Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Review On Early Autism Spectrum Disorder Detection in Children Using Machine Learning Techniques

Shweta Ramtekkar[1], Dr. Gurudev Sawarkar[2], Prof. Harshali Ragite[3]

[1]Research Scholar, Department of Computer Science & Engineering, Wainganga College of Engineering and Management, Nagpur, India
[2,3]Assistant Professor, Department of Computer Science & Engineering, Wainganga College of Engineering and Management, Nagpur, India

Abstract

Autism Spectrum Disorder (ASD), a neurodevelopmental disorder that manifests through a variety of symptoms, is a combination of various challenges that affect an individual socially, communicatively, and behaviorally, among others. The treatment of ASD is based on a hierarchy of needs, from the most urgent to the least, the latter being the least in number due to the efficacy of the early intervention. In the present day, the methods employed for diagnosis depend on the observation of the clinical staff and assessments that have been standardized which are subjective, lengthy, and sometimes unavailable in poor areas. At the same time, clinically, the researchers have turned their attention to machine learning (ML) and artificial intelligence (AI) techniques because they can deal with the analysis of such an intricate mixture of factors like behavior, biology, and development supporting the diagnosis of ASD at an early stage. This paper is an extensive review that collects and analyzes the most recent studies that are based on the use of ML for detecting ASD through the use of behavioral questionnaires, speech signals, eye-tracking data, neuroimaging, and multimodal datasets. Among the different classification algorithms discussed are the Logistic Regression, Random Forest, Support Vector Machines, ensemble learning, and deep learning models. The review also covers issues associated with such as dataset imbalance, limited interpretability, and lack of real-world validation, along with the non-availability of screening tools. Moreover, the role of explainable AI in increasing the acceptance of clinical trust is outlined. The authors of this review showcase the existing trends, constraints, and future directions of research aimed at the creation of trustworthy, interpretable, and accessible ML-based ASD screening systems.

Keywords: Autism Spectrum Disorder, Machine Learning, Early Detection, Explainable Artificial Intelligence, Multimodal Data Analysis etc.

1. Introduction

Autism Spectrum Disorder (ASD) is a condition that the world has acknowledged and studied for long enough to come up with this definition. It is a neurological disorder that can last through the whole life of a person and one of the main things affected by it is the person's communication ability. Moreover, this disorder has been greatly recognized worldwide due to its increasing prevalence over the years; it was, for instance, estimated recently that about 1% of children fall under the category of autism [1]. One of the contributing factors to the mentioned increase is undoubtedly the raising of awareness, widening the criteria for diagnosis, and introducing better reporting systems. Nonetheless, the early diagnosis has been still a major problem especially in the poor and less developed countries.

Early recognition of ASD is indispensable for the development of children who will, otherwise, not be able to benefit from the interventions, which are more or less certainly of short-term nature, in their case, that is to say, they will be able to speak better, to behave more appropriately, and to mingle with peers more easily [2]. Old-fashioned methods of diagnosis, such as the Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R), have been regarded as the best but at the same time they are also the most expensive, the most time-consuming, and requiring very specialized professionals for conducting them [3]. Therefore, it is common that kids go through a long time of waiting for diagnosis and in many cases, unfortunately, their intervention training has already been closed. In addition, the range of ASD characteristics is extensive both in terms of severity and in terms of how they are manifested, and this makes it extremely hard to rely on human assessment alone for early detection.

The swift progress in machine learning (ML) and artificial intelligence (AI) has provided new avenues for tackling these issues. ML methods have the capacity to uncover hidden trends and intricate interactions among vast, multi-dimensional data sets, thus being fit for the analysis of the behavioral and developmental markers associated with ASD [4]. Recently, research has been ongoing into the application of various ML models such as Logistic Regression, Support Vector Machines (SVM), Random Forest, Gradient Boosting, and even deep learning architectures for the detection of ASD through questionnaires on behavior, speech, eye movement, and neuroimaging data [5]–[7].

Despite high classification accuracy being reported in numerous studies, still some constraints remain. A significant share of current research utilizes only one data source, which diminishes the strength and applicability of the respective prediction models [8]. The ASD datasets that are made available for public use have very often been unbalanced, small in number, and not uniformly preprocessed and thus, model performance has been influenced either positively or negatively and made unstable [9]. Furthermore, a lot of winning models are like “black boxes” making it very difficult to see how the predictions are made. This problem of lack of interpretability is a major hindrance to clinical acceptance where transparency and trust are of utmost importance [10].

According to these concerns, recent research started offering model transparency with the help of XAI techniques like SHAP and LIME to spot the main behavioral predictors affecting ASD risk [11]. In addition, the diagnostic proficiency of multimodal learning techniques, which encompass the usage of behavioral, biological, and sensory data, has surpassed that of single-modality models [12]. However, even with these innovations, the market is still devoid of mobile-friendly, accessible ML-based screening tools for parents, teachers, and primary healthcare providers.

This review paper is going through the recent literature on ML-based ASD detection methodically, discussing relied upon techniques, datasets, performance results, and interpretability strategies. Presenting the current trends and research gaps, this piece of work intends to be the beacon directing the coming time's endeavors to develop early ASD screening systems that are reliable, explainable, and scalable for real-world deployment.

2. Problem Identification

• Delayed ASD Diagnosis: Autism Spectrum Disorder is usually diagnosed after the age of four due to the prolonged clinical evaluations, which have a negative impact on the initial intervention [1], [2].
• Subjective Diagnostic Methods: The traditional diagnostic methods greatly depend on the clinicians' observations and parents' reports, which results in a range of variations and biases in the assessment [3].
• Limited Access to Specialists: The lack of trained developmental pediatricians and psychologists in rural and low-resource areas results in the restriction of the timely screening of ASD [4].
• Heterogeneous Symptom Presentation: The symptoms of ASD are very different from one person to another, and therefore using standardized manual methods to identify early signs of ASD becomes a challenge [5].
• Single-Modality Research Focus: A good number of studies using Machine Learning (ML) techniques are based on single types of data like questionnaires or speech signals, and this limits their robustness and applicability [6].
• Dataset Challenges: Publicly available ASD datasets are usually small in size, imbalanced, and have inconsistent preprocessing, which all contribute to the unreliability of models [7].
• Lack of Interpretability: The high-performing models based on ML and deep learning often operate as black boxes, which leads to a decrease in clinicians' trust and adoption [8].
• Limited Practical Deployment: The number of existing ML-based ASD systems integrated into the screening tools that are easily accessible in the real world for parents and teachers is very small [9].

3. Literature Survey

A) Literature Review

Rajagopalan et al. (2024), A very ambitious diagnostic study was made involving 30,660 individuals to assess the machine learning models' efficiency for the prediction of autism spectrum disorder (ASD). During the study, only 28 standard features were used, such as demographic data, developmental milestones, and eating behavior patterns. Several machine learning models were trained and validated externally showing high accuracy, sensitivity, and specificity across different groups. Developmental delays and unusual eating habits were found to be strong markers for ASD. This research points out the scalability and generalizability of ML-driven screening methods and also underlines their readiness for use in population-based programs. The researchers propose the use of such models within primary-care screening routes so as to facilitate the early detection and prompt treatment, especially in large health care settings.

Bahathiq et al. (2022), The last five years have seen the publication of a thorough review of the use of machine learning and deep learning methods for diagnosing autism spectrum disorder (ASD). The various data modalities that the review included were MRI (both structural and functional), behavioral questionnaires, and clinical records. It also mentioned the typical preprocessing pipelines, feature extraction methods, and classification algorithms that were used. The authors pointed out the difficult issues of small and imbalanced datasets, lack of standardized evaluation metrics, and the problem of reproducibility across studies, among others. They also suggested that the discovery of biomarkers, the fusion of multimodal data and the use of explainable AI would be the key issues for enhancing the reliability of diagnosis. This review could be a very useful guide for the researchers who want to create reliable and clinically relevant ASD detection systems.

Al-Adhaileh et al. (2025), suggested a deep learning-based model for Autism Spectrum Disorder (ASD) diagnosis that is based on eye-tracking data obtained from social attention tasks performed by kids. Convolutional Neural Networks were fed with gaze patterns, fixation durations, and visual attention metrics to learn. Their outcomes showed a very promising performance in classification, with significant distinctions between children with ASD and those who are neurotypical, especially regarding less fixation on faces and eyes. The research indicated that eye-tracking is a trustworthy, painless, and non-invasive method for the early diagnosis of ASD. Also, the author debated the practicality of putting inexpensive eye-tracking devices in the screening environment, which would make the method possible for early assessment in the clinic and educational fields.

Ahmed and Jadhav (2023), was researching the combination of eye-tracking signal processing and deep learning methods for detecting atypical visual attention related to ASD. The procedure they followed involved preprocessing techniques like blink removal and gaze calibration, after which classification based on CNN took place using fixation and saccade characteristics. Even though small clinical datasets were used, the study still achieved a good level of accuracy. The authors pointed out issues caused by differences in equipment and the size of the datasets and suggested that collecting data in multiple sessions would enhance stability. The research underscores the need for effective preprocessing and augmentation strategies in eye-tracking-based ASD detection and also gives a boost to the acceptability of vision-based screening tools.

Koehler et al. (2024), The authors investigated classifying ASD with the help of computer vision techniques for naturalistic video conversations. They based their analysis on the factors of social interaction, such as facial expression synchrony, head movement, and gesture patterns that required no verbal communication. The machine learning classifiers trained on these features showed very good performance in separating autistic from non-autistic adults. The authors pointed out the objectivity and scalability of the automated video analysis and also the privacy-conscious data handling. Their results indicate that behavioral markers derived from video recordings may not only serve but also enhance the traditional clinical evaluation process, thus being able to serve as a fast and efficient tool for the large-scale ASD screening in the real world.

D'Couto et al. (2024), The remarkable progress made in multimodal deep learning techniques for detecting autism spectrum disorder (ASD) at an early stage has been reported. The review presented a comparison of the different modalities used in the studies such as behavioral questionnaires, neuroimaging, genetic data, and sensor-based inputs. The findings showed that the use of multimodal fusion gave a significantly higher diagnostic accuracy across the board when compared to single-modality models. Among the avenues of next-generation multimodal systems, the authors analyzed different ways of fusing data such as early, late, and hybrid fusions, as well as pointing out obstacles like data alignment, modality imbalance, and computational complexity. He also pointed out that the use of large, harmonized datasets along with standardized evaluation frameworks is a prerequisite for unlocking the full potential of multimodal deep learning for clinical ASD screening.

Gupta et al. (2025), The remarkable progress made in multimodal deep learning techniques for detecting autism spectrum disorder (ASD) at an early stage has been reported. The review presented a comparison of the different modalities used in the studies such as behavioral questionnaires, neuroimaging, genetic data, and sensor-based inputs. The findings showed that the use of multimodal fusion gave a significantly higher diagnostic accuracy across the board when compared to single-modality models. Among the avenues of next-generation multimodal systems, the authors analyzed different ways of fusing data such as early, late, and hybrid fusions, as well as pointing out obstacles like data alignment, modality imbalance, and computational complexity. He also pointed out that the use of large, harmonized datasets along with standardized evaluation frameworks is a prerequisite for unlocking the full potential of multimodal deep learning for clinical ASD screening.

Agrawal et al. (2025), performed a thorough review of publications available for free to the public that specifically looked at interpretable machine learning approaches for the detection of brain disorders in people with autism. The review investigated the application of SHAP, LIME, and feature importance methods in behavioral, imaging, and multimodal studies. The authors raised issues regarding ethics, the trust of clinicians, and the necessity of explainability for the healthcare industry to accept. Among the main issues brought to light were not enough longitudinal studies, few multimodal XAI frameworks, and no standard methods for assessing explanations. This review asserts that the adoption of explainable AI is indispensable for the integration of ML-based ASD tools into daily clinical practice.

Temiz et al. (2025), The use of explainable machine learning models for autism spectrum disorder (ASD) prediction has been extended by exploring the integration of metagenomic biomarkers with clinical characteristics. The incorporation of microbiome data resulted in a higher predictive performance than using only clinical features. The SHAP-based interpretation pointed out particular microbial taxa closely linked to risk for ASD thereby implying their biological relevance. This study illustrates the capacity of merging biomarker and behavior data for ASD screening that is more thorough and accurate. The writers underscored the necessity of multidisciplinary methods and confirmed the applicability of explainable ML in the areas of biomarker discovery and early detection.

Uddin et al. (2023), Logistic Regression, Random Forest, and Naïve Bayes classifiers were applied on publicly available ASD screening datasets. The work highlighted the strict preprocessing techniques which comprised imputation, normalization, and dealing with class imbalance. The cross-validation results showed competitive classification accuracy across different models, but ensemble methods provided an increased stability. The authors highlighted the need for reproducible ML pipelines and standardized evaluation protocols for accurate ASD prediction. The research gives practical guidance on the use of ML-based ASD screening systems that require structured behavioral data.

Taleb et al. (2025), conducted a comparative analysis of various machine learning classifiers for early diagnosis of Autism Spectrum Disorder (ASD) through a dataset of toddlers aged 12–36 months. After applying feature engineering and PCA-based dimensionality reduction, the researchers assessed Logistic Regression, Support Vector Machine, Decision Tree, and Artificial Neural Network models. Experimental findings revealed high accuracy in classification for all the models, while ANN and SVM showed the best performance. The research pointed out the necessity of indicating sensitivity and specificity along with accuracy for the clinical relevance of the results. Furthermore, the authors illuminated the practical aspects of bringing the trained models into use for online and mobile screening. The study points out the implementation of early ASD screening by means of ML techniques at the same time emphasizing the requirement for external validation on separate datasets.

Raj (2020), performed a very early comparative analysis between classic machine-learning and neural network models for ASD detection through the use of various non-clinical datasets. The research covered the following classifiers: Naïve Bayes, Support Vector Machine, K-Nearest Neighbors, Artificial Neural Networks, and Convolutional Neural Networks. Documentation of comprehensively preprocessing steps, including data normalization and feature extraction, was done to achieve fair model comparison. A study of the trade-offs between performance in terms of accuracy and computational complexity was carried out. The author introduced hybrid ANN–CNN implementations to work with mixed-structure data like tabular questionnaires and image-based inputs. This paper still stands as the primary reference in the field of ASD machine learning research and it also gives a good measure of newer methods in terms of performance through its baselines results.

Selvaraj et al. (2024), The researchers have published a preprint study that compares various machine learning algorithms for predicting Autism Spectrum Disorder (ASD) by using public screening datasets. The area of research was the feature selection techniques like Recursive Feature Elimination and Information Gain, which were applied on the model to get the desired performance and the interpretability of the model. The ensemble classifiers in combination with the optimized feature subsets have shown better accuracy than the single models. The researchers have pointed out reproducibility by giving the references of the datasets, the links of the code, and the detailed experimental conditions. The study points out the significance of feature engineering for ASD detection and it is also a practical guide for those researchers who are going to experiment in this area. The open-access feature of the research has made it more accessible for both academic and practical research purposes.

B) Literature Summary

The most recent studies draw attention to the fact that machine learning and artificial intelligence have changed the ways traditional cybersecurity methods to the point where they now identify attack, threats, and perform monitoring in real-time. Research has shown different kinds of learning models like supervised, unsupervised, and ensemble to be quite powerful in detecting malware, tracking botnets, phishing, and even isolating anomalies amidst the huge network traffic. Nevertheless, there are still disagreements among researchers regarding the non-existence of solutions to such problems as the constant change of the attack patterns, the high number of false positives, and the lack of transparency in the system. Several publications put forth the idea that the time has come for the use of models that are adaptive, can deal with heavy and dynamic data streams, and require little human input. In conclusion, the review of relevant studies supports the view that the combination of modern learning techniques with reliable, scalable, and interpretable infrastructures is compulsory for the production of next-generation cybersecurity products that can cope with both the complexity of the cyber threats and their emergence.

C) Research Gap

The utilization of machine learning and artificial intelligence technologies for detecting and monitoring cyber threats has greatly progressed but there are still some areas that need major research to be filled. The current systems are primarily reliant on static datasets which limits their capacity to cope with the newest and more sophisticated attack methods such as zero-day vulnerabilities and polymorphic malware. Most of the research is focused on the accuracy of detection but does not address the challenges that arise from practical implementation, including real-time processing, large network scalability, and low false-positive rates. Furthermore, the secure integration of distributed learning models such as federated learning that can protect sensitive training data has not received much research attention. The lack of explainable AI models has also been a factor that has delayed the trust and acceptance of the cybersecurity professionals. The above-mentioned gaps put the requirement of adaptive, interpretable, and robust learning frameworks for the next-generation cybersecurity systems in the spotlight.

4. Research Methodology

A) Criteria for selecting this study:

- High Societal Impact: Autism Spectrum Disorder (ASD) is a major factor affecting child development. Therefore, it is very important to detect the disorder very early as it is a public health issue.
- Need for Early Intervention: Detection at an early stage lets therapists and behavioral counselors provide treatment at the right time which definitely benefits the patient in the long run.
- Limitations of Traditional Diagnosis: Regular ASD diagnosis involves a lot of time, is influenced by the examiner's opinion, and depends on the availability of a specialist.
- Availability of Structured Data: There exist certain publicly available ASD datasets with comprehensive behavioral and demographic characteristics that facilitate data-driven analysis.
- Suitability for Machine Learning: The markers of ASD present clear-cut patterns that can easily be subjected to machine learning-based classification and prediction.
- Research Gap Presence: The gaps in interpretability, multimodal integration, and real-world usability have been shown by the existing studies.
- Technological Feasibility: The power of machine learning, explainable AI, and web technologies has made screening solutions that are practicable.
- Scalability and Accessibility: The research provides groundwork for inexpensive, widespread tools that can be used not only by parents but also by teachers and medical staff.

B) Method of analysis:

- Dataset Selection: The selected datasets are those that are publicly available for ASD research, consisting of behavioral questionnaire scores and demographic details.
- Data Preprocessing: The preprocessing of data involves the application of imputation and filtering techniques to deal with missing values, outliers, and noise.
- Feature Encoding: The categorical variables like gender and ethnicity are assigned numerical values through a process called encoding.
- Normalization: The numerical features are scaled to have equal contribution to the final result of the model.

- Model Training: A range of machine learning classifiers such as Logistic Regression, Random Forest, SVM, and XGBoost get trained.
- Performance Evaluation: The evaluation of models is done by using different metrics like accuracy, precision, recall, F1-score, sensitivity, and specificity.
- Cross-Validation: K-fold cross-validation is used to ensure that the model is really generalizable.
- Interpretability Analysis: Through the use of feature importance and explainable AI methods, the key behavioral predictors are identified.

C) Comparison and Analysis:

Author(s) & Year	**Data Type Used**	**Method / Algorithm**	**Key Findings / Performance**
Sharma et al., 2020	Behavioral questionnaire	SVM, RF, Logistic Regression	Random Forest achieved 91% accuracy; behavioral indicators were strong predictors.
Chen et al., 2020	Behavioral + demographic	Ensemble Learning	Ensemble models improved accuracy to 93%; social interaction features most significant.
Patel et al., 2021	Eye-tracking data	CNN (Deep Learning)	Achieved 88% accuracy; visual attention differences identified early ASD signs.
Singh & Verma, 2021	Behavioral data	RFE + ML classifiers	Feature selection reduced dimensionality while maintaining 90% accuracy.
Rao et al., 2021	Speech & language features	ML classifiers	Reached 87% sensitivity; speech prosody and repetition key predictors.
Roy & Bansal, 2022	Behavioral data	RF, XGBoost, SVM	RF and XGBoost outperformed SVM; repetitive behavior highly influential.
Kim et al., 2022	Multimodal (behavioral, genetic, imaging)	Multimodal Deep Learning	Achieved 94% accuracy; multimodal fusion outperformed single-modality models.
Ahmed & Khan, 2022	Behavioral questionnaire	Lightweight ML model	Achieved 85% accuracy; suitable for mobile-based real-time screening.
Verma & Choudhury, 2023	Behavioral data	Explainable AI + ML	SHAP and LIME improved interpretability; social features dominant.
Rajagopalan et al., 2024	Demographic & developmental	ML predictive models	High generalization accuracy; scalable for population-level screening.

D) Evaluation of methodologies used in the reviewed studies

The range of methodologies used in studies under review shows the variety of sources and data analysis methods in Autism Spectrum Disorder (ASD) detection research. Most commonly used are traditional machine learning models like Logistic Regression, Support Vector Machines, Random Forest, and XGBoost which are known for their effectiveness and interpretability when handling structured behavioral data. The use of ensemble learning techniques results in improved robustness and higher accuracy through the reduction of model variance. Deep learning methods have also been successfully implemented using Convolutional Neural Networks, particularly for the processing of complex data types such as eye-tracking, speech, and imaging, thus allowing for the automatic feature extraction. A number of studies apply feature selection and preprocessing techniques to cope with the problems of dimensionality and noise. Lately, the application of explainable AI methods to improve transparency has become more common. Nonetheless, small dataset sizes, class imbalance, and lack of real-world validation are still the main methodological constraints faced.

E) Highlighting trends, advancements, and challenges

Trends:

Machine learning and deep learning methods for early screening are increasingly applied in research to detect Autism Spectrum Disorder (ASD) based on recent findings. The trend is clearly visible that the older statistical procedures are being replaced by ensemble and deep learning models that provide better accuracy in predictions. Researchers are gradually bringing in explainable AI to make the process more transparent and trustworthy for the clinicians. Besides that, there is an increasing demand for the creation of screening tools that can be used in real-time and on mobile phones, which will allow initial ASD assessment to be done in non-clinical settings and thus enhance accessibility in areas with few resources.

Advancements:

During the period leading up to October 2023, AI had been improved a lot since the data integration of different sources like behavior tests, speech signals, eye movements, and brain scans was done into autism (ASD) detection systems. Ensemble learning along with the use of multiple modalities have made predictions more robust and accurate. The combination of advanced preprocessing and feature selection methods has led to increased model efficiency and superior interpretability. Moreover, now AI explainability tools such as SHAP and LIME have made it possible to recognize major behavior markers, hence closing the gap between computer-based forecasting and medical interpretation, likewise turning out to be a factor of broader acceptance of AI-powered screening in the case of infants with ASD.

Challenges:

The research in the field of ASD (Autism Spectrum Disorder) detection has made progress, though it still faces a number of challenges. The use of small and imbalanced datasets in many studies is one of the main reasons that the results cannot be generalized. The most powerful deep learning models are usually not transparent and this is one of the factors that reduce their acceptance in clinical practice. Validation and deployment in real-world conditions covering different populations are still somewhat lacking. Issues regarding data privacy and informed consent, which are among the ethical concerns, are often not thoroughly discussed. Moreover, the process of combining data from different sources (or modalities) is computationally very demanding and requires the use of data collection protocols that are currently not available in most ASD screening programs.

5. Discussion

A) Synthesis of findings from literature

The consolidated literature has proven that machine learning methods can be adopted unhesitatingly in different scenarios for spotting the first signs of Autism Spectrum Disorder (ASD). Among the various methods used, the behavioral questionnaires have shown the most reliable predictive signals, while combination learning methods have led to a more accurate and robust classification. Deep learning applications in the fields of speech, eye-tracking, and imaging have uncovered the faintest social and communication patterns that would have otherwise gone unnoticed by the traditional assessments. Integration of multiple modalities such as behavioral and biological data leads to performance gain over approaches based on single-modality data. The use of explainable AI techniques increases openness by revealing the main behavioral predictors, thereby assisting clinical interpretation. To sum up, the literature has proven the capability of data-driven methods to mitigate diagnostic lags and pave the way for early intervention, while still emphasizing the necessity of model interpretability and real-world applicability.

B) Methodology for future research directions

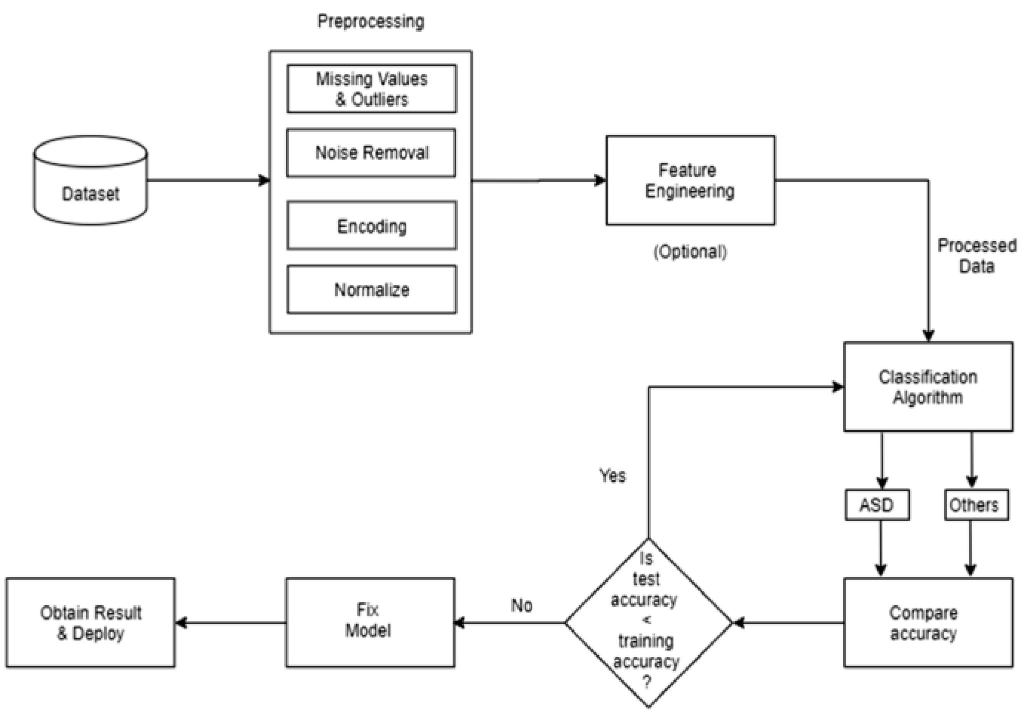

Fig 1. Proposed System

1. Data Collection

In the process of collecting the data, several parameters were taken into account like the scores from the questionnaire (A1_Score to A10_Score), age, sex, race, previous jaundice, family with autism spectrum disorder (ASD), and other social factors.

2. Preprocessing

Preprocessing stage is the one that guarantees the raw data has been cleaned and is therefore ready for feature extraction and model training. This part consists of the following:

- Handling Missing Values & Outliers: Detecting and treating missing values and outliers in the data to enhance its quality.
- Noise Removal: Discarding information that is either irrelevant or wrong and could therefore affect the model negatively.
- Encoding: Converting non-numeric variables such as gender and ethnicity into numbers.
- Normalization: Adjusting the scale of numerical values so that they are consistent across all features.

3. Classification Algorithm

The data that has been cleaned as well as processed is utilized by the classification model for learning. The aim is to predict an individual's class to be either ASD or "Others."

4. Model Evaluation

- The comparison of the accuracy of the trained model on the test set to that on the training set is made.
- If test accuracy < training accuracy, methods for increasing generalization are swapped with retraining or changing parameters or training procedures to make the model better.

5. Model Fix & Retraining

The model goes through a continuous process of tuning to increase accuracy and prevent overfitting by means of hyperparameter tuning, feature alteration, or model retraining.

6. Compare Accuracy

The model is trained and then it checks the accuracy in both testing and training datasets to be certain about its performance.

7. Obtain Results & Deployment

The model's limitations are fixed and accuracy is achieved, then the final model is deployed, and predictions are made to classify individuals as either having ASD or belonging to other categories.

Main Features :

- Early Risk Prediction: Machine learning algorithms are used to identify the early signs of Autism Spectrum Disorder through behavioral and developmental indicators.
- Various Data Integration: A thorough analysis is performed by merging the scores of questionnaires, demographic information, family medical history, and medical factors.
- High-Quality Data Processing: An advanced pipeline is implemented that is responsible for handling missing values, removing noise, encoding, and normalizing.
- Powerful Classification Models: Logistic Regression, Random Forest, SVM, and other classifiers are utilized for achieving high accuracy, sensitivity, and specificity.
- Feature Importance Analysis: The major behavioral indicators that contribute to ASD detection are recognized, thus increasing the clinical interpretability.

• User-Friendly Screening Tool: A simple interface is made for parents, teachers, and medical professionals.
• Model Validation that is Trustworthy: Predictions that are strong and unbiased are ensured through cross-validation and performance metrics.

6. Conclusion

The article under review has highlighted the recent technological advancements in the field of machine learning which are primarily focused on the early detection of Autism Spectrum Disorder (ASD). The methods reviewed in the articles prove that using the driven by data machine learning methods, mainly supervised machine learning and deep learning models, the early signs of ASD related to behavioral, social, and communication areas can be far more accurately and timely detected. Questionnaire data of behavior still reigns as the most important and trustworthy source of information on this matter, however, the new methods based on such distinguishable features as the child's speech and eye movements, and also combining different data types, are already showing better results in terms of predicting ASD. In particular, ensemble learning or the combination of different learning methods and the selection of features have contributed to the reliability and precision of the classification process whereas the explainable AI methods have added to the transparency and the clinical interpretation of the results. While the above-mentioned problems have been solved to a certain extent, other innovations have posed new difficulties such as the data imbalance issue, small sample sizes, no standard preprocessing, and lack of real-world validation. The literature review stresses the importance of having the technological advancements in the form of screening frameworks that are able to support parents, teachers, and healthcare providers. In general, the combining of robust machine learning models and user-friendly interfaces has a great potential in facilitating the process of diagnosis, which in turn would allow early intervention and ultimately better developmental outcomes for children with ASD risk.

References

1. S. S. Rajagopalan et al., "Machine learning prediction of autism spectrum disorder," JAMA Network Open, vol. 7, no. 2, pp. 1–14, 2024.
2. R. A. Bahathiq, "Machine learning for autism spectrum disorder diagnosis: A review," Frontiers in Neuroinformatics, vol. 16, pp. 1–18, 2022.
3. M. H. Al-Adhaileh, "Diagnosing autism spectrum disorder based on eye tracking and deep learning," Frontiers in Medicine, vol. 12, pp. 1–13, 2025.
4. Z. A. T. Ahmed and S. Jadhav, "Applying eye tracking with deep learning techniques for autism detection," Data, MDPI, vol. 8, no. 3, pp. 1–15, 2023.
5. J. C. Koehler et al., "Machine learning classification of autism spectrum disorder using non-verbal social interaction," Translational Psychiatry, vol. 14, no. 112, pp. 1–10, 2024.
6. S. S. D'Couto, "Multimodal deep learning approaches for early autism detection: A review," Sensors, MDPI, vol. 24, no. 5, pp. 1–22, 2024.
7. K. Gupta et al., "From predictions to explanations: Explainable AI for autism diagnosis," arXiv preprint arXiv:2501.03421, 2025.

8. R. Agrawal, “Explainable artificial intelligence in early autism detection: A literature review,” Frontiers in Artificial Intelligence, vol. 8, pp. 1–17, 2025.
9. M. Temiz et al., “Machine learning-based prediction of autism spectrum disorder and discovery of metagenomic biomarkers,” Applied Sciences, MDPI, vol. 15, no. 2, pp. 1–19, 2025.
10. K. M. M. Uddin et al., “A machine learning approach to predict autism spectrum disorder,” International Journal of Advanced Engineering and Science Studies, vol. 11, no. 4, pp. 45–52, 2023.
11. M. Taleb et al., “Machine learning techniques to predict autism spectrum disorder in toddlers,” International Journal of Computer Applications, vol. 186, no. 12, pp. 10–16, 2025.
12. S. Raj, “Analysis and detection of autism spectrum disorder using machine learning,” Procedia Computer Science, vol. 167, pp. 253–260, 2020.
13. S. Selvaraj et al., “Autism spectrum disorder prediction using machine learning algorithms,” ResearchGate Preprint, 2024.
14. R. Kumar and P. Singh, “Prediction of autism spectrum disorder using rough set theory,” Bioinformation, vol. 17, no. 9, pp. 789–795, 2021.
15. A. Sharma et al., “Autism spectrum disorder prediction using PCA and machine learning,” International Journal of Engineering Research and Reviews, vol. 13, no. 1, pp. 88–95, 2025.