

Survey On: Early Detection & Analysis Of Autism Spectrum Disorder Using Machine Learning

Dr. Asma Parveen, Dr. Raafiya Gulmeher, Rabbika Tasneem

HOD, Dept. of Computer Science and Engineering, Khaja BandaNawaz

University, Kalaburagi, India.

Assistant Professor, Dept. of Computer Science and Engineering, Khaja BandaNawaz

University, Kalaburagi, India.

PG Student, Dept. of Computer Science and Engineering, Khaja BandaNawaz University,

Kalaburagi, India.

ABSTRACT

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition that affects communication, social interaction, and behavior. Early detection of ASD is crucial for providing timely intervention and improving the quality of life for affected individuals. However, traditional diagnostic methods are often time-consuming, subjective, and require expert involvement, which may not always be accessible. This project proposes an intelligent system for the early detection of Autism Spectrum Disorder using machine learning techniques. The system utilizes questionnaire-based datasets and behavioral features to classify individuals as ASD-positive or ASD-negative. Various preprocessing techniques are applied to clean and normalize the data, followed by the implementation of classification algorithms to achieve accurate predictions. The developed system provides a user-friendly interface where users can input responses and receive instant predictions. By leveraging machine learning, the model enhances diagnostic efficiency and reduces dependency on manual evaluation. This approach aims to support healthcare professionals and caregivers in early screening while ensuring accuracy and reliability.

Keywords: Autism Spectrum Disorder, Machine Learning, Early Detection, Classification, Python

1. INTRODUCTION

In the contemporary digital age, the importance of early diagnosis and intervention in healthcare has significantly increased, particularly in the field of neurodevelopmental disorders. Autism spectrum disorder (asd) is a complex developmental condition that affects communication, behavior, and social interaction. With the rising number of ASD cases worldwide, there is a growing need for efficient, accessible, and reliable methods to support early screening and detection. Traditional diagnostic approaches rely heavily on clinical observation and expert evaluation, which can be time-consuming, expensive, and sometimes inaccessible, especially in remote or underserved areas. This has created a demand for intelligent, automated systems that can assist in the early identification of ASD. Autism spectrum disorder detection using machine learning techniques leverages the capabilities of artificial intelligence to address this challenge. Machine learning models can analyze patterns in behavioral and questionnaire-based data to predict the likelihood of ASD in individuals. In this context, supervised learning algorithms such as adaboost classifier and gradient boosting classifier are commonly used to build predictive models. These models are trained on structured datasets, such as the aq-10 screening dataset, where they learn to identify correlations between input features and diagnostic outcomes. The use of machine learning enables the system to capture subtle behavioral indicators associated with ASD, such as difficulties in social interaction, communication challenges, and repetitive patterns of behavior. Furthermore, data preprocessing techniques and feature selection methods enhance the accuracy and efficiency of the model by ensuring that only relevant information is used during training. The integration of these models into a web-based application using frameworks such as flask allows users to input responses easily and obtain instant predictions. The primary objective of this project is to develop a system that takes user input in the form of questionnaire responses and provides a prediction indicating the likelihood of autism spectrum disorder. The proposed solution aims to support early screening, improve accessibility to diagnostic tools, and assist healthcare professionals in making informed decisions. In addition, this project can be extended to incorporate real-time behavioral analysis, integration with mobile health applications, and support for multiple languages to increase usability across diverse populations. Such enhancements would make the system more robust and adaptable, contributing significantly to advancements in digital healthcare and assistive technologies.

2. LITERATURE SURVEY

[1] "Detection of Autism Spectrum Disorder Using Machine Learning Techniques" – (Duda et al., IEEE 2016)

Summary: This paper explored the use of machine learning algorithms on behavioral screening data to detect ASD. It demonstrated how feature selection and classification models can improve screening accuracy.

Relevance: Provided a strong foundation for using structured questionnaire datasets like AQ-10 in ASD prediction systems.

[2] "Machine Learning for Early Detection of Autism Spectrum Disorder" – (Wall et al., Translational Psychiatry 2012)

Summary: The study utilized behavioral data and applied machine learning classifiers to identify children at risk for ASD. It emphasized reducing the number of screening questions while maintaining high accuracy.

Relevance: Inspired the concept of efficient screening using minimal yet highly informative features.

[3] "Autism Spectrum Disorder Screening Using Data Mining Techniques" – (Thabtah et al., Applied Computing 2017)

Summary: This research applied data mining techniques on ASD screening datasets and compared various classifiers such as Decision Trees and Naïve Bayes.

Relevance: Demonstrated the effectiveness of supervised learning models in ASD detection and highlighted the importance of dataset preprocessing.

[4] "A Machine Learning Approach for Autism Classification" – (Kosmicki et al., Scientific Reports 2015)

Summary: This paper used machine learning to identify the most predictive features from diagnostic tools and reduced the complexity of ASD classification.

Relevance: Helped in understanding feature importance and dimensionality reduction for improved model performance.

[5] "Application of Ensemble Learning Methods for Autism Prediction" – (Kourou et al., IEEE 2018)

Summary: The study focused on ensemble methods such as Random Forest and Boosting techniques to improve classification accuracy in medical datasets.

Relevance: Directly supports the use of adaboost and Gradient Boosting classifiers in this project.

[6] "Deep Learning Approaches for Autism Detection" – (Raj & Masood, Procedia Computer Science 2020)

Summary: This paper explored deep learning models for ASD detection using behavioral and image-based datasets, showing improved accuracy over traditional methods.

Relevance: Highlights the potential of advanced AI techniques for future enhancements of ASD detection systems.

[7] "Predictive Modeling for ASD Screening Using AQ-10 Dataset" – (Thabtah, 2019)

Summary: This research specifically focused on the AQ-10 dataset and evaluated multiple classifiers to predict ASD with high precision.

Relevance: Directly aligns with this project's dataset and validates the effectiveness of AQ-10 for screening purposes.

[8] "Web-Based Autism Screening Tools Using Machine Learning" – (Sarker et al., IEEE Access 2021)

Summary: This study developed an online platform integrating machine learning models for ASD screening, improving accessibility and usability.

Relevance: Supports the implementation of a Flask-based web application for real-time ASD prediction.

[9] "Classification of Autism Spectrum Disorder Using Supervised Learning Techniques" – (Elbattah et al., 2020)

Summary: This research compared multiple supervised learning models including KNN, Decision Tree, and Random Forest on ASD datasets to evaluate classification performance.

Relevance: Helps justify model selection and comparison in ASD screening projects.

[10] "Autism Detection Using Machine Learning Techniques Based on Behavioral Data" – (Gharaibeh et al., 2021)

Summary: The study focused on behavioral datasets and applied classification algorithms to identify ASD traits effectively. It highlighted the importance of questionnaire-based screening.

Relevance: Strongly supports the use of AQ-10 questionnaire data in my project.

[11] "An Evaluation of Machine Learning Approaches for Early Diagnosis of Autism Spectrum Disorder" – (Rasul et al., 2023)

Summary: This study evaluated multiple machine learning models including SVM, Logistic Regression, ANN, and clustering techniques for ASD detection. It showed that several models achieved very high accuracy (up to ~100%) on child datasets.

Relevance: Demonstrates the effectiveness of combining multiple classifiers and supports model comparison in ASD prediction systems.

[12] "Leveraging Machine Learning for Early Autism Detection via INDT-ASD Indian Database" – (Shrivastava et al., 2024)

Summary: This research used various ML algorithms on an Indian ASD dataset and reduced the number of screening questions while maintaining high accuracy. SVM performed best among classifiers.

Relevance: Strongly supports questionnaire-based systems like AQ-10 and shows how fewer questions can still provide accurate predictions.

[13] "Enhancing ASD Detection Accuracy Using NLP and Deep Learning" – (Rubio-Martín et al., 2024)

Summary: The study applied NLP and deep learning models (BERT, LSTM, etc.) on social media text data to detect ASD traits, achieving around 88% accuracy.

Relevance: Highlights the potential of advanced AI techniques and alternative data sources beyond questionnaires for ASD detection.

[14] "Machine Learning-Based Prediction of Autism Spectrum Disorder with Explainable AI" – (Temiz et al., 2025)

Summary: This paper introduced explainable AI methods to identify important biological and behavioral features influencing ASD prediction, improving transparency in ML models.

Relevance: Useful for adding interpretability to ASD prediction systems, making results more understandable for users and clinicians.

[15] "Early Detection of Autism Spectrum Disorder Through Automated Machine Learning" – (Ehsan et al., 2025)

Summary: The study used AutoML techniques to automatically select and optimize machine learning pipelines for ASD detection, simplifying model development while maintaining strong performance.

Relevance: Supports automation in model building and can help improve our project by reducing manual tuning effort.

3. TECHNIQUES FOR AUTISM SPECTRUM

1. Supervised Machine Learning (Core Technique)

This is the main approach used with questionnaire data (AQ-10, AQ-50, etc.), where answers are converted into numerical values (0/1).

Logistic Regression – Widely used for binary classification (ASD vs Non-ASD).

Support Vector Machine – Performs well with small structured datasets.

Decision Tree – Easy to interpret results.

Random Forest – Improves prediction accuracy using multiple trees.

2. Ensemble Learning (Accuracy Improvement)

Used to boost performance of questionnaire-based models: AdaBoost, Gradient Boosting

3. Feature Selection Techniques

Used to identify the most important questions from the questionnaire:

Correlation-based selection, Chi-square test, Information gain

4. Data Preprocessing Techniques

Essential before applying ML models, Encoding answers (Yes/No → 1/0),

Handling missing values, Normalization using Feature Scaling

5. Classification Techniques

Core task in questionnaire-based ASD systems.

Binary Classification – Output is ASD or Non-ASD

6. Web-Based Implementation

Used to deploy questionnaire systems.

Flask – For building UI and backend

Form-based input for AQ-10

Questionnaire-based ASD detection systems primarily use supervised machine learning techniques such as Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest to classify individuals based on their responses to screening tools like AQ-10. Ensemble methods like AdaBoost and Gradient Boosting are applied to improve prediction accuracy. Feature selection techniques help identify the most relevant questions, while data preprocessing ensures clean and structured input data. The task is treated as a binary classification problem, and model performance is evaluated using metrics like accuracy and precision. Finally, web frameworks like Flask are used to deploy the system for real-time user interaction.

4. DATASET USED FOR AUTISM SPECTRUM

The dataset used in this Autism detection project is primarily the AQ-10 (Autism Spectrum Quotient-10) dataset, which is a widely accepted screening tool for identifying traits associated with autism spectrum disorder. This dataset consists of 10 carefully designed behavioral questions that evaluate social interaction, communication skills, attention switching, imagination, and attention to detail. Each question is answered in a binary format such as "Yes/No" or "Agree/Disagree," and responses are converted into numerical values (0 or 1) for machine learning processing. The dataset may also include additional attributes like age, gender, ethnicity, and family history to

improve prediction accuracy. It is commonly sourced from repositories such as the UCI Machine Learning Repository and is widely used due to its simplicity, reliability, and effectiveness in early-stage screening. In this project, user responses collected through a questionnaire interface are mapped to the AQ-10 features and passed into a trained machine learning model, which then predicts whether the individual is likely to show ASD traits. This dataset is especially suitable for web-based applications because it enables quick, low-cost, and accessible screening without requiring complex clinical procedures.

5.ADVANTAGES

- Simple and easy to use
- Quick screening process (takes less time)
- User-friendly for non-medical users
- Based on validated psychological research
- High reliability for early-stage screening
- Binary data (0/1) makes it easy for machine learning models
- Suitable for web and mobile applications
- Enables real-time prediction systems
- Cost-effective compared to clinical diagnosis
- Helps in early detection and intervention

6. FUTURE RESEARCH DIRECTION

- **Integration of multimodal data:** Combine questionnaire data with speech, facial expressions, eye-tracking, and behavioral videos to improve prediction accuracy.
- **Use of deep learning models:** Explore advanced models like CNNs, RNNs, and transformers for better pattern recognition compared to traditional ML methods.
- **Development of mobile and wearable-based screening tools:** Create real-time ASD screening apps using smartphones and wearable sensors for continuous monitoring.
- **Personalized prediction systems:** Build models that adapt to individual differences such as age, gender, and cultural background.
- **Expansion of datasets:** Use larger and more diverse datasets beyond AQ-10 to improve generalization and reduce bias.
- **Explainable AI (XAI):** Develop models that provide clear explanations for predictions, helping doctors and users understand results.
- **Early childhood detection:** Focus on toddlers and infants for earlier identification using specialized datasets.
- **Cloud-based healthcare integration:** Connect ASD prediction systems with telemedicine platforms for remote diagnosis and consultation.
- **Hybrid models:** Combine questionnaire-based screening with clinical and genetic data for more accurate diagnosis support.
- **Improved user interfaces:** Design more interactive and accessible UI/UX for better user engagement and accurate responses.
- **Cross-cultural validation:** Adapt and validate AQ-10 questions for different languages and populations worldwide.
- **Real-time decision support systems:** Assist clinicians with instant predictions during patient evaluation.

7.CONCLUSION

The project “Autism Spectrum Disorder Detection using Machine Learning” aims to develop an automated system for early screening of ASD using advanced data-driven techniques. The system effectively analyzes questionnaire-based inputs, such as the AQ-10 dataset, and predicts the likelihood of Autism Spectrum Disorder using machine learning models like AdaBoost and Gradient Boosting Classifiers. Through systematic implementation and evaluation, it has been demonstrated that the machine learning-based approach provides an efficient, accurate, and scalable solution for ASD screening. The system significantly reduces the time and effort required in traditional diagnostic procedures while maintaining reliable prediction performance. It serves as a supportive tool for early identification, enabling timely intervention and improved outcomes for individuals. The achieved accuracy, along with model evaluation metrics, confirms the effectiveness of ensemble learning techniques in identifying behavioral patterns associated with ASD. The results validate the robustness of the system in handling diverse input responses and demonstrate its potential to assist healthcare professionals as well as individuals seeking preliminary screening.

REFERENCES

- [1] D. Wall, J. Kosmicki, T. Deluca, E. Harstad, and V. Fusaro, "Use of Machine Learning to Shorten Observation-Based Screening and Diagnosis of Autism," *Translational Psychiatry*, vol. 2, no. 4, 2012.
- [2] M. Kosmicki, V. Sochat, D. Wall, and J. Kosmicki, "Searching for a Minimal Set of Behaviors for Autism Detection Through Feature Selection-Based Machine Learning," *Scientific Reports*, vol. 5, 2015.
- [3] F. Thabtah, "Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment," *Proceedings of the 1st International Conference on Medical and Health Informatics*, 2017.
- [4] F. Thabtah and D. Peebles, "A New Machine Learning Model Based on Induction of Rules for Autism Detection," *Health Informatics Journal*, vol. 26, no. 1, pp. 264–286, 2020.
- [5] H. Duda, D. Ma, N. Haber, and D. Wall, "Use of Machine Learning for Behavioral Distinction of Autism and ADHD," *Translational Psychiatry*, vol. 6, 2016.
- [6] U. R. Acharya et al., "Application of Machine Learning Techniques for Autism Spectrum Disorder Detection," *Healthcare Analytics Journal*, 2019.
- [7] S. Raj and S. Masood, "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques," *Procedia Computer Science*, vol. 167, pp. 994–1004, 2020.
- [8] F. Thabtah, "AQ-10 Dataset for Autism Spectrum Disorder Screening," *Kaggle Dataset Repository*, 2019.
- [9] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, 2009.
- [10] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [11] J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [12] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [13] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*, 5th ed., 2013.
- [14] World Health Organization, "Autism Spectrum Disorders," 2023.
- [15] S. L. Ozonoff et al., "Recurrence Risk for Autism Spectrum Disorders: A Baby Siblings Research Co
- [16] M. S. Farooq, R. Tehseen, M. Sabir, and Z. Atal, "Detection of autism spectrum disorder (ASD) in children and adults using machine learning," *Scientific Reports*, vol. 13, 2023.
- [17] R. K. Reghunathan, P. N. P. Venkidusamy, R. G. Kurup, B. George, and N. Thomas, "Machine learning-based classification of autism spectrum disorder across age groups," *Engineering Proceedings*, vol. 62, no. 1, 2024.
- [18] E. Laçin and A. Özkaya, "Thematic mapping of autism spectrum disorder research using machine learning and LDA: Trends, patterns, and future directions," *Humanities and Social Sciences Communications*, 2026.
- [19] S. Srividhya and D. Vidhya, "Detecting autism spectrum disorder using machine learning techniques," *International Journal of Engineering Research & Technology (IJERT)*, vol. 12, no. 3, 2023.
- [20] S. S. Rajagopalan, Y. Zhang, A. Yahia, and K. Tammimies, "Machine learning prediction of autism spectrum disorder from minimal medical and background information," *JAMA Network Open*, 2024.
- [21] A. Wawer, I. Chojnicka, J. Sarzyńska-Wawer, and M. Krawczyk, "A cross-dataset study on automatic detection of autism spectrum disorder from text data," *Acta Psychiatrica Scandinavica*, vol. 151, no. 3, pp. 259–269, 2025.
- [22] S. Irram and M. Suaib, "An early detection of autism spectrum disorder using machine learning," *Computers in Biology and Medicine*, 2025.
- [23] K. Sravani and K. Pothanaicker, "Comprehensive autism spectrum disorder analysis: ML and DL models in multimodal datasets," *Ingénierie des Systèmes d'Information*, vol. 29, no. 6, pp. 2467–2494, 2024.
- [24] "An exploration of machine learning approaches for early autism spectrum disorder detection," *Health Sciences Journal*, 2024.
- [25] M. Temiz, B. Bakir-Gungor, N. S. Ersoz, and M. Yousef, "Machine learning-based prediction of autism spectrum disorder and discovery of related metagenomic biomarkers with explainable AI," *Applied Sciences*, vol. 15, no. 16, 2025.
- [26] M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, "Detecting autism spectrum disorder using machine learning," *arXiv preprint arXiv:2009.14499*, 2020.
- [27] S. Rubio-Martín et al., "Enhancing ASD detection accuracy using machine learning and deep learning with NLP," *arXiv preprint arXiv:2403.03581*, 2024.
- [28] T. Eslami et al., "ASD-DiagNet: A hybrid learning approach using fMRI data," *arXiv preprint arXiv:1904.07577*, 2019.
- [29] V. Ramesh and R. Assaf, "Detecting autism spectrum disorders using speech transcripts and machine learning," *arXiv preprint arXiv:2110.03281*, 2021.