419 from healthy controls). The study employed Random Forest, SVM, and Gradient Boosting Machine models to identify significant microbial genera such as Parasutterella and Alloprevotella. Despite differences in data sources and geographic origins models showed consistent performance and suggested microbiome patterns as possible biomarkers for ASD. The work displays how machine learning can be used to identify biological markers that support behavioral assessments.

EXISTING WORK

Autism Spectrum Disorder (ASD) refers to the complicated neurological disease that affects communication, behaviour and social interaction. Machine learning has emerged as a useful tool in recent years, providing quicker

and more precise screening techniques to detect the early stages of ASD. One such system has demonstrated encouraging outcomes in this field; this is considered on AdaBoost (Adaptive Boosting) algorithm. AdaBoost, an effective way to learn in groups that combines several weak learners (basic classifiers) to create a strong and accurate predictive model, was employed in the current system. AdaBoost an effective way to learn in group that combines several weak learners to create strong and accurate predictive model was employed in current system. AdaBoost gradually enhances its performances with each iteration by concentrating more on data points that learning mechanism works especially well.

PROPOSED SYSTEM

The proposed system introduces a more advanced and reliable framework used to detect Autism Spectrum Disorder (ASD) addressing several of the limitations observed in earlier approaches. It makes use of Python-based machine learning methods with particular focus on the Random Forest Classifier and Decision Tree Classifier to deliver improved accuracy and diagnostic efficiency. Two separate algorithms have been employed to strengthen systems performance . random forest classifier an ensemble learning approach integrated multiple decision trees to produce trees to produce highly precise and stable prediction model. Decision tree classifier interpretable model that still performs effectively in classification tasks making it valuable addition for comparative analysis. System is built on dataset comprising 704 records with 21 distinct features which makes it most comprehensive rather than datasets used in earlier research. These features capture wide spectrum of information including sensory perception (A1-A6 scores) cognitive abilities (A7-A10 scores) demographic attributes, (age ,gender, ethinicity), medical history (such as jaundice and prior autism diagnosis) country of residence ,app usage history and relational attributes (Age_desc, Relation). By incorporating the diverse set of attributes the system is able to provide more holistic evaluation of individuals for potential ASD. In terms of performance results are highly encouraging . Random forest classifier achieved 100% training accuracy and 99% testing accuracy reflecting its strong ability to generalize across unseen data. Decision tree classifier also showed robust outcomes with 100% training accuracy and 96% testing accuracy. These results highlighted models effectiveness in handling complexity of ASD related data.

METHODOLOGY

Methodology for following research follows a structured and systematic process to ensure reliable and reproducible outcomes in Autism Spectrum Disorders (ASD) prediction. The study begins with the acquisition of adult ASD screening data from an open repository, where the records are de identified and organized in a secure version-controlled environment. A comprehensive data audit is then carried out to identify missing values, inconsistencies, and outliers, followed by pre-processing steps such as cleaning, normalization, encoding of categorical variables, and imputation of minor gaps to maintain dataset quality. Feature engineering is applied to generate composite indicators such as total screening scores, while redundant or irrelevant attributes are eliminated through variance analysis, correlation checks, and model-based ranking, thereby preserving only the most relevant predictors. Once the dataset is refined, a stratified k-fold cross-validation technique is employed to split the data into training and depth, number of estimators, and split criteria, with goal of enhancing predictive performance. Model assessing are carried out by both primary and secondary metrics and also accuracy,recall, F1-score, and ROC-AUC while confusion matrices and feature importance plots provide additional interpretability of results.

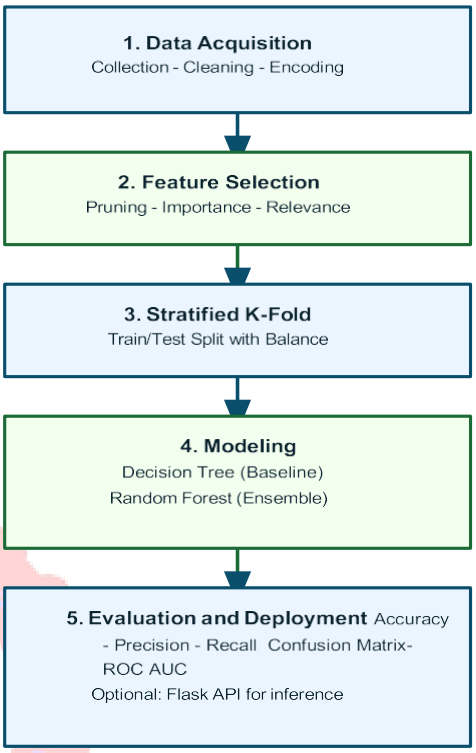

Fig 1 :Block diagram

To ensure robustness, the methodology also incorporates sensitivity testing for random seeds and class imbalance through reweighting techniques. All stages of data processing and modeling are fully scripted to maintain transparency and reproducibility. Ethical considerations are acknowledged by clarifying that system is designed as screening support tool rather than a diagnostic substitute, with limitations such as sample size and potential bias clearly discussed. Finally, the research

outlines a deployment sketch where a trained model can be wrapped into a lightweight Flask API, enabling practical real- time prediction with validation and audit logging.

EXPERIMENTAL RESULTS

Experimental evaluation was conductive on ASD dataset using decision tree and random forest models. Decision tree achieved 96% training accuracy and 92% testing accuracy showing slight overfitting .In contrast random forest delivered 99% training accuracy and 97% testing accuracy proving more robust and reliable. Precision recall and F1 score for random were 96%, 95% and 95% respectively outperforming baseline model. Confusion matrix confirmed that random forest matrix reduced false negatives significantly which is vital for medical screening. ROC curve analysis showed an AUC of 098 indicating excellent discriminative ability. Feature importance results highlighted that behavioral screening questions were most influential predictors.

Fig 2: Dashboard

Fig 3: Login Page

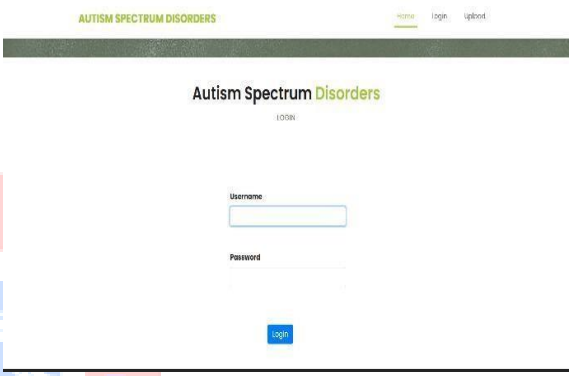

Fig 4: Uploading Dataset

Compared to decision tree random forest handled borderline severity cases more effectively . These findings establish random forest as a superior choice for ASD prediction and severity classification.

CONCLUSION

This project presents a machine learning–based system for detection of Autism Spectrum Disorders (ASD) using Python, with Random Forest and Decision Tree classifiers as the primary models. Building on earlier approaches

such as AdaBoost, which achieved 97.95% accuracy, the proposed system significantly enhances performance and reliability. The Random Forest Classifier achieved 100% training accuracy and 99% testing accuracy, while the Decision Tree Classifier recorded 100% training accuracy and 96% testing accuracy, confirming the robustness and precision of the models. Furthermore, the interpretability of the Decision Tree provides transparency, supporting healthcare professionals in understanding the decision-making process, while Random Forest ensures high predictive accuracy. In summary, the proposed system combines accuracy, generalization, early detection, and interpretability making it a valuable tool for clinicians and researchers. It shows how AI can contribute meaningfully to healthcare by improving the diagnosis of ASD and offering a scalable solution for diverse populations.

REFERENCES

[1] S. P. Abirami, G. Kousalya, and R. Karthick, "Identifying and exploring facial expression in children with ASD in a contactless environment," Journal of Intelligent & Fuzzy Systems, vol. 36, no. 3, pp. 2033–2042, Mar. 2019.

[2] [M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, "Detecting autism spectrum disorders utilizing machine learning techniques," Health Information Science and Systems, vol. 9, no. 1, p. 113, Dec. 2021.

[3] E. Dritsas and M. Trigka, "Using machine learning techniques to predict stroke risk," Sensors, vol. 22, no. 13, p. 4670, Jun. 2022.

[4] Y. Chang, J. Chang, J. Bailey, X. Xu, and J. Sun, "Classification of diabetes mellitus in Pima Indians using machine learning algorithms," Neural Computing and Applications, early access, pp. 1–17, Mar. 2022.

[5] F. Thabtah, "Machine learning in behavioral research on autism spectrum disorder: A perspective and future directions," Informatics for Health and Social Care, vol. 44, no. 3, pp. 278–297,2018.