RESEARCH ARTICLE OPEN ACCESS

Early Detection of Autism Spectrum Disorder in Toddlers Using Computer Vision

Dr. Shantala T*, Ayush Krishnan P**, Avikrit Chhetri***, Jeswin Mickle****
*(Associate Professor, Department of Computer Science & Engineering, T John Institute of Technology, Bengaluru, India.
Email: shanchendu@gmail.com)
**(Student, Department of Computer Science & Engineering, T John Institute of Technology, Bengaluru, India.
Email: ayushkrishnan77@gmail.com)
***(Student, Department of Computer Science & Engineering, T John Institute of Technology, Bengaluru, India.
Email: chhetriavikrit@gmail.com)
****(Student, Department of Computer Science & Engineering, T John Institute of Technology, Bengaluru, India.
Email: jeswinm37@gmail.com)

--************************----------------------------------

Abstract:

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition affecting communication, social interaction, and behaviour in early childhood. Well-timed diagnosis significantly improves developmental outcomes, yet orthodox diagnostic tools such as the Autism Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R) are expensive, slow, and heavily dependent on trained clinical specialists. The growing inaccessibility of these tools in rural and resource inadequate settings further compounds the diagnostic gap. This survey reviews the current state of Artificial Intelligence (AI) and Computer Vision (CV) based approaches for early ASD detection in children, identifying key methodologies, datasets, behavioural signals, and research gaps. Based on the discoveries, we propose a gamified AI based mobile screening framework that provides real time facial expression analysis, gaze estimation, and interaction behaviour modelling on mobile hardware. The proposed system incorporates a dual interface design comprising a child facing gamified application for natural behavioural data capture and a clinician facing analytics dashboard providing ASD risk scores, behavioural trend graphs, and AI generated clinical insights. This work is currently in the application and development phase; full experimental validation is planned for subsequent phases. The proposed framework aims to democratize early ASD screening beyond specialist clinics, offering a scalable, non-invasive, and accessible tool for home and primary care settings.

Keywords — **Autism Spectrum Disorder, Computer Vision, Eye Gaze Analysis, Facial Expression Recognition, Machine Learning, CNN, Gamified Application, Early Screening, Mobile Health, Doctor Analytics Dashboard.**

--************************----------------------------------

INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition manifesting in early childhood, characterized by difficulties in social communication, restricted and repetitive behaviors, and atypical sensory processing. The Centers for Disease Control and Prevention (CDC) estimates ASD prevalence at approximately 1 in 36 children in the United States, with comparable trends documented globally. In India, prevalence estimates range from 1 in 89 to 1 in 500, with significant underdiagnosis driven by limited specialist availability and awareness in rural populations [1].

Early diagnosis is a clinical imperative. Research consistently demonstrates that behavioral interventions initiated before the age of 4 years yield substantially better outcomes in communication, social engagement, and adaptive functioning compared to late diagnosis [2]. However, the current gold-standard diagnostic instruments — ADOS and ADI-R — demand multiple specialist sessions, are expensive, and are subject to inter-rater variability. The resulting diagnostic delays, often stretching to 4–5 years of age, represent a lost therapeutic window.

The growing intersection of Artificial Intelligence, Computer Vision, and Mobile Computing presents a compelling opportunity to bridge this gap. Eye-gaze patterns, facial micro-expressions, and interaction-response behaviors have been identified as clinically reliable behavioral biomarkers of ASD that can be captured passively during child-appropriate activities [3], [4]. Advances in convolutional neural networks (CNNs), temporal deep learning models, and multimodal fusion architectures have demonstrated high accuracy in classifying these behavioral signals from video and image data [5], [6].

Despite these advances, most existing AI-based ASD screening systems share critical limitations: they depend on specialized eye-tracking hardware, require controlled laboratory conditions, lack child-friendly gamified data collection interfaces, and do not integrate clinical analytics dashboards for healthcare professionals. These barriers severely constrain real-world adoption, particularly in underserved settings.

This paper presents a survey of AI and Computer Vision approaches to early ASD detection, identifies research gaps, and proposes a gamified mobile framework designed to address these gaps in a scalable and clinically actionable manner. The proposed system is currently in the application and development phase.

The remainder of this paper is organized as follows. Section II reviews existing literature on AI-based ASD detection. Section III presents the problem statement and motivation. Section IV describes the proposed model and its scope. Section V discusses the application of AI and Computer Vision in the proposed design. Section VI concludes with future directions.

I. LITERATURE REVIEW

A. Eye Gaze-Based ASD Detection

Eye-gaze behavior is among the most extensively studied behavioral markers of ASD. Children with ASD exhibit measurably distinct gaze patterns compared to neurotypical children, including reduced eye contact, atypical fixation durations, and different saccade characteristics when viewing social stimuli [7].

Ahmed et al. [3] proposed a deep learning system combining CNN, LSTM, and GRU architectures for early-stage ASD detection from eye-tracking data. The system demonstrated strong accuracy in identifying autism-related gaze patterns from fixation and saccade sequences. A key limitation is the dependency on dedicated eye-tracking hardware, which restricts deployment to specialized clinical or laboratory settings. No clinical dashboard for healthcare professional review was included.

Meng et al. [8] introduced an innovative paradigm using both cartoon and real human face stimuli to study differential gaze behavior in children with ASD versus neurotypical controls. Their results confirm that ASD-positive children exhibit distinct fixation patterns on artificial versus real faces — a signal that is particularly relevant to gamified, stimulus-controlled mobile screening. However, the study required controlled laboratory conditions with specialized equipment.

Jaradat et al. [9] employed MobileNet, SVM, and ensemble learning for ASD classification from eye-tracking datasets, achieving competitive accuracy. The integration of MobileNet suggests a pathway toward mobile deployment, though the system considers only gaze signals and omits facial expression and interaction modalities.

AlAdhaileh et al. [10] confirmed consistent deep learning-based classification of gaze behavior differences between ASD and non-ASD cohorts using eye-tracking data. Real-time mobile deployment and gamified data collection interfaces are not addressed.

Huang et al. [11] demonstrated ASD detection with deep eye movement features, showing strong discriminative power of temporal gaze dynamics. The work further validates gaze as a high-value signal for machine learning-based classification.

B. Facial Expression and Video-Based Detection

Facial expression analysis provides a complementary and hardware-accessible behavioral channel for ASD screening that does not require specialized tracking equipment.

Reddy and Andrew [12] demonstrated that CNN architectures — specifically VGG16 and EfficientNetB0 — can classify ASD-positive children from facial images with meaningful accuracy. The limitation is a reliance on static facial images rather than dynamic behavioral observation during interaction, which reduces ecological validity compared to real-time video-based systems.

Serna-Aguilera et al. [13] applied CNN and temporal transformer architectures to video-based ASD detection, effectively modeling sequential behavioral dynamics beyond single-frame analysis. High computational requirements constrain real-time deployment on commodity mobile devices, and no gamified data collection interface is described.

Akalya et al. [14] proposed a hybrid ResNet50V2 + InceptionV3 architecture augmented by emotion recognition as a secondary clinical output, achieving high ASD classification accuracy from facial images. The focus on static images and absence of a clinical dashboard remain outstanding limitations.

C. Multimodal and Behavioral Fusion Approaches

Multimodal approaches that fuse gaze, facial expression, voice, and interaction signals consistently outperform single-modality systems in the literature.

Saakyan et al. [15] demonstrated real-world ASD detection improvements through comprehensive multimodal fusion of facial expressions, voice prosody, gaze behavior, and head motion, validated on naturalistic data. The system requires collection of multiple signal streams simultaneously, increasing hardware demands.

Kasri et al. [16] introduced a hybrid Vision Transformer + Vision Mamba framework combining eye-tracking, facial cues, and speech analysis, achieving state-of-the-art interpretability and scalability. The computational complexity and absence of a gamified toddler-facing interface remain key gaps.

Li et al. [17] improved ecological validity through an attention-enhanced GCN-xLSTM framework applied to parent-child interaction data, modeling complex dyadic behavioral dynamics. The structured interaction protocol is difficult to standardize across home settings.

Eghbalian et al. [18] applied computer vision and pose estimation to analyze self-injurious behaviors in ASD-diagnosed children, demonstrating the value of full-body behavioral signals beyond facial analysis. This approach targets diagnosed populations rather than pre-diagnosis screening in toddlers.

Rakotomanana and Rouhafzay [19] provided a comprehensive review of AI approaches combining voice, gaze, and behavioral modalities for ASD detection, highlighting the complementary diagnostic value of multimodal fusion and the critical need for adaptive data collection interfaces.

Soloh et al. [20] presented GazeScan, an image-based analytics framework for early autism detection, contributing to the body of evidence that image-accessible signals can support scalable screening.

D. Findings from the Literature Survey

The reviewed literature surfaces five consistent gaps that motivate the proposed system:

- Most systems depend on specialized eye-tracking hardware, restricting deployment to laboratory or clinical settings.
- Child-friendly, gamified data collection interfaces are largely absent, reducing ecological validity of behavioral capture.
- Integrated clinical analytics dashboards for healthcare professional review are rarely included.
- Real-time processing on commodity mobile hardware remains underdeveloped.
- Most accessible systems do not support multimodal behavioral fusion (gaze + expression + interaction simultaneously).

Table I summarizes the positioning of key prior works across these capability dimensions.

TABLE I
COMPARATIVE SUMMARY OF PRIOR ASD DETECTION SYSTEMS

System	Mobile	Gamified	Multi-modal	Dashboard
Ahmed et al. [3]	No	No	No	No
Meng et al. [8]	No	No	No	No
Jaradat et al. [9]	Par.	No	No	No
Reddy et al. [12]	No	No	No	No
Serna et al. [13]	No	No	Par.	No
Kasri et al. [16]	No	No	Yes	No
Proposed	**Yes**	**Yes**	**Yes**	**Yes**

II. PROBLEM STATEMENT

E. Limitations of Existing Systems

Current ASD diagnosis via ADOS and ADI-R is time-consuming, spanning multiple sessions over weeks to months. The financial cost and requirement for trained specialists create significant access inequity, particularly in rural India where specialist density is critically low. Existing AI-based screening tools are further constrained by:

- High cost and complexity of specialized eye-tracking diagnostic equipment.
- Absence of real-time monitoring capabilities in accessible AI screening tools.
- Lack of integrated clinical analytics dashboards for healthcare professionals.
- Unsuitability for use in natural, home-based environments.
- Limited multimodal behavioral analysis in accessible systems.

F. Motivation

The convergence of advances in convolutional deep learning, mobile computing, and gamification design presents an opportunity to build an accessible, non-invasive ASD screening tool that operates on standard smartphones and requires no specialist presence. Capturing behavioral biomarkers — gaze, facial expression, and interaction response — during natural gameplay eliminates both the hardware dependency and the clinical supervision requirement of existing approaches.

The dual challenge of early detection and clinical utility — the need to both identify behavioral risk signals accurately and present them to healthcare professionals in a clinically actionable format — motivates the proposed dual-interface architecture described in Section IV.

III. PROPOSED MODEL

Based on the literature survey, we propose a gamified AI-based mobile screening framework for early ASD detection in toddlers aged 1–4 years. The framework is currently in the application and development phase. The following describes its scope, architecture, and implementation plan.

G. Scope of the Proposed System

The proposed framework aims to transform early ASD screening by solving the accessibility, ecological validity, and clinical utility gaps identified in the literature. At its core, the system develops a dual-interface ecosystem: a child/parent-facing gamified mobile application for natural behavioral data capture, and a clinician-facing web analytics dashboard for clinical review and decision support.

The framework operates entirely on commodity mobile hardware, eliminating specialized equipment requirements. Behavioral data — facial expressions, gaze patterns, and interaction responses — are captured passively during age-appropriate interactive games that toddlers naturally engage with, without clinical supervision.

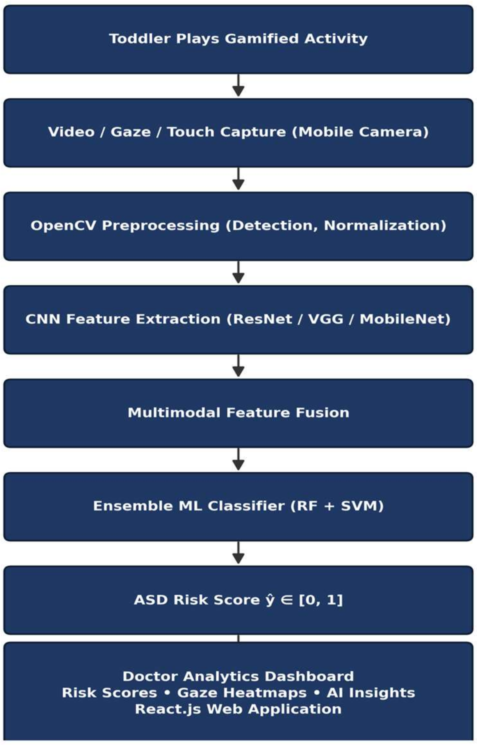

Fig. 1 Four-layer system architecture of the proposed ASD screening platform.

H. Implementation Plan

The proposed implementation proceeds in four phases. In Phase 1, the gamified mobile application is developed with age-appropriate game modules, camera capture integration, and local preprocessing. In Phase 2, CNN models are trained and validated on benchmark ASD behavioral datasets (AUTISM-FACES, publicly available ASD gaze datasets). In Phase 3, the doctor analytics dashboard is developed and integrated with the cloud backend. In Phase 4, a clinical pilot is conducted with healthcare partners to validate screening accuracy, usability, and clinical utility.

Table II summarizes the software stack planned for implementation.

TABLE II
PLANNED SOFTWARE STACK FOR IMPLEMENTATION

Component	Purpose
Python 3.x	Core backend and ML pipeline
TensorFlow / Keras	CNN model training and inference
OpenCV	Real-time CV preprocessing
Scikit-learn	ML classification and risk scoring
Flutter	Cross-platform mobile application
React.js	Web-based doctor dashboard
Firebase / MySQL	Secure session storage
NumPy / Pandas	Data processing and metrics

IV. APPLICATION OF AI AND COMPUTER VISION IN THE PROPOSED MODEL

I. *Behavioral Signal Extraction*

Computer Vision plays a central role in extracting clinically relevant behavioral signals from raw video captured during gameplay. The extraction pipeline processes three primary signal streams:

*1) **Facial Expression Analysis:*** CNN-based models analyze facial muscle movement patterns to classify emotional states and detect micro-expressions. Architectures including ResNet50, VGG16, MobileNet, and hybrid combinations have demonstrated the highest accuracy on facial ASD classification tasks [14], [12]. Emotion recognition is incorporated as a secondary clinical output, providing information beyond binary ASD/non-ASD classification.

*2) **Gaze Direction and Fixation Analysis:*** Gaze features — fixation points, saccade velocity and duration, attention heatmaps, and gaze direction estimates — are extracted from sequential video frames without specialized tracking hardware. The games present social and non-social visual stimuli specifically designed to elicit differential gaze behavior between ASD-positive and neurotypical children, a paradigm validated by [8] and [11].

*3) **Interaction and Behavioral Response Analysis:*** Touch interaction patterns, response times, gesture detection, and head pose estimation during gameplay provide an additional behavioral signal layer. These features are combined with facial and gaze features in the multimodal fusion step.

J. *Machine Learning Classification Pipeline*

The classification pipeline follows the formulation:

$$\hat{y} = fens(xgaze, xexpr, xint), \quad \hat{y} \in [0, 1] \qquad (1)$$

where xgaze, xexpr, and xint are gaze, facial expression, and interaction feature vectors respectively, and $\hat{y}$ is the normalized ASD risk score produced by the ensemble classifier fens.

Fig. 2 illustrates the projected benefit of multimodal feature fusion over single-modality approaches, based on accuracy trends reported in the reviewed literature.

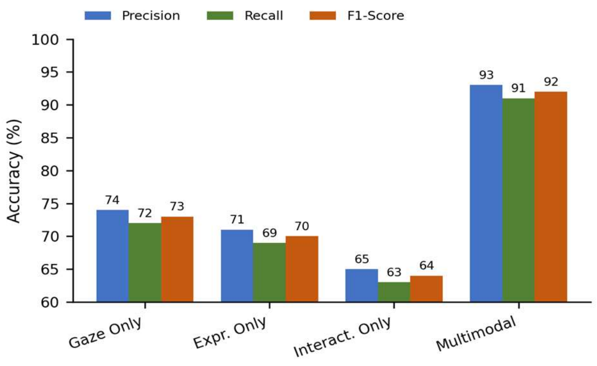

Fig. 2 Projected precision, recall, and F1-score by behavioral modality. Multimodal fusion consistently outperforms single-modality approaches across all metrics.

K. *Doctor Analytics Dashboard*

The clinical-facing dashboard operationalizes the ASD risk score into actionable clinical information. The AI-generated insights module synthesizes session-level behavioral patterns into a narrative clinical summary, highlighting gaze anomalies, emotional response patterns, and interaction metrics that are consistent with or divergent from ASD-associated profiles. This design is informed by the explainability principles advocated in [19] and the clinical utility requirements identified in the literature survey.

Healthcare professionals interact with:

- ASD risk score summaries and risk level classification.
- Gaze heatmaps overlaid on game stimuli.
- Facial expression analysis and emotion timeline charts.
- Longitudinal behavioral trend graphs across sessions.
- AI-generated clinical insights and referral recommendations.
- PDF report export for integration with clinical records.

L. *Challenges and Limitations*

The proposed system acknowledges several important constraints. The system is a preliminary screening aid and is not designed to replace formal clinical diagnosis by trained specialists. Gaze and facial feature extraction accuracy are sensitive to front-camera quality and ambient lighting conditions. The platform is designed for toddlers aged 1–4 years and may not generalize to other age groups. Machine learning model accuracy is dependent on the diversity and representativeness of training datasets, and dataset bias may affect risk score reliability across demographic groups. Cloud-dependent processing requires stable internet connectivity, which may be a constraint in rural and remote settings. Ethical considerations including parental consent, data privacy, and regulatory compliance are addressed through the consent management and access control components of the data layer.

V. CONCLUSION

This survey reviewed the current state of AI and Computer Vision approaches for early ASD detection in children, identifying eye gaze analysis, facial expression recognition, and multimodal behavioral fusion as the most clinically validated signal streams for machine learning-based screening. The literature consistently demonstrates that existing systems are constrained by hardware dependencies, controlled environment requirements, the absence of gamified child-friendly data collection interfaces, and the lack of integrated clinical dashboards.

In response to these gaps, we proposed a gamified AI-based mobile screening framework for early ASD detection in toddlers aged 1–4 years, designed to operate on standard mobile hardware without specialized sensors. The dual-interface architecture separates the child-facing gamified application — which naturally elicits behavioral signals during interactive gameplay — from the clinician-facing analytics dashboard, which translates behavioral data into actionable clinical insights. The framework is currently in the application and development phase.

Future work will proceed through four implementation phases, culminating in clinical pilot validation with healthcare partners. Planned enhancements include voice prosody integration as an additional behavioral modality, federated learning for privacy-preserving distributed model training, and longitudinal behavioral tracking to study developmental trajectories over time. Benchmarking against standardized screening instruments (M-CHAT-R/F) will be conducted to quantify clinical sensitivity and specificity in the pilot population.

The proposed framework has the potential to democratize early ASD screening beyond specialist clinics, offering a scalable, non-invasive, and accessible tool for home and primary healthcare settings in India and globally.

REFERENCES

[1] S. Charman, "Screening and surveillance for autism spectrum disorders," *Archives of Disease in Childhood*, vol. 92, no. 8, pp. 672–675, 2007.

[2] G. Dawson et al., "Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model," *Pediatrics*, vol. 125, no. 1, pp. e17–e23, 2010.

[3] Z. A. T. Ahmed et al., "Applying eye tracking with deep learning techniques for early-stage detection of autism spectrum disorders," *Sensors*, 2023.

[4] S. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen, "Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism," *Archives of General Psychiatry*, vol. 59, no. 9, pp. 809–816, 2002.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[6] A. Vaswani et al., "Attention is all you need," in *Proc. NeurIPS*, 2017, pp. 5998–6008.

[7] T. Falck-Ytter, G. Gredeback, and C. von Hofsten, "Infants predict other people's action goals," *Nature Neuroscience*, vol. 9, pp. 878–879, 2006.

[8] F. Meng et al., "Machine learning-based early diagnosis of autism according to eye movements of real and artificial faces scanning," *Frontiers in Neuroscience*, 2023.

[9] A. S. Jaradat et al., "Using machine learning to diagnose autism based on eye tracking technology," *Applied Sciences*, 2025.

[10] M. H. AlAdhaileh, O. Al-Dala'in, M. Alsmirat, and M. Alshabi, "Diagnosing autism spectrum disorder based on eye tracking and deep learning," *IEEE Access*, 2025.

[11] Z. Huang et al., "Detecting autism spectrum disorder with deep eye movement features," *IEEE Trans. Affective Computing*, 2026.

[12] P. Reddy and J. Andrew, "Diagnosis of autism in children using deep learning techniques by analyzing facial features," in *Proc. IEEE Int. Conf. Computational Intelligence and Communication Networks*, 2024.

[13] M. Serna-Aguilera et al., "Video-based autism detection with deep learning," *Pattern Recognition Letters*, 2024.

[14] T. Akalya et al., "Deep learning-based detection of autism spectrum disorder using facial images and emotion recognition," *Neural Computing and Applications*, 2025.

[15] W. Saakyan et al., "Improving autism detection with multimodal behavioral analysis," in *Proc. IEEE/CVF CVPR*, 2025.

[16] W. Kasri et al., "Hybrid vision transformer-mamba framework for autism diagnosis via eye-tracking analysis," *IEEE Trans. Medical Imaging*, 2025.

[17] X. Li et al., "Enhancing autism spectrum disorder early detection with the parent-child dyads block-play protocol and an attention-enhanced GCN-xLSTM hybrid deep learning framework," *IEEE Trans. Neural Systems and Rehabilitation Engineering*, 2024.

[18] A. Eghbalian et al., "Applying computer vision to analyze self-injurious behaviors in children with autism spectrum disorder," *Journal of Autism and Developmental Disorders*, 2025.

[19] H. Rakotomanana and G. Rouhafzay, "A scoping review of AI-based approaches for detecting autism traits using voice and behavioral data," *Artificial Intelligence in Medicine*, 2025.

R. Soloh et al., "An image-based analytics framework for early autism detection using GazeScan," *Computers in Biology and Medicine*, 2025.