



A Systematic Review On Autism Spectrum Disorder Detection Using Machine Learning Techniques

Abhishek Umre
MTech Scholar
CSE, Department
NIIST, Bhopal

Vaibhav Patel
Associate Professor,
CSE, Department
NIIST, Bhopal

Anurag Shrivastav
Associate Professor
CSE, Department
NIIST, Bhopal

Abstract: Autism Spectrum Disorder (ASD) is a complex neurodevelopment condition characterized by challenges in social interaction, communication, and repetitive behaviors. Early and accurate diagnosis of ASD is crucial for effective intervention and treatment. Traditional diagnostic methods rely on behavioral assessments, which can be time-consuming and subjective. In recent years, machine learning (ML) techniques have emerged as powerful tools for automating and improving ASD detection by analyzing diverse data sources such as clinical reports, neuroimaging scans, genetic data, and behavioral patterns. This systematic review explores the latest advancements in ML-based ASD detection, highlighting various algorithms, feature selection methods, and performance metrics used in recent studies. We examine supervised, unsupervised, and deep learning approaches, emphasizing their strengths and limitations in ASD diagnosis. Additionally, we discuss the challenges associated with dataset availability, model interpretability, and generalization across diverse populations. The review aims to provide insights into the current trends and future directions of ML applications in ASD detection, ultimately contributing to the development of more efficient, accurate, and accessible diagnostic tools.

Keywords— Autism Spectrum Disorder, Machine Learning, Early Diagnosis, Neurodevelopment Disorder, Deep Learning

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopment condition that affects communication, social interactions, and behavior. It manifests in early childhood and varies in severity, making early and accurate diagnosis essential for effective intervention. Traditional ASD diagnosis primarily relies on behavioral assessments, such as the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R). These methods, while widely used, are time-consuming, subjective, and require specialized expertise, often leading to delayed diagnosis and intervention. With the increasing prevalence of ASD worldwide, there is a growing need for more efficient, objective, and scalable diagnostic methods. Machine Learning (ML) has emerged as a promising approach for ASD detection by leveraging computational models to analyze large datasets and identify patterns that may not be easily recognizable by human evaluators. ML techniques can process a variety of data sources, including genetic information, neuroimaging scans, eye-tracking data, speech patterns, and behavioral attributes. By automating the analysis of these complex data types, ML models offer the potential for faster, more accurate and cost-effective ASD diagnosis. Several ML techniques have been explored for ASD detection, including supervised learning algorithms such as Support Vector Machines (SVM), Random Forest (RF), and Artificial Neural Networks (ANN). Deep learning, particularly Convolutional Neural Networks (CNN)

and Recurrent Neural Networks (RNN), has demonstrated significant potential in analyzing medical imaging and sequential behavioral data for ASD classification. Additionally, unsupervised learning methods and clustering techniques have been applied to identify hidden patterns in ASD-related data, contributing to improved diagnostic accuracy.

Despite the promising advancements, ML-based ASD detection faces several challenges. The availability of high-quality, diverse datasets remains a significant hurdle, as most studies rely on limited sample sizes that may not generalize well to broader populations. Additionally, model interpretability and explainability are crucial concerns in medical applications, as clinicians require transparent decision-making processes to trust AI-assisted diagnostics. Ethical considerations, such as data privacy and potential biases in ML models, also need to be addressed to ensure fair and responsible deployment of these technologies. This systematic review aims to provide a comprehensive analysis of the current state of ML applications in ASD detection, highlighting key methodologies, challenges, and future research directions. By exploring existing literature, this study seeks to contribute to the development of reliable, AI-driven diagnostic tools that can aid in early ASD detection and improve patient outcomes. Recent studies have explored hybrid models that combine multiple ML techniques to improve ASD detection accuracy. Additionally, advancements in natural language processing (NLP) have enabled the analysis of speech and text-based interactions to identify ASD-related linguistic markers. Transfer learning and federated learning are also gaining attention to address challenges related to limited datasets and data privacy.

Despite the promising advancements, ML-based ASD detection faces several challenges. The availability of high-quality, diverse datasets remains a significant hurdle, as most studies rely on limited sample sizes that may not generalize well to broader populations. Additionally, model interpretability and explainability are crucial concerns in medical applications, as clinicians require transparent decision-making processes to trust AI-assisted diagnostics. Ethical considerations, such as data privacy and potential biases in ML models, also need to be addressed to ensure fair and responsible deployment of these technologies.

II. BACKGROUND AND RELATED WORK

The literature review explores existing research on Autism Spectrum Disorder (ASD) detection using machine learning techniques, highlighting key methodologies, datasets, and performance metrics. It critically analyzes various approaches, including supervised, unsupervised, and deep learning models, to identify strengths, limitations, and research gaps in the field.

This work [1] utilizes a variety of machine learning techniques to identify key traits associated with ASD, with the goal of improving and automating the diagnostic process. Eight advanced classification models are analyzed to assess their effectiveness in ASD detection. These models are evaluated based on accuracy, precision, recall, specificity, F1-score, AUC, kappa, and log loss metrics to identify the most suitable classifier for binary datasets. For the children's dataset, SVM and LR models achieve a perfect accuracy of 100%, while the LR model attains the highest accuracy of 97.14% for the adult dataset. Additionally, our proposed ANN model, when optimized with carefully tuned hyperparameters, achieves a peak accuracy of 94.24% on a newly combined dataset. Given the high accuracy of classification models that rely on true labels, we further explore five widely used clustering algorithms to analyze model behavior in the absence of true labels. The best clustering models are identified using Normalized Mutual Information (NMI), Adjusted Rand Index (ARI), and Silhouette Coefficient (SC) metrics.

This study [2] proposes a machine-learning framework for ASD detection across different age groups, including toddlers, children, adolescents, and adults. The findings highlight the effectiveness of predictive models as valuable tools for healthcare professionals in diagnosing ASD. The research provides insights into key screening features, though its limitation lies in the small dataset, restricting model generalization.

Despite this, the study contributes to developing an automated diagnostic model for ASD in youngsters. Future work will focus on expanding the dataset to enhance model generalization and improve ASD detection across all age groups, with potential applications in identifying other neurodevelopmental disorders.

The purpose of this work [3] is to detect autism from facial images using a deep learning model. To accurately identify autism in children, we used three pre-trained CNN models, VGG16, VGG19 and, EfficientnetB0, as feature extractors and binary classifiers. The suggested models were trained using a publicly available dataset from Kaggle that included 3014 images of children characterized as autistic and non-autistic. The models yielded accuracies of 84.66%, 80.05%, and 87.9%, respectively.

In this work [4], detection of Autism Spectrum Disorder was attempted using various machine learning and deep learning techniques. Various performance evaluation metrics were used to analyze the performance of the models implemented for ASD detection on non-clinical dataset from three sets of age groups viz. Child, Adolescents and the Adult. When comparing the result with another recent study [3] on this problem got a better result of the CNN classifier instead of SVM with including all its features attributes after handling missing values. In this work after handling missing value, both the SVM and CNN based models show the same accuracy of prediction of about 98.30 % for ASD Child dataset. However for the remaining two other datasets, the CNN based model was able to achieve highest accuracy result than all the other considered model building techniques, These results strongly suggest that a CNN based model can be implemented for detection of Autism Spectrum Disorder instead of the other conventional machine learning classifier suggested in earlier researches.

This research work [5] attempts towards investigate potential use about machine learning algorithms for predicting & analyzing ASD problems in children, adults & adolescents. On three separate publicly accessible, non-clinically relevant ASD datasets, suggested approaches are assessed. There are 292 instances & 21 attributes in first dataset relating towards screening for ASD in children. Adult individuals make up second dataset for ASD screening, which has a total about 704 instances & 21 attributes. There are 104 cases & 21 attributes in third dataset, which is focused on ASD screening in adolescent individuals. Convolutional neural network based models had higher accuracy about 99.53 percent, 98.30 percent, & 96.88 percent for the three datasets respectively.

Autism Spectrum Disorder (ASD) encompasses a range of neurodevelopmental conditions that, while incurable, can be improved with early intervention. In this study [6], we collected early-detected ASD datasets covering toddlers, children, adolescents, and adults. Various feature transformation techniques, including logarithmic, Z-score, and sine functions, were applied to these datasets. Subsequently, multiple classification techniques were implemented and evaluated for their performance. Our findings revealed that SVM achieved the best accuracy for the toddler dataset, Adaboost performed best for the children and adult datasets, and Glmboost yielded the best results for the adolescent dataset. Among the feature transformations, the sine function was most effective for toddlers, while Z-score normalization produced optimal results for children and adolescents. Following these analyses, we employed various feature selection techniques on the Z-score-transformed datasets to identify key ASD risk factors across all age groups. These results suggest that, with proper optimization, machine learning models can effectively predict ASD status, potentially enabling early detection and intervention.

III. DATASET USED

The dataset used in this research was sourced from the publicly accessible UCI Repository. The four ASD datasets—toddlers, adolescents, children, and adults—were obtained from publicly available repositories, specifically Kaggle and UCI ML. These repositories serve as valuable resources for research and analysis related to ASD

IV. TABLE 3.1 DETAIL ASD DATASET

Category	Details
Source	Kaggle, UCI ML Repository
Age Group	Toddlers (1.5 – 3 years), Children (4 – 11 years), Adolescents (12 – 18 years), Adults (18+ years)
Features	<ul style="list-style-type: none"> - Demographic Information (Age, Gender, Ethnicity) - Screening Results (AQ Test Scores, ASD-related Questions) - Family & Medical History (ASD in Family, Jaundice at Birth) - Social & Behavioral Patterns (Eye Contact, Social Interactions, Repetitive Behavior) - Diagnostic Label (ASD Positive/Negative)
Data Collection	Screening tests and standardized ASD questionnaires (e.g., Autism Spectrum Quotient - AQ)
Dataset Format	CSV format with structured features and labels

V. FINDINGS OF THE SURVEY

The survey on Autism Spectrum Disorder (ASD) detection using machine learning (ML) techniques reveals significant advancements in automated diagnosis, offering promising alternatives to traditional behavioral assessments. Various ML models, including Support Vector Machines (SVM), Random Forest (RF), Artificial Neural Networks (ANN), and deep learning-based architectures like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have demonstrated high accuracy in identifying ASD traits from diverse data sources. These include neuroimaging scans, eye-tracking patterns, speech analysis, and genetic information, providing a multidimensional approach to ASD diagnosis. One key finding is that deep learning models, particularly CNNs and hybrid frameworks, outperform traditional ML techniques in processing high-dimensional medical imaging and behavioral data. Additionally, Natural Language Processing (NLP)-based models show potential in detecting linguistic patterns associated with ASD. The integration of multiple data modalities, such as combining neuroimaging with behavioral data, enhances diagnostic accuracy and reduces false positives. However, challenges remain in dataset quality, as most studies rely on limited, imbalanced, or non-diverse datasets, impacting generalizability. The need for explainable AI is another crucial finding, as black-box models hinder clinical trust and adoption. Ethical concerns, including data privacy, bias in models, and the lack of standardized ML-based ASD detection frameworks, pose barriers to widespread implementation.

Future research should focus on enhancing dataset diversity, developing interpretable ML models, and integrating multi-modal approaches to improve diagnostic reliability. Advancements in transfer learning, federated learning, and real-time ASD screening tools hold promise for more accessible and scalable solutions. Overall, ML-driven ASD detection presents transformative potential but requires further refinement for clinical acceptance and widespread application.

CONCLUSION

The systematic review highlights the growing role of machine learning (ML) in Autism Spectrum Disorder (ASD) detection, offering promising alternatives to traditional diagnostic approaches. Various ML techniques, including Support Vector Machines (SVM), Random Forest (RF), Artificial Neural Networks (ANN), and deep learning models like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have demonstrated significant potential in improving the accuracy and efficiency of ASD diagnosis. The integration of multiple data sources, such as neuroimaging, eye-tracking, speech patterns, and genetic data, has further enhanced diagnostic capabilities. Despite these advancements, challenges remain in dataset availability, model interpretability, and ethical concerns such as bias and data privacy. Many ML models struggle with generalizability due to limited, imbalanced datasets, and the black-box nature of deep learning techniques hinders their clinical adoption. Addressing these issues requires further research into explainable AI, diverse dataset collection, and robust validation strategies. Future advancements in transfer learning, federated learning, and real-time ASD screening applications could significantly improve early detection and intervention. With continued research, ML-driven ASD detection can become a more reliable, scalable, and accessible tool for healthcare professionals and families.

REFERENCES

- [1] Rownak Ara Rasul et. al “An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder” <https://doi.org/10.1016/j.health.2023.100293> Published by Elsevier 2024
- [2] Mahmoud M. Abdelwahab et. al “Autism Spectrum Disorder Prediction in Children Using Machine Learning” DOI: 10.57197/JDR-2023-0064, Journal of Disability Research 2024
- [3] Pranavi Reddy and Andrew J “Diagnosis of Autism in Children Using Deep Learning Techniques by Analyzing Facial Features” 2023, 59, 198. <https://doi.org/10.3390/engproc2023059198>
<https://www.mdpi.com/journal/engproc>
- [4] Suman raj, sarfaraj masood “Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques” International Conference on Computational Intelligence and Data Science (ICCIDS 2019) Elsevier 2020
- [5] P. Kiran Kumar¹, Dr. K. Santhi Sree² “analysis & detection of autism spectrum disorder using machine learning techniques” 2022 IJCRT | Volume 10, Issue 9 September 2022 | ISSN: 2320-2882
- [6] Tania Akter et. al. “Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders” IEEE 2019, 0.1109/ACCESS.2019.2952609
- [7] Mengyi Liao et al. “Application of Machine Learning Techniques to Detect the Children with Autism Spectrum Disorder” Hindawi Journal of Healthcare Engineering Volume 2022, Article ID 9340027, 10 pages <https://doi.org/10.1155/2022/9340027>
- [8] S.R. de Boer, D.E. Griswold, B.S. Myles, S.E. Byrd, J.B. Ganz, K.T. Cook, K. Otten, J. Ben-Arieh, S.A. Kline, L.G. Adams, et al., Autism Spectrum Disorders: Interventions and Treatments for Children and Youth, Corwin Press, 2004.
- [9] M. Rutter, A. Le Couteur, C. Lord, et al., Autism Diagnostic Interview-Revised, Vol. 29, Western Psychological Services, Los Angeles, CA, 2003, p. 30.
- [10] C. Lord, S. Risi, L. Lambrecht, E.H. Cook, B.L. Leventhal, P.C. DiLavore, A. Pickles, M. Rutter, The autism diagnostic observation schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism, J. Autism Dev. Disord. 30 (2000) 205–223.

- [11] R.J. Landa, Diagnosis of autism spectrum disorders in the first 3 years of life, *Nat. Clin. Pract. Neurol.* 4 (3) (2008) 138–147.
- [12] S.R. Shahamiri, F. Thabtah, Autism AI: a new autism screening system based on artificial intelligence, *Cogn. Comput.* 12 (4) (2020) 766–777.
- [13] Y.-Y. Song, L. Ying, Decision tree methods: applications for classification and prediction, *Shanghai Arch. Psychiatry* 27 (2) (2015) 130.
- [14] C.-W. Hsu, C.-C. Chang, C.-J. Lin, et al., *A practical guide to support vector classification*, 2003.
- [15] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.* 20 (1987) 53–65.
- [16] N.X. Vinh, J. Epps, J. Bailey, Information theoretic measures for clustering's comparison: is a correction for chance necessary? in: *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, pp. 1073–1080.
- [17] L. Hubert, P. Arabie, Comparing partitions, *J. Classification* 2 (1) (1985) 193–218.
- [18] J. MacQueen, Classification and analysis of multivariate observations, in: *5th Berkeley Symp. Math. Statist. Probability*, 1967, pp. 281–297.
- [19] W.E. Donath, A.J. Hoffman, Lower bounds for the partitioning of graphs, *IBM J. Res. Dev.* 17 (5) (1973) 420–425, <http://dx.doi.org/10.1147/rd.175.0420>.
- [20] Kaggle. (2022). Autism Spectrum Disorder Detection Dataset for Toddlers. [Online]. Available: <https://www.kaggle.com/fabdelja/autism-screeningfor-toddlers>
- [21] UCI. (2022). UCI Machine Learning Repository: Autistic Spectrum Disorder Screening Data for Adolescent Data Set. [Online]. Available: <https://shorturl.at/fhxCZ>
- [22] UCI. (2022). UCI Machine Learning Repository: Autism Screening Adult Data Set. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult>
- [23] UCI. (2022). UCI Machine Learning Repository: Autistic Spectrum Disorder Screening Data for Children Data Set. [Online]. Available: <https://shorturl.at/fiwLU>