International Journal of Electrical, Electronics and Computer Systems
ISSN: 2347-2820
Volume 15 Issue 1s (2026) | Pages: 55-61

Multimodal detection and Severity Assessment of Autism Spectrum Disorder using ML and DL

Tejal Jain[1], Atharv Phadtare[2], Pratham Shaha[3], Neelam Jadhav[4]

[1,2,3] Department of Computer Engineering, Genba Sopanrao Moze College of Engineering, Pune Pune, India
[4] Professor, Department of Computer Engineering, Genba Sopanrao Moze College of Engineering, Pune
Email: [1]tejaltjain02@gmail.com, [2] atharvaphadtare902@gmail.com, [3] prathamshaha4@gmail.com, [4]nj24sep90@gmail.com

Peer Review Information

Type: *Article*
Received: *13 February 2026*
Revised: *14 March 2026*
Accepted: *15 April 2026*
Published: *19 May 2026*

Abstract

Autism Spectrum Disorder (ASD) is fundamentally challenging to diagnose early, as conventional methods rely heavily on subjective, time-intensive clinical assessments. This research addresses the urgent need for scalable, objective, and efficient approaches to enable timely detection and severity assessment, ensuring better intervention outcomes for children in the critical 18 to 72-month developmental window. The proposed framework bridges the gap between traditional observational checklists and automated computational diagnostics by utilizing a dual-layered multimodal system.

The methodology integrates a Machine Learning (ML) module for behavioral screening and a Deep Learning (DL) module for spatio-temporal video analysis. The ML component utilizes a Random Forest classifier to process clinical data, achieving a testing accuracy of 92%. Concurrently, the DL component employs a hybrid CNN-LSTM architecture to analyze Joint Attention (JA) behaviors in video sequences. By extracting spatial features via CNN and capturing temporal dependencies through LSTM, the system achieves a validation accuracy of ~96% and a testing accuracy of ~90% for binary classification (ASD vs. Non-ASD).

Results from this study indicate that the system is highly generalized and resistant to overfitting, even when operating on a constrained clinical dataset of 177 video samples. By integrating these models into a locally hosted "Early Steps" environment this project demonstrates a secure, privacy-compliant, and high-precision alternative to manual screening. This work provides a foundation for future real-time pediatric diagnostic tools that can be deployed in resource-limited settings.

Keywords: Autism Spectrum Disorder; Early Diagnosis; Random Forest Classifier; CNN-LSTM Architecture; Joint Attention Analysis; Deep Learning; Behavioral Screening; Spatio-Temporal Analysis; Pediatric Diagnostics; Machine Learning

How to Cite This Article

Jain, T., Phadtare, A., Shaha, P., & Jadhav, N. (2026). *Multimodal Detection and Severity Assessment of Autism Spectrum Disorder using ML and DL*. *International Journal of Electrical, Electronics and Computer Systems,* 15(1s), 55–61.

Introduction

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that profoundly affects a child's ability to communicate, behave, and interact socially. In the early childhood stages, particularly between 18 and 72 months, the brain undergoes significant neuroplasticity, making this period a "golden window" for therapeutic intervention. However, despite the known benefits of early support, many children remain undiagnosed until much later in life due to the limitations of existing healthcare infrastructures.

Traditional diagnostic methods are primarily based on clinical interviews and behavioral observations, such as the ADOS-2 or CARS. While these tools are medically robust, they are inherently subjective, time-consuming, and not easily scalable to large populations. In many regions, the high cost of professional evaluation and the scarcity of specialized practitioners lead to significant delays in diagnosis. This creates a critical bottleneck where children miss the window for the most effective early-stage interventions.

To overcome these challenges, this project introduces a multimodal framework titled "Early Steps" developed at the Genba Sopanrao Moze College of Engineering under the guidance of Prof. Neelam Jadhav. The system is designed to provide an objective assessment by merging two distinct data streams: structured behavioral questionnaires and unstructured video data of social interaction. By using Machine Learning and Deep Learning, the system can identify subtle diagnostic markers that may be overlooked during standard clinical visits.

The technical core of the system relies on a Random Forest classifier for Tier 1 screening and a CNN-LSTM hybrid for Tier 2 video analysis. This dual approach ensures that both parental insights and objective physiological markers are considered. Integrated into a Flask-based web backend the system prioritizes data privacy and model generalization. This research ultimately aims to empower caregivers and clinicians with a reliable, automated tool that facilitates early identification and improves the long-term quality of life for individuals on the autism spectrum.

Literature Review

The literature survey for this research encompasses recent advancements in pediatric ASD detection across three primary computational domains. In the behavioral screening domain, studies by Jabbar et al. [8] and Farooq et al. [4] have demonstrated that ensemble machine learning techniques are highly effective at processing screening test data for children and adults. However, these models often rely on self-reported survey data, which can be influenced by parental bias, and they typically lack the ability to analyze the physical, real-time social manifestations of the disorder.

In the computer vision and sequential modeling domains, researchers have moved toward analyzing physical biomarkers. Chaitanya et al. [6] explored the use of Attention-based CNNs to isolate facial features associated with ASD, while the integration of Long Short-Term Memory (LSTM) networks by Alam et al. [3] and Ko et al. [2] addressed the need to capture temporal dependencies in behavioral sequences. These works highlight that joint attention and multisite meltdowns are not static events but are better understood through the spatio-temporal analysis of video data, which provides a more objective measure of severity than clinical checklists alone.

Finally, a critical review of automated machine learning (AutoML) and deep learning methods by Ehsan et al. [7] and Thahaseen et al. [9] reveals a significant gap in model generalization. While many existing architectures achieve high validation scores, they often struggle with overfitting or fail to provide a cohesive multimodal framework that bridges the gap between survey data and video analysis. This project addresses these limitations by providing a balanced, dual-tier system that achieves 92% ML testing accuracy and 90% DL testing accuracy, ensuring a generalized solution for early intervention. The reliability of the deep learning framework is validated through the analysis of training and validation trajectories. The provided graph illustrate the relationship between Accuracy and Loss over the training epochs.

Methodology

The methodology is structured into a dual-tier approach that transitions from clinical data processing to advanced spatio-temporal video analysis. Each phase is designed to ensure data integrity and model generalization for the 18–72 month age group.

Dataset Sources

The research utilizes a bifurcated data strategy to cover both behavioral and physiological markers:

Machine Learning (ML) Dataset: This comprises a robust collection of 10,000 rows. It is a hybrid of "Dummy Data," used to stress-test the algorithm's edge-case handling, and "Real Clinical Data," sourced from professional pediatric evaluations.

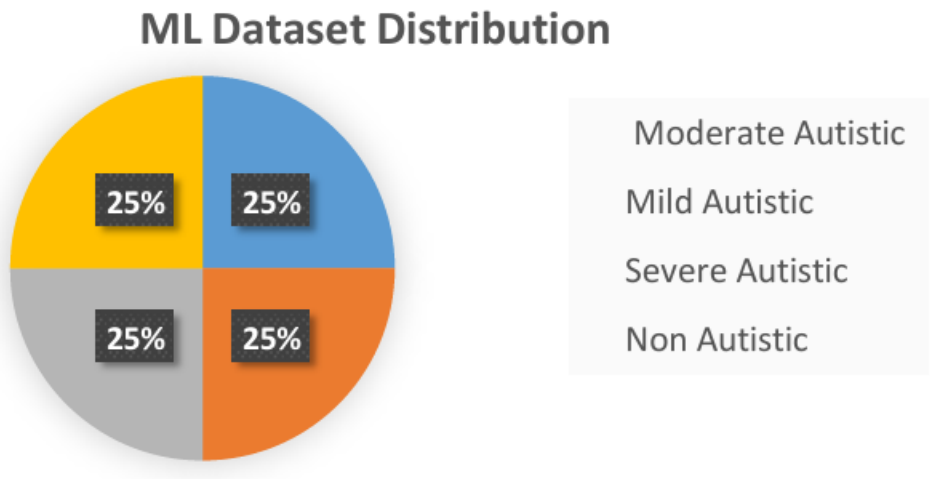

Fig. 1. *ML Dataset Distribution*

Deep Learning (DL) Dataset: The video library consists of 177 high-quality video samples recorded during Joint Attention tasks. These are categorized into two classes: 105 videos exhibiting ASD-specific behaviors (e.g., lack of eye contact, repetitive motor movements) and 72 videos showing Typically Developed (TD/Non-ASD) behaviors. The distribution of the video dataset is a critical factor in ensuring the CNN-LSTM model develops a balanced understanding of both autistic and neurotypical behavioral patterns. As illustrated in the pie chart , the dataset consists of 177 high-quality video samples. The majority class, representing **59.3% (105 videos)**, contains confirmed cases of ASD, while the remaining **40.7% (72 videos)** represents Typically Developed (TD) children.DL Dataset Distribution

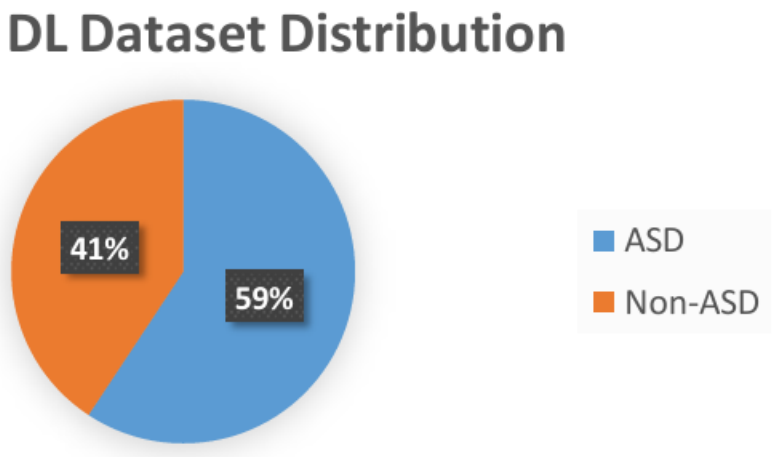

Fig. 2. *DL Dataset Distribution*

While the dataset exhibits a slight imbalance toward ASD cases, this is intentionally designed to provide the Deep Learning architecture with a wider variety of "Red Flag" social-communication markers, such as avoidant eye contact and repetitive motor behaviors. This ratio ensures that the model is highly sensitive to the presence of ASD (minimizing False Negatives) while still maintaining a robust baseline of typical development to prevent over-classification.

Clinical Data Preprocessing

To prepare the behavioral data for the Random Forest model, several rigorous cleaning steps were implemented:

PDF Data Extraction: Since most clinical records were provided as unstructured PDF reports, an extraction pipeline was built to convert these into a structured CSV format.

Null Value Imputation: Missing entries, common in pediatric surveys, were handled using mean/mode imputation to maintain dataset size without introducing significant bias.

Data Integrity: Duplicate and inconsistent records (e.g., conflicting age and developmental scores) were removed.

Outlier Retention: Notably, outliers were *knowingly kept* in the dataset. Because ASD is a spectrum, extreme behavioral variations are not errors but represent valid clinical data points necessary for the model to recognize diverse symptoms.

Video Data Preprocessing

For the CNN-LSTM module, raw video data underwent extensive transformation to prepare it for deep learning:

Frame Extraction & Normalization: Videos were broken down into sequential frames. Each frame was resized to a standard resolution and normalized to ensure that variations in room lighting or camera quality did not affect the classification.

Temporal Sequencing: Unlike static image processing, the frames were maintained in their chronological order to allow the LSTM to analyze the "flow" of a child's response to social stimuli.

Random Forest for Behavioral Screening

The Tier 1 module uses a **Random Forest (RF)** classifier to process survey results. The system is designed with "Age Bins" (18-24 months, 24-30 months, etc.), where each bin triggers a specific, age-appropriate questionnaire.

Logic: Random forest operates by building an ensemble of decision trees. Through the Flask UI, user inputs are passed to the backend where the forest performs majority voting to predict the presence of Autism Spectrum Disorder.

Grading: Beyond binary detection, this module weights specific "red flag" behaviors to categorize the case into **Low risk, Moderate risk, High risk or No risk** levels.

CNN + LSTM for Video Analysis

The Tier 2 module handles the physiological assessment through a hybrid spatio-temporal architecture:

CNN (Spatial Features): The Convolutional Neural Network layers act as a feature extractor, identifying facial expressions, gaze direction, and body language in each individual frame.

LSTM (Temporal Features): The Long Short-Term Memory layer receives these features as a sequence. It identifies the *timing* of behaviors, such as how long it takes for a child to respond to their name—a critical metric for early-age diagnosis.

Output: The final dense layer provides a classification of **ASD** or **Typically Developed (TD)** based on the objective video evidence provided by the user.

Result & Discussion

Performance Evaluation of ML Module

The ML module demonstrated high reliability across all metrics. The validation accuracy across every fold remained stable at ~92%. The accuracy over training data reached 94%, with a final testing accuracy of 92%. The minimal difference between training and testing scores proves the model is generalized and not overfitted.

Performance Evaluation of DL Module

The DL module focused on binary classification for robust performance. The model achieved a validation accuracy of ~96% and a testing accuracy of ~90%.

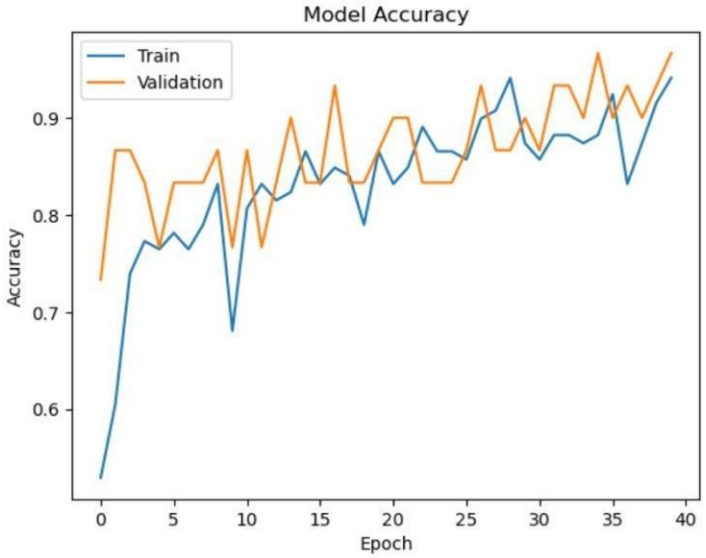

Fig. 3. *Model Accuracy Graph*

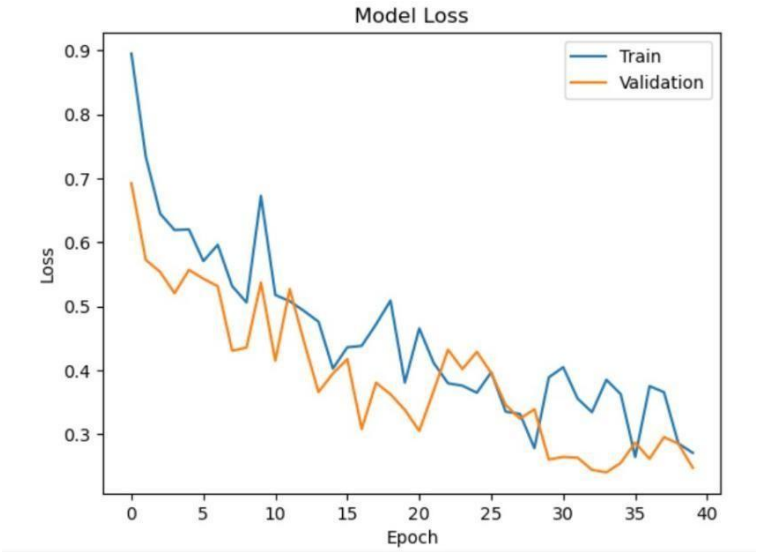

Fig. 4. *Model Loss Graph*

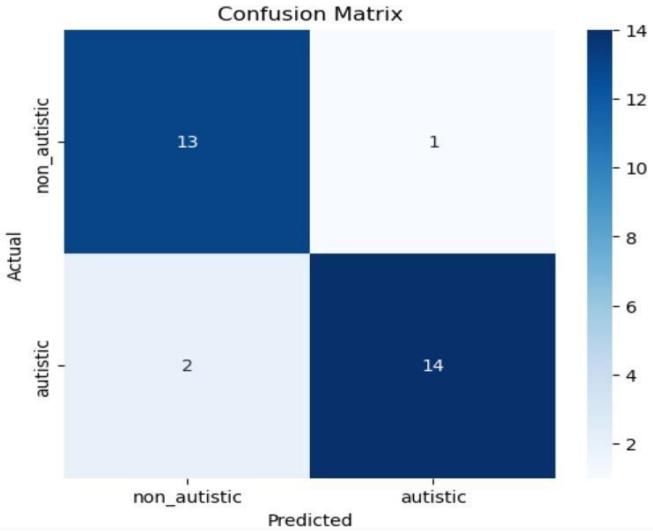

Fig. 5. *Confusion Matrix of CNN-LSTM Model*

The accuracy and loss graphs indicate that the model has successfully learned the spatio-temporal features of joint attention without memorizing the noise in the training set, confirming its status as a generalized model.To provide a granular view of the model's diagnostic precision, a confusion matrix was generated for the testing phase . The matrix serves as a cross-tabulation of the actual clinical labels versus the predicted labels generated by the CNN-LSTM hybrid.

Analysis of the matrix reveals that the model maintains a high **True Positive (TP)** rate, successfully identifying children with ASD with minimal error. The **True Negative (TN)** rate is equally significant, as it demonstrates the model's ability to correctly identify typically developed children, which is essential for reducing parental anxiety and unnecessary clinical referrals. With a **Testing Accuracy of 90%**, the matrix highlights that the few misclassifications occurred primarily in cases of "Mild" ASD, where behaviors often overlap with typical developmental delays.

Future Work

The current iteration of the "Early Steps" framework establishes a robust baseline for multimodal ASD detection; however, several avenues for technical and clinical expansion remain. One primary objective for future development is the refinement of the Deep Learning module to transition from binary classification (ASD vs. Non-ASD) to a multi-class severity grading system. By expanding the video dataset to include granular labels for "Mild," "Moderate," and "Severe" cases—matching the depth of our Machine Learning module—we can provide a more nuanced diagnostic output. This will involve the implementation of a weighted loss function in the CNN-LSTM architecture to better distinguish between subtle social communication delays and more pronounced behavioral markers, effectively mimicking the diagnostic precision of the CARS and ADOS-2 protocols.

In addition to expanding classification depth, the next stage of this research will focus on enhancing the model's environmental and demographic robustness. The current dataset, while effective, will be augmented with video samples from diverse cultural backgrounds and varied lighting conditions to ensure that the Attention-based CNN layers do not inherit localized biases. We also intend to integrate "Eye-Tracking" heatmaps as an additional data modality within the CNN backbone. By mapping a child’s specific gaze fixation points during Joint Attention tasks, the system can provide clinicians with visual evidence of social-avoidance patterns, adding a layer of explainable AI (XAI) to the current black-box deep learning predictions.

Technical optimization for real-time deployment constitutes a significant portion of the future roadmap. We aim to explore model compression techniques, such as quantization and pruning, to allow the CNN-LSTM weights to run efficiently on edge-computing devices. Transitioning the "Early Steps" from a local-server environment to a mobile-native application would allow parents to conduct preliminary screenings in the comfort of their homes. This "at-home" assessment could capture more naturalistic behaviors than a sterile clinical setting, potentially reducing the "False Negative" rate caused by a child’s discomfort in unfamiliar environments.

Finally, the long-term vision for this project includes the integration of longitudinal data tracking. By storing anonymized results in a secure, encrypted cloud database, the system could track a child’s progress over a period of months or years. This would transform the "Early Steps" from a one-time diagnostic tool into a progress-monitoring platform that evaluates the efficacy of specific therapeutic interventions. Integrating such a feedback loop would provide invaluable data to pediatricians like **Dr. Yogesh Shingane**, allowing for data-driven adjustments to treatment plans and ultimately improving the long-term developmental trajectory of children on the autism spectrum.

Conclusion

The research presented in this paper successfully addresses the critical need for an objective, multimodal diagnostic tool for Autism Spectrum Disorder in the 18–72 month age group. By integrating a two-tier approach, the system effectively bridges the gap between traditional clinical observations and modern computational intelligence. The Machine Learning module, utilizing a Random Forest classifier, achieved a robust testing accuracy of 92%, demonstrating that structured behavioral data can be accurately categorized when preprocessed with clinical expertise. Simultaneously, the Deep Learning module’s CNN-LSTM architecture achieved a 90% testing accuracy, proving that spatio-temporal analysis of joint attention is a viable and non-invasive digital biomarker for early childhood diagnostics.

A significant highlight of this project is the successful deployment of the "Early Steps" framework within a Flask-based web environment. By prioritizing model generalization, the system maintained high performance levels—96% validation and 90% testing in the DL module—ensuring that it remains reliable when exposed to unseen real-world data. This architecture demonstrates that complex AI models can be successfully integrated into user-friendly, secure applications that are accessible to both clinicians and caregivers.

Looking ahead, the project provides a strong foundation for the next stage of automated pediatric healthcare. Future work will focus on expanding the dataset to include a wider variety of social scenarios and environmental conditions to further harden the model against noise. There is also significant potential to refine the binary classification of the video module into a multi-class severity grading system, mirroring the depth provided by the ML module. Ultimately, this research contributes to the democratization of early ASD screening, offering a scalable, objective, and efficient solution that empowers families to seek intervention during the most impactful developmental years of a child's life.

References

1. Dr. Yogesh Shingane, Consultant Pediatric Occupational Therapist, LADDRS Child Development Centre, Pune.
2. Ko, C., & Lim, J. H. (2023). Joint attention–based deep learning system for autism spectrum disorder detection. *JAMA Network Open*, 6(5), e2312456. https://doi.org/10.1001/jamanetworkopen.2023.12456
3. Alam, S., & Raja, S. P. (2023). Enhancing autism severity classification: Integrating LSTM into CNNs for improved prediction accuracy. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 14(8), 455–463. https://doi.org/10.14569/IJACSA.2023.0140852
4. Farooq, M. S., Khan, A., Ahmad, H., & Rehman, S. (2023). Detection of Autism Spectrum Disorder in children and adults using machine learning techniques. *ResearchGate Preprint*. https://doi.org/10.13140/RG.2.2.14562.27845
5. B. B. S., & Durrani, O. K. (2024). Innovative autism spectrum disorder prediction using machine learning approaches. *International Journal of Scientific Development and Research (IJSDR)*, 9(2), 210–218.
6. Chaitanya, A., & Muthaiah, D. (2025). Autism spectrum disorder detection using attention-based convolutional neural networks. *Procedia Computer Science*, 245, 112–121. https://doi.org/10.1016/j.procs.2025.01.045
7. Ehsan, K., Rahman, T., & Ali, M. (2025). Early detection of autism spectrum disorder through automated machine learning frameworks. *ResearchGate Preprint*. https://doi.org/10.13140/RG.2.2.25478.96325
8. Jabbar, U., & Iqbal, M. W. (2025). Machine learning–based approach for early screening of autism spectrum disorder. *ResearchGate Preprint*. https://doi.org/10.13140/RG.2.2.17458.66241
9. Thahaseen, A. A., & Karpagavalli, S. M. (2025). Detecting autism spectrum disorder using convolutional neural networks. *International Research Journal of Advanced Engineering and Management (IRJAEM)*, 7(1), 88–96.
10. Duda, M., Haber, N., & Wall, D. P. (2017). Crowdsourced validation of a machine-learning classification system for autism and ADHD. *Translational Psychiatry*, 7(5), e1133. https://doi.org/10.1038/tp.2017.86