iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81371

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Autisense: An Intelligent Web System for Early Detection of Autism

Amshula M R[1], Thejas R M[2], Ayyan Gowda E A[3], Vigneshwar Nagesh Murki[4], Prof. Brunda S[5]
[1, 2 ,3, 4] Student, [5]Assistant Professor, Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, India

***Abstract:** It is vital to find Autism Spectrum Disorder (ASD) early so that proper intervention may happen and developmental outcomes can get better. Traditional screening approaches depend on expert observation and standardized tests, but this takes a lot of time and is hard to get to. This research presents a multi-modal autism screening method that integrates video-based behavioral analysis with questionnaire-based assessment for children aged 16 to 48 months. MediaPipe is used to get features including face land marks, gaze, and body positions. These features are then used to look at behavioral indications like eye contact, gestures, and emotional responses. A questionnaire is included to get information from parents and help with proper risk assessment. A Retrieval-Augmented Generation (RAG)-based chatbot makes it easier to use by giving context-aware explanations, automatic PDF reports, and location-based suggestions to assist users find autism support organizations near them. The proposed framework combines several modalities without needing big labeled datasets, which makes it good for early autism screening that is both useful and respects privacy.*
***Keywords—** Autism Spectrum Disorder (ASD), Computer Vision, MediaPipe, Pose Estimation, Eye Contact Detection, Machine Learning, Retrieval-Augmented Generation (RAG), Chatbot*

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that impacts social interaction, communication, and behavior. Early identification is crucial, especially in children aged 16 to 48 months when behavioral markers become evident. However, traditional screening procedures rely on expert evaluation and standardized clinical tools, making them time-consuming, subjective, and less accessible. Recent research has investigated artificial intelligence and computer vision methodologies for automated ASD screening. Pose estimation techniques have been employed to detect repetitive behaviors, including hand movements [1], while gaze analysis and joint attention are acknowledged as primary indicators [2]. Deep learning techniques have yielded encouraging outcomes in behavioral classification; however, they require extensive annotated datasets and substantial computational resources [3]. Markerless motion analysis methods have also demonstrated the effectiveness of video-based behavioral assessment [4]. Despite these advancements, most current approaches focus on a single type of behavior and lack integration into a unified system, limiting their applicability in real-world scenarios. To address these limitations, this paper presents a multi modal autism screening approach that combines video-based behavioral analysis with questionnaire-based evaluation. MediaPipe is used for real-time feature extraction, and a structured questionnaire improves screening reliability. The system also includes a Retrieval-Augmented Generation (RAG)-based chatbot, automated report generation, and location-based recommendations for nearby autism support centers. This provides a practical and privacy-aware solution for early autism screening.

II. LITERATURE SURVEY

Recent advancements in artificial intelligence and computer vision have significantly contributed to the early detection of Autism Spectrum Disorder (ASD). Several studies have explored video-based behavioral analysis as an effective approach for identifying autism-related patterns. Alessandro Zunino et al. proposed a deep learning-based approach for autism detection using video gesture analysis
[1]. Their method utilized temporal modeling of motion patterns to distinguish between autistic and typically developing children, demonstrating that even simple gestures can provide meaningful diagnostic cues. S. Tariq et al. developed an AI-based classification system using convolutional neural networks (CNN) and long short term memory (LSTM) models for autism detection from video sequences

[2]. Their approach effectively captured both spatial and temporal behavioral features, achieving improved accuracy. However, it requires large annotated datasets and significant computational resources. H. Jiang et al. emphasized the importance of joint attention and gaze behavior as key indicators for early autism detection

[3]. Their work analyzed gaze shifts and attention patterns between objects and individuals, highlighting the role of eye contact in identifying ASD. However, the approach is limited to a single behavioral modality. Mackenzie Mathis et al. introduced DeepLabCut, a marker less pose estimation framework used to analyze motor patterns such as movement coordination and periodicity in infants

[4]. While this method provides detailed motion analysis, it is computationally intensive and not suitable for real-time applications. Dennis P. Wall et al. proposed a machine learning-based approach that combines questionnaire data with home video analysis for early autism screening

[5]. Their system improves screening accuracy by integrating multiple data sources; however, it relies on manual video annotation and lacks full automation.

A. Gap Analysis

Despite the significant progress in ASD detection, several limitations remain in existing approaches:

- Single-modality limitation: Most existing works focus on individual aspects such as gesture, gaze, or questionnaire-based screening, rather than integrating multiple behavioral indicators.
- High dependency on large datasets: Deep learning models require extensive labeled datasets, which are difficult to obtain due to privacy concerns related to children's data.
- Lack of real-time and scalable solutions: Many approaches are computationally intensive or designed for controlled environments, limiting their real-world applicability.
- Limited automation: Some systems rely on manual annotation or expert involvement, reducing usability and scalability.
- Absence of intelligent user support: Existing methods primarily focus on detection and do not provide interactive guidance or explanation for users.
- Lack of practical assistance features: Existing systems do not provide actionable outputs such as automated report generation or recommendations for nearby health care and support centers, limiting their usefulness for caregivers.

III. SYSTEM ARCHITECTURE

The proposed system adopts a multi-layered architecture to ensure modularity, scalability, and efficient interaction among components. The architecture consists of five layers: User Interface Layer, Application Layer, Processing Layer, Intelligence Layer, and Data Layer, as illustrated in Fig. 1.

A. User Interface Layer

The User Interface Layer enables interaction between users and the system. It is implemented using React and supports functionalities such as user registration, login, child profile management, questionnaire input, video upload, and result visualization. It also allows report viewing and interaction with. The interface is designed to be responsive and user-friendly, ensuring smooth and efficient interaction.

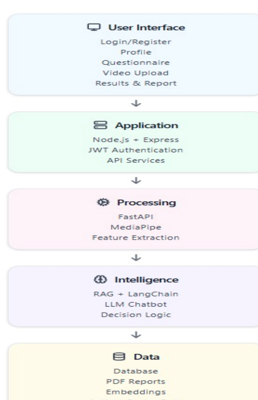

Fig. 1. Proposed Multi-Layer System Architecture

B. Application Layer

The Application Layer manages core business logic and API communication. It is developed using Node.js and Express.js. This layer handles authentication using JSON Web Tokens (JWT), processes incoming user requests, and coordinates communication between frontend components and backend services, including machine learning modules, report genera tion services, chatbot services, and location-based recommendation services.

C. Processing Layer

The Processing Layer is responsible for data processing and feature extraction. It is implemented using FastAPI to enable efficient interaction with machine learning services. Video data uploaded by users is processed using MediaPipe and OpenCV to extract facial landmarks, eye contact patterns, body pose features, and behavioural indicators. Questionnaire responses are also processed to compute intermediate risk scores for further analysis.

D. Intelligence Layer

The Intelligence Layer performs advanced analysis and inference. It incorporates a Retrieval-Augmented Generation (RAG) module implemented using LangChain for orchestration and ChromaDB for embedding storage. A language model generates context-aware explanations, recommendations, and chatbot responses. This layer integrates outputs from video analysis and questionnaire processing to provide interpretable outputs and assist in generating detailed screening reports.

E. Data Layer

The Data Layer manages system data, including user in formation, questionnaire responses, processed video features, generated PDF reports, embeddings, and autism center data. Structured data is stored in a database, while embeddings used in the RAG module are maintained in ChromaDB for efficient retrieval. Additionally, this layer supports location based processing to identify nearby autism support centers and provide relevant recommendations based on user location.

IV. METHODOLOGY (PROPOSED SYSTEM)

The proposed system uses a multi-modal and microservices based approach to early screening of Autism Spectrum Dis order (ASD) by combining video-based behavioral analysis, questionnaire-based assessment, and intelligent response generation. The methodology emphasizes objective behavioral observation, starting with video data analysis, followed by questionnaire-based validation and machine learning-based risk estimation.

A. Data Acquisition

The first step in the screening process is for the user to upload a pre-recorded video of the child's behavior. This ensures that natural and unbiased behavior patterns are captured. After that, the user fills out a structured 20-question screening questionnaire. Both inputs are securely sent to backend services through REST APIs.

B. Video-Based Multi-Feature Extraction

The uploaded video is processed frame by frame using a computer vision pipeline built with MediaPipe and OpenCV. The system extracts six key behavioral features:

- Eye contact detection
- Hand gesture detection
- Head stimming detection
- Hand stimming detection
- Social reciprocity assessment

Each feature is calculated using temporal analysis across frames, enabling reliable detection of behavioral patterns. Public datasets such as SSBD are used as references; however, large-scale autism datasets are limited due to privacy concerns

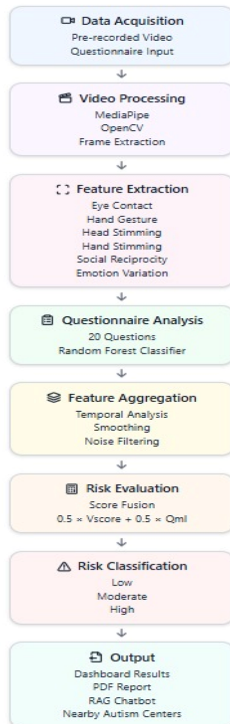

Fig. 2. Workflow of the Proposed System

C. Questionnaire-Based Machine Learning Assessment

After analyzing the video, the system gathers responses to a 20-question screening questionnaire designed to evaluate social communication and behavior. A Random Forest classifier predicts autism risk based on response patterns.

D. Behavioural Analysis and Feature Aggregation

The extracted video features are analyzed to identify indicators of ASD. Each feature contributes a normalized score, and temporal aggregation ensures stability. Confidence filtering is applied to reduce noise.

E. Risk Evaluation and Multi-Modal Fusion

The final risk score is computed as:

$$Risk = 0.5 \times Vscore + 0.5 \times Qml \quad (1)$$

where *Vscore* is the video analysis score and *Qml* is the questionnaire-based score. The results are categorized as:

- Low Risk: $< 40\%$
- Moderate Risk: 40%–70%
- High Risk: $\geq 70\%$

F. Intelligent Response Generation

A Retrieval-Augmented Generation (RAG) module uses a language model to provide context-aware explanations. It also allows comparison between past and current screening results.

G. Report Generation and Support Services

The system generates a structured PDF report and recommends nearby autism support centers based on user location, facilitating access to professional support.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

H. Ethical Considerations

Due to privacy concerns in autism datasets, the system uses lightweight feature extraction instead of data-intensive models, ensuring ethical and practical deployment

V. IMPLEMENTATION

The proposed autism screening system uses a microservices based architecture that integrates web technologies, machine learning services, and retrieval-based intelligent systems. The system enables efficient real-time processing of multiple inputs, such as video and questionnaire data, through components communicating via RESTful APIs.

A. Frontend

The user interface is built using React with a component based architecture. It supports user login, child profile management, video upload, questionnaire input, and result visualization. The MediaRecorder API is used for capturing video input, and Axios is used for API communication. React hooks manage state and navigation, while Tailwind CSS ensures a responsive design.

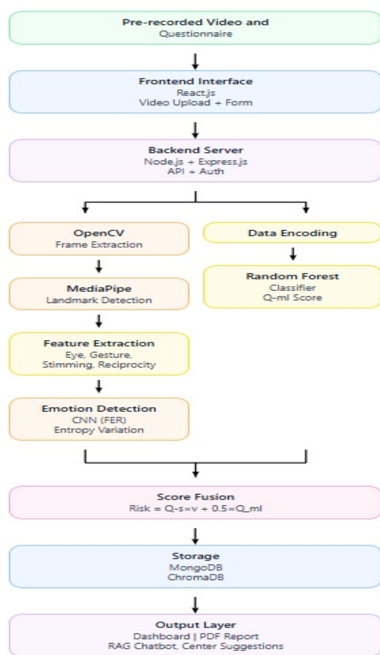

Fig. 3. Implementation Workflow of the Proposed System

B. Backend

The backend is implemented using Node.js and Express.js, acting as the central orchestration layer. It handles authentication using JSON Web Tokens (JWT), processes API requests, manages video uploads using Multer, and enables communication with machine learning and RAG services. It also computes risk scores and generates reports using PDFKit.

C. Video Analysis and Feature Extraction

A FastAPI-based service processes video data using OpenCV and MediaPipe. Facial, hand, and pose landmarks are extracted to compute behavioral features such as eye contact, gestures, and stimming patterns. Temporal smoothing and confidence filtering are applied to improve stability and reduce noise.

D. Emotion Detection

A CNN model implemented in PyTorch and trained on the FER dataset is used to analyze emotion variation. Frame-level predictions are aggregated, and entropy-based analysis is used to measure changes in emotions over time

E. Questionnaire Model

A Random Forest classifier implemented using Scikit-learn processes questionnaire responses. The model learns non linear relationships between behavioral indicators and ASD likelihood to generate a probabilistic risk score.

F. RAG-Based Chatbot

The system includes a Retrieval-Augmented Generation (RAG) module using LangChain and ChromaDB. Text embeddings are generated from reports, and a language model provides context-aware explanations and guidance.

G. Integration and Storage MongoDB

is used to store user data, screening results, and chat history. Video files and reports are stored locally, while embeddings are managed in ChromaDB. REST APIs enable communication between all components, supporting scalable and asynchronous processing.

H. Workflow

The system workflow begins with video upload and questionnaire input. The video is processed for feature extraction and emotion analysis, while questionnaire responses are evaluated using the Random Forest model. The outputs are combined to compute the final risk level, which is stored and displayed through the interface. The RAG module provides additional explanations and recommendations to assist users in decision-making.

VI. RESULTS AND DISCUSSION

The proposed system was evaluated using both questionnaire-based classification and video-based behavioural analysis. The results demonstrate the effectiveness of the multi-modal approach for early autism risk screening.

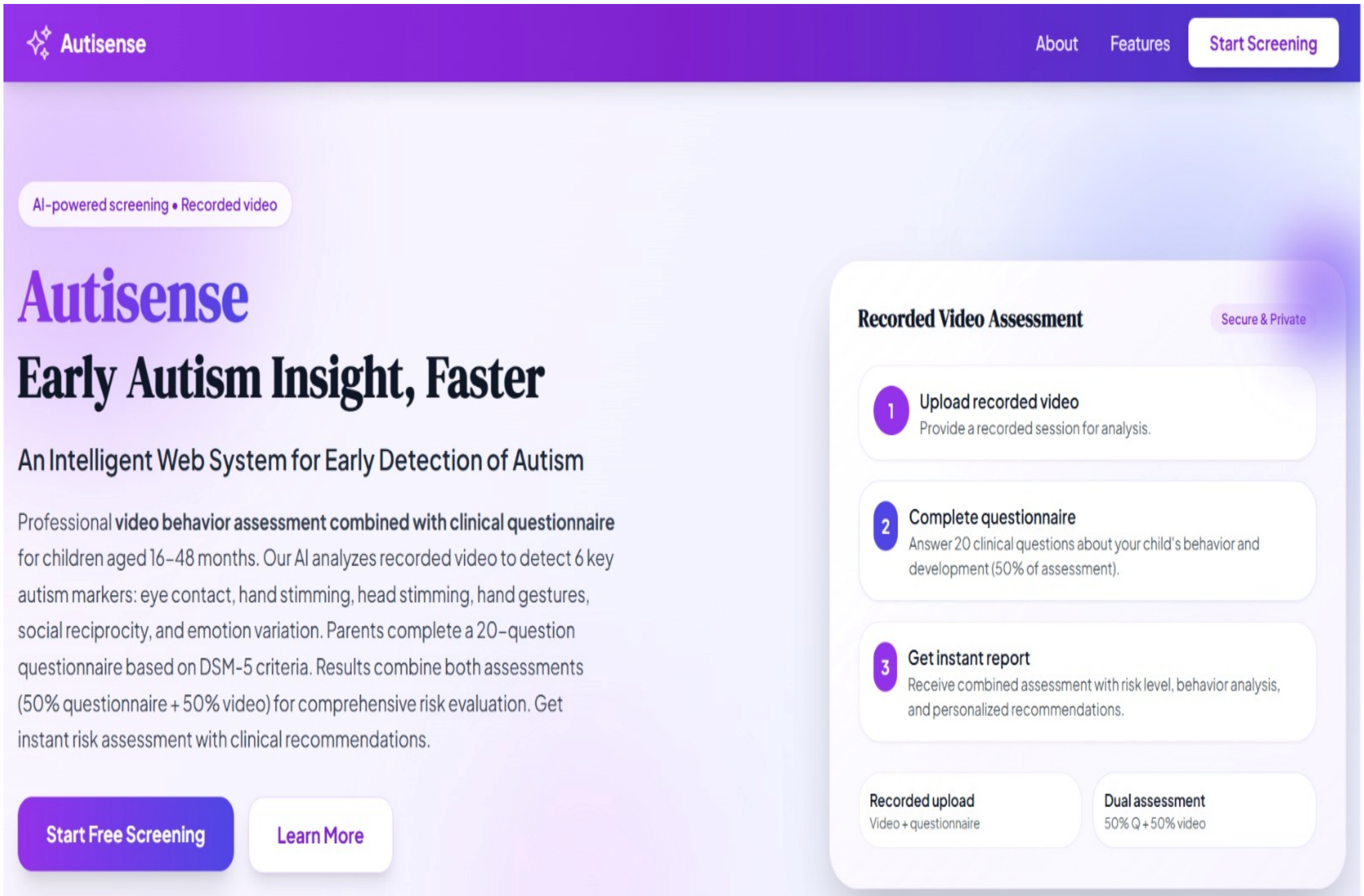

Fig. 5. Home Page Interface

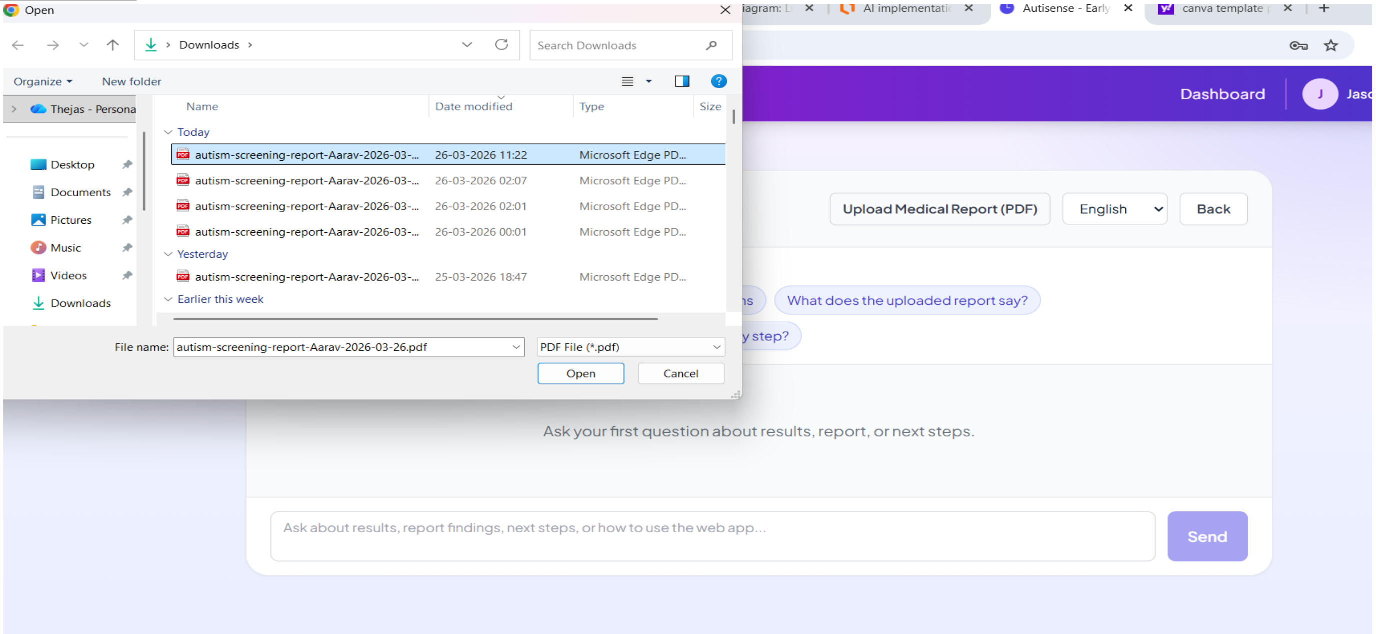

Fig 6. Video Upload Interface

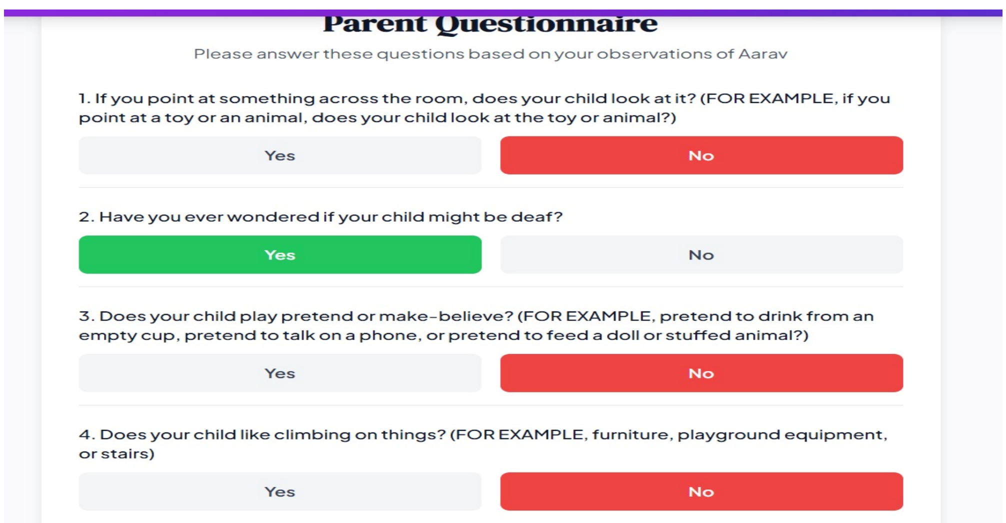

Fig. 7. Questionnaire Module

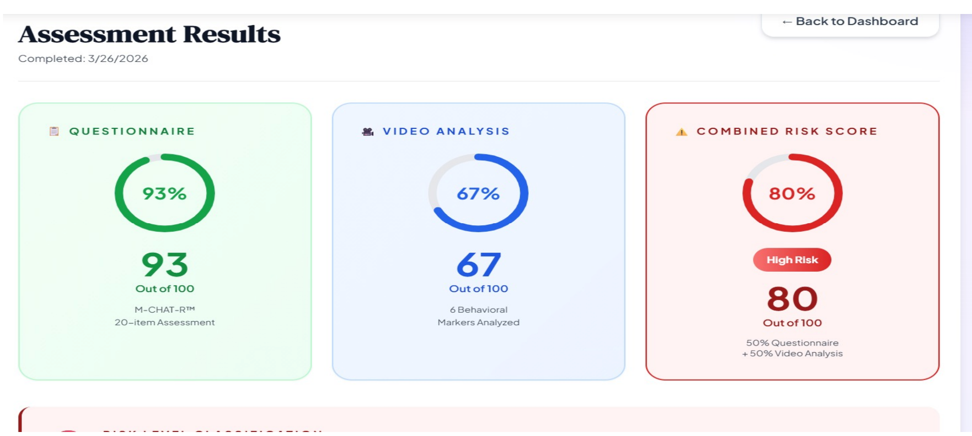

Fig. 8. Final Risk Assessment Dashboard

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

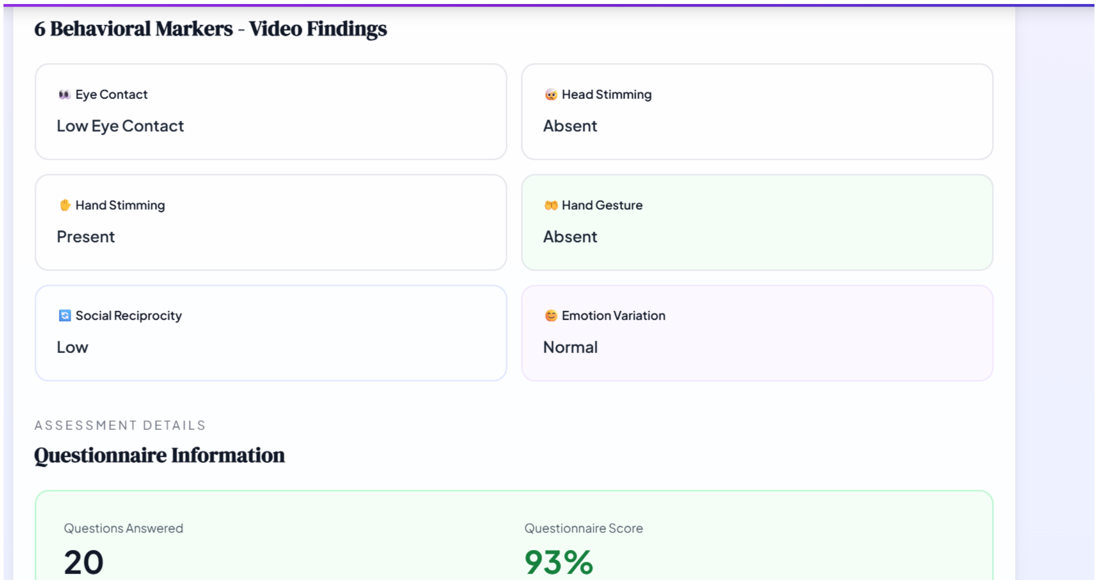

Fig. 9. Behavioural Feature Output

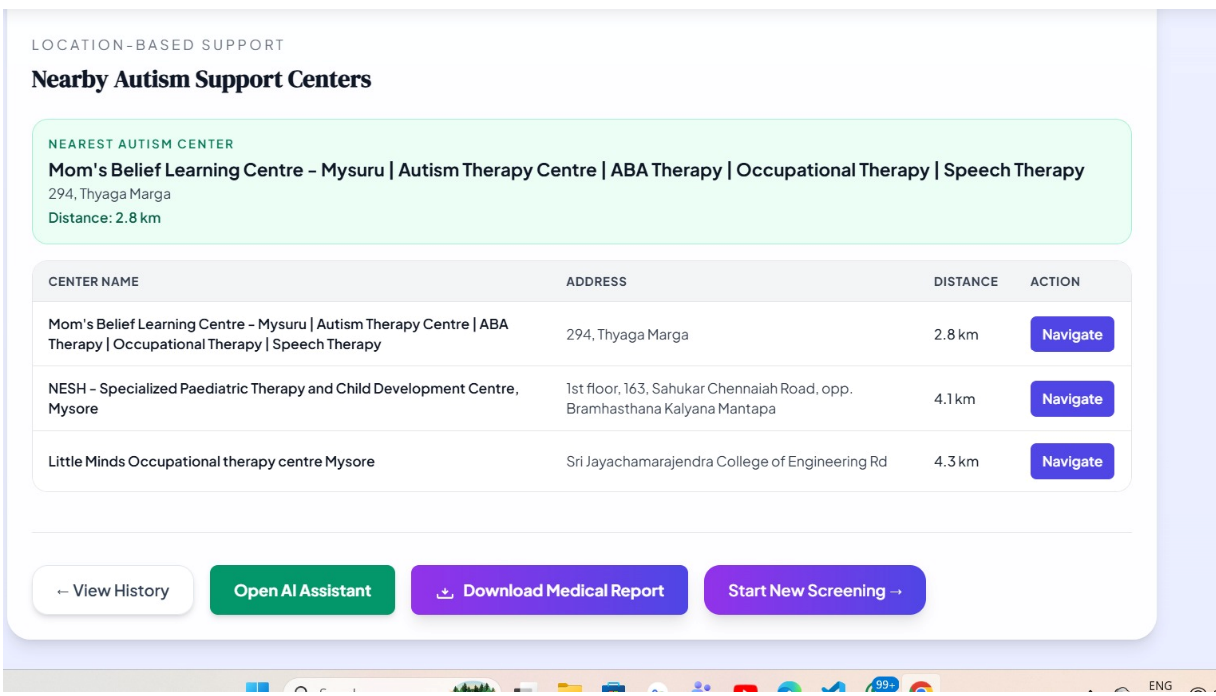

Fig. 10. Nearby Autism Center Recommendations

A. Questionnaire Model Performance

The Random Forest classifier achieved an overall accuracy of 97%. The model demonstrated strong performance with precision of 0.97 for both classes, recall of 0.99 for non ASD and 0.93 for ASD cases, and F1-scores of 0.98 and 0.95 respectively. These results confirm the reliability of questionnaire-based screening in capturing behavioural indicators.

B. Video Behavioural Analysis

The video processing module extracted key behavioural features including eye contact, hand gestures, head stimming, hand stimming, social reciprocity, and emotion variation. These features were computed using MediaPipe and OpenCV with temporal aggregation, enabling consistent and interpretable outputs.

Fig. 4 shows a sample frame from the input video used for behavioural analysis. The frame illustrates observable hand movement and facial expression patterns, which are analysed by the system to detect behavioural indicators associated with ASD.

C. Multi-Modal Risk Evaluation

The final risk score was computed by combining questionnaire and video analysis scores using equal weighting. This fusion improves robustness by integrating both subjective (questionnaire) and objective (video) inputs. The system classifies results into low, moderate, and high-risk levels for easy interpretation.

D. System Interface and Outputs

The system provides an interactive and user-friendly interface .The interface enables seamless interaction for video upload, questionnaire input, and result visualization. The outputs are designed to be clear, interpretable, and useful for caregivers.

E. Discussion

The results demonstrate that the questionnaire model provides high accuracy, while video analysis enhances behavioural understanding. The multi-modal integration improves overall system reliability and reduces dependence on a single data source. The system is lightweight and avoids heavy deep learning requirements by leveraging feature-based analysis, making it suitable for real-time and scalable deployment. The inclusion of explainable outputs and autism center recommendations improves usability and practical relevance for caregivers.

F. Limitations

Despite promising results, the proposed system has several limitations:

- The performance of the video analysis module is dependent on input video quality, including lighting conditions, camera angle, occlusions, and motion clarity.
- Limited availability of publicly accessible autism-related video datasets restricts large-scale validation and model generalisation.
- The emotion detection module is based on a general FER dataset and may not fully capture child-specific behavioural expressions.
- The equal weighting strategy used for multi-modal fusion may not be optimal for all cases and could be improved with adaptive weighting.
- The system functions as a screening tool and does not replace professional clinical diagnosis.
- Real-world deployment may require further validation with medical experts and integration with healthcare systems.

VII. CONCLUSION

This paper presented a multi-modal autism screening system that combines video-based behavioural analysis and questionnaire-based assessment to enable early detection of Autism Spectrum Disorder (ASD). The system utilises Medi aPipe for real-time feature extraction, a Random Forest model for questionnaire-based prediction, and a fusion strategy to generate reliable risk scores without relying on large annotated datasets. The integration of multiple components, including emotion analysis, automated PDF report generation, and a Retrieval Augmented Generation (RAG)-based chatbot, enhances both accuracy and interpretability. Furthermore, the inclusion of a location-based recommendation module for suggesting nearby autism support centers improves the practical applicability of the system by guiding caregivers toward professional assis tance. Overall, the proposed system provides a scalable, privacy aware, and user-friendly solution for early autism screening, making it suitable for real-world deployment and assisting in timely intervention

VIII. FUTURE WORK

Future enhancements of the proposed system can further improve its performance and usability:

- Incorporating larger and more diverse datasets to improve model accuracy and generalisation.
- Extending behavioural analysis using advanced deep learning models for more precise feature extraction.
- Enabling real-time video processing for immediate screening feedback.
- Developing a mobile application to improve accessibility and usability for a wider audience.
- Enhancing the recommendation module with personalized suggestions based on user history and clinical data.
- Integrating with healthcare systems for clinical validation and professional support

IX. ACKNOWLEDGMENT

The authors want to dedicate this work by presenting thanks to the people and bodies below for providing great encouragement and contributions during this study:

Project Mentor: Brunda S is kindly acknowledged as she guided them constantly by giving good feedback on her part and encourage her with various aspects throughout this course of work. We extend our gratitude to JSS Science and Technology University, with special thanks to the Computer Science Department, for giving us the necessary resources and infrastructure to conduct this research. Thank you, faculty, and staff, for your support.

REFERENCES

[1] A. Zunino, J. Cavazza, and V. Murino, "Video gesture analysis for autism spectrum disorder detection," in Proc. IEEE Int. Conf. Comput. Vis. Workshops, 2017, pp. 1–9.

[2] S. Tariq, S. Shah, M. U. Akram, and A. Khan, "Detection of autism spectrum disorder using deep learning techniques on video data," IEEE Access, vol. 7, pp. 148331–148342, 2019.

[3] H. Jiang, J. Zhou, and Y. Liu, "Gaze behavior analysis for early autism detection using computer vision," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 28, no. 2, pp. 1–10, 2020.

[4] M. Mathis, A. Mamidanna, K. M. Cury, et al., "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning," Nature Neuroscience, vol. 21, pp. 1281–1289, 2018.

[5] D. P. Wall, J. Kosmicki, T. F. DeLuca, et al., "Use of machine learn ing to shorten observation-based screening and diagnosis of autism," Translational Psychiatry, vol. 2, no. 4, pp. e100, 2012.

[6] E. A. Di Martino, C. Yan, Q. Li, et al., "The autism brain imaging data exchange: towards a larg-scale evaluation of intrinsic brain architecture in autism," Molecular Psychiatry, vol. 19, pp. 659–667, 2014.

[7] N. D. Lane, S. Bhattacharya, P. Georgiev, et al., "DeepEar: robust smartphone audio sensing in unconstrained acoustic environments using deep learning," in Proc. ACM Int. Joint Conf. Pervasive and Ubiquitous Computing, 2015, pp. 283–294.

[8] Y. Wang, X. Li, and Q. Ji, "Early detection of autism using facial expression analysis and machine learning," in Proc. IEEE Int. Conf. Image Processing, 2018, pp. 1–5.

[9] A. Voss, K. Washington, J. Kline, et al., "Automated autism screening using mobile-based facial expression analysis," Sci. Transl. Med., vol. 11, no. 491, 2019.

[10] R. Rudovic, J. Lee, M. Dai, et al., "Personalized machine learning for robot-assisted autism therapy: a preliminary study," IEEE Trans. Human Machine Systems, vol. 48, no. 1, pp. 1–12, 201