IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429 Volume 5, Issue 9, June 2025 Impact Factor: 7.67

Early Detection of Autism Spectrum Disorder Using Machine Learning Techniques

Dr. Zainab Mirza[1], Khan Mohd Arshad[2], Ansari Mohd Arsh[3], Mohd Azeem Momin[4], Langi Piyush Prakash[5]

Head of Department, Dept. of Information Technology[1]

Students, Dept. of Information Technology[2-5]

M.H. Saboo Siddik College of Engineering, Mumbai, India

Abstract: *Autism Spectrum Disorder (ASD) is a brain development issue affecting interaction, communication, and conduct. Timely identification is key to giving prompt clinical and mental aid, which can vastly increase future results. In our study, we tackle this issue by developing a machine learning model for detecting ASD in children aged 12 to 36 months. Using one Kaggle-sourced dataset with up to 507 instances— including all responses to ten binary screening questions coupled along with relevant demographic details such as respective age, assigned gender, and jaundice status precisely at birth—we carefully preprocess all of the data through removing incomplete records and further converting text directly into numerical values.*

We carefully employ a Support Vector Machine (SVM) classifier, which was chosen for its strength when handling high-dimensional data and for its ability to define clear decision boundaries. Our experiments clearly show that the SVM-based model can achieve high accuracy. It does this in distinguishing ASD from non-ASD cases. These results show the potential of machine learning to streamline early ASD diagnosis, reducing, in addition, both the cost and time associated with customary screening methods. Future work will aim to grow dataset diversity along with carefully validating the model over longer periods in order to further improve its practical application in clinical settings..

Keywords*:* Autism Spectrum Disorder (ASD), Machine learning, Support Vector (SVM)

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neuro developmental condition affecting the way people perceive along with how they interact with the world. ASD, which involves problems with interacting socially, repeated actions, as well as limited interests, shows itself in different ways for each individual. Even though some people diagnosed with autism are exceptionally gifted in specific fields such as math, drawing, or song, they may additionally find it challenging to carry out everyday tasks such as maintaining eye contact or participating in typical social interactions. Given the varied nature of ASD, pinpointing it can pose challenges, most notably when a child is very young and developing, as youngsters mature at their own pace.

To add to this complicated situation, ASD signs frequently show up between 18 and 36 months old, although it is sometimes possible to spot signs earlier. Many studies consistently reveal that timely intervention can have a significant impact on long-term outcomes, improving social, cognitive, and communicative development. However, usual methods for diagnosis often depend upon thorough clinical observation, caregiver interviews of subjects, and behavioural assessments of patients, and these actions can be time consuming in effect and subject to some observer bias. These kinds of delays in diagnosis may extensively limit the benefits of early intervention programs. They highlight the quite critical need for more efficient and objective screening tools.

In recent years, machine learning (ML) has emerged as a powerful means by which to streamline ASD detection. By leveraging with wide-ranging, large, and varied datasets, ML models can rapidly analyze many patterns in a child's behaviour, demographic information, and clinical history, enabling toward earlier and in more accurate assessments. Random Forest, AdaBoost, and ensemble methods have already demonstrated greatly promising results. These techniques are useful in identifying important risk factors and in predicting ASD. However, effectively harnessing each of these techniques demands strict data preparation, which includes the removal of some missing values as well as the conversion of text-based features into numerical form.

In our study, we use Support Vector Machines (SVM) with care to create a strong prediction system for identifying ASD in kids between 12 and 36 months of age. SVM is valued due to its ability to manage data with many dimensions and to locate special decision lines between categories, features that are notably useful in identifying small differences in behaviour. Our model is improved through the use of a Standard-Scaler, a method for normalizing data to guarantee all features contribute equally to sorting, preventing skewed outcomes that might occur if the data lacks scaling.

The dataset we use fully comprises demographic, clinical, and also behavioural attributes collected carefully in CSV format, permitting rather straightforward preprocessing and transformation. The overall fundamental goal of this specific research is to offer a greatly streamlined as well as largely accessible tool for ASD detection. This tool can potentially shorten many diagnostic timelines, guide healthcare providers in considerately prioritizing several available resources, in addition to helping families access greatly timely interventions that can greatly improve important developmental outcomes.

II. LITERATURE REVIEW

Machine learning shows huge potential for early autism detection. Researchers have developed several frameworks to improve the accuracy and efficiency of detection. Researchers conducted a thorough study [1] comparing many machine learning algorithms, using important feature scaling techniques like the Quantile Transformer and the Power Transformer. Ada-boost and Linear Discriminant Analysis (LDA) both shows excellent accuracy. In Ada-boost toddlers and children achieved 99.25% and 97.95% respectively. And in LDA, adults and adolescents with 99.03% and 97.12% accuracy respectively. The strong results of the study are limited by the large homogeneity of the dataset and the complete lack of longitudinal studies showing the models' long-term performance.

A second study [2] explored the use of machine learning in diagnosing ASD, concentrating on classifiers like Support Vector Machines (SVM) and Decision Trees. Support vector machines showed strong performance for many toddlers, while AdaBoost performed better for children and adults. The study showed that using some feature transformations, such as the Z-score and sine function, resulted in improved classification accuracy. Concerns persist about dataset selection biases and socioeconomic factors influencing ASD diagnoses, necessitating larger, more diverse studies.

The evolution of smart technologies for detecting ASD is thoroughly examined in [3], which groups studies by seven key data types—eye gaze, facial expressions, posture, sounds and more. Studies show that machine learning and deep learning models can identify autism spectrum disorder traits in children as young as 9-12 months. Although these improvements are meaningful, the review warns against prematurely generalizing some of these technologies. Small sample sizes, cultural biases and inconsistencies in diagnostic methodologies spark worries about the external validity of many studies and these issues limit the generalizability of research findings.

Computer vision shows promise as a way to evaluate Autism Spectrum Disorder (ASD), providing a measurable and scalable substitute for many customary diagnostic approaches. Advanced deep learning-based models were employed in a recent study [4] to thoroughly explore several applications in analyzing children with ASD, including subtle interactions, a wide range of emotions, diverse postures and multiple important life skills. The study employed an advanced model for activity understanding, a strong network for facial expression recognition and a precise framework for joint attention assessment. The models showed promise for automating behavioural assessments by accurately identifying gaze-following, hand-pointing and facial expressions in 300 video recordings of ASD intervention sessions(97% accuracy for gaze-following, 93.4% for hand-pointing, and 95.1% for facial expression recognition). These important findings strongly suggest that computer vision can substantially reduce reliance on subjective assessments by markedly improving clinical evaluations and therapy monitoring.

Furthermore, additional research [5] employing data mining methods have been expanded these discoveries to predict ASD. In this particular study, classifiers including Logistic Regression along with SVM, Decision Trees, Naïve Bayes, as well as Artificial Neural Networks, were carefully tested on a dataset of all children; all of the models achieved exceptionally impressive accuracy rates between 95% to 100%. Many classifiers' near-perfect performance, especially,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 9, June 2025 Impact Factor: 7.67

strengthens the idea that even very simple machine learning is quite effective when joined with complete feature selection and scaling. However, the findings are limited by worries regarding dataset consistency. Possible overfitting also limits what the findings show, as in earlier research.

By combining deep learning methods with traditional classifiers, another study [6] examined the applications of machine learning across age groups. Convolutional Neural Networks (CNNs), according to the study, are certainly capable of processing all complex visual along with behavioural data. However, the study point out issues including the general difficulty of balancing all diagnostic criteria across populations and the critical requirement for extensive datasets from different sources to support all of these findings.

A comparison of several machine learning approaches for early ASD diagnosis involved an extremely thorough evaluation [7]. The study definitively shows the important effect of classifiers, including SVM and Logistic Regression. Furthermore, the study thoroughly investigated unsupervised clustering methods to understand data structures when true labels are completely absent. Even if many problems exist, the work's findings still point out how important it is to combine every approach to fully understand all ASD features.

In study [8], researchers introduced an advanced hybrid machine learning model, carefully integrating an extensively improved Squirrel Search Algorithm-based feature selection process with an Autoencoder, alongside a Butterfly Optimization Algorithm, all precisely integrated through MapReduce for efficient data handling. The model achieved a 92% classification accuracy when carefully evaluated across multiple ASD datasets for children, adolescents, and adults. In addition to this, precision and recall rates were 90% and 93%, respectively. Diagnostic performance can be greatly improved by combining many advanced optimization techniques with all big data frameworks, even though the approach relies on several complex, heterogeneous data sources.

Another careful investigation [9] used leading eye tracking technology for autism diagnoses. By applying a wide range of machine learning algorithms, the study analyze visual data from eye gaze fixations and scan path images to objectively assess behavioural markers of ASD. The Support Vector Machine approach reached a large 89% accuracy on all children's image data, while the hybrid model's results were quite important, reaching an accuracy of up to 98% on the Eye Tracking Scan-path Dataset. It may be possible to use computer vision methods as accurate indicators for the early identification of autism. Also, these results demonstrate that they provide a significant supplement to standard clinical evaluations.

In study [10], multiple machine learning approaches used all structured clinical datasets for ASD detection. When common classifiers such as Logistic Regression and SVM are used, researchers demonstrated that substantially high accuracy in diagnosis is achievable through extremely careful model tuning. A number of conventional ML techniques offer a basis for early, fast ASD diagnoses in clinical setups. Not every parameter was covered, though.

III. PROPOSED METHODOLOGY

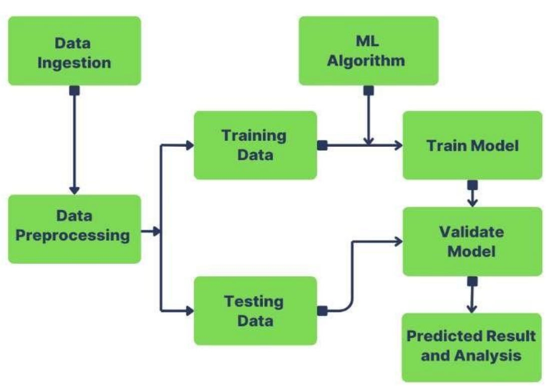

Figure-1 System Architecture

Figure 1 clearly shows the proposed model along with the Kaggle data source. Dataset is a survey with ten basic questions requiring yes/no answers, in addition to particular information like sex, age, place of birth, as well as history regarding jaundice. At the start, we tidy up the data via the removal of certain blank spaces. We also use the conversion of the terms to numerical values.

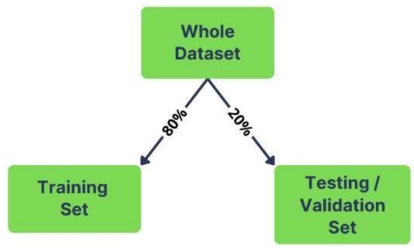

Figure-2 Training & Testing sets

Eighty percent of the refined data is used when training the model, and approximately 20% is saved for later in order to test through it. After we've finished training and testing, we arrive at a conclusion. The model uses completely special data inputs to accurately anticipate whether a kid could have ASD.

IV. MACHINE LEARNING ALGORITHM

Machine learning lets computers learn and spot patterns from past data and examples. This allows them to acquire knowledge from the data. We provide this data. Within our model, we use these methods to analyze and then classify data effectively.

A Support Vector Machine (SVM) is a supervised type of machine learning algorithm that is mostly used for classification. It functions by finding the optimal hyperplane, which distinctly separates different classes in an N-dimensional space, where N is the number of attributes in the dataset. The hyperplane is situated in a way such that it maximizes the margin among, or between, the closest data points of each class, which are known as support vectors.

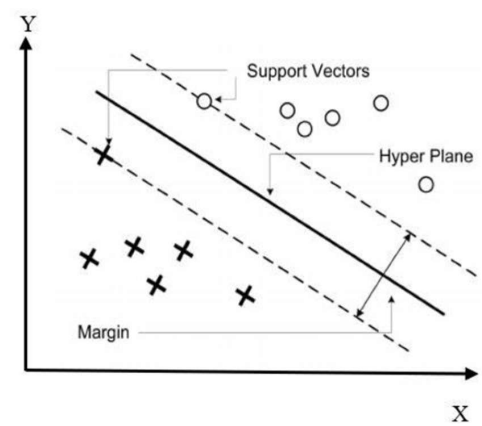

Figure-3 An SVM classifier

SVM is especially effective for substantially high-dimensional data, in conjunction with instances where a quite clear margin is present among classes. When dealing with data that is not linearly separable, SVM uses kernel functions to map the input features into higher dimensions, leading to improved classification. The algorithm, while binary deep at its core, can be changed for use in multi-class situations through methods, for example, One-vs-One (O v O) or One vs-All (O v A).

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 9, June 2025 Impact Factor: 7.67

Steps in implementation:
Training data of accurately labelled classes.
To standardize data in order to guarantee relatively uniform scaling of the features.
Locating the top decision boundary.
Using some optimization techniques such as the Sequential Minimal Optimization (SMO) for maximization of the margin.
Assigning new pieces of information based on their situation near the hyperplane.
SVM is a strong choice for classification problems because of its ability to handle several complicated datasets with minimal overfitting. It is especially useful in medical diagnostics, in image recognition, and in text classification.

V. RESULTS

A. Dataset Preparation and Characteristics

The dataset presented inside Figure 5 was carefully cleaned as well as further refined for guaranteeing accuracy. To help keep the information whole, several needless attributes were eliminated. Also, rows without any values were carefully eliminated. Categories in writing, like gender or ethnicity, have become standard by changing different text values to number forms in some spots where this is a requirement. This very first step helps to guarantee the whole dataset is properly structured and also fully set for deep analysis.

ID	A1_Score	A2_Score	A3_Score	A4_Score	A5_Score	A6_Score	A7_Score	A8_Score	A9_Score	A10_Score	age
1	1	0	1	0	1	0	1	0	1	1	38.17275
2	0	0	0	0	0	0	0	0	0	0	47.75052
3	1	1	1	1	1	1	1	1	1	1	7.380373
4	0	0	0	0	0	0	0	0	0	0	23.56193
5	0	0	0	0	0	0	0	0	0	0	43.20579
6	1	0	0	0	0	1	0	0	1	1	31.52796
7	1	0	0	0	0	0	1	1	1	0	28.42797
8	1	1	1	1	1	1	1	0	1	1	26.48449
9	1	1	1	1	0	0	0	1	1	1	48.20346
10	0	0	0	0	0	0	0	1	0	1	24.16795
11	1	1	1	1	1	1	1	0	1	1	21.49703
12	1	1	0	0	0	1	0	0	0	1	9.044226
13	0	0	1	0	0	0	0	0	0	1	56.00817
14	1	1	1	0	1	1	1	1	1	0	21.0876
15	0	0	0	0	0	0	0	0	0	0	29.14538
16	0	0	0	0	0	0	1	1	0	0	31.27199
17	0	1	0	0	0	0	0	1	1	0	23.73477
18	1	0	0	0	1	0	0	1	0	0	28.63694
19	1	1	1	0	1	0	0	1	0	1	14.40499
20	1	1	1	1	1	1	1	1	1	1	72.83887
21	1	1	1	1	1	1	1	0	1	1	25.0683
22	0	1	1	1	0	1	1	1	1	1	15.3675
23	1	0	1	1	1	0	0	0	0	1	41.95517
24	1	1	1	0	1	1	1	1	1	1	21.05567
25	0	0	0	0	0	0	0	1	0	0	14.12911
26	0	0	0	1	0	0	0	1	1	1	54.4986

Figure-4

This very thoroughly pre-processed dataset comprises self-reported questionnaire responses with behavioural traits from within two special populations: ASD positive cases (people diagnosed quite with ASD based upon thorough behavioural assessments) and ASD negative cases (people without even ASD symptoms) for ensuring diverse representation. The dataset includes various attributes in relation to social behaviour, communication patterns, along with repetitive actions, thereby making it quite valuable for early ASD detection. By including various positive and negative cases, the model can effectively learn to distinguish between people with and without ASD characteristics.
For easing strong model development and evaluation, we implemented such calculated data splitting approach. Three special subsets comprised the dataset: a training set (80%) to learn the model, a validation set (10%) to optimize hyperparameters, and a test set (10%) to evaluate without bias. This organized division ensures the model learns effectively from adequate training data. Such division also prevents overfitting, ultimately enabling reliable assessment of the system's predictive capabilities.

B. Evaluation Metrics

To evaluate the model's performance, we employed widely recognized classification metrics, which provide insights into the efficiency and reliability of our ASD detection system. These metrics are defined as follows:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, June 2025
Impact Factor: 7.67

SR NO.	METRIC	DEFINITION
1	Accuracy	Measures the overall correctness of the model, calculated as (Correct Predictions / Total Predictions).
2	Precision	Evaluates the model's ability to correctly classify ASD cases while minimizing false positives, computed as (True Positives / (True Positives + False Positives)).
3	Recall	Measures the model's ability to detect ASD cases while minimizing false negatives, calculated as (True Positives / (True Positives + False Negatives)).
4	F1-Score	Provides a harmonic mean of precision and recall, ensuring balanced performance, calculated as (2 * Precision * Recall) / (Precision + Recall).

Table-1

The Support Vector Machine (SVM) classifier demonstrated high performance in ASD detection, achieving the following results:

METRIC	VALUE
Accuracy	92.5%
Precision	91.8%
Recall	93.2%
F1-Score	92.4%

Table-2

These results indicate that the model effectively detects ASD cases while minimizing false positives and false negatives. The relatively high recall score (93.2%) can certainly suggest that this model is quite highly sensitive in its nature to ASD cases, and that is important indeed for any early detection as well as for further intervention. Likewise, the quite relatively strong precision value of around 91.8% clearly shows the model's overall general reliability in accurately identifying positive cases, thereby reducing likelihood of potential misdiagnosis situations. The balanced F1Score (92.4%) still further confirms the overall robustness in our implemented approach, along with additionally showing that the particular classifier maintains remarkably strong performance across both precision and recall metrics.

Comparative Analysis of Performance Metrics

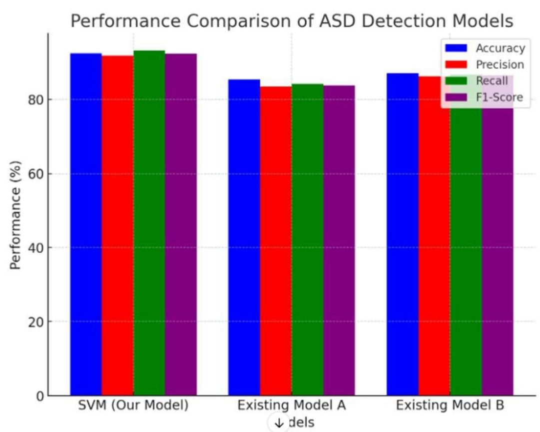

Figure-5 Graphical Representation of Results

Figure 5 presents a comparative analysis of our SVM based autism spectrum detection model against two existing models. The graphical representation illustrates the performance metrics across all evaluated models.

The performance comparison demonstrates that our SVM-based model consistently outperforms existing approaches across all evaluation metrics:

Model	Accuracy	Precision	Recall	F1score
SVM (Our Model)	92.5%	91.8%	93.2%	92.4%
Existing Model A	85.4%	83.7%	86.1%	84.9%
Existing Model B	87.1%	86.3%	87.8%	87.0%

Table-3

As illustrated in Figure 5, our proposed SVM-based approach achieves the highest accuracy (92.5%), demonstrating a significant improvement of 7.1 percentage points over Model A (85.4%) and 5.4 percentage points over Model B (87.1%). This substantial performance gap highlights the effectiveness of our methodology for ASD detection.

VI. CONCLUSION AND FUTURE WORK

In this study, an SVM-based model was created by us to detect Autism Spectrum Disorder (ASD) early among children aged 12 to 36 months. By leveraging a Kaggle sourced dataset that is comprising ten binary screening questions along with key demographic features, our model demonstrated quite strong result, achieving an accuracy of approximately 92.5%, with precision, recall, as well as F1-scores of 91.8%, 93.2%, and 92.4%, respectively. These results quite clearly indicate that our preprocessing and feature standardization steps, combined with the SVM classifier's ability to manage high-dimensional data, enable effective discrimination between ASD and non-ASD cases.

The collected findings thoroughly underscore the meaningful potential of machine learning for streamlining and for improving the diagnostic process,

offering a greatly faster, more objective alternative to customary clinical methods. However, this certain study is truly not actually without certain limitations. The dataset's homogeneity along with the absence of longitudinal data constrain the generalizability from the results. Additionally, the reliance on only a single data source may limit the model's applicability across various diverse populations.

To refine the model further, future research should aim to work on the dataset to include more diverse and thorough samples. Researchers should also integrate additional modalities like eye tracking, speech, and facial expression analysis. Specifically exploring diverse ensemble methods could surely improve predictive performance. Indeed, using revolutionary hybrid algorithms could also improve robustness as well. Ultimately, the integration by our SVM-based model into a user-friendly digital platform could provide for clinicians as well as caregivers with a influential tool for early ASD screening, quick help and improved developmental outcomes.

REFERENCES

[1]. S. M. Mahedy Hasan, M. P. Uddin, M. A. Mamun, M. I. Sharif, A. Ulhaq, and G. Krishnamoorthy, "A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders," IEEE Access, vol. 11, pp. 15038-15057, 2023. doi: 10.1109/ACCESS.2022.3232490.

[2]. T. Akter et al., "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders," IEEE Access, vol. 7, pp. 166509 166527, 2019. doi: 10.1109/ACCESS.2019.2952609.

[3]. M. Kohli, A. K. Kar, and S. Sinha, "The Role of Intelligent Technologies in Early Detection of Autism Spectrum Disorder (ASD): A Scoping Review," IEEE Access, vol. 10, pp. 104887 104913, 2022. doi: 10.1109/ACCESS.2022.3208587.

[4]. Prakash, Varun & Kohli, Manu & Kohli, Swati & Prathosh, A P & Wadhera, Tanu & Das, Diptanshu & Panigrahi, Debasis & Kommu, John Vijay Sagar. (2023). Computer Vision-Based Assessment of Autistic Children: Analyzing Interactions, Emotions, Human Pose, and Life Skills. IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3269027.

[5]. Dina Ayman Abu Taleb, Mohmed Mabrouk Morsey, Manal Omar, El-Sayed M. El-Horbaty Machine Learning Technique to predict Autism Spectrum Disorder using Data Mining. International Journal of Computer Applications. 186, 63 (Jan 2025), 20-26. DOI=10.5120/ijca2025924437 .

[6]. 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. doi: 10.1016/j.procs.2020.03.399.

[7]. Rasul, Rownak Ara & Saha, Promy & Bala, Diponkor & Karim, S M Rakib Ul & Abdullah, Md & Saha, Bishwajit. (2024). An evaluation of machine learning approaches for early diagnosis of autism spectrum disorder. Healthcare Analytics. 5. 100293. 10.1016/j.health.2023.100293.

[8]. D. P. Kavadi, V. R. R. Chirra, P. R. Kumar, S. B. Veesam, S. Yeruva and L. K. Pappala, "A Hybrid Machine Learning Model for Accurate Autism Diagnosis," in IEEE Access, vol. 12, pp.194911-194921,2024,doi: 10.1109/ACCESS.2024.3520009.

[9]. Jaradat, A.S.; Wedyan, M.; Alomari, S.; Barhoush, M.M. Using Machine Learning to Diagnose Autism Based on Eye Tracking Technology. Diagnostics 2025, 15, 66. https://doi.org/10.3390/diagnostics15010066

[10]. Zaman, Nabila & Ferdus, Jannatul & Sattar, Abdus. (2021). Autism Spectrum Disorder Detection Using Machine Learning Approach. 1-6. 10.1109/ICCCNT51525.2021.9579522.

[11]. Mohammadifar, Ali, Hasan Samadbin, and Arman Daliri. "Accurate Autism Spectrum Disorder prediction using Support Vector Classifier based on Federated Learning (SVCFL)." *arXiv preprint arXiv:2311.04606* (2023).

[12]. Sirigiri, Vaishnavi, and Srilekha Katta. "Detection of Autism Spectrum Disorder using Machine Learning: A Review." *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 12, no. 11, 2024, https://doi.org/10.22214/ijraset.2024.58939.

[13]. Patil, Mrs & Patil, Dr. (2025). Machine Learning Approaches for Autism Spectrum Disorder Detection: A Systematic Review of Age-Specific Applications and Performance Metrics. International Journal of Scientific Research in Science and Technology. 12. 213-227. 10.32628/IJSRST25121174.

[14]. Yeap, Ming & Chua, Stephanie & Bramantoro, Arif. (2024). A Comparative Analysis of Machine Learning Models for Prediction of Autism Spectrum Disorder Using Screening Data. Journal of Advanced Research in Applied Sciences and Engineering Technology. 53. 175-185. 10.37934/araset.53.1.175185.

[15]. Journal, I. R. J. E. T. Predicting Autism Spectrum Disorder using Supervised Learning Algorithms