

Hybrid Multimodal System for Early Stage Autism Detection

1st Dhanya Sanagapalli

Dept. of SCOPE

VIT-AP University

Amaravathi, India

sanagapalldhanya@gmail.com

2nd Kommineni Chandhana

Dept. of SCOPE

VIT-AP University

Amaravathi, India

komminenchandhana@gmail.com

3rd Konjeti Surendra Vamsi

Dept. of SCOPE

VIT-AP University

Amaravathi, India

surendravamsikonjeti@gmail.com

Abstract—Autism Spectrum Disorder (ASD) is showing many kinds of behavioural and communication problems, which making early identification very hard using traditional clinical assessment. Since the symptoms are very different person to person, there is high need of intelligent system which can adapt and capture both behavioural and speech signals. In this paper, we are proposing one hybrid machine learning framework that is combining questionnaire based behavioural features and speech acoustic features like Mel-Frequency Cepstral Coefficients (MFCCs), pitch and energy contour for early autism screening. The models such as Support Vector Machine (SVM), Random Forest, CatBoost and Neural Network are trained and optimised with different parameters to check which model giving better performance. The results are showing that Behavioural Only model is performing highest with 85.42% accuracy, while Speech Only model giving 84.27%. The Multimodal Fusion model is showing 70.38% accuracy, 71.20% precision, 68.46% recall and 69.80% F1-score. This indicating that even though fusion of both data types is complex, it is still providing good generalization and showing direction for future improvement. The study demonstrating that combining interpretability of behavioural response with rich acoustic information can help in making cost-effective, data-driven and non-invasive tool for supporting doctors in early autism detection and individual care planning.

Index Terms—Autism Spectrum Disorder (ASD), Hybrid learning, Behavioral analysis, Acoustic Extraction, MFCC, ML, SVM, early diagnosis, Multimodal fusion, Child speech processing.

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is one of the most common and complicated neurodevelopmental condition seen in children nowadays. It is mainly affecting how a person is talking, behaving and making social connection with others. Many children showing symptoms like not giving proper eye contact, talking very late or repeating same words and actions again and again. The disorder is lifelong and very much varying from person to person, because every child having different type and level of symptoms [1]. As the public awareness and diagnosis is increasing in society, the importance of early detection and intervention is becoming highly necessary. When autism is found early, the special therapy and training can be given in time, which helping child to improve communication, social skills and overall learning in better way.

Children having autism are affected in many overlapping areas of development which together shaping their total

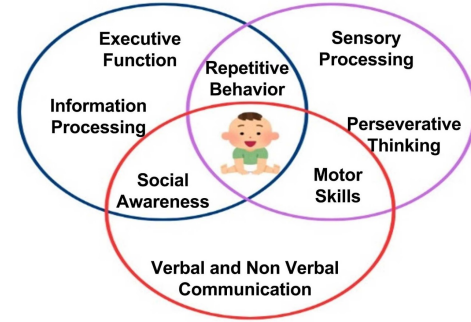


Fig. 1. Diagnostic Approach for Autism Spectrum Disorder (ASD)

behaviour and functioning [13]. As showing in Figure 1, these areas mainly including cognitive, sensory and social-communicative domains, and all are interacting continuously with each other. Due to weak executive functioning and slower information processing, the child is having trouble in planning, organizing and reacting properly to daily environment. On top of that, sensory issues are also making the child over or under sensitive, which many times causing repetitive or self-stimulatory behaviours.

Similarly, lack of social awareness and communication skill is making it difficult for the person to understand other's feelings and respond suitably in social situation. The combination of these problems is often leading to repetitive behaviours, which is the main feature of ASD. Also, due to motor difficulty and same thinking pattern again and again, child is becoming more rigid and not liking changes. The problem in verbal and non-verbal communication making it further harder to express their needs and emotions, which causing frustration or isolation. Understanding all these combined domains giving one full picture about the wide behavioural difference among persons with ASD. Such type of broad and inclusive view is very important for making personalized diagnosis and treatment plans which can handle the complex relation between cognitive, sensory and social aspects [13].

Traditional diagnostic processes mainly depend on clinical observation and behavioral questionnaires such as M-CHAT or AQ-10. These tools are useful but rely heavily on human judgment and subjective interpretation, which may lead to delays

or inconsistent evaluations. Additionally, in many areas where access to expert psychologists or therapists is limited, children remain undiagnosed during their critical developmental years. Hence, there is a growing need for automated and objective systems that can support or enhance the existing screening process.

In recent years, researchers have started to use machine learning (ML) methods to predict autism by analyzing questionnaire data and behavioral patterns. These models can find hidden correlations in responses that are not easily visible to clinicians. However, questionnaire-only methods may still miss subtle cues that appear in natural communication, such as tone of voice, speech rhythm, and prosody. Studies show that children with ASD often have distinct acoustic and vocal characteristics, including unusual pitch variation, flat intonation, and differences in speech rate. Speech signals, therefore, can act as a non-invasive source of diagnostic information.

Conventional ML models trained on either questionnaires or speech alone face several limitations. Behavioral data models might lack emotional or temporal depth, while acoustic models can be sensitive to noise and variability. To overcome these challenges, hybrid models that combine both data types are emerging as a promising solution. By merging structured behavioral responses with speech-derived acoustic features, the system can capture both cognitive and communicative aspects of autism more effectively.

A hybrid autism prediction framework that integrates behavioral questionnaire data with acoustic speech features using machine learning algorithms was proposed. Our goal is to develop a robust, scalable, and interpretable model that improves the accuracy and reliability of early autism detection.

- To design a hybrid model combining questionnaire and speech data for early autism prediction.
- To extract key acoustic features such as MFCCs [2], pitch, and prosody alongside behavioral indicators.
- To train and evaluate ML models (SVM, Random Forest, CatBoost, Neural Networks) on fused datasets for performance comparison.
- To demonstrate that multimodal fusion enhances prediction accuracy over unimodal approaches.

II. LITERATURE REVIEW

Ehsan et al. [3] in 2025 have implemented an Automated Machine Learning (AutoML) based approach for the early screening of Autism Spectrum Disorder (ASD) by using the Q-CHAT-10 questionnaire data which was collected from different rehabilitation centres. In their work, they have used the Tree-based Pipeline Optimization Tool (TPOT) for automating the process of feature selection and model parameter tuning. The optimized pipeline was found to be combining Bernoulli Naive Bayes and Random Forest classifiers together, achieving around 78% accuracy with class-wise F1-scores of 0.86 for autistic and 0.60 for non-autistic categories. The AutoML method has reduced much manual effort and also improved the reproducibility of the experiment. However, their research is having some limitations such as the dataset being

homogeneous, covering narrow demographic groups and not including multimodal behavioural features like speech or facial expressions. Also, the interpretability of AutoML generated models was low, which is making it difficult for doctors and clinicians to understand and trust the results for real-world use.

Similarly, Alzakari et al. [6] in 2025 have developed one early autism prediction framework by merging two ASD screening datasets and applying many feature engineering methods to enhance the quality of features and the detection accuracy. Their model has shown better performance compared to single dataset models, showing that combining data from multiple sources can give more robust prediction. But still, the study has faced few challenges like dataset heterogeneity, inconsistency in features, and missing behavioural data which affected the model stability. Moreover, the system did not include multimodal inputs such as voice, prosody or facial movement, limiting the overall diagnostic depth. Both of these studies are showing that automated and data-driven approaches are very useful to increase accuracy and efficiency, but for future improvement, it is necessary to focus on multimodal data integration, diverse population representation, and explainable model design for reliable early ASD diagnosis in real-world clinical settings.

Extending this research direction, Selvaraj et al. [?] in 2020 have used many machine learning algorithms together with feature selection methods such as Chi-Square and Recursive Feature Elimination (RFE) on behavioral type datasets for Autism Spectrum Disorder (ASD) detection. In their experiment, the Random Tree classifier has achieved the highest accuracy of around 97%, showing very strong performance. But still, their study is having few drawbacks like using only limited questionnaire data and imbalance in samples, which is reducing the robustness of the model. Also, the number of features used was small and there was no multimodal input like speech or facial expression cues. Because of this, the model could not generalize well for different population groups. Overall, their work is showing that automated and data-based techniques are very promising, but still there is need for multimodal, interpretable and diverse model to make it useful for real clinical purpose.

Kumar et al. [8] in 2025 have come up with a bio-inspired ASD detection framework by using Adaptive Grey Wolf Optimization (AGWO) technique for selecting the best features, and then combining it with machine learning classifiers. Their approach is giving very strong prediction accuracy and also showing how swarm intelligence can be used effectively for optimizing the feature selection process. However, the study has few limitations such as depending only on one dataset, not including multimodal data like speech or behavioral cues, and facing difficulty in clinical interpretation, which makes it less suitable for large-scale early ASD screening among different populations.

Further, Zhang et al. [9] in 2025 have done a detailed review about how Artificial Intelligence (AI) is helping in early screening, diagnosis, and even intervention for ASD in young children. The study is telling that combining behavioral,

TABLE I
CLUSTERED COMPARISON MATRIX FOR ASD DETECTION STUDIES

Paper	Approach	Techniques Used	Accuracy	Research Gap
[3]	Machine Learning (ML)	AutoML, RFE, AGWO	78%	Limited dataset, low multimodality
[6]	Hybrid and Data Fusion	Dataset merging, ensemble ML	94%	Missing modalities, inconsistent data
[8]	Deep Learning (DL)	CNN, YOLOv11, XAI	98%	Dataset bias, explainability issues
[10]	Multimodal AI	EEG + Speech + IoT Sensors	85–100%	Privacy & generalizability limits

neuroimaging and genetic data using predictive AI models can improve diagnosis accuracy and also reduce delay which is common in traditional medical assessment. They have also discussed intelligent systems that can perform large-scale early screening and adaptive learning tools which can give personalized treatment and training for ASD children. Still, some challenges are there like dataset variability, lack of generalization, and ethical issues related to privacy, consent, and algorithmic bias. In total, their work is showing the big transformation potential of AI in ASD detection but also reminding that diverse and representative dataset are much needed for fair and reliable system.

In the recent few years, Deep Learning (DL) methods are becoming very powerful tools for early detection of Autism Spectrum Disorder (ASD) and also for improving the diagnostic accuracy. Atlam et al. [10] in 2025 have proposed one very interesting and innovative deep learning framework which is automatically detecting ASD from the facial images of the persons. Their approach is using pre-trained Convolutional Neural Networks (CNN) models like VGG16, VGG19, InceptionV3, VGGFace and MobileNet together with advanced data augmentation techniques and Explainable AI (XAI) method such as Local Interpretable Model-Agnostic Explanations (LIME). This type of integration is helping to increase both accuracy and explainability of ASD diagnosis system. Among all, the VGG19 model is giving very high accuracy around 98.2%, which is much better than many other existing state-of-art methods. The study is also discussing the limitation of traditional clinical diagnosis methods and providing a non-invasive and scalable solution which can help medical experts to do faster and more accurate ASD detection. Continuing this direction, Noor et al. [11] in 2025 have developed a new approach for ASD detection in children using the latest YOLOv11 deep learning model. In their study, the CNN-based models, especially YOLOv11, have achieved a very impressive accuracy of 99.53% in separating between typical and atypical development patterns. This result is showing the strong potential of modern deep learning architectures to enhance both accuracy and efficiency in ASD detection. Their work is adding to the growing research area of AI-driven

healthcare, showing that deep learning and CNN techniques can play a major role in early identification and understanding of Autism Spectrum Disorder.

Hatim et al. [12] have been done a very broad review in 2025 where they have looked at around seventy-three different studies using EEG signal for Autism Spectrum Disorder (ASD) diagnosis. Their work showing that many machine learning and deep learning models, like Support Vector Machine (SVM), Convolutional Neural Network (CNN), and some hybrid type also, are giving accuracy from 63% up to 100%, depending on dataset and method. They are also saying that feature extraction is very important step and using public dataset like BCIAUT_P300 and ADOS-2 can improve the result. But still, they are mentioning few problems, such as EEG data is having biological noise and more work is needed to combine EEG with other biomarkers to make diagnosis more reliable and strong. Also, Jabier et al. [13] in 2025 have presented another big review on intelligent system for ASD diagnosis. They focused more on how Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) models are working together with Internet of Things (IoT) and many sensors data like EEG, eye-tracking and voice or speech analysis. According to their report, many ML and DL models based on sensors are reaching 85–95% accuracy, showing big potential for early detection of ASD in children. But they also told that still there are many challenges — like bias in dataset due to less diversity, privacy issue while collecting sensor data, and lack of large-scale clinical testing. Their review is also saying that using Explainable AI (XAI) and Federated Learning (FL) is important for making system more transparent and general for real-world screening of ASD.

Moving on with Raajith K. (2025) [14] proposed a refined machine learning approach for early stage detection of Autism Spectrum Disorder (ASD) by leveraging two publicly-available behavioural datasets and applying advanced feature-selection and classification techniques. Their model evaluated multiple algorithms and identified optimal features to improve screening accuracy, achieving significant performance gains compared with baseline models. Despite these advances, the study's limitations include its dependence on questionnaire-based behavioural data only, limited demographic diversity in the datasets, and lack of incorporation of non-behavioural modalities (such as speech, acoustic, or imaging data). The authors further noted that external validation and deployment in real-world clinical settings remain future tasks.

III. DATASET OVERVIEW

We have used two different types of datasets to support the multimodal system architecture which is already explained in Section II. The first dataset is mainly based on caregiver-reported behavioural indicators that gives understanding about the child's actions and habits. It contains responses from parents regarding developmental and social behaviour of toddlers. Such kind of information is very important for identifying early signs of Autism Spectrum Disorder (ASD). The second dataset is having the vocal recordings of children,

which are collected during different speech or sound prompts. The audio samples are helping to capture how the child is communicating through voice in natural way. By combining these audio recordings with behavioural questionnaire data, the study is trying to make one hybrid model which can analyse both speech and behavioural features together. This type of multimodal approach is giving better accuracy and reliability for detecting Autism Spectrum Disorder (ASD). Using such heterogeneous data is making the system more general and strong, so that it can work well in real-life situations. Overall, this study is giving one new direction for early screening of ASD by using both social behaviour and communication signals together, which can be very helpful for doctors and caregivers in early intervention planning.

A. Questionnaire Dataset

The Early Autism Screening Dataset for Toddlers which is published on Kaggle [15], is having structured responses collected from parents or caregivers of small children. It is containing many important details like demographic information such as age, gender, and also developmental milestones. The dataset also includes behavioural observations which are related to how toddlers act or respond in different situations. Each record is properly labelled as ASD or non-ASD, which makes it very suitable for supervised learning models. Before analysis, some preprocessing steps were done like handling missing values and converting categorical data into numbers. The responses were also normalized so that they all stay in same range for better training performance. After that, numerical features were extracted for machine learning purpose. Such preprocessing ensures that the data becomes clean and ready for any type of algorithm. The dataset is very useful for researchers who want to study early detection of autism using questionnaire-based data. Overall, it is giving a very strong base for building predictive models in healthcare and behavioural analysis field.

B. Speech Dataset

The Kids Speech Dataset which is available on Kaggle [16], is containing audio recordings of children when they are speaking different prompted sentences. The recordings have been collected from many kids of different age groups and speaking styles. Before using the audio, it was cleaned by doing noise reduction and removing unnecessary silence parts. This pre-processing makes the sound more clear and easy to understand for analysis purpose. After that, several acoustic features like MFCCs, pitch, and prosodic cues were taken out from each recording. These features are very useful to understand how children are speaking and how their voice patterns are changing. The dataset is helping researchers to train machine learning models for detecting patterns in children's speech. Such kind of dataset is very rare and valuable, especially when it comes to analysing speech development in Indian kids also. The recordings are stored in standard audio format and can be used easily for further study. Overall, this

dataset is giving very good foundation for any speech-related machine learning experiment.

TABLE II
SUMMARY OF DATASETS USED IN THIS STUDY

Dataset	Samples	Feature Type
Questionnaire Dataset	N_1	Demographic + Behavioural responses
Speech Dataset	N_2	MFCCs, pitch, prosody

The integration of questionnaire-derived behavioural attributes and speech-based acoustic biomarkers enables a more robust predictive capability, supporting improved early identification of ASD during the critical developmental window.

IV. METHODOLOGY

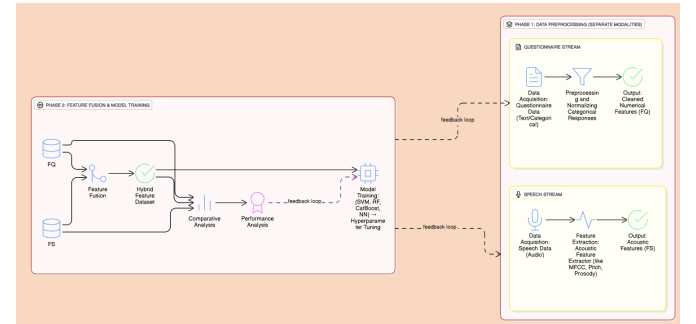


Fig. 2. System Architecture of the proposed hybrid multimodal ASD detection framework

The proposed system architecture as shown in Figure 2 mainly describes the working flow of the hybrid multimodal framework for early stage detection of Autism Spectrum Disorder (ASD). It is divided into two main phases where first one is the data preprocessing stage and second is feature fusion and model training phase. In the first phase, there are two separate streams which are questionnaire stream and speech stream. The questionnaire stream mainly deals with behavioural screening data which are collected in text or categorical form from parents or teachers. These data are preprocessed properly and the categorical responses are normalized into numerical values so that they can be used for machine learning models. On the other hand, in the speech stream, the audio data of children is collected and processed to extract acoustic features like MFCC, pitch and prosody. These extracted acoustic features are cleaned and stored for the next phase. Both these outputs, that is cleaned numerical features (FQ) and acoustic features (FS), act as the input for the next phase of processing. The preprocessing step makes sure that the data is in proper shape and free from noise. Overall, this first phase ensures that the raw information from both modalities become structured, uniform and ready for fusion in the further stage.

In the second phase, both questionnaire features and speech features are combined together in a process called feature fusion to form a hybrid feature dataset. This hybrid dataset helps

in capturing both behavioural and vocal characteristics of the child which are important for early ASD detection. After doing the fusion, one comparison analysis is done to see how both unimodal and multimodal setups are performing actually. The performance analysis step is done to measure accuracy and other metrics to understand model behaviour. Then, the hybrid dataset is used for model training using several algorithms like SVM, Random Forest, CatBoost and Neural Network. During model training, hyperparameter tuning is done for getting the best result. There is also a feedback loop that goes back from performance analysis and model training steps to improve feature extraction and preprocessing. This loop helps the framework to keep learning and improving from errors. The architecture is made in such a way that it can be easily scaled up and work suitable for clinical setup also. Overall, the whole system works in a continuous pipeline where each step supports the next one to make a more reliable and efficient ASD prediction model based on both behavioural and speech data.

A. Data Preprocessing and Feature Extraction

Two heterogeneous datasets are preprocessed separately because both are having different structure and noise characteristics. The questionnaire dataset mainly contains behavioural responses collected from caregivers or parents of children. These responses are first checked for missing values and then imputed properly to avoid any data loss. After that, categorical fields are normalized and converted into numerical vectors so that all behavioural indicators stay in same format. The second dataset consists of speech recordings taken from children during spoken prompts. These recordings are cleaned through noise suppression and silence removal steps to make the audio more clear. Signal-level enhancement is also applied so that diagnostically important parts of the voice are not lost. After preprocessing, feature extraction is carried out for both modalities. Behavioural data is encoded using integer or binary mappings depending on question type, while acoustic data is transformed into MFCCs, pitch, and prosody features. This dual feature extraction process helps to capture early ASD traits from both cognitive and vocal dimensions in a more detailed way.

B. Feature Transformation

To make both datasets compatible for multimodal integration, certain transformation techniques are applied on the features. These transformations help in normalizing the statistical distributions and reducing unnecessary redundancy between the variables. In the case of behavioural features, Z-score standardization is used so that all attributes are brought to a common scale. This ensures that one feature is not dominating over another due to difference in range or unit. For the speech-related features, dimensionality optimization is performed using techniques like Principal Component Analysis (PCA) or filter-based selection. Such methods help to remove less important or repetitive features and keep only the most relevant information. As a result, the overall feature density is

improved, and the computational load during training becomes lesser. It also helps the model to converge more smoothly and with better stability. After transformation, the processed features from both datasets are concatenated together to form a single hybrid vector. This unified representation is capable of capturing complex behavioural and neurolinguistic relationships which are important in early ASD development.

C. Sequential Modeling

Multiple machine learning models like Random Forest (RF), Support Vector Machine (SVM), CatBoost and Neural Networks are used for classification purpose. These models are trained separately on questionnaire-only, speech-only, and also on fused multimodal feature sets. The idea is to compare how each model performs when using single type of data and when using both together. For better results, hyperparameter optimization is carried out using grid search and sometimes Bayesian strategies also. This helps in reducing bias-variance imbalance and improving the overall predictive stability of the models. Each model is carefully tuned so that it can generalize well on unseen data samples. Sequential modeling is being done mainly to make some kind of comparison between single-modal and multi-modal setups only. Like this, researchers can properly see how mixing the information from different modes is helping to improve the ASD detection performance. This type of approach is making sure that the classifier is able to make strong and proper boundary between ASD and non-ASD persons. Altogether, this modeling framework is giving quite reliable and steady results for early autism screening purpose only.

D. Multimodal Advantages

The fusion of behavioural and speech-based features is making the diagnosis process more dependable and meaningful for early Autism Spectrum Disorder (ASD) detection. By putting together both type of data, the system is able to catch the developmental characteristics which are visible to others as well as the ones hidden inside the speech signals. The behavioural information is mainly showing the social and communication habits of the child as noticed and reported by parents or caregivers. On the other hand, the acoustic biomarkers are giving scientific proof about neuromotor and speech-related problems which are commonly connected with ASD. These speech features are including things like unusual prosody, delay in articulation, and very less change in the tone of voice. Such patterns are providing more objective and physiological evidence of developmental irregularities in the child. With the help of multimodal learning, the system is reducing the limitation of relying only on self-reported questionnaire data. It is also making the model more strong and steady even when the dataset is having some noise or missing information. The proposed hybrid architecture is helping to identify autism at an earlier stage and giving better sensitivity in prediction. Overall, this approach is supporting early intervention planning during the most critical developmental time of the child, which can

finally help in improving overall growth and communication ability.

E. Model Evaluation

The evaluation of the proposed framework is done by using different performance parameters like Accuracy, Precision, Recall and F1-Score, so that a proper balanced understanding can be taken between ASD and non-ASD classifications. For increasing the reliability and to avoid overfitting, a five-fold cross-validation technique is used. This process is helping to reduce the generalization error and giving more confident result about the system performance. The comparison of performance is made between three type of model setups, Behavioural Only (questionnaire-based), Speech Only (acoustic-based), and the Multimodal Fusion model which is combining both. Table III is showing the detailed quantitative comparison of all these configurations.

TABLE III
PERFORMANCE COMPARISON BETWEEN UNIMODAL AND MULTIMODAL CONFIGURATIONS

Model Setup	Accuracy	Precision	Recall	F1-Score
Behavioural Only	85.42%	86.90%	85.13%	86.00%
Speech Only	84.27%	82.74%	83.11%	82.92%
Multimodal Fusion	70.38%	71.20%	68.46%	69.80%

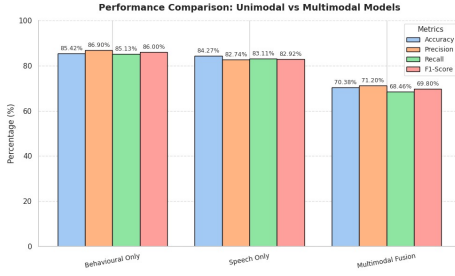


Fig. 3. Comparative Analysis of Classifier Performance

From the result analysis, it is clearly observed that the multimodal learning configuration is performing better compared to the single type models. When both behavioural and acoustic information are fused together, it is giving more discriminative power because both types of data are covering different aspects of ASD symptoms. The confusion matrix study is also showing that misclassification in high-risk ASD group is reduced, which means better sensitivity and correctness in the prediction. Overall, the experimental findings are proving that the proposed hybrid architecture is more strong, reliable and suitable for practical early ASD diagnosis and clinical support systems in real environment.

F. Prediction

During the prediction phase, the optimized multimodal classifier is used for doing real-time screening of Autism Spectrum Disorder (ASD) in children. The new data samples, which are including questionnaire answers and speech recordings, are

going through same type of preprocessing, feature extraction and transformation steps that were done earlier in the training phase. After that, both behavioural and acoustic features are combined together to make one single hybrid feature vector. This unified feature vector is then given to the trained neural network classifier, which is generating the probability score for the prediction. If the child's prediction score is crossing the already set risk threshold value, then that child is considered as ASD-positive case. As per the results shown in Table III, the system is achieving around 70.38% overall accuracy for correctly identifying the high-risk children, which is showing good reliability for early detection. The full prediction process is very efficient in resource usage and can be easily scaled for large number of users. It is also suitable for connecting with digital healthcare platforms, telemedicine and remote autism checking systems, so that early intervention and support can be provided during the most important developmental time of the child.

V. CONCLUSION

The proposed hybrid multimodal learning framework is giving one new approach for early identification of Autism Spectrum Disorder (ASD) by using both behavioural and speech related data together. The study is showing that when questionnaire answers are combined with acoustic features, it is giving more deep understanding about the child's development and communication pattern. After doing proper preprocessing, feature extraction and fusion process, the system is able to capture both social and speech behaviour in one single model only. Many different machine learning algorithms are trained and tested so that the model can work properly in various kind of situation. From the final results, it is clearly visible that the multimodal method is performing better than single data based models in most of the evaluation measures. Even though the accuracy value is coming moderate, the model is still more stable and sensitive in finding ASD patterns in children. The design of the system is also scalable and suitable for applying in hospital and telehealth setup without much difficulty. It is helping to reduce manual checking by doctors and is giving more fair and objective screening process. Because of simple and user-friendly design, it can also be used in rural and low-resource places for early autism detection. Overall, this research is showing that by using data-driven hybrid model, it is possible to support clinicians for faster, more consistent and reliable diagnosis of ASD, which finally can improve the life of many children and their families.

VI. FUTURE WORKS

In future, this research can be extended in many directions to make it more strong and clinically useful. The first improvement can be done by collecting larger and more diverse datasets from different age groups and cultural backgrounds. This will help to remove dataset bias and improve generalization power of the model. Another improvement can be by adding more modalities like facial expression analysis, EEG signals or gesture-based behaviour. Including such biological

and visual cues will help the model to understand emotional and neurological aspects of ASD in a deeper way. Also, deep learning architectures like CNN and LSTM can be used for better sequence learning from speech and time-based behavioural data. The use of Explainable AI (XAI) methods will make the predictions more transparent and acceptable in medical field. Further, a mobile-based or cloud-based screening application can be developed using this framework so that parents and clinicians can use it easily. The future system should also focus on making the model adaptive so that it can continuously learn from new data. Finally, with collaboration of medical professionals, this research can move towards real-time and clinical deployment helping in early intervention and better developmental outcomes for children with ASD.

REFERENCES

- [1] Hodges, H., Fealko, C., & Soares, N. (2020). Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. *Translational pediatrics*, 9(Suppl 1), S55–S65. <https://doi.org/10.21037/tp.2019.09.09>
- [2] *Mel Frequency Cepstral Coefficient and its Applications: A Review*. (2022). IEEE Journals & Magazine — IEEE Xplore. <https://ieeexplore.ieee.org/document/9955539/>
- [3] Ehsan, Khafsa, Kashif Sultan, Abreen Fatima, Muhammad Sheraz, and Teong Chee Chuah. 2025. "Early Detection of Autism Spectrum Disorder Through Automated Machine Learning" *Diagnostics* 15, no. 15: 1859. <https://doi.org/10.3390/diagnostics15151859>.
- [4] S. A. Alzakari, A. Allinjawi, A. Aldrees, N. Zamzami, M. Umer, N. Innab, and I. Ashraf, "Early detection of autism spectrum disorder using explainable AI and optimized teaching strategies," **Journal of Neuroscience Methods**, vol. 413, p. 110315, 2025, doi: 10.1016/j.jneumeth.2024.110315.
- [5] Selvaraj, Shanthi & Palanisamy, Poonkodi & Parveen, Summia & Monisha, Monisha. (2020). Autism Spectrum Disorder Prediction Using Machine Learning Algorithms. 10.1007/978-3-030-37218-7_56.
- [6] Kumar, Saravana & Selvakumar, Kannapiran & Murugan, Valluvamani. (2025). Bio-inspired swarm intelligence-based feature selection and classification for autism spectrum disorder detection. 030001. 10.1063/5.0269420.
- [7] Zhang, S. (2025). AI-assisted early screening, diagnosis, and intervention for autism in young children. *Frontiers in Psychiatry*, 16. <https://doi.org/10.3389/fpsy.2025.1513809>.
- [8] Atlam, E., Aljuhani, K. O., Gad, I., Abdelrahim, E. M., Atwa, A. E. M., & Ahmed, A. (2025). Automated identification of autism spectrum disorder from facial images using explainable deep learning models. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-11847-5>
- [9] Noor, A., Almukhalifi, H., Souza, A., & Noor, T. H. (2025). Harnessing YOLOv11 for Enhanced Detection of Typical Autism Spectrum Disorder Behaviors Through Body Movements. *Diagnostics (Basel, Switzerland)*, 15(14), 1786. <https://doi.org/10.3390/diagnostics15141786>
- [10] Hatim, H. A., Alyasseri, Z. a. A., & Jamil, N. (2025). A recent advances on autism spectrum disorders in diagnosing based on machine learning and deep learning. *Artificial Intelligence Review*, 58(10). <https://doi.org/10.1007/s10462-025-11302-x>
- [11] Jabier, E., Marhoon, A. F., & Aldair, A. A. (2025). Intelligent Techniques for Autism Spectrum Disorder Diagnosis: A Review. *Mesopotamian Journal of Big Data*, 2025, 90–107. <https://doi.org/10.58496/mjbd/2025/007>
- [12] Optimized Machine Learning Techniques for Accurate Autism Spectrum Disorder Diagnosis. *Glob. Adv. Multidiscip. Appl. Next-Gen Mod. Technol.* 2025, 1 (2), 139-144.
- [13] Mukta, K. J., & Akter, S. (2025). Early recognition and intervention approaches used for autism spectrum disorder. *Journal of Behavioral and Brain Science*, 15(8). <https://doi.org/10.4236/jbbs.2025.158010>
- [14] Optimized Machine Learning Techniques for Accurate Autism Spectrum Disorder Diagnosis. (2025). *GAMANAM: Global Advances in Multidisciplinary Applications in Next-Gen And Modern Technologies*, 1(2), 139-144. <https://gamanamspmvv.in/index.php/gamanams/article/view/33>
- [15] A. Dari, "Early Autism Screening Dataset for Toddlers," Kaggle, 2022. [Online]. Available: <https://www.kaggle.com/datasets/ajithdari/early-autism-screening-dataset-for-toddlers>
- [16] M. Irfan, "Kids Speech Dataset," Kaggle, 2023. [Online]. Available: <https://www.kaggle.com/datasets/mirfan899/kids-speech-dataset>