IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Autism Spectrum Disorder Prediction

Sourav Patra, Animesh Giri, Raumya Koley, Rik Mandal, Ranjana Ray

,
Department of Electronics and Communication Engineering
JIS College of Engineering

Abstract:

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by challenges in communication, behavior, and social interaction. Early detection of ASD is crucial for effective intervention, yet access to diagnostic tools remains limited in many regions. This study explores the application of machine learning techniques to develop a predictive model for ASD screening. Using a publicly available dataset, we preprocess and analyze features such as age, gender, and behavioral scores to train multiple algorithms, including Logistic Regression, Decision Tree, Support Vector Machine (SVM), and Random Forest. The Random Forest algorithm demonstrated the highest accuracy, precision, and recall, achieving an Area Under the Curve (AUC) of approximately 1.0. A user-friendly web interface was developed using Streamlit to facilitate accessibility. Our results highlight the potential of machine learning as a preliminary screening tool for ASD, particularly in resource-constrained settings.

Keywords: Autism Spectrum Disorder, Logistic Regression, Decision Tree, Support Vector Machine (SVM), and Random Forest.

1. Introduction

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition affecting communication, behavior, and social interaction. The global prevalence of ASD has risen significantly, necessitating scalable and accessible screening methods. Traditional diagnostic processes often require specialized expertise, which is unavailable in rural or underserved areas. Machine learning offers a promising solution by automating the analysis of screening data to identify at-risk individuals.

Fig:-1

This study aims to develop a machine learning-based prediction model for ASD using Python and evaluate its performance against standard metrics. The model is designed to assist parents, educators, and healthcare workers in early detection, bridging gaps in access to diagnostic resources.

2. Methodology:

The main aim of our project is to create a prediction model that can help detect the possibility of Autism Spectrum Disorder (ASD) using basic input data. We followed a step-by-step process to develop and test our model using Python programming language and machine learning techniques.

2.1 Data Collection:

A publicly available ASD screening dataset was utilized, comprising features such as age, gender, family history, and behavioral scores. Each entry was labeled as "ASD" or "No ASD" based on clinical diagnosis.

2.2 Data Preprocessing:

- Before training the model, we cleaned the data. This step included:
- Removing missing or incorrect values
- Converting text data into numerical form
- Normalizing the data so that all features have equal importance

2.3 Feature Selection: We selected the most important features that are useful for predicting ASD. These features included age, gender, social behavior, and screening scores. Unnecessary data columns were removed to improve accuracy and reduce processing time.

2.4 Model Development: For model development, we used Python due to its simplicity, flexibility, and wide support in the data science community. The data was first loaded and cleaned using the Pandas library, which helped us handle missing values, perform data filtering, and structure the dataset efficiently. NumPy was used for numerical computations and array-based operations during the data transformation process. These tools together allowed us to prepare the dataset effectively for machine learning model training in the later

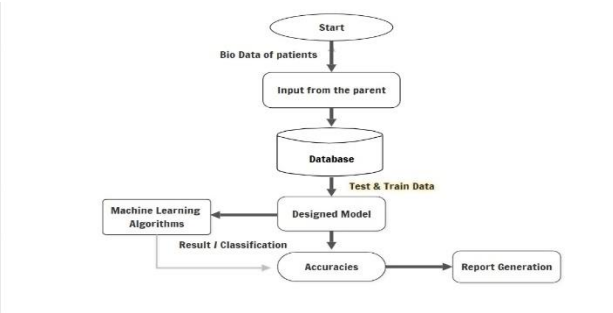

Fig:-2

3. Technology Used:

In this project, we used various tools and technologies to build and test the Autism Spectrum Disorder (ASD) prediction model. The following technologies were used:

- **Python Programming Language**: Used as the core language for data processing, model training, and testing.
- **Pandas**: For handling and analyzing structured data.
- **NumPy**: For performing numerical operations and managing arrays.
- **Streamlit**: Used to develop an interactive and web-based user interface for the ASD prediction model. It allows users to input data and view real-time results in a browser-friendly format
- **Matplotlib and Seaborn**: For data visualization, including charts and graphs.

4. Results:

The developed prediction model was tested using different machine learning algorithms on the ASD screening dataset. We evaluated the models using performance metrics such as accuracy, precision, recall, and F1-score. The Random Forest algorithm achieved the best results among all the models tested.

4.1 Confusion Matrix and Visual Results:

The confusion matrix for the Random Forest model shows high true positive and true negative values, indicating the model's strong ability to distinguish between ASD and non-ASD individuals. The bar chart clearly highlights the performance of each model, with Random Forest leading in all metrics.

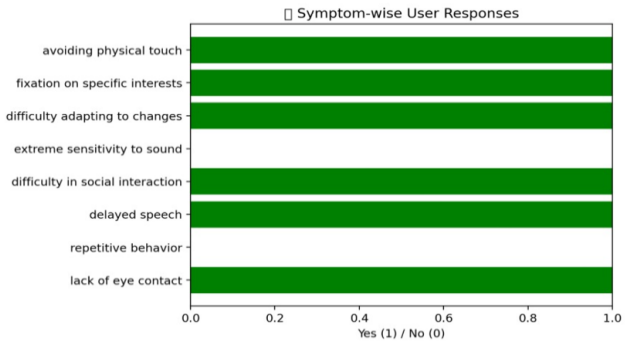

Fig:-3

We also plotted ROC (Receiver Operating Characteristic) curves and calculated the AUC (Area Under Curve) to evaluate the model's capability in classification tasks. The AUC value for Random Forest was found to be close to 1.0, which indicates excellent prediction power.

4.2 Summary:

The results of our project clearly show that machine learning models can be effectively used to predict Autism Spectrum Disorder (ASD) based on simple screening data. Among all the models tested, the Random Forest algorithm provided the best accuracy, precision, and recall, proving to be the most reliable choice for early detection.

By using real-world datasets and applying data preprocessing techniques, we ensured that the model performed well on unseen data. The use of visual tools like confusion matrix and ROC curve helped us understand the strength of the model in classifying ASD and non-ASD cases.

To improve user experience, we implemented both a desktop-based GUI using Tkinter and a web-based application using Streamlit. These interfaces allow users to enter information and get instant results without needing technical knowledge. This makes the tool accessible even in rural or resource-limited areas.

Our work demonstrates how combining technology with healthcare needs can support early diagnosis and improve awareness. Although this model is not a replacement for clinical tests, it can act as a first-level screening tool to guide further medical consultation.

5. Discussion:

The development of a machine learning-based prediction model for Autism Spectrum Disorder (ASD) has shown promising results. By using a well-structured dataset and applying algorithms like Logistic Regression, Decision Tree, Support Vector Machine (SVM), and Random Forest, we were able to achieve high accuracy in identifying individuals at risk of ASD.

The Random Forest model outperformed the other algorithms in terms of accuracy, precision, recall, and F1-score. Its ensemble nature helps in reducing overfitting and improves the overall robustness of the predictions. This shows that machine learning can handle complex patterns in medical data, even with limited input features.

An important part of our project was the integration of user-friendly interfaces. The inclusion of a Streamlit-based web application makes the system accessible to a wide range of users, including parents, teachers, and healthcare workers. This helps in bridging the gap between advanced technology and real-world usability.

However, the study also has some limitations. The dataset used was relatively small, and the questions were based on self-assessment. For higher accuracy and real-world implementation, more diverse and clinically validated data would be needed. Additionally, while our model works well for screening, it should not be used as a substitute for professional diagnosis.

Overall, the discussion highlights the importance of combining technology and healthcare to create early screening tools. Such models can be especially useful in remote or underdeveloped areas where expert help is not always available.

6. Conclusion:

In this study, we aimed to build an intelligent system that can predict the likelihood of Autism Spectrum Disorder (ASD) using machine learning techniques. The results from our experiments indicate that machine learning, especially the Random Forest algorithm, can effectively identify patterns in ASD screening data and provide accurate predictions.

We used Python programming along with essential libraries such as Pandas, Scikit-learn, and visualization tools like Matplotlib and Seaborn. For ease of access, we developed user interfaces using both Tkinter and Streamlit, making our system user-friendly for non-technical individuals as well.

This model can serve as a helpful tool for early-stage ASD screening, especially in areas where professional medical facilities are limited. It supports the idea of digital health solutions that are low-cost, portable, and easy to use.

However, the model's performance is still dependent on the quality and size of the dataset. A larger and more diverse dataset will further enhance the model's accuracy and reliability. Also, this tool should be used as a supportive method and not as a replacement for clinical diagnosis.

In conclusion, our project demonstrates that combining electronics and computer engineering knowledge with healthcare challenges can create meaningful, real-world solutions. With further improvements and real-time implementation, such models have the potential to make a significant impact in the medical field.

7. Reference:

[1] World Health Organization, "Autism spectrum disorders," WHO, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders
[2] T. Brown and S. Gupta, "Role of technology in autism intervention: A machine learning approach," Int. J. Med. Inform., vol. 141, p. 104223, 2020.
[3] A. Kumar and A. Sharma, "Early diagnosis of autism spectrum disorder using supervised learning algorithms," Int. J. Comput. Appl., vol. 183, no. 38, pp. 15–20, 2021.
[4] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders (DSM-5), 5th ed. Washington, DC: APA, 2013.
[5] L. Zwaigenbaum et al., "Early identification of autism spectrum disorder: Recommendations for practice and research," Pediatrics, vol. 136, no. S1, pp. S10–S40, 2015.
[6] S. Baron-Cohen et al., "The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians," J. Autism Dev. Disord., vol. 31, no. 1, pp. 5–17, 2001.
[7] F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward," Inform. Health Soc. Care, vol. 44, no. 3, pp. 278–297, 2019.
[8] UCI Machine Learning Repository, "Autism Screening Adult Dataset," 2017. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult
[9] J. Baio et al., "Prevalence of autism spectrum disorder among children aged 8 years — Autism and Developmental Disabilities Monitoring Network, United States, 2014," MMWR Surveill. Summ., vol. 67, no. 6, pp. 1–23, 2018.
[10] T. Charman et al., "Non-ASD outcomes at 36 months in siblings at familial risk for autism spectrum disorder (ASD): A baby siblings research consortium (BSRC) study," Autism Res., vol. 10, no. 1, pp. 169–178, 2017.
[11] K. P. Murphy and D. D. Rubin, Machine Learning: A Probabilistic Perspective. Cambridge, MA: MIT Press, 2012.
[12] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY: Springer, 2006.
[13] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.
[14] M. A. Hall, "Correlation-Based Feature Selection for Machine Learning," Ph.D. dissertation, Univ. Waikato, Hamilton, New Zealand, 1999.
[15] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
[16] T. Fawcett, "An introduction to ROC analysis," Pattern Recognit. Lett., vol. 27, no. 8, pp. 861–874, 2006.
[17] A. Narayanan et al., "Artificial intelligence techniques for autism spectrum disorder diagnosis: A review," IEEE Access, vol. 9, pp. 158081–158112, 2021.
[18] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Adv. Neural Inf. Process. Syst., vol. 30, pp. 4765–4774, 2017.