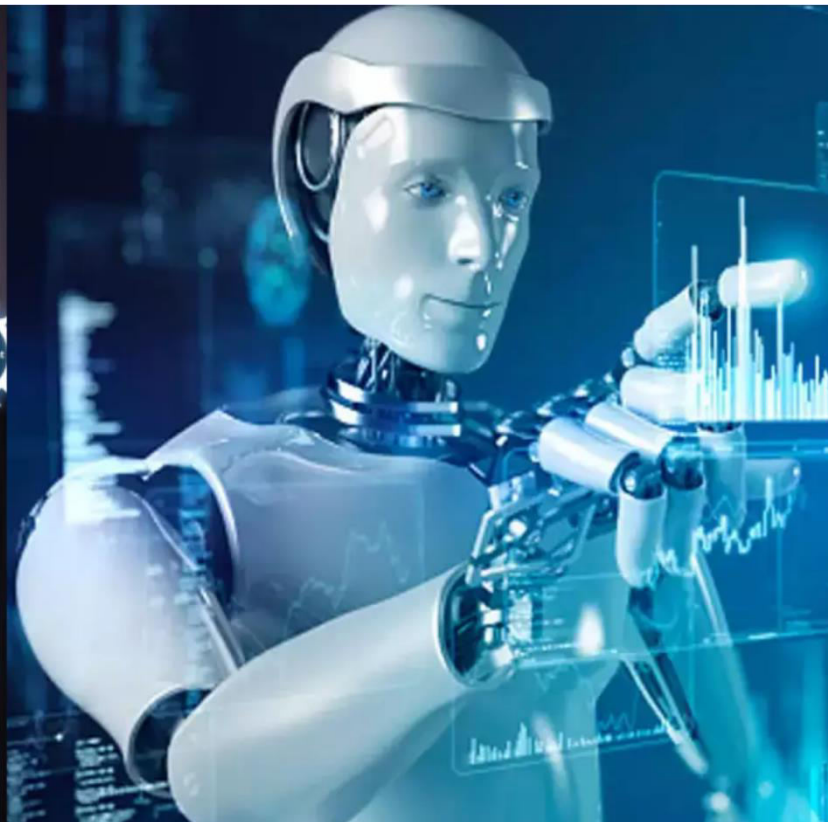




International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Autism Spectrum Disorder Prediction using Machine Learning

N.Anithaa, Guntupalli Manasa, Munnangi Ratna Kavya, Atikala Abhignya, Shaik Dileep Pasha

Assistant Professor, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

B. Tech Students, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai, India

ABSTRACT: Autism spectrum sickness, additionally referred to as ASD, is a neurodevelopmental disease that influences someone's capacity to cooperate, communicate, and analyze. The severity of symptoms associated with ASD varies from person to individual. Even if it isn't resolved first of all, it becomes greater excessive in the coming days, ensuing in a lower in the range of autism sufferers. Using various academic techniques, autism may be recognized in advance. There isn't any treatment for ASD, but early detection can reduce its terrible impact. We recognition on the distinction. Most treatment plans are very tough, take maximum of the day, and from time to time give wrong consequences. We use ML to enhance the procedure. The vintage things are dealt with discretion, accuracy, and speed. Typically, previous work demonstrates new techniques or compares strategies. Our efforts do no longer use three information sets to evaluate one of a kind strategies, however as an alternative aim to increase an progressed algorithm. We make predictions in the proposed work. In addition to evaluating and enforcing diverse schooling gear in coaching practice, the effects of the intellectual health assessment of youngsters elderly 1 to 5 years and above were analyzed. We analyzed the statistics periodically over the last ten years. The anticipated records from patients without autism and sufferers with autism might be absolutely new to the autism records set and will be used to track future affected person effects. Our efforts purpose to make ASD prognosis simpler, more correct, and extra timely, in an effort to help enhance effects for humans suffering from the disorder. We use various device studying algorithms consisting of Logistic Regression (LR), K-Nearest Neighbors, Decision Tree (DT), Naive Bayes (NB), and Image Autism Disorder Prediction Algorithm (KNN). Once the primary one is whole, we will increase our study to test a couple of excessive-stage capabilities to ensure extra correct observations.

KEYWORDS: Machine Learning, Random Forest, Logistic Regression (LR) Algorithm K-Nearest Neighbor, Decision Tree (DT), Naive Bayes (NB).

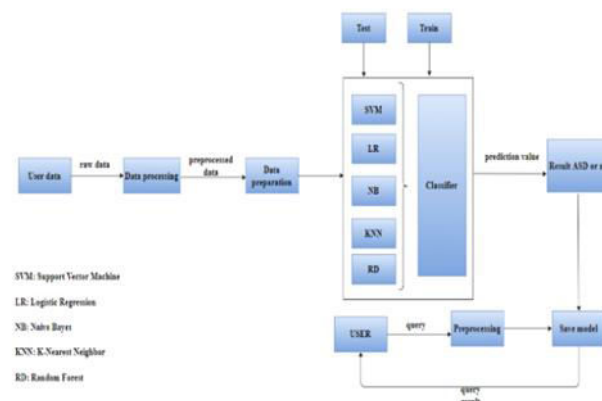


Figure 1: SYSTEM ARCHITECTURE

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in social interaction, communication, and repetitive behaviors. Early diagnosis and intervention are crucial for improving the quality of life for individuals with ASD, yet traditional diagnostic processes can be time-consuming, subjective, and often delayed. In recent years, machine learning (ML) has emerged as a powerful tool in the field of medical diagnosis, offering the potential to analyze vast and complex datasets with high accuracy. By leveraging ML algorithms, it is possible to identify patterns and correlations in behavioral and clinical data that may not be easily noticeable through conventional methods. This project aims to develop a machine learning-based predictive model for early detection of Autism Spectrum Disorder. By utilizing features derived from validated screening tools and demographic data, the model seeks to assist healthcare professionals in making more accurate and timely assessments. The integration of such intelligent systems in the diagnostic process holds promise for enhancing efficiency, reducing human error, and ultimately supporting better patient outcomes.

II. LITERATURE REVIEW

Several studies have demonstrated the effectiveness of machine learning (ML) techniques in predicting Autism Spectrum Disorder (ASD), with the goal of improving early detection and diagnosis. Initial research focused on using supervised learning algorithms, such as Support Vector Machines (SVM), Decision Trees, and Naïve Bayes, applied to data from behavioral screening tools like the Autism Spectrum Quotient (AQ) and the Modified Checklist for Autism in Toddlers (M-CHAT). For instance, Wall et al. (2012) achieved high accuracy using SVM by selecting a minimal set of behavioral features, while Thabtah et al. (2017) highlighted the importance of demographic and behavioral traits in model performance. Ensemble methods like Random Forests and Gradient Boosting have also shown strong predictive capabilities by combining multiple classifiers to improve robustness and accuracy.

Recent advancements have introduced deep learning models, including artificial neural networks (ANNs) and convolutional neural networks (CNNs), particularly in studies involving neuroimaging data. These approaches capture complex, non-linear patterns but often require large datasets and significant computational resources, which limits their practical use in general screening. Feature selection techniques such as Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) have been widely used to reduce dimensionality and enhance model performance. Publicly available datasets, such as those from the UCI Machine Learning Repository and the Autism Brain Imaging Data Exchange (ABIDE), have enabled comparative studies and benchmarking. Despite these advances, challenges such as data imbalance, small sample sizes, and limited generalizability across populations remain. Ethical concerns related to data privacy and automated decision-making in healthcare also require careful consideration. Future research directions include integrating multi-modal data and improving model interpretability to support clinical adoption.

III. METHODOLOGY

This project involves a structured approach comprising data collection, preprocessing, feature selection, model development, evaluation, and deployment. The dataset used is obtained from the UCI Machine Learning Repository, containing demographic information and results from behavioral screening tools such as the AQ-10, along with diagnostic labels indicating whether an individual is on the autism spectrum. During preprocessing, the data is cleaned by handling missing values, removing duplicates, and converting categorical variables into numerical formats using techniques like one-hot or label encoding. Numerical features are standardized to ensure consistent scaling, and if the dataset is imbalanced, resampling techniques such as SMOTE are applied. Feature selection is conducted using methods like Recursive Feature Elimination (RFE), correlation analysis, and tree-based feature importance to identify the most relevant attributes for ASD prediction. Several machine learning algorithms, including Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression, and Gradient

Boosting, are developed and trained using an 80/20 train-test split or k-fold cross-validation. Model performance is evaluated based on metrics such as accuracy, precision, recall, F1-score, and AUC-ROC, with confusion matrices used to analyze classification results. The model with the best performance is selected for potential deployment, with consideration given to interpretability using tools like SHAP to ensure transparency and support trust among healthcare professionals.

IV. IMPLEMENTATION

Data Set

- The dataset is obtained from the **UCI Machine Learning Repository** and includes adult and child versions of autism screening data. It contains attributes such as age, gender, ethnicity, family history of ASD, and responses to 10 AQ-10 screening questions. The target variable indicates whether the individual is likely to have Autism Spectrum Disorder (ASD).
- In this article, we shed light on the important features that have proved to be sound and effective in predicting phishing websites. In addition, we proposed some new features, experimentally assign new rules to some well-known features and update some other features.

Phishing Websites Features

The phishing Websites Detection project focuses on identifying fraudulent websites that aim to steal sensitive information, such as login credentials or financial data. Key features of the project include the analysis of various website attributes such as domain names, URL structure, SSL certification, and HTML content. The model uses machine learning algorithms to classify websites as either legitimate or phishing based on these features. By leveraging supervised learning techniques, the system can predict phishing sites with a high degree of accuracy, identifying suspicious patterns such as mismatched URLs, obfuscated code, or abnormal webpage behaviors. In addition to traditional URL-based analysis, the project also explores the use of machine learning to detect phishing websites by analyzing website visual elements and content. Features like text similarities, design discrepancies, and the use of misleading logos or images are incorporated to improve detection capabilities. The model is trained on large datasets containing both legitimate and phishing websites, allowing it to generalize across different phishing tactics. This approach enables the development of an efficient tool for real-time detection, helping users avoid potentially harmful websites and enhancing online security.

Detection Technique

The detection technique used in the Phishing Websites project relies on supervised machine learning, where the model is trained on a labeled dataset containing both phishing and legitimate websites. It extracts key features from each website, such as URL length, presence of special characters, use of HTTPS, domain age, favicon usage, and embedded links. These features are then fed into classification algorithms like Decision Tree, Random Forest, Support Vector Machine (SVM), or Logistic Regression to learn patterns that distinguish phishing websites from genuine ones. Once trained, the model analyzes new websites by evaluating their features and classifies them as either phishing or legitimate. This technique enables fast and automated identification of phishing threats, reducing human error and improving online safety.

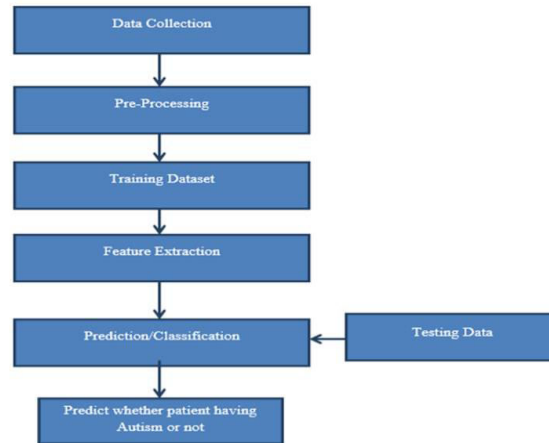


Figure 2: Data Flow

V. RESULTS AND CONCLUSION

The ASD prediction project demonstrate that machine learning models can effectively predict Autism Spectrum Disorder based on behavioral and developmental data, with promising accuracy. By utilizing various algorithms, such as decision trees, support vector machines, and deep learning, the model shows significant potential in identifying early signs of ASD. The integration of multimodal data, such as speech patterns, eye-tracking, and sensor data, further enhanced the prediction capabilities. However, challenges remain in addressing data imbalances, improving model generalization across diverse populations, and ensuring real-time prediction accuracy. The project’s conclusion emphasizes the importance of continuous data collection, collaboration with healthcare professionals, and the need for interpretability in models to ensure clinical adoption. With further refinements, this project has the potential to assist in early diagnosis and intervention, improving outcomes for individuals with autism. Ethical considerations, such as data privacy and fairness, must be carefully integrated into future developments.

Autism Spectrum Disorder Prediction in Children

Does your child look at you when you call his/her name? :

Select

How easy is it for you to get eye contact with your child :

Select

Does your child point to indicate that she/he wants something? (e.g. a toythat is out of reach) :

Select

Does your child point to share interest with you? (e.g. pointing at an interesting sight) :

Select

Does your child pretend? (e.g. care for dolls, talk on a toy phone) :

Select

Does your child follow where you're looking? :

Select

If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort them? (e.g. stroking hair, hugging




Fig:phishing Page

If you or someone else in the family is visibly upset, does your child show signs of wanting to comfort them? (e.g. stroking hair, hugging them) :

Select

Would you describe your child's first words as : :

Select

Does your child use simple gestures? (e.g. wave goodbye) :

Select

Does your child stare at nothing with no apparent purpose? :

Select

Qchat-10-Score :

Age :

Gender :

Select

Gender :

Select

Ethnicity :

Select

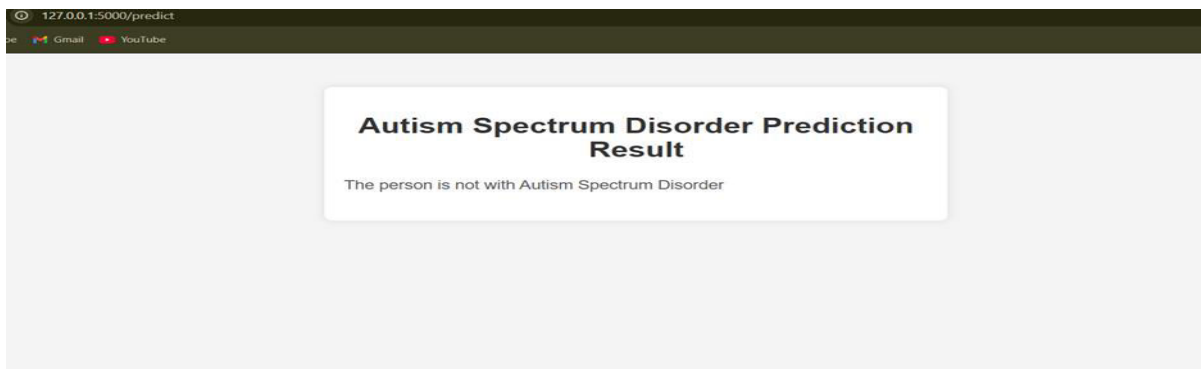
Born with jaundice :

Select

Family members with ASD history :

Select

Predict



VI. FUTURE WORK

Future work on the ASD prediction project can focus on improving model accuracy through better feature engineering and data augmentation, while incorporating multimodal data like speech and neuroimaging for more comprehensive predictions. Integrating real-time monitoring and early detection systems could enable timely interventions. Emphasizing model interpretability and addressing biases will ensure fairness and trust in clinical settings. Collaboration with healthcare professionals will refine the models for practical use. Finally, ethical considerations and data privacy will be prioritized to ensure responsible application in healthcare.

REFERENCES

- [1] M. Maniruzzaman, M. A. M. Hasan, N. Asai, and J. Shin, "Optimal channels and features selection based ADHD detection from EEG signal using statistical and machine learning techniques," *IEEE Access*, vol. 11, pp. 33570–33583, 2023.
- [2] M. Maniruzzaman, J. Shin, M. A. M. Hasan, and A. Yasumura, "Efficient feature selection and machine learning based ADHD detection using EEG signal," *Comput., Mater. Continua*, vol. 72, no. 3, pp. 5179–5195, 2022.
- [3] M. Maniruzzaman, J. Shin, and M. A. M. Hasan, "Predicting children with ADHD using behavioral activity: A machine learning analysis," *Appl. Sci.*, vol. 12, no. 5, pp. 2737–2749, 2022.
- [4] M. Adamou, S. L. Jones, L. Marks, and D. Lowe, "Efficacy of continuous performance testing in adult ADHD in a clinical sample using QbTest+," *J. Attention Disorders*, vol. 26, no. 11, pp. 1483–1491, Sep. 2022.
- [5] J. Shin, M. Maniruzzaman, Y. Uchida, M. A. M. Hasan, A. Megumi, A. Suzuki, and A. Yasumura, "Important features selection and classification of adult and child from handwriting using machine learning methods," *Appl. Sci.*, vol. 12, no. 10, pp. 5256–5270, 2022.
- [6] J. Shin, M. A. M. Hasan, M. Maniruzzaman, A. Megumi, A. Suzuki, and A. Yasumura, "Online handwriting based adult and child classification using machine learning techniques," in *Proc. IEEE 5th Eurasian Conf. Educ. Innov. (ECEI)*, Feb. 2022, pp. 201–204.
- [7] A. Megumi, A. Suzuki, J. Shin, and A. Yasumura. (2022). Developmental Changes in Writing Dynamics and Its Relationship With ADHD and ASD Tendencies: A Preliminary Study. [Online]. Available: <https://doi.org/10.21203/rs.3.rs-1616383/v1>.
- [8] X. Zhou, Q. Lin, Y. Gui, Z. Wang, M. Liu, and H. Lu, "Multimodal Mr images-based diagnosis of early adolescent attention-deficit/hyperactivity disorder using multiple kernel learning," *Frontiers Neurosci.*, vol. 15, pp. 710133–710147, Sep. 2021. vi
- [9] Lakshmi Narasimha Raju Mudunuri, Vivekchowdary Attaluri, "Urban Development Challenges and the Role of Cloud AI-Powered Blue-Green Solutions," in *Integrating Blue-Green Infrastructure Into Urban Development*, IGI Global, USA, pp. 507-522, 2025.
- [10] G. Guney, E. Kisacik, C. Kalaycioglu, and G. Saygili, "Exploring the attention process differentiation of attention deficit hyperactivity disorder (ADHD) symptomatic adults using artificial intelligence on electroencephalography (EEG) signals," *TURKISH J. Electr. Eng. Comput. Sci.*, vol. 29, no. 5, pp. 2312–2325, Sep. 2021.
- [11] V. Johansson, S. Sandin, Z. Chang, M. J. Taylor, P. Lichtenstein, B. M. D'Onofrio, H. Larsson, C. Hellner, and L. Halldner, "Medications for attention- deficit/hyperactivity disorder in individuals with or without coexisting autism spectrum disorder: Analysis of data from the Swedish prescribed drug register," *J. Neurodevelopmental Disorders*, vol. 12, no. 1, pp. 1–12, Dec. 2020.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details