Autism Prediction using Machine Learning

Abhishek Madhu Vidya  
*Department of Electronics and Computer Engineering, Amrita Vishwa Vidyapeetham* Bengaluru,India  
<bl.en.u4eac21003@bl.students.amrita.edu>A J Sreekar  
*Department of Electronics and Computer Engineering, Amrita Vishwa Vidyapeetham* Bengaluru,India  
[bl.en.u4eac21011@bl.students.amrita.edu](mailto:bl.en.u4eac21011@bl.students.amrita.edu)

Mohith D M  
*Department of Electronics and Computer Engineering, Amrita Vishwa Vidyapeetham* Bengaluru,India  
[bl.en.u4eac21044@bl.students.amrita.edu](mailto:bl.en.u4eac21044@bl.students.amrita.edu)Doradla Kaushik   
*Department of Electronics and Computer Engineering, Amrita Vishwa Vidyapeetham* Bengaluru,India  
[bl.en.u4eac21020@bl.students.amrita.edu](mailto:bl.en.u4eac21020@bl.students.amrita.edu)

**Abstract—Autism spectrum disorder (ASD), a neurodevelopmental syndrome, is commonly accompanied by sensory issues, such as an excessive or inadequate sensitivity to sound, smell, or touch. Machine learning (ML) is becoming more and more important in our daily lives. A person with an ASD struggle with social interaction and communication their entire lives. The progression of this illness starts in childhood and continues throughout adulthood. As a result, this condition has a profound impact on a person's life. An early diagnosis is crucial for reducing the symptoms of autism spectrum disorders and enhancing the quality of life for persons with autism. The early detection method will significantly aid in managing the subject's physical and mental health.**

Keywords—Autism Spectrum Disorder (ASD), Machine Learning, Artificial Intelligence, Support Vector Machine, Naïve Bayes, Neural Network, CNN.

# INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder characterized by challenges in social interaction, speech and nonverbal communication, and repetitive behaviors. Early detection and intervention are critical for improving the quality of life for individuals with ASD. Machine Learning (ML) has emerged as a powerful tool in automating the detection and diagnosis of ASD, providing a data-driven approach to augment clinical assessments [1].

Few applications that can implemented in autism prediction are developing a Web Application [2], Developing a Mobile Application [4], Speech and Language Analysis, Behavioral Pattern Recognition

The general phases and steps used in predicting autism are Data cleaning- The raw data, which comes from primary sources, must undergo extensive preparation before we can draw any inferences from it or perform any modelling, exploratory data analysis here data is analysed utilising visual methods which is known as "EDA." Using statistical summaries and graphical depictions, it is used to identify common trends and patterns as well as to test hypotheses.

The motivation for this projects ASD is difficult to diagnose, and the rise in ASD diagnoses around the world has motivated medical professionals and scientists to develop better screening techniques [3]. These investigations have not produced any solid results regarding the capacity to predict autism features in terms of a number of age parameters. Therefore, our main aim for the project is to find a machine learning model which covers the above problem.

# LITERATURE SURVEY

K. Mahalakshmi et al. [1] suggested that early diagnosis can save a person's life, disease prediction models are crucial to the delivery of healthcare services. Techniques like exploratory data analysis and data pre-processing are crucial for creating machine learning models that work better. They also said that RF (Random Forest) classifier was able to give highest accuracy when compared to KNN, SVM (Support Vector Machine) and DT (Decision Tree)

Shiva Kumar R et al. [2] used machine learning techniques and algorithms like KNN, Logistic regression, decision trees, random forests, the Naive Bayes and XGB classifier to propose a system to diagnose autism spectrum disorder based on various characteristics presented by the user on the front end. They put their theory into practice, and it was discovered that the algorithm had a 98.2% accuracy rate in detecting autism spectrum condition.

V. Kavita et al. [3] have presented the accuracy rates of various techniques, including LR (Logistic Regression), SVM, NB (Na¨ıve Bayes), and the proposed PSO-CNN. But concluded that on all four datasets, the suggested PSO-CNN method achieves an accuracy rate of 98.53%. The proposed model offers accurate performance rates in toddlers, kids, teens, and adults that is 98.9%, 98.8%, 98.9%, and 99.8%, respectively. This investigation reveals that the recommended PSO-CNN approach performs better in terms of accuracy than other ways.

Miguel Ângelo Lellis Moreira et al. [4] has said that Modern life has been increasingly influenced by technology, which has compelled an unavoidable shift away from traditional physical services to online ones. In cases when assessing bank fraud, ML technologies are useful for determining whether predictive analysis is feasible. They may be helpful in identifying potential fraud attempts in a system for financial transactions.

P. Nivedhitha et al. [5] have presented that their research has revealed: first, a model was developed to anticipate the characteristics of autism. Using the AQ-10 dataset, the proposed method can identify autism with 95% accuracy. After obtaining results from various supervised learning techniques, including Logistic Regression, Ada boost and Random Forest were found to be very viable and to be more accurate than the other algorithms.

Dr. R. Siva et al. [6] have suggested that in the study they conducted, For the detection of ASD, fMRI, and EEG signals, as well as intensive training techniques and machine learning models, were offered. A variety of neural network types are used in the in-depth study model to increase classification accuracy. The drawback of the Deep Learning approach is that it needs more data for training to create good models. In conclusion, there are several techniques for diagnosing ASD which are covered in this document's conclusion, along with actionable suggestions for each model or algorithm.

Vaibhav Vishal et al. [7] focused on the identification of specific qualities that aid in automating the diagnosis process and further test and compare K-Nearest Neighbour, Logistic Regression, SVM, and Nave Bayes machine learning algorithms to predict the occurrence of autism. The experimental investigation demonstrates that Nave Bayes outperforms the other models for predicting autistic disorder. In the future, the system can be enhanced to increase precision and decrease mistake rates. A 98% prediction was made.

Ryoichi Kojima et al. [8] employed an effective and efficient method to identify autism tendencies for various age groups is to apply the mobility characteristics of human groupings, where each group is classified according to demographics, which according to our EDA are universal throughout places. A distance representation between zones was also attempted, along with its combination with other geographical data like population and trip destination prediction using closest neighbor search.

 Shaikhah Almana et al. [9] used NBC, KNN, and LR to analyze ASD datasets for kids, teens, and adults. The three datasets are utilized to evaluate the models using a 10-fold cross-validation technique. The features are then reduced using the Info gain evaluator and the Attribute selection method. KNN, logistic regression and Naïve Bayes are the algorithms used in this study.

Linda Ejlskov et al. [10] proposed that risk for autism spectrum disorder (ASD) has been associated with a family history of illnesses, such as autism, depression, or epilepsy. This study investigates the possibility of predicting ASD risk using family history information.

Gurusubramani S et al.[11], this study suggests autism spectrum disorder emotion detection can watch a child's facial expression and anticipate their movements using facial expression recognition (FER).

Liangqi Chen et al. [12] has said that Data cleaning is the starting point of data research and one of the crucial ways to enhance data quality. Other tasks include describing the data, visualizing the distribution, comparing relationships between them, developing intuition from the data, and summarizing the outcomes.

Reshma Shaji [13] has presented that social networking sites are becoming increasingly popular, which has increased the amount of user-generated material and insightful data. According to an analysis of the first 3000 entries, non-political subjects predominate, followed by political conversations. The classification accuracy of Logistic Regression and SVM is greater, however KNN is less effective. The greatest results are obtained when post title, self-text, and comments are used as characteristics for classification, highlighting the significance of high-quality data and rich content. Language analysis will be explored in greater detail in future studies.

Tania Akter et. al [14] have proposed recent CDC data reports from a nationwide survey finding ASD in 1 in 59 children, with a higher incidence in boys. Early diagnosis and intervention are crucial. In this study, data was consolidated, preprocessed, and divided by gender, with the female dataset showing better classification results. The most successful model (MLP) achieved high accuracy (98.25%), precision (98.33%), recall (98.16%), AUROC (98.25%), and F-measure (98.25%) for the female dataset. This work emphasizes the importance of early ASD diagnosis, offering potential benefits for therapy and drug development, with the future to expand using deep learning and more data.

Sajeev Ram Arumugam et. al [15] has presented that the original system was created using machine learning methods, and it was later changed to include neural networks. Convolution neural networks performed well on the system, which was later improved to improve the classification outcomes. The layer counts and weights were manually modified to prevent the system from performing too well or too poorly. Loss value and accuracy metrics were evaluated to determine the system's performance, which was determined to be 91% and 0.53, respectively.

Chongruo Wu et al. [16], Chongruo’s team experimented on ASD diagnosed using a machine learning (ML) approach that looks for behaviors in films of infants between the ages of 6 and 36 months. Directed gaze toward people or items of interest, pleasant affect, and vocalization are among the interesting activities.

Suman Raj et al. [17] sought to investigate the potential application of Nave Bayes, Support Vector Machine, Logistic Regression, KNN, Neuronal Network, and Convolution Neural Network for analysis and prediction of ASD issues in children, adolescents, and adults. Both the SVM and the CNN-based models in this work exhibit the same accuracy of prediction of 98% after accounting for missing values.

Prodipta Mondal et al. [18] This paper aims to propose a useful prediction model based on ML technique and to develop a mobile application for predicting ASD for individuals of any age. The proposed model can predict autism with 92.26%, 93.78%, and 97.10% accuracy in case of child, adolescent, and adult persons, respectively. The outcome of this research provides an effective and efficient approach to detect autism traits for different age groups.

Thy Nguyen et al. [19] provides that they used the methodology to analyze the phenotypes of ASD using various normalizing methods and clustering algorithms. Due to the configuration and intensity of the behavioral symptoms, there is a substantial variation in the phenotypic presentation of ASD.

In a different study, Zunino et al. [20] attempted to identify ASD by examining video motions. Twenty of the 40 youngsters in the trial had an ASD diagnosis, while the other twenty were healthy, according to the authors. All the kids were videotaped making the same basic motion of holding a bottle. To determine if the patient has ASD or not, the experiment used video recordings and a recurrent deep neural network model. The model performed most accurately when the threshold was set to 0.5.

This research provides, Early diagnosis of autism spectrum disorder (ASD) can have a substantial impact on how well a person with ASD is treated and supported, it is frequently emphasised throughout the declarations. Several machine learning techniques are frequently used in the creation of prediction models for the diagnosis of ASD, including Random Forest (RF), K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), Naive Bayes (NB), and Convolutional Neural Network (CNN). Some of these models achieve rates as high as 98%, however they frequently aim for high accuracy rates. Additionally, the study takes into account several age brackets, including kids, teens, and adults, and it emphasises the significance of good data pretreatment, including feature selection and data cleaning. Prediction models take into account demographic and behavioural information, including age, gender, and behavioural traits like gaze and vocalisation patterns. For better classification results, deep learning approaches, particularly neural networks like CNN, are being investigated. The use of machine learning extends beyond ASD diagnosis into other industries including financial fraud detection, with some research doing comparison evaluations of several machine learning algorithms to identify the best efficient classifier for ASD prediction.

The following are the common research gaps and drawback found from the above literature survey conducted:

* None of the papers used Exploratory data Analysis technique to predict autism.
* Lack of Clinical Data Integration
* Heterogeneity of ASD
* Insufficient Detail on Feature Selection
* Overemphasis on Accuracy

# REFERENCES

[1] K. Mahalakshmi and P. Sujatha, "The Role of Exploratory Data Analysis and Pre-processing in the Machine Learning Predictive Model for Heart Disease," *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, Chennai, India, 2023, pp. 1-8, doi: 10.1109/ACCAI58221.2023.10199714.

[2] S. K. R. Naik, D. M, R. P B, S. Prakash and U. J. Royal, "Determination and Diagnosis of Autism Spectrum Disorder using Efficient Machine Learning Algorithm," *2023 3rd International Conference on Intelligent Technologies (CONIT)*, Hubli, India, 2023, pp. 1-5, doi: 10.1109/CONIT59222.2023.10205718.

[3] V. Kavitha and R. Siva, "Classification of Toddler, Child, Adolescent and Adult for Autism Spectrum Disorder Using Machine Learning Algorithm," *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2023, pp. 2444-2449, doi: 10.1109/ICACCS57279.2023.10112932.

[4] Miguel Ângelo Lellis Moreira, Claudio de Souza Rocha Junior, Diogo Ferreira de Lima Silva, Marcos Alexandre Pinto de Castro Junior, Igor Pinheiro de Araújo Costa, Carlos Francisco Simões Gomes, Marcos dos Santos,Exploratory analysis and implementation of machine learning techniques for predictive assessment of fraud in banking systems, Procedia Computer Science, Volume 214, 2022, Pages 117-124, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2022.11.156.

[5] K. S. Tejaswi, K. Meghavarshini and P. Nivedhitha, "Autism Prediction using ML Algorithms," *2022 1st International Conference on Computational Science and Technology (ICCST)*, CHENNAI, India, 2022, pp. 1-6, doi: 10.1109/ICCST55948.2022.10040300.

[6] V. Kavitha and R. Siva, "Review of Machine Learning Algorithms for Autism Spectrum Disorder Prediction," *2022 International Conference on Automation, Computing and Renewable Systems (ICACRS)*, Pudukkottai, India, 2022, pp. 608-613, doi: 10.1109/ICACRS55517.2022.10029201.

[7] V. Vishal, A. Singh, Y. B. Jinila, K. C, S. P. Shyry and J. Jabez, "A Comparative Analysis of Prediction of Autism Spectrum Disorder (ASD) using Machine Learning," *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India, 2022, pp. 1355-1358, doi: 10.1109/ICOEI53556.2022.9777240.

[8] R. Kojima, R. Legaspi and S. Wada, "Trip Destination Prediction by Cross-City Exploratory Data Analysis Approach in People Flow Data," *2022 IEEE International Conference on Big Data (Big Data)*, Osaka, Japan, 2022, pp. 6547-6552, doi: 10.1109/BigData55660.2022.10020611.

[9] S. Almana and M. Hammad, "Predicting ASD Using Optimized Machine Learning," *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, Chiangrai, Thailand, 2022, pp. 1598-1602, doi: 10.1109/DASA54658.2022.9765019.

[10] Linda Ejlskov, Jesper N. Wulff, Amy Kalkbrenner, Christine Ladd-Acosta, M. Danielle Fallin, Esben Agerbo, Preben Bo Mortensen, Brian K. Lee, Diana Schendel,Prediction of Autism Risk From Family Medical History Data Using Machine Learning: A National Cohort Study From Denmark,Biological Psychiatry Global Open Science,Volume 1, Issue 2,2021,Pages 156-164,ISSN 2667-1743,https://doi.org/10.1016/j.bpsgos.2021.04.007.

[11] G. S, D. B and S. M, "Artificial Intelligence and Face Emotion Prediction Based Training for Autism Kids," *2022 1st International Conference on Computational Science and Technology (ICCST)*, CHENNAI, India, 2022, pp. 344-349, doi: 10.1109/ICCST55948.2022.10040418.

[12] L. Chen, B. Wang, Y. Wang and X. Wang, "Exploratory Data Analysis on the Usage of COVID-19 Vaccine," *2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT)*, Changsha, China, 2021, pp. 502-506, doi: 10.1109/ISCIPT53667.2021.00107.

[13] R. Shaji, "Exploratory data analysis on Reddit data: An efficient pipeline for classification of flairs," *2021 IEEE Seventh International Conference on Multimedia Big Data (Big MM)*, Taichung, Taiwan, 2021, pp. 65-68, doi: 10.1109/BigMM52142.2021.00018.

[14] T. Akter and M. H. Ali, "Predicting Autism Spectrum Disorder Based On Gender Using Machine Learning Techniques," *2021 3rd International Conference on Electrical & Electronic Engineering (ICEEE)*, Rajshahi, Bangladesh, 2021, pp. 185-188, doi: 10.1109/ICEEE54059.2021.9718798.

[15] S. R. Arumugam, R. Balakrishna, R. Khilar, O. Manoj and C. S. Shylaja, "Prediction of Autism Spectrum Disorder in Children using Face Recognition," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2021, pp. 1246-1250, doi: 10.1109/ICOSEC51865.2021.9591679.

[16] C. Wu *et al*., "Machine Learning Based Autism Spectrum Disorder Detection from Videos," *2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM)*, Shenzhen, China, 2021, pp. 1-6, doi: 10.1109/HEALTHCOM49281.2021.9398924.

[17] Suman Raj, Sarfaraz Masood,Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques,Procedia Computer Science,Volume 167,2020,Pages 994-1004,ISSN 1877-0509,https://doi.org/10.1016/j.procs.2020.03.399

[18] K. S. Omar, P. Mondal, N. S. Khan, M. R. K. Rizvi and M. N. Islam, "A Machine Learning Approach to Predict Autism Spectrum Disorder," 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox'sBazar, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ECACE.2019.8679454.

[19] T. Nguyen, K. Nowell, K. E. Bodner and T. Obafemi-Ajayi, "Ensemble validation paradigm for intelligent data analysis in autism spectrum disorders," *2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, St. Louis, MO, USA, 2018, pp. 1-8, doi: 10.1109/CIBCB.2018.8404960. 

[20] A. Zunino et al., "Video Gesture Analysis for Autism Spectrum Disorder Detection," 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 2018, pp. 3421-3426, doi: 10.1109/ICPR.2018.8545095.