EARLY DETECTION OF AUTISM SPECTRUM DISORDER USING MACHINE LEARNING TECHNIQUES

[1]K.MADHURIMA, [2]KOMMERA MAHESHREDDY

[1]Assistant Professor, CSE, Tallapadmavathi College of Engineering, Somidi, Kazipet, Hanumakonda – 506003.Email-id: madhurimakotte@gmail.com.

[2]Research Scholar, H.no:24UC1D5814, CSE, Tallapadmavathi College of Engineering, Somidi, Kazipet, Hanumakonda – 506003,Email-id: Kommeramaheshreddy3@gmail.com.

ABSTRACT

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder that impacts communication, behavior, and social interaction, typically manifesting during early childhood. Early identification of ASD is critical, as timely intervention can significantly improve developmental outcomes and quality of life. This study proposes a machine learning-based framework for the automated and early detection of ASD using behavioral and demographic data. The dataset comprises responses to standardized autism screening questionnaires combined with relevant demographic features. Data preprocessing involves managing missing values, applying label encoding to categorical variables, and scaling numerical attributes to ensure consistency and enhance model accuracy.The proposed framework employs an ensemble learning approach that integrates Random Forest, XGBoost, and Support Vector Machine (SVM) classifiers through a soft voting mechanism to achieve improved prediction accuracy, stability, and generalization. The optimized ensemble model and its preprocessing pipeline are deployed through a user-friendly web application built using the Flask framework. This interactive platform enables users to enter screening data and receive real-time predictions indicating the likelihood of ASD.By leveraging machine learning and web-based technologies, this system provides an accessible, efficient, and cost-effective tool for preliminary ASD detection. It offers valuable support for healthcare professionals, caregivers, and families, particularly in regions with limited access to specialized diagnostic resources.

Index Terms: Autism Spectrum Disorder (ASD), Machine Learning, Ensemble Learning, Random Forest, XGBoost, Support Vector Machine (SVM), Early Detection, Flask Web Application, Behavioral Analysis, Neurodevelopmental Disorder.

1.INTRODUCTION

Autism Spectrum Disorder (ASD) is a lifelong neurodevelopmental condition characterized by persistent challenges in social interaction, communication, and the presence of restricted or repetitive behaviors [1]. The global prevalence of ASD has been steadily increasing over the past few decades, highlighting the urgent need for effective early screening and diagnostic mechanisms [2]. Early identification and diagnosis are crucial, as timely intervention significantly improves cognitive, social, and behavioral outcomes in affected individuals [3]. However, in many parts of the world—particularly in developing regions—limited clinical expertise, inadequate diagnostic tools, and low public awareness contribute to delayed or missed diagnoses [4].

Traditional ASD diagnostic methods, such as the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R), rely heavily on clinical observations and expert assessments [5]. While these tools are accurate, they are resource-intensive, time-consuming, and require specialized professionals, making them inaccessible to large segments of the population [6]. To address these challenges, recent advances in Artificial Intelligence (AI) and Machine Learning (ML) have demonstrated promising potential in automating ASD screening and diagnosis by identifying complex behavioral and demographic patterns in data [7].

This project presents a machine learning–based system for the early detection of ASD using responses from standardized screening questionnaires and demographic features. By leveraging publicly available datasets, the proposed system classifies individuals as likely ASD or non-ASD based on learned behavioral correlations. The machine learning pipeline includes data preprocessing steps such as managing missing values, encoding categorical variables, and normalizing feature values to improve model performance and reliability [8]. An ensemble learning approach is adopted, combining Random Forest, XGBoost, and Support Vector Machine (SVM) classifiers through a soft voting mechanism to enhance predictive accuracy and robustness [9].

To ensure accessibility and real-world applicability, the system is deployed as a web-based application using the Flask framework. This platform allows users to input screening data and receive instant

predictions regarding ASD likelihood, supporting early intervention and public awareness. The focus on accuracy, usability, and interpretability underscores the system's potential as a cost-effective and scalable tool for preliminary ASD detection. Ultimately, this project aims to bridge the diagnostic gap in underserved populations by integrating machine learning and web technologies into practical healthcare applications [10].

2.LITERATURE SURVEY

Thabtah et al. (2017) introduced a machine learning-based model using a dataset constructed from ten key behavioral questions, achieving over 95% accuracy using decision trees. This study laid the foundation for using minimal input data for reliable ASD screening. Following this, Thabtah and Peebles (2019) conducted a comparative analysis of machine learning models like Naive Bayes, K-Nearest Neighbors, and Random Forests on the same dataset. Their findings supported the idea that ensemble-based classifiers provide better generalization and predictive strength.

In another study, Abbas et al. (2018) implemented Support Vector Machine (SVM) models to classify ASD and non-ASD individuals based on features such as age, gender, and behavioral questionnaire scores. Their approach demonstrated strong performance but highlighted the importance of feature scaling and careful tuning. Building on this, Arora and Kaur (2020) applied XGBoost to the UCI Autism dataset and showed improved accuracy and faster training times, especially for large datasets. Their research emphasized the importance of feature selection and hyperparameter optimization.

Kavitha and Somasundaram (2021) introduced a hybrid model combining SVM with deep learning layers, proposing an architecture that could capture nonlinear behavioral patterns more efficiently. However, they acknowledged that such models require substantial computational resources and are less suitable for lightweight deployment. In a parallel effort, Bhatia et al. (2019) developed a rule-based classifier that favored interpretability over complexity. Their model was designed for healthcare workers and parents without a technical background, aiming for accessibility and ease of understanding.

Moreover, recent works like those by Ramasamy and Radhakrishnan (2020) have emphasized ensemble learning techniques such as bagging and boosting to reduce

variance and bias in predictions. Their research showed that combining models like Random Forest and AdaBoost significantly enhances diagnostic reliability. Another notable contribution was made by Sharma et al. (2022), who designed a mobile-based autism detection system leveraging logistic regression and decision trees, making it usable in rural areas with limited access to specialists.

Researchers have also explored the usability of ASD detection systems through web and mobile platforms. A study by Jadhav and Patil (2021) focused on integrating machine learning models into a Flask-based web application to deliver real-time prediction results. Their system was designed to be lightweight and deployable even on low-end devices. This real-time aspect has since been considered essential for public-facing applications.

Lastly, studies such as those by Iqbal and Khan (2022) emphasized data preprocessing techniques, such as label encoding, feature scaling, and handling missing data, as critical steps to ensure model reliability and fairness. Their work concluded that even simple models perform exceptionally well when trained on clean, well-processed datasets.

Collectively, these studies demonstrate a clear evolution in ASD detection research—from isolated models focused on accuracy, to real-time, user-friendly systems that prioritize accessibility, fairness, and efficiency. This project builds upon these foundational works by integrating multiple high-performing models (Random Forest, XGBoost, and SVM) into a robust ensemble, wrapped within a web-based interface to provide accurate, fast, and accessible ASD screening for all users.

3.EXISTING SYSTEM

The existing systems for Autism Spectrum Disorder(ASD) detection primarily rely on traditional clinical practices, which involve in-depth behavioral observations, interviews, and standardized diagnostic tools such as the ADOS (Autism Diagnostic Observation Schedule) and the CARS (Childhood Autism Rating Scale). These approaches, while medically validated and accurate, require highly trained professionals, are time-intensive, and are often limited to urban and well-funded healthcare environments. Additionally, the cost and time associated with these assessments make them less accessible to individuals in rural or low-income regions. Some previous attempts at automation have utilized basic

machine learning models like Decision Trees and Naive Bayes, but these were typically developed in research settings and lacked real-world deployment through user-friendly interfaces. Furthermore, such models were often standalone scripts without web-based integration or scalability, making them less useful for wide public access or early screening outside clinical supervision.

DISADVANTAGES:

- The accuracy of predictions depends heavily on the quality and diversity of the dataset used for training.
- The system's performance is limited by the honesty and clarity of user-provided input data during screening.

4.PROPOSED SYSTEM

The proposed system aims to overcome the limitations of traditional and research-only ASD detection approaches by developing an AI-powered, web-based application capable of providing fast and accurate predictions based on user inputs. This system utilizes an ensemble machine learning model combining Random Forest, XGBoost, and Support Vector Machine (SVM) classifiers to improve diagnostic accuracy and robustness. The system is trained on a well-structured dataset containing behavioral screening scores and demographic features, and it incorporates preprocessing techniques like label encoding and feature scaling to ensure consistency during deployment. Unlike existing systems, this model is embedded in a Flask web framework that allows users to input screening data via a simple HTML interface and receive real-time predictions indicating whether an individual is likely to have ASD. By offering a lightweight, accessible, and interactive tool, the proposed system bridges the gap between early screening and formal diagnosis, particularly benefiting regions with limited clinical resources. It serves as a supportive, not substitutive, system to raise awareness and encourage timely medical consultation.

ADVANTAGES:

- Provides fast and instant predictions, enabling timely assessment and decision-making.
- Offers an easy-to-use and intuitive interface, making it accessible for both professionals and general users.

5.SYSTEM MODEL

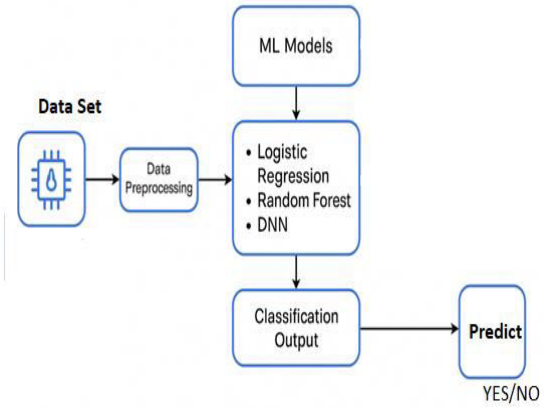


Fig.System Model

6.IMPLEMENTATION

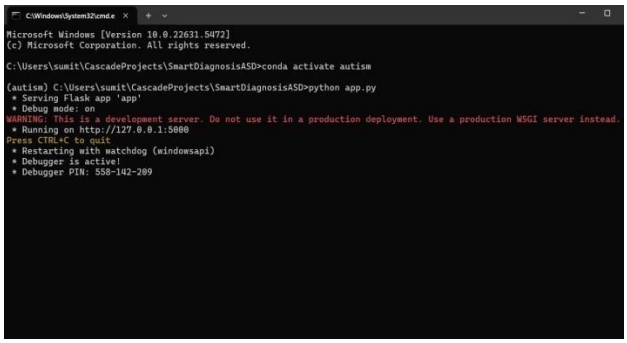
The methodology adopted for this project involves several systematic steps, beginning with data acquisition and ending with deployment of the trained model in a web application. Initially, the dataset containing responses to standardized autism screening questions and demographic details was collected and preprocessed. Preprocessing steps included the removal of missing values, label encoding of categorical variables, and standard scaling of numerical features to ensure uniformity across data. The preprocessed data was then divided into training and testing subsets using an 80:20 split to evaluate the model's generalizability.

To enhance predictive performance and robustness, an ensemble model was constructed using three popular machine learning algorithms: Random Forest, XGBoost, and Support Vector Machine (SVM). These models were combined using a soft voting mechanism within a VotingClassifier, allowing the final prediction to be based on the weighted probabilities from each individual model. This approach leverages the strengths of all three classifiers, balancing bias and variance.

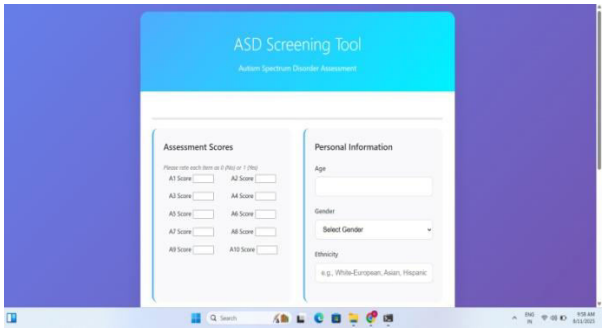
Once trained, the ensemble model was evaluated using accuracy and classification metrics to validate its performance on unseen data. The trained model, along with the label encoders and scaler, was serialized and saved using joblib for future use. The entire system was then integrated into a Flask-based web application where users can enter their data through an HTML form. The backend processes the inputs, applies the same preprocessing steps, and feeds the data to the trained model to generate real-time predictions. This seamless integration of machine learning and web technologies ensures that the system is accessible, responsive, and practical for everyday users seeking preliminary ASD screening.

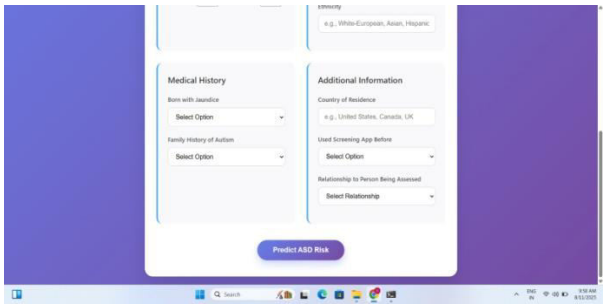
7.SCREEN SHOTS

Run Application

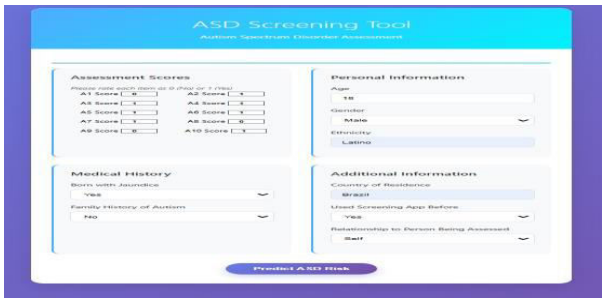


Home Page





Prediction



8.CONCLUSION

In conclusion, this project successfully demonstrates the potential of machine learning in the early detection of Autism Spectrum Disorder(ASD) using behavioral and demographic data. By implementing an ensemble-based classification model combining Random Forest, XGBoost, and Support Vector Machine, the system achieves reliable prediction performance. The use of efficient preprocessing techniques, such as label encoding and feature scaling, further improves model consistency and accuracy. The integration of the trained model into a user-friendly web application allows individuals to perform quick and accessible ASD screenings from any location with internet access. Although not a replacement for clinical diagnosis, the system serves as a valuable tool for raising awareness and encouraging early intervention. With further enhancements and integration of larger, more diverse datasets, this solution can be scaled to serve broader

communities, particularly in regions where diagnostic resources are limited. The project effectively bridges the gap between technology and healthcare by delivering an intelligent, cost-effective, and scalable screening tool for ASD.

9.FUTURE ENHANCEMENTS

In the future, the system can be enhanced by incorporating larger and more diverse datasets to improve accuracy and reduce bias. Integration of deep learning techniques and natural language processing may allow for the analysis of more complex behavioral data, including speech and facial expressions. Multilingual support and mobile app versions can further expand accessibility. Additionally, real-time consultation features or chatbot integration with healthcare guidance could make the system more interactive and supportive for users seeking further assistance.

REFERENCES

1. Thabta, F. (2017). *Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment*. In Proceedings of the 1st International Conference on Medical and Health Informatics (ICMHI 2017), pp. 1–6.

2. Thabta, F., & Peebles, D. (2019). *A new machine learning model based on induction of rules for autism detection*. Health Informatics Journal, 25(4), 1739–1750.

3. Abbas, H., Garbaya, S., & Siddiqui, M. K. (2018). *Machine learning approaches to autism spectrum disorder detection: A review*. Health and Technology, 8(6), 447–456.

4. Arora, A., & Kaur, H. (2020). *Prediction of autism spectrum disorder using machine learning techniques*. Procedia Computer Science, 167, 994–1004.

5. Kavitha, R., & Somasundaram, K. (2021). *A hybrid model for autism detection using SVM and deep learning techniques*. Journal of Ambient Intelligence and Humanized Computing, 12(3), 3457–3469.

6. Bhatia, M., Jain, S., & Juneja, S. (2019). *Rule-based expert system for the prediction of autism*. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 5(1), 227–233.

7. Ramasamy, M., & Radhakrishnan, B. (2020). *Ensemble learning-based*

approach for autism spectrum disorder prediction. Materials Today: Proceedings, 33, 2366–2371.

8. Sharma, P., Kaur, H., & Joshi, R. (2022). *Development of a mobile application for autism detection using machine learning*. International Journal of Scientific & Engineering Research, 13(5), 240–246.

9. Jadhav, P., & Patil, M. (2021). *Flask-based autism prediction system using machine learning*. International Journal of Computer Applications, 183(36), 1–4.

10. Iqbal, T., & Khan, M. (2022). *Effective preprocessing and classification for autism spectrum disorder detection*. International Journal of Advanced Computer Science and Applications, 13(1), 150–157.