

Machine Learning for Autism Spectrum Disorder Prediction: A Comprehensive Review

Vidyashree R

Mtech, Dept of CSE, RLJIT

Dr. Mamatha C M

Professor, Dept of CSE, RLJIT

Abstract

Advancements in machine learning (ML) and deep learning (DL) are transforming Autism Spectrum Disorder (ASD) diagnostics by enabling a shift from traditional behavioral observations to objective, algorithmic approaches. These models effectively leverage diverse data modalities—such as neuroimaging, behavioral cues (e.g., facial expressions, eye movements), and electronic health records—to achieve high diagnostic precision. The integration of multimodal data sources and the rise of interpretable AI systems are fostering greater accuracy and clinician confidence.

Despite promising research outcomes, significant barriers to clinical translation persist, including data heterogeneity, reproducibility challenges, a lack of standardization, and the crucial need for model interpretability. Furthermore, critical ethical considerations, particularly concerning algorithmic bias and data privacy, remain paramount. The field is actively progressing towards precision diagnosis, leveraging advanced multimodal data fusion techniques, sophisticated interpretable AI, and the potential of Generative AI for personalized insights. Overcoming existing barriers and realizing the full clinical utility of these technologies necessitates deep interdisciplinary collaboration among researchers, clinicians, and ethicists.

Introduction

1. Introduction: The Imperative for AI in ASD Diagnosis

1.1. Understanding Autism Spectrum Disorder (ASD)

Autism Spectrum Disorder (ASD) is a complex neuro developmental condition marked by persistent deficits in social interaction and communication, alongside restricted, repetitive behaviors. In India, the prevalence of Autism Spectrum Disorder (ASD) is estimated to be around 1 in 68 children, with boys being more commonly affected than girls. ASD presents with significant phenotypic and genotypic heterogeneity and often co-occurs with conditions like gastrointestinal or movement disorders.

1.2. Limitations of Traditional Diagnostic Methods

Traditional ASD diagnosis relies on subjective behavioral observations, making it time-intensive, inconsistently accessible, and prone to delays. Such delays can lead to more pronounced impairments, limit access to interventions, and increase healthcare costs.

1.3. The Transformative Potential of Machine Learning (ML) in ASD

Machine learning offers an objective, data-driven approach to ASD diagnosis by analyzing large, complex datasets and uncovering patterns beyond human capabilities. This can significantly enhance ASD understanding and improve screening. However, high accuracy in research settings does not guarantee real-world clinical feasibility. Factors like integration into workflows, cost-effectiveness, and ease of use are crucial for successful clinical translation. The primary goal of ML research in this area is early and accurate diagnosis, which is vital for better developmental outcomes and improved quality of life for individuals with ASD.

2. Data Modalities and Key Datasets for ASD Prediction

2.1. Diverse Data Modalities

ML algorithms can analyze complex, high-dimensional data from various sources for ASD prediction, with multimodal integration being crucial for robust diagnostic models.

Neuroimaging Data: Provides objective insights into brain structure and function.

Structural MRI (sMRI) assesses structural brain changes, though its standalone success in ASD diagnosis is still evolving.

Resting-state functional MRI (rs-fMRI) captures functional connectivity (FC), with altered patterns being a key neurobiological correlate of ASD.

Task-based fMRI amplifies differences by collecting data during specific tasks.

Other modalities like PET, MEG, EEG, iEEG, and NIRS are also being explored.

Behavioral Data: Observable behaviors offer accessible data.

Eye-tracking data is a promising early indicator due to its ability to capture atypical visual attention patterns; it is fast, inexpensive, and applicable across ages.

Facial images serve as a physiological identifier, especially for developing nations, with deep learning models showing high accuracy.

Questionnaire data from standardized assessments like the Autism Quotient (AQ) provide a robust framework.

Exploratory studies also investigate **micro-expressions**.

Genetic Data: Over 200 risk regions and genes have been identified. Neuroimaging genetics integrates AI to link ASD-risk genes to brain variations.

Clinical and Background Information: Minimal sets of medical and developmental information are important predictors.

Novel Biomarkers: Research includes factors like intestinal microbial composition in risk prediction models.

Fusing these diverse data types is crucial for robust and comprehensive ML models, addressing the multidimensional nature of ASD and leveraging complementary information for more accurate and generalizable predictions.

2.2. Prominent Datasets

Several large-scale datasets drive ML research for ASD prediction:

Autism Brain Imaging Data Exchange (ABIDE): A widely used, freely available database of functional and structural brain imaging from over 24 international sites. ABIDE I (2012) includes 539 ASD and 573 control subjects (ages 7-64). ABIDE II (2016-2017) comprises 1114 datasets (521 ASD, 593 control). Its multi-site nature introduces heterogeneity, impacting model reproducibility and generalizability. Gold standard diagnostic instruments like ADOS and ADI-R are used across most sites.

Kaggle Dataset: Contains 2,940 facial images of autistic and non-autistic children, but has limitations in clinical history and demographic representation (10:1 white to non-white children).

SPARK Database: A large diagnostic study (30,660 participants in v8, 14,790 in v10) with extensive medical and background information used for ML model development and validation.

Open Neuro: A free platform for sharing BIDS-compliant neuroimaging data (MRI, PET, MEG, EEG, iEEG, NIRS), crucial for standardization and reproducibility.

3. Machine Learning Approaches and Their Performance

3.1. Traditional Machine Learning Algorithms

Various traditional classifiers like Logistic Regression, Decision Tree, Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Naïve Bayes (NB), and ensemble methods (AdaBoost, XGBoost) have been applied. Reported accuracies vary widely; some studies show RF at 100%, Neural Networks at 99%, and Decision Tree at 100% with complete feature sets. Logistic Regression achieved 97.541% with feature selection, and XGBoost showed strong performance on the SPARK database with an AUROC of 0.895. The wide range of accuracies highlights that performance is context-dependent, influenced by dataset, feature engineering, preprocessing, and evaluation setup.

3.2. Deep Learning (DL) Models

Deep learning methods, popular since 2017, generally outperform traditional ML with large datasets. Common architectures include CNNs, RNNs, LSTMs, DNNs, Autoencoders, DBNs, and GCNs. DL models on facial images show high accuracy (e.g., VGG19 at 87.9% and 98% ; a proposed framework at 95.6% accuracy). For eye-tracking data, FFNNs and ANNs achieved 99.8% accuracy, and CNNs 97.41%. A meta-analysis of DL for childhood ASD reported overall sensitivity of 0.95, specificity of 0.93, and AUC of 0.98.

3.3. Hybrid and Ensemble Approaches

Combining deep learning (e.g., GoogleNet, ResNet-18) with traditional ML (e.g., SVM) has shown superior diagnostic ability. Ensemble models, like GCN classifiers, achieved 72.2% accuracy and an AUC of 0.77 on ABIDE data, outperforming individual models. Multimodal fusion frameworks, integrating diverse data types, have achieved very high accuracies (e.g., 99.2%). This approach leverages complementary information from genetic, behavioral, and neurological aspects, which is crucial for capturing ASD's complexity.

3.4. Performance Metrics and Interpretation

ML models for ASD prediction are evaluated using metrics like accuracy, precision, recall (sensitivity), specificity, F1-score, and AUROC. While high accuracies are reported, dataset choice, feature engineering, and evaluation pipelines significantly influence these figures. Studies with heterogeneous datasets like ABIDE often show similar accuracies around 70% across different models. This suggests that data quality, meticulous preprocessing, effective feature engineering, and standardized evaluation are often more critical than the specific ML algorithm for real-world data.

<i>Model</i>	Reported Accuracy (%)	Strengths	Weaknesses
<i>Logistic Regression</i>	85–95%	Simple, interpretable	Poor performance on non-linear data
<i>Decision Tree (DT)</i>	90–100%	Easy to understand, quick training/testing	Prone to overfitting
<i>Random Forest (RF)</i>	96–100%	High accuracy, reduces overfitting via ensemble	Less interpretable than single tree
<i>K-Nearest Neighbors</i>	80–95%	No training phase, intuitive	Sensitive to noise and irrelevant features
<i>Naïve Bayes (NB)</i>	80–93%	Fast, effective with small datasets	Assumes feature independence
<i>Support Vector Machine (SVM)</i>	85–98%	Effective in high-dimensional space	Requires careful kernel and parameter tuning
<i>XGBoost / AdaBoost</i>	94–99%	Excellent for imbalanced/structured data	Computationally intensive, harder to interpret

4. Methodological Considerations in ML for ASD

4.1. Data Preprocessing and Quality

The effectiveness of ML models hinges on data quality and preparation. Raw data requires extensive preprocessing, including outlier removal, imputation of missing values, and resolving discrepancies. Data cleaning removes errors, duplicates, and structures the dataset. Class imbalance, common in ASD datasets, where ASD cases are fewer than controls, can hinder accurate positive class prediction. Techniques like Safe-Level SMOTE are used to balance data, improving generalization.

4.2. Feature Engineering and Selection

Feature engineering involves deriving new, valuable features from existing ones to boost model performance and provide insights. Examples include summing clinical scores or grouping ages. For high-dimensional data, feature selection methods like Chi-square, LASSO, and Boruta reduce dimensionality, enhance accuracy, and speed up classification. Deep learning models, like CNNs, can automatically extract features, reducing reliance on manual engineering.

4.3. Cross-Validation and Model Validation

Rigorous validation ensures model robustness and generalizability. K-fold cross-validation (e.g., 10-fold) assesses performance across data subsets, providing a more reliable estimate than a single split. Validating models on independent datasets confirms generalizability to new data.

4.4. Impact of Dataset Properties

Dataset properties significantly influence ML model performance. Large, multi-site datasets like ABIDE introduce heterogeneity from varying scanners, protocols, and populations, often leading to lower accuracies than homogeneous, single-site datasets. However, selecting specific subsets can mitigate heterogeneity and improve accuracy. Larger, well-curated datasets can also improve performance.

5. Challenges in Clinical Translation of ML/DL Models for ASD

1. One of the major barriers is the lack of standardized datasets and consistent methodologies, which makes it difficult to compare results across studies. Additionally, models often fail to generalize well due to

demographic imbalances and the inherent heterogeneity of ASD. While they may perform well on small, curated datasets, their accuracy often drops significantly on more diverse, real-world datasets.

2. Many ML and DL models operate as "black boxes," providing little insight into how decisions are made. Clinicians require transparency and explainability to trust these systems in practice. Without clear rationale behind predictions, integration into clinical workflows becomes difficult. Tools like SHAP and Monte Carlo Dropout are helping address this by making model decisions more understandable.
3. High research accuracy doesn't guarantee clinical readiness. Models that excel in controlled environments may struggle with practical challenges such as variability in patient populations, data quality, and the need to address the full range of ASD subtypes and comorbidities. These limitations hinder real-world deployment and impact.

6. Future Directions and Research Opportunities

Advancing Autism Spectrum Disorder (ASD) diagnostics will rely heavily on multimodal data fusion, integrating neuroimaging, genetic, and behavioral data to capture the disorder's full complexity and improve diagnostic accuracy. Equally important is the development of interpretable and explainable AI models, using techniques like SHAP and Monte Carlo Dropout to clarify predictions and build clinical trust. Generative AI, including large language models and multimodal systems, offers new opportunities for personalized diagnostics by synthesizing clinical, textual, and visual data to generate tailored assessments and recommendations. Additionally, future research should focus on precision diagnosis by identifying biologically distinct ASD subtypes, moving beyond the traditional monolithic view and enabling individualized monitoring and targeted treatment strategies.

Future ML models should aim to identify specific ASD subtypes, leading to more targeted developmental monitoring, precise treatment strategies, and tailored support based on an individual's unique biological and clinical profile. This approach explains why past genetic studies often fell short, as they were mixing different puzzles together.

8. Conclusions

The application of machine learning and deep learning in Autism Spectrum Disorder prediction holds immense promise for transforming diagnostic processes from subjective behavioral observations to objective, data-driven methodologies. Current models demonstrate high accuracy across diverse data modalities, including neuroimaging, behavioral cues, and clinical records. The trend towards multimodal data fusion, integrating genetic, behavioral, and imaging data, is particularly compelling, as it aligns with the multidimensional nature of ASD and shows superior predictive performance. The development of interpretable AI frameworks and the exploration of Generative AI for personalized diagnostics represent critical advancements aimed at enhancing clinical trust and utility.

However, significant challenges persist that impede the full clinical translation of these technologies. These include pervasive issues of data heterogeneity, the difficulty in reproducing research findings across different settings, and a fundamental lack of standardization in data collection and methodologies. The generalizability of models across diverse populations, especially given historical biases in datasets, remains a concern, as models performing exceptionally well in controlled environments may not translate effectively to real-world clinical variability. Furthermore, the "black box" nature of many advanced ML models necessitates a strong emphasis on interpretability to foster clinician trust and facilitate evidence-based decision-making.

Ethical considerations are paramount, particularly concerning algorithmic bias that could exacerbate existing disparities in patient care, as well as stringent requirements for data privacy, security, transparency, and accountability. The critical importance of interdisciplinary collaboration between computational scientists, clinicians, and behavioral experts cannot be overstated. Ignoring the clinical context can lead to misinformed conclusions and hinder the development of truly effective and equitable diagnostic tools. Ultimately, realizing the full potential of machine learning for ASD prediction requires not only continued technical innovation but also a concerted effort to address these methodological and ethical challenges through collaborative, standardized, and transparent research practices, paving the way for more precise, accessible, and equitable diagnostic pathways.

References

1. Analysis & Detection of ASD Using ML Techniques (Procedia Comp. Sci.,2020)
Raj & Masood, 2020 — Compares NB, SVM, LR, KNN, Neural Networks over child/adolescent/adult datasets; CNN outperformed ML, but traditional ML still central
2. Detection of ASD in Children Using ML Techniques.(SN Comput.Sci.,2021)
Vakadkar et al., 2021 — Benchmarks SVM, RF, NB, LR, KNN; Logistic Regression performed best for this dataset
3. A machine learning–based diagnostic model for ASD with intellectual disability (children)
Song et al., 2022 — Evaluates logistic regression, SVM, Random Forest, and XGBoost on behavioral and socio-demographic features; best AUC scores around 0.835–0.858 and SVM accuracy ~0.836
4. Prediction of ASD across age groups (MDPI, 2023)
R.K. Raj et al., 2023 — Public UCI ASD screening datasets (child, adolescent, adult); multiple classifiers evaluated for robust predictions
5. Prediction and analysis of ASD using ML (PMC,2023)
Anonymous, 2023 — Applied SVM, NB, RF, MLP on balanced vs. imbalanced datasets (SMOTE); RF reached ~94.7% accuracy on balanced data.
6. Machine learning differentiation of ASD sub-classifications (SPARK dataset, 2023)
Unknown authors, 2023 — Multiclass ML to distinguish ASD subtypes under DSM-IV vs non-spectrum; AUROCs 0.863–0.980, overall ~80.5% correct classification.
7. A Prediction Model of Autism Spectrum Diagnosis from Well-Baby Electronic Data Using ML. Ben-Sasson et al., Children (Apr 2024)
Applied gradient boosting on wellness records of over 780,000 infants (including 1,163 ASD cases), achieving AUROC ≈ 0.86 . Identified key medical and developmental milestone predictors using SHAP
8. Early Identification of Autism Spectrum Disorder Based on Machine Learning with Eye-tracking-data
Journal of Affective Disorders (Aug 2024)
Evaluated ML classifiers (e.g., RF, SVM, XGBoost) on eye-tracking data in 449 children. Random Forest delivered AUROC ≈ 0.849 , with social- and non-social cognition features both contributing significantly.
9. Machine Learning-Based Classification of Autism Spectrum Disorder across Age Groups
Reghunathan et al., Engineering Proceedings (Mar 2024)
Benchmarked classifiers (LR, SVM, RF, KNN) on child/adolescent/adult UCI datasets using feature reduction via Cuckoo Search. Logistic Regression achieved highest accuracy across age groups
10. Combining Radiomics and Machine Learning Approaches for Objective ASD Diagnosis: Verifying White Matter Associations with ASD
Song et al., arXiv (May 2024) Focused on white matter MRI radiomics features. SVM achieved ~89.5% accuracy; RF, LR, and KNN also exceeded 80%. Demonstrated structural biomarkers predictive of ASD using purely ML methods