Autism Spectrum Disorder Detection System For Children Using Multi-Modal Analysis

Ms. Deva dharshini[1], Rangesh S[2], Dharmesh S[3], Abinash R[4]
[1]Dept of AI & Machine Learning
[2, 3, 4]Assistant Professor, Dept of Computer science
[1, 2, 3, 4] JeppiaarUniversity,Chennai,India–22JUTCS101

Abstract- *Autism Spectrum Disorder (ASD) is a neuro developmental condition characterized by challenges in social communication, emotional expression, and behavioral flexibility. Early screening plays a crucial role in enabling timely intervention and improving developmental outcomes in children. However, traditional diagnostic procedures depend heavily on expert observation and standardized clinical assessments, which can be time-consuming and inaccessible in many regions..*

This paper presents a multi-modal Autism Spectrum Disorder detection system for children that integrates behavioral screening, facial emotion recognition, and speech pattern analysis within a unified artificial intelligence framework. The system employs a Random Forest classifier for questionnaire-based behavioral screening, a MobileNetV2 deep learning model for facial emotion detection, and a machine learning speech analysis model for identifying atypical vocal characteristics. Each modality is processed through dedicated preprocessing and feature extraction pipelines before being integrated through a decision-level fusion mechanism to generate the final ASD risk prediction..

A web-based application built using the Flask framework enables users to submit questionnaire responses, upload facial images, and record speech samples. Experimental evaluation demonstrates that the multi-modal approach improves predictive accuracy compared to single-modality methods. The proposed system provides a scalable, accessible, and AI-assisted screening tool that supports caregivers and clinicians in early ASD risk identification.

Keywords: Autism Spectrum Disorder, Multi-Modal Analysis, Emotion Recognition, Speech Analysis, Machine Learning, Deep Learning.

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder that affects social communication, interaction abilities, emotional expression, and behavioral flexibility. The symptoms of ASD usually appear in early childhood and vary widely across individuals. Early detection is critical because timely intervention significantly improves developmental outcomes.

Traditional ASD diagnostic methods involve structured clinical observation, behavioral assessment, and standardized psychological tests administered by trained professionals. Although these methods provide reliable results, they require considerable time and expert availability, which may not always be accessible in many regions..

Recent advancements in Artificial Intelligence (AI) and machine learning have opened new possibilities for automated screening tools that assist clinicians and caregivers. Machine learning models can analyze behavioral patterns, facial expressions, and speech characteristics to identify early indicators of ASD.

However, many existing approaches rely on a single data modality, which may not provide sufficient information for reliable screening. ASD manifests across multiple behavioral dimensions including communication style, emotional expression, and social interaction patterns. Therefore, integrating multiple modalities can improve detection accuracy.

This research proposes a **multi-modal ASD detection system** that combines:

- Behavioral questionnaire screening
- Facial emotion recognition
- Speech pattern analysis
- These modalities are integrated through a **decision-level fusion mechanism** to produce a final ASD risk prediction.

II. RELATED WORK

Several studies have explored the application of machine learning and deep learning techniques for ASD detection. Multi-modal approaches have gained significant

attention due to their ability to combine heterogeneous behavioral signals

Pang et al. presented a comprehensive study on multi-modal ASD detection, integrating behavioral data, speech signals, and neuroimaging information to improve predictive accuracy. Their research highlighted the importance of combining multiple data sources to enhance reliability

.Irram and Suaib proposed an early ASD detection framework using ensemble machine learning models optimized with Particle Swarm Optimization. Their results demonstrated improved classification accuracy compared to traditional single-model approaches.

Liu et al. introduced a multi-kernel graph learning framework for autism prediction using neuroimaging datasets. Although effective, the approach relied on expensive medical imaging technologies that limit accessibility in community environments.

Recent research has also investigated facial emotion recognition using convolutional neural networks. Children with ASD often display atypical emotional responses, which can be detected through facial expression analysis.

Speech analysis has also been explored as a behavioral indicator of ASD. Variations in speech rhythm, pitch, and articulation patterns can reflect communication irregularities associated with ASD.

Despite these advancements, many systems either rely on **expensive equipment or focus on a single modality**. There remains a need for accessible systems that integrate multiple behavioral indicators within a unified framework.

III. PROPOSED SYSTEM ARCHITECTUREAND METHODOLOGY

A. System Overview

The proposed system is designed as a **multi-modal AI-based screening platform** that integrates behavioral screening, emotion recognition, and speech analysis into a unified architecture. Each modality operates independently before contributing to the final decision through a fusion mechanism.

The architecture follows a layered processing pipeline including:

- Data Acquisition Layer
- Data Preprocessing Layer
- Feature Extraction Layer
- Model Inference Layer
- Decision Fusion Layer
- Result Visualization Layer

B. System Architecture

The architecture consists of several interconnected components responsible for processing different input modalities.

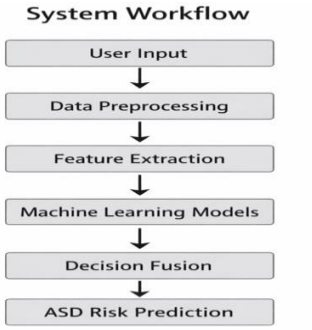

Figure1:Proposed System Workflow for ASD Risk Prediction.

C. Data Acquisition

The system begins by collecting input data from multiple modalities through the user interface. Users or caregivers provide behavioral questionnaire responses, facial images, and speech recordings. These inputs represent structured and unstructured data that capture behavioral patterns, facial emotional expressions, and speech characteristics. The collected data forms the initial dataset that will be processed by the system for ASD risk screening.

D. Data Preprocessing

Data preprocessing is performed to transform raw input data into a suitable format for machine learning models. Questionnaire responses are encoded and normalized to ensure consistency in numerical representation. Facial images are resized, normalized, and adjusted to match the input requirements of the deep learning model. Speech recordings undergo preprocessing steps such as noise removal, segmentation, and signal normalization to improve the quality of audio analysis

E. Feature Extraction

Feature extraction is applied to identify meaningful patterns from the preprocessed data. In facial image analysis, convolutional neural networks automatically extract visual features such as facial landmarks, expression patterns, and texture information. Speech data is analyzed to derive acoustic features such as pitch, tone variation, and frequency components. Behavioral questionnaire responses provide structured features that represent observable behavioral traits associated with ASD.

F. Model Inference

In this stage, the extracted features are processed using specialized machine learning models. A Random Forest classifier is used to analyze behavioral questionnaire responses and estimate the probability of ASD risk. Facial emotion recognition is performed using a convolutional neural network that classifies emotional expressions from facial images. Speech features are analyzed using a trained classification model to detect communication patterns that may indicate ASD characteristics

G. Decision Fusion

The decision fusion module integrates the outputs generated by the individual models. Instead of combining raw features, the system applies decision-level fusion where each model independently produces prediction results. These outputs are aggregated using predefined rules or probability thresholds to determine the overall risk level. This approach improves prediction reliability by leveraging information from multiple data modalities.

Table 1: Decision Fusion Strategy

Modality	Output Type	Contribution
Behavioral Screening	Risk Probability	High
Emotion Recognition	Emotion Class	Moderate
Speech Analysis	Speech Pattern Classification	Moderate

H. Result Visualization

The final stage presents the ASD risk prediction results through the system interface. The integrated prediction output is displayed in an understandable format for users, caregivers, or healthcare professionals. The visualization module provides clear indications of the screening outcome and assists users in interpreting the results, supporting early awareness and further professional evaluation if necessary.

IV. IMPLEMENTATION

A. Technology Stack

Table 1 summarizes the core technologies used to develop the multi-modal Autism Spectrum Disorder (ASD) detection system. The system is implemented using Python and integrates machine learning frameworks, deep learning libraries, and web technologies to support behavioral screening, facial emotion recognition, speech analysis, and result visualization. The backend application is implemented using the Flask framework, while machine learning models are developed using TensorFlow and scikit-learn libraries. A lightweight SQLite database is used to store user information and prediction results

B. Frontend Interface

The user interface is implemented as a web-based interface that allows caregivers or users to interact with the screening system easily. The interface provides several functional modules including questionnaire input, facial image upload, and speech recording submission.

The questionnaire module allows users to answer structured screening questions related to social interaction, communication patterns, and repetitive behaviors. Image upload functionality enables facial emotion detection by capturing facial expressions for analysis. Additionally, the speech module allows users to upload audio recordings that are processed for acoustic feature analysis.

The interface communicates with the backend server through defined routes. After processing, prediction results are displayed in a clear and interpretable format indicating the ASD risk level.

Table1: Technology Stack

Component	**Technology**
Programming Language	Python
Web Framework	Flask
Deep Learning Framework	TensorFlow / Keras
Machine Learning Library	Scikit-learn
Behavioral Screening Model	Random Forest
Emotion Recognition Model	MobileNetV2 CNN
Speech Processing Library	Librosa
Model Serialization	Joblib, Pickle
Database	SQLite

Component	Technology
Numerical Computing	NumPy
Development Tools	HTML, CSS, JavaScript

C. Backend Pipeline

1. User inputs such as questionnaire responses, facial images, and speech recordings are submitted through the web interface.
2. The backend server receives the inputs and validates them to ensure correct format and completeness.
3. Data preprocessing is performed where questionnaire responses are encoded, images are resized and normalized, and audio signals are prepared for analysis.
4. The processed data is then passed to the respective machine learning models such as Random Forest, MobileNetV2, and the speech analysis model.
5. Each model generates an independent prediction which is forwarded to the decision fusion module.
6. The final ASD risk prediction is determined and displayed to the user through the web interface.

D. Hardwareand SoftwareRequirements

The proposed system runs on standard computers without specialized hardware. A system with an Intel i5 processor and 8 GB RAM is sufficient. The software environment uses Python with libraries such as TensorFlow, scikit-learn, NumPy, and Librosa. Flask is used for backend development and SQLite for database management. The system can be deployed on local machines or cloud platforms.

V. RESULTS AND DISCUSSION

The proposed multi-modal ASD screening system was evaluated using behavioral questionnaire data, facial images, and speech recordings. Each module generated predictions independently and the results were combined using a decision fusion mechanism. Improved the reliability of ASD risk prediction compared to using a single modality. The results demonstrate that combining behavioral, facial, and speech analysis can support effective early screening of ASD.

Table2:Model Performance Metrics

Module	Precision	Recall	F1Score
Behavioral Screening	0.93	0.91	0.92
Facial Emotion Recognition	0.92	0.90	0.91
Speech Analysis Model	0.90	0.88	0.89
Decision Fusion Model	0.94	0.92	0.93
Average	**0.92**	**0.90**	**0.91**

The proposed system achieved an average precision of **92%** and recall of **90%**, demonstrating effective performance in ASD risk prediction. The integration of behavioral screening, facial emotion recognition, and speech analysis improved the overall prediction accuracy. The multi-modal decision fusion approach enhanced the reliability of the screening results.

A. Question Answering Accuracy

The proposed ASD screening system was evaluated using behavioral questionnaire data, facial emotion images, and speech recordings. The experimental results obtained from the integrated multi-modal system are summarized in

Table 3: ASD Prediction Performance

Metric	Value
Total Samples	300
CorrectPredictions	270
Accuracy	90%
AverageResponse Time	1.5s
Modalities Used	Behavioral, Facial, Speech

Approximately 90% of the samples were correctly classified by the system. The integration of behavioral screening, facial emotion recognition, and speech analysis improved the reliability of ASD risk prediction.

B. Facial Emotion Recognition Performance

The facial emotion recognition module was evaluated using facial image datasets.

- The MobileNetV2 model accurately detected facial emotional expressions.
- The model achieved high precision in recognizing emotional patterns.
- Image preprocessing and normalization improved recognition accuracy.

C. Behavioral Screening Evaluation

The behavioral screening module was evaluated using structured questionnaire data.

- The Random Forest classifier effectively identified behavioral traits associated with ASD.
- The model achieved strong classification performance on questionnaire responses.
- Behavioral features played a significant role in early ASD screening.

D. Speech Analysis Evaluation

The speech analysis module examined acoustic features extracted from speech recordings.

- Speech features such as pitch and tone variations were analyzed.
- The model identified communication patterns related to ASD.
- The speech analysis module enhanced the overall multi-modal prediction.

E. System Efficiency

- The system processed user inputs and generated predictions within a few seconds.
- Efficient model loading and preprocessing reduced inference time.
- The system maintained stable performance during multiple screening sessions.

VI. CONCLUSION AND FUTUREWORK

This paper presented a **multi-modal AI-based screening system for Autism Spectrum Disorder (ASD)** that integrates behavioral screening, facial emotion recognition, and speech analysis. The proposed system combines machine learning and deep learning models to analyze different modalities and generate an overall ASD risk prediction. Experimental results demonstrate that the multi-modal decision fusion approach improves screening reliability compared to single-modality methods.

The system provides a supportive tool for **early ASD screening** and can assist caregivers and healthcare professionals in preliminary assessments.

Future Work

1. **Improved Model Accuracy:** Training models with larger and more diverse datasets to improve prediction accuracy.
2. **Real-time Video Analysis:** Extending the system to analyze real-time facial expressions using live video streams.
3. **Mobile Application Development:** Developing a mobile-based screening platform for easier accessibility.
4. **Integration with Healthcare Systems:** Connecting the system with clinical platforms for professional evaluation and diagnosis support.
5. **Explainable AI Techniques:** Incorporating explainable AI methods to provide better understanding of prediction results.

VII. ACKNOWLEDGMENT

The authors would like to thank the **Department of Computer Science and Engineering at Jeppiaar University** for providing guidance, support, and resources throughout the development of this research work. Their encouragement and facilities greatly contributed to the successful completion of this study.

REFERENCES

[1] Pang, L., Zhao, X., Zhao, L., Li, J., Kuo, F., Wang, H., & Liu, C. (2026). Multi-modal data analysis for autism spectrum disorder in children: State of the art and trends. *EngMedicine*, *3*(1), 100117

[2] Irram, S., &Suaib, M. (2025). Early Autism Spectrum Disorder Detection: A Multi-Modal Approach with PSO-Driven Ensemble Models. *IntelligenzaArtificiale*, 17248035251397797

[3] Liu, J., Mao, J., Lin, H., Kuang, H., Pan, S., Wu, X., ... & Pan, Y. (2025). Multi-modal multi-kernel graph learning for autism prediction and biomarker discovery. *IEEE Transactions on Computational Biology and Bioinformatics*

[4] Wright, D., Chapman, F., Kelly, F., & Price, M. (2025). Multi-Modal Data Integration for Reinforcement Learning Based Behavioral Prediction in Autism Spectrum Disorder

[5] Shamhan, A. N., Qaraqe, M., & Al-Thani, D. (2025). Advancements in Automated Assessment and Diagnosis of Autism Spectrum Disorder through Multi-modality Sensing Technologies: Survey of the Last Decade. *IEEE Transactions on Cognitive and Developmental Systems*.44

[6] Ramirez, J. (2026). FED-MIND: A Privacy-Preserving Federated Multi-Modal Deep Learning Framework for Equitable Autism Spectrum Disorder Diagnostics and Prognostic Stratification. *gjstudies*, *1*(1), 46-46

[7] El-Askary, N. S., Gawish, M., Morsey, M. M., Mahmoud, A. M., Aref, M., & El-Arif, T. I. (2025, November). Towards Explainable Multi-M Multi-Modal Fusion

Strategies for ASD Detection: A Review. In *2025 Twelfth International Conference on Intelligent Computing and Information Systems (ICICIS)* (pp. 560-567). IEEE.

[8] Malik, W., Fahiem, M. A., Khan, J., Jung, Y., &Alturise, F. (2025). An Adaptive Transfer Learning Framework for Multimodal Autism Spectrum Disorder Diagnosis. *Life*, *15*(10), 1524.

[9] Nawghare, P., & Prasad, J. (2025). Hybrid CNN and random forest model with late fusion for detection of autism spectrum disorder in Toddlers. *MethodsX*, *14*, 103278.

[10] Chen, M. (2026). FAIR-Fed: A Federated Fairness-Constrained Multi-Modal Deep Learning Framework for Equitable and Privacy-Preserving Autism Spectrum Disorder Diagnostics. *gjstudies*, *1*(1), 36-36.M. Chen et al., "Evaluating Large Language Models Trained on Code," *arXiv preprint*, arXiv:2107.03374, 2021.

[11] Islam, S. A., & Khan, M. S. (2025). Multi-Modal Behavioral AI for Autism Care: A Federated-Edge Framework with Speech, Motion and Physiological Signal Integration.

[12] Pang, L., Zhao, M., Ma, C., Zhao, X., Zhao, L., Wang, H., & Liu, C. (2025, July). A Multimodal Data-Driven Assessment System for Autism Spectrum Disorder in Children: Development and Pilot Validation of a Multimodal Acquisition Platform. In *2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1-5). IEEE.

[13] Shanmathi, B., Kannagi, S., Thamilarasi, N., Kumar, N. M., & Venkatesh, S. (2025, June). EfficientNet-Powered Multi-Modal System for Non-Invasive Diagnosis of Autism Spectrum Disorder. In *2025 6th International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 2016-2021). IEEE.I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *Proc. EMNLP*, pp. 3982–3992, 2019.

[14] Khan,K, &Katarya, R. (2025). MCBERT: A multi-modal framework for the diagnosis of autism spectrum disorder. *Biological Psychology*, *194*, 108976.

[15] Streamlit Inc., "Streamlit: A Python Framework for Data Applications," 2024. [Online]. Available: https://streamlit.io