IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received 6 August 2024, accepted 27 August 2024, date of publication 2 September 2024, date of current version 10 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3453440

RESEARCH ARTICLE

Enhanced Special Needs Assessment: A Multimodal Approach for Autism Prediction

SUSEELA SELLAMUTHU AND SHARON ROSE
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu 600127, India
Corresponding author: Suseela Sellamuthu (suseela.s@vit.ac.in)

ABSTRACT Autism spectrum disorder (ASD) poses significant challenges in early detection, necessitating innovative approaches for accurate identification. In this study, we propose a novel method utilizing machine learning models trained on a diverse dataset comprising facial images and behavioral ADOS scores. Employing cutting-edge convolutional neural network (CNN) architectures such as MobileNetV2, ResNet50, and InceptionV3, alongside a bespoke CNN model tailored for ASD detection, we explore the efficacy of our approach. Additionally, we introduce a multimodal concatenation model that integrates image features with behavioral scores to enhance predictive performance. Our results showcase promising outcomes, with the multimodal concatenation model achieving a remarkable accuracy of 97.05%. Furthermore, our models demonstrate competitive precision, recall, F1 score, and area under the ROC curve (AUC), underscoring their potential to facilitate early ASD diagnosis. These findings signify the significance of leveraging multimodal data fusion techniques to augment ASD detection accuracy, thereby contributing to advancements in early intervention strategies.

INDEX TERMS Autism detection, MobileNetV2, RestNet50, InceptionV3, multimodal concatenation, convolution neural network.

I. INTRODUCTION

The term Autism Spectrum Disorder (ASD) refers to a collection of neurodevelopmental disorders marked by issues with speech and social contact in addition to recurrent behaviors and narrow interests. With varying degrees of severity, people with ASD may find it difficult to comprehend social cues, participate in everyday activities, or communicate successfully. In addition to sensory sensitivity that might interfere with day-to-day functioning, repetitive behaviors and a strong focus on particular subjects are prevalent [1]. To support people with ASD and improve their communication abilities, social relationships, and general quality of life, early intervention and specialized therapies are crucial. The study aims to use machine learning methods to detect autism spectrum disorder (ASD) in children at an early age. ASD is marked by difficulties with behavior, social interaction, and communication. Emerging research aims to overcome this by combining behavioral scores and face image analysis as all-encompassing ASD screening techniques. This study is driven by the urgent need for reliable and effective techniques to identify ASD, especially in the early stages of life when treatment can greatly enhance the lives of those who are impacted. The time-consuming and subjective assessment-based nature of many current ASD detection techniques might cause delays in diagnosis and intervention. Through the utilization of machine learning and data-driven methodologies, this project aims to create novel solutions that can improve the efficiency, precision, and usability of ASD screening. The ultimate objective is to aid in the creation of technology-driven, scalable solutions that can improve the lives of people with ASD and their families. Traditional ASD detection techniques frequently encounter issues with scalability, accuracy, and dependability.

Furthermore, the efficacy of current screening methods may be constrained by their dependence on single-modal data

The associate editor coordinating the review of this manuscript and approving it for publication was S. M. Abdur Abdur Razzak.

sources. The development of ASD detection tools is further complicated by ethical problems surrounding privacy and data collection. Resolving these issues is critical to improving the precision and effectiveness of ASD screening techniques. The main goal of the research is to develop and evaluate ML models that use facial image analysis and behavioral scores to accurately detect ASD. The project aims to create a strong hybrid deep learning model that can detect autism spectrum disorder (ASD) by combining facial image analysis and behavioral score assessment. Additionally, it seeks to evaluate the model's effectiveness using a large dataset of behavioral scores and facial images from peers with ASD diagnoses and those without.

The initiative specifically seeks to accomplish the following goals: (1) Create innovative CNN architectures specifically designed to identify ASD (2) Investigate how behavioral scoring and facial image analysis might be combined for better detection accuracy (3) Use relevant metrics, such as accuracy, precision, recall, F1 score, and area under the ROC curve (AUC), to assess the performance of the presented models. The scope of the project encompasses the collection and pre-processing of relevant datasets comprising facial images and behavioral scores. In this project, cutting-edge methods like CNNs and multimodal data fusion will be used to construct and train deep learning models. The performance of the model will be assessed by experimentation and comparison with current techniques. The research will also tackle ethical issues related to data confidentiality and privacy.

II. LITERATURE REVIEW

According to current estimates, the prevalence of ASD has been rising significantly, with 1 in 54 children in the US thought to have an ASD diagnosis. According to current estimations, 1 in 160 children worldwide are thought to have an ASD. However, because of its variety and the variation in symptoms throughout individuals, diagnosing ASD can be challenging. Thus, meeting the needs of people with ASD and their families requires a grasp of prevalence rates, diagnostic difficulties, and the significance of early detection and intervention.

The substantial problem that learning impairments (LD) pose in the educational system has led researchers to investigate a range of detection and intervention strategies. Numerous investigations have used a range of techniques and technology to look into the occurrence and detection of LD and autism as presented in Table 1.

Al-Qadri et al. [1] offer important insights into the landscape of learning problems (LD) in educational contexts by providing an observation approach meant to ascertain the widespread nature of LD in academic settings. Their observational research provides insight into the complex nature of kids' academic difficulties and establishes the groundwork for additional study and intervention efforts. The work of Al-Qadri et al. emphasizes how crucial it is to comprehensively evaluate the incidence of LD to guide focused therapies and support systems. Developing a guidebook specifically for low- and middle-income countries, Hayes et al. [2] further the subject by building on the findings of Al-Qadri et al. This guide provides an organized method for identifying and assessing learning disabilities that takes into consideration the particular contextual elements common in settings with limited resources. Hayes et al.'s study addresses the urgent need for accessible and culturally relevant therapies for LD by offering useful recommendations and tools for evaluation.

Technological developments, especially in the domains of machine learning (ML) and deep learning (DL), have created new opportunities for LD detection. Using EEG data and handwriting analysis, Vilasini et al. [3] and Seshadri et al. [4] investigate the use of DL approaches in LD categorization. The promise of non-invasive and effective LD assessment tools is demonstrated by this paper by combining state-of-the-art computational approaches with conventional evaluation methodologies. In the meanwhile, LD in school-age children has been promisingly predicted by machine learning techniques. This is demonstrated by David and Balakrishnan [5], who use ML techniques to identify learning impairments early on. Their research demonstrates the value of predictive analytics in identifying kids who are at risk and enabling prompt interventions to meet their requirements.

Similarly, by creating a diagnostic model for kids with complicated autism spectrum disorders (ASD) and intellectual impairments, Song et al. [6] expand the use of ML in LD identification. Song et al. emphasize the value of customized assessment strategies in meeting the various demands of students with numerous learning challenges by utilizing advanced machine learning techniques. One frequent LD that has been the focus of ML-based identification studies is dyslexia. Saminathan and Kanimozhiselvi [7] examine dyslexia diagnosis with machine learning techniques on a variety of datasets, emphasizing the need for data-driven methods for the precise identification of learning disabilities. Their research highlights how crucial it is to use a variety of data sources to improve diagnosis accuracy and guide focused actions. Chakraborty [8] and Kaisar [9] have conducted survey studies that offer thorough summaries of the state-of-the-art approaches for ML-based LD prediction. Future study paths and intervention tactics are guided by the insightful information provided by these surveys, which highlight industry best practices and new trends. Additionally, Poornappriya and Gopinath [10] highlight how technology-driven therapies may improve the course of LD. Poornappriya and Gopinath support customized intervention methods based on individual learning profiles to maximize educational achievements for kids with learning disabilities (LD). They do this by utilizing machine learning techniques. Simultaneously, prognostic instruments created with ML technology provide chances for LD early identification. Prognostic methods using machine learning (ML) are proposed by Loizou and Laouris [11] to proactively identify learning disabilities in youngsters. Their research emphasizes how

TABLE 1. Literature survey.

Ref	Methodology	Advantages	Limitations	Ref	Methodology	Advantages	Limitations
[1]	Observation tool	Provides a systematic assessment method for the prevalence of academic learning difficulties	Reliance on subjective observation may introduce bias	[14]	Diagnostic and classification system for dyslexia	Offers a structured diagnostic approach	Performance may vary based on dataset quality
[2]	Screening and evaluation guide	Tailored for low- and middle-income countries, comprehensive approach	Implementation may require adaptation to specific contexts	[15]	ML-based models for early-stage ASD detection	Addresses early intervention needs	Generalizability may be limited by dataset characteristics
[3]	Deep learning techniques for handwriting analysis	Utilizes advanced computational methods, potentially high accuracy	Dependency on quality and quantity of handwriting samples	[16]	EEG-based dyslexia detection with novel predictor selection	Provides non-invasive detection method	Interpretation of EEG signals can be complex
[4]	EEG-based classification using shallow and deep neural networks	Non-invasive, objective assessment	Interpretation of EEG signals can be complex	[17]	EEG-based identification of learning disability	Offers a non-invasive diagnostic approach	Performance may vary based on EEG signal quality
[5]	Machine learning approach	Utilizes ML algorithms for prediction	Relies on the availability and quality of input data	[18]	Prediction of ADHD diagnosis using clinical measures	Utilizes low-cost clinical measures for prediction	Performance may vary based on feature selection
[6]	Diagnostic model for ASD with intellectual disability	Addresses complexity of ASD diagnosis and incorporates ML	Performance may vary based on dataset diversity	[19]	Nationwide deep learning approach for ADHD onset prediction	Addresses population-level prediction needs	Generalizability may be limited by dataset characteristics
[7]	Dyslexia detection using ML techniques	Utilizes multiple datasets, comprehensive approach	Generalizability may be limited by dataset characteristics	[20]	Behavioural activity-based prediction of ADHD	Utilizes behavioral data for prediction	Performance may vary based on behavioral data quality
[8]	Survey paper on learning disability prediction	Provides an overview of existing research	Limited to summarizing existing literature	[21]	Multimodal affect recognition adaptive learning system	Integrates multiple modalities for affect recognition	The complexity of integrating multimodal data
[9]	Survey on developmental dyslexia detection using ML	Offers insights into various ML techniques	May not cover the latest advancements in the field	[22]	Predicting behavior change in students with special needs using multimodal	Addresses behavior prediction needs in special education	Performance may vary based on data quality and model complexity

TABLE 1. ***(Continued.)*** **Literature survey.**

					learning analytics		
[10]	Application of ML techniques for improving learning disabilities	Addresses practical implementation of ML in education	Specific techniques and applications may vary	[23]	Multimodal approach for identifying ASD in children	Utilizes multiple modalities for ASD identification	The complexity of integrating multimodal data
[11]	Prognosis tools using ML technologies	Addresses prognosis aspect, utilizes ML	Performance may be influenced by feature selection	[24]	Intelligent multimodal framework for identifying ASD	Integrates multiple modalities for ASD identification	Performance may vary based on feature selection and model complexity
[12]	Deep learning approach for prediction of learning disability	Utilizes deep learning techniques for prediction	The complexity of deep learning models may hinder interpretability	[25]	Multimodal machine learning system for early screening of ASD	Utilizes response to name across modalities for early screening	Generalizability may be limited by dataset characteristics
[13]	Fuzzy logic K means clustering in ML	Incorporates fuzzy logic for prediction	Performance may vary based on parameter selection				

crucial early intervention is to reducing the negative effects of LD on well-being and academic achievement.

Furthermore, Dhamal and Mehrotra [12] present DL approaches for LD prediction, emphasizing how LD detection methods are developing. Dhamal and Mehrotra show how cutting-edge computational techniques may be used to better diagnose LD patients by utilizing deep learning models to capture intricate patterns. Adding to the conversation, Mary et al. [13] offer a unique method of LD prediction based on machine learning's fuzzy logic and K-means clustering. Their research demonstrates how flexible hybrid approaches may be in accurately predicting learning disabilities in youngsters enrolled in school. Mary et al. show how hybrid models may capture complicated correlations in LD datasets improve diagnosis accuracy and guide focused intervention methods by combining fuzzy logic with clustering approaches.

Similar to this, Khan et al. [14] further the area by employing machine learning to create a diagnostic and classification system (DCS) for kids with learning difficulties. Their work focuses on using machine learning (ML) algorithms to diagnose and categorize learning disabilities (LD), allowing for customized therapies based on unique learning profiles. The work of Khan et al. emphasizes the value of tailored strategies for LD evaluation and intervention, emphasizing the potential of ML-based systems to maximize learning results for children with LD.

To improve diagnostic efficiency and accuracy, Akter et al. [15] provide machine-learning models for the early-stage diagnosis of autism spectrum disorders. These models make use of cutting-edge computational techniques. Their research highlights how machine learning (ML) can help people with ASD receive early intervention and assistance. By presenting an early dyslexia diagnosis method based on EEG signals and employing cutting-edge predictor extraction and selection techniques, Parmar and Paunwala [16] make a significant contribution to the discipline. With this method, dyslexia can be detected non-invasively in its early stages, allowing for prompt treatments to support those who are impacted.

Furthermore, Ahire et al. [17] concentrate on the creation of machine learning algorithms for the EEG-based diagnosis of learning difficulties, thereby augmenting the range of non-invasive diagnostic instruments for neurodevelopmental diseases. Their research emphasizes how crucial it is for machine learning specialists and neuroscientists to collaborate across disciplines to advance diagnostic techniques. A competitive model evaluation for ADHD diagnosis prediction utilizing quick, low-cost clinical measures is presented by Mooney et al. [18]. Their research highlights how important it is to use affordable diagnostic methods to increase the number of people with ADHD who have access to early intervention treatments.

Additionally, a countrywide deep learning technique for predicting the emergence of ADHD in children and adolescents is proposed by Garcia-Argibay et al. [19]. Through the utilization of sophisticated computational methods and extensive datasets, their study provides a valuable understanding of the intricate interactions between hereditary and environmental elements that impact the development of ADHD. Maniruzzaman et al. [20] further advance the research by creating a machine learning analysis that uses behavioral activity to predict children with ADHD. Their research highlights how behavioral data can be used to improve ADHD prediction models' accuracy and enable more individualized treatment approaches.

A Multimodal Affect Recognition Adaptive Learning System designed for people with intellectual disabilities was introduced by Bhatti et al. [21]. By identifying affective states using a variety of data modalities, this cutting-edge system offers individualized learning experiences that improve engagement and learning outcomes.

By utilizing Multimodal Learning Analytics to create a predictive model for behavior modification in kids with special education needs, Chan et al. [22] contribute to the field. Using the amalgamation of several data sources, including behavioral and academic performance data, their methodology allows the prompt identification of students who are at risk and the execution of targeted intervention strategies. Furthermore, Han et al. [23] offer a multimodal method that combines behavioral, physiological, and neuroimaging information to identify autism spectrum disorders in children. Their research demonstrates how multimodal evaluations can enhance diagnostic precision and help comprehend the diversity of ASD presentations. Furthermore, Chen et al. [24] integrate information from several sources, including behavioral observations, clinical evaluations, and neuroimaging, to provide a smart multimodal framework for diagnosing children with autism spectrum disorder. Their method provides a thorough diagnostic tool that takes into account the various ways that ASD manifests itself in different people.

Additionally, Zhu et al. [25] created a multimodal machine-learning system based on name recognition to evaluate toddlers for autism spectrum disorders early. Their technique achieves great sensitivity and specificity in detecting children who are at risk, allowing for early intervention and assistance. This is made possible by the combination of behavioral and physiological data. When summed up, these findings highlight how multimodal techniques can help us better understand neurodevelopmental problems and develop better strategies for diagnosis and treatments. Researchers can obtain a thorough understanding of individual variations and customize interventions to fit the unique requirements of people with ASD and intellectual disabilities by incorporating a variety of data modalities.

III. METHODOLOGY

A. DATASET DESCRIPTION AND PRE-PROCESSING

The dataset employed in this study is a comprehensive collection consisting of real-time clinical images and Autism Diagnostic Observation Schedule (ADOS) assessment scores. This dataset is pivotal for training and evaluating our deep-learning models. It includes images of both autistic and non-autistic children, supplemented with detailed ADOS scores that provide crucial behavioral context.

The images were sourced from clinical environments, ensuring they accurately represent the data encountered in practical applications. Each image is associated with an ADOS score that includes several key parameters: Social Affect (assessing social communication and interaction), Repetitive and Restricted Behaviors (evaluating the presence and severity of repetitive behaviors and restricted interests), and an overall ADOS Severity Score. These scores are integral to the dataset, offering insights into the behavioral characteristics associated with autism spectrum disorder (ASD).

For model training and evaluation, the dataset was limited to 2,536 samples due to memory constraints. This dataset was then split into training and testing sets with an 80-20 ratio, resulting in 2,029 images for training and 507 images for testing.

An essential step in getting the dataset ready for model training is data pre-processing. To make sure the data is appropriate for the model, this entails doing procedures like data cleaning, and normalization. Pre-processing of facial images will involve procedures such as scaling and normalization to improve the dataset's variability. The preprocessing steps for the images included resizing to a uniform dimension of 100 × 100 pixels and normalizing pixel values to a range between 0 and 1. This preprocessing ensures consistency and enhances the models' ability to generalize from training data to real-world scenarios.

In parallel, ADOS scores were extracted from an Excel file, containing separate sheets for autistic and non-autistic children. The relevant columns for social affect, repetitive and restricted behaviors, total ADOS score, and severity were selected. Labels were created for autistic (1) and non-autistic (0) children, and the scores and labels were concatenated to form a unified dataset of numerical scores.

During model training, images and scores were handled as separate inputs. For the transfer learning models (MobileNetV2, ResNet50, and InceptionV3) and the novel CNN model, only the image data was used. In contrast, the multimodal concatenation model incorporated both image data and numerical scores. This model extracted features from the images using a CNN and then concatenated these features with the numerical scores before passing them through dense layers for binary classification.

All data collection processes adhered to stringent ethical guidelines, ensuring the privacy and confidentiality of the participants. Informed consent was obtained, and data was anonymized to protect personal information.

The dataset, with its combination of real-time clinical images and detailed ADOS scores, provides a robust resource for developing and evaluating ASD detection models. Its comprehensive and representative nature ensures that models trained on this dataset can effectively generalize to real-world scenarios, making them highly relevant for practical applications in ASD diagnosis and assessment.

B. BENEFITS OF MULTIMODAL APPROACH

Much potential can be seen in a multimodal strategy that combines behavioural scores with picture data. The objective of the multimodal method [29] is to improve the accuracy and resilience of ASD detection models by incorporating data from several sources, including facial photos and ASD evaluation scores. This method recognizes the complexity of diagnosing ASD, which frequently entails several variables and symptoms. A more thorough understanding of a person's condition can be achieved by utilizing both the quantitative data from standardized ASD screening methods and the visual indicators collected in face photographs. Additionally, the model can gather complementary data due to the multimodal integration of image and score data, which may enhance classification performance and give physicians more dependable diagnostic support. Additionally, the model's discriminative power is improved by capturing a wider variety of ASD-related traits thanks to the merging of data from several modalities. This comprehensive approach increases the model's interpretability and robustness by offering insights into the several aspects that affect ASD classification. All things considered, the multimodal approach is a promising trend for ASD detection, providing a more thorough and nuanced evaluation of each person's neurodevelopmental state.

C. PROPOSED METHODOLOGY

The proposed approach for ASD diagnosis combines new CNN architectures with transfer learning models to capitalize on each technology's advantages in feature extraction and classification. Using the learned features from the ImageNet dataset, transfer learning models like ResNet50 [26], InceptionV3 [27], and MobileNetV2 [28] can extract rich hierarchical features from input photos. These trained models provide a strong basis for feature extraction and are used in ASD classification tasks. Furthermore, to directly extract pertinent characteristics from input photos, a novel CNN architecture designed for ASD detection is created. By extracting relevant information from the input pictures, this specially built CNN architecture improves the model's capacity to differentiate between ASD and non-ASD instances. Moreover, a multimodal concatenation model is shown that combines picture characteristics taken from CNNs with scores from ASD evaluation instruments to enhance classification performance. This method allows for more accurate ASD identification by augmenting the discriminative capacity of the model with input from different modalities.

Overall, to accomplish reliable and efficient ASD detection, the suggested methodology makes use of a variety of transfer learning models, cutting-edge CNN architectures, and multimodal fusion techniques as illustrated in Fig 1.

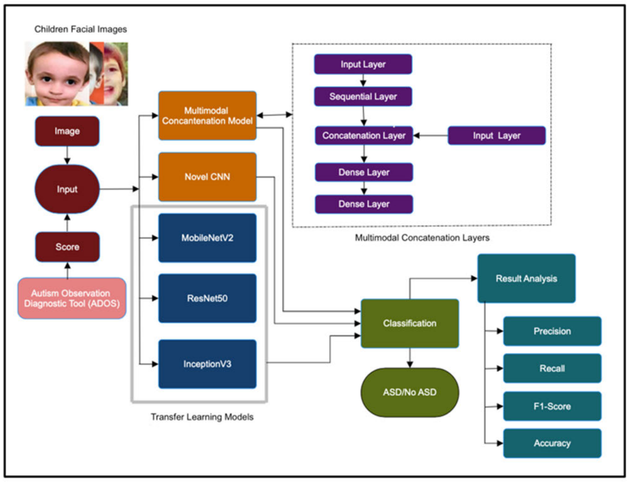


FIGURE 1. Proposed methodology.

IV. MODEL ARCHITECTURE

A. MODEL ARCHITECTURE–NOVEL CNN AND MULTIMODAL CONCATENATION MODEL

Transfer learning models and a unique CNN architecture make up the two primary parts of the model architecture. The trained convolutional layers in the transfer learning models are followed by dense layers and global average pooling for classification. Many convolutional layers with max-pooling (MP) and dense layers make up the unique CNN architecture, which enables features to be extracted directly from the input pictures.

Furthermore, to increase classification accuracy, a multimodal concatenation model is presented that incorporates characteristics derived from both picture data and ASD evaluation scores. The two primary input branches of this model are designed to process picture data and ASD evaluation scores, respectively. It consists of dense layers for processing concatenated information and convolutional neural network (CNN) layers for extracting visual features. The binary cross-entropy loss function is used to train the model, and the Adam optimizer is used to optimize it. During training, early stopping is employed to prevent overfitting.

To extract spatial characteristics from the input photos, the image data branch usually includes convolutional layers followed by max-pooling layers. However, to handle the numerical input, the branch for ASD evaluation scores can use fully connected layers or other suitable architectures. The feature representations from both branches are processed independently for each input modality, then concatenated or merged before being fed into successive dense layers for classification. The amalgamation of data allows the model to assimilate further insights from both picture data and ASD evaluation scores, consequently enhancing its capacity to differentiate between persons with and without autism. Dropout layers may also be incorporated into the model architecture to prevent overfitting and guarantee generalization to new data. All things considered, the multimodal concatenation model, by utilizing the advantages of both image-based and numerical evaluations, provides a thorough method for ASD identification.

In the novel CNN model to extract features from input images, the architecture includes convolution layers followed by MP layers. The spatial dimensions are flattened and sent via fully connected layers (FCL) for classification following several convolution and pooling layers. The architecture is as follows:

(i) Convolution layers with ReLU activation

- Conv2D**:** A 2D convolutional layer that applies 32, 64, and 128 filters of size 3 × 3 respectively.
- ReLU Activation**:** Applies the Rectified Linear Unit function to introduce non-linearity in the model, which helps in learning complex patterns.
- Input Shape**:** Specifies the dimensions of the input images, including the number of color channels (3 for RGB images).

(ii) MP layers: A max-pooling layer that reduces the spatial dimensions of the feature maps by selecting the maximum value from each 2 × 2 pool, effectively down-sampling the input.

(iii) Flatten layer: Converts the 2D feature maps obtained after the convolution and pooling layers into a 1D vector, preparing it for the fully connected (dense) layers.

(iv) Dense layer: A fully connected layer with 64 units/ neurons. Applies the Rectified Linear Unit function to introduce non-linearity.

(v) Dropout layer for regularization: A regularization technique that randomly sets 50% of the input units to 0 during training, which helps prevent overfitting by ensuring the model does not become too reliant on specific neurons.

(vi) Output layer with sigmoid activation for binary classification (autistic or non-autistic): A fully connected layer with a single unit/neuron. Sigmoid activation outputs a probability value between 0 and 1, which is ideal for binary classification tasks such as determining whether a subject is autistic or non-autistic.

Features from both picture data and ASD diagnostic scores are integrated using the Multimodal Concatenation Model [30]. The architecture and explanation of this model's layers is provided below:

(i) Image input branch:

- Convolution layers with ReLU activation
- MP layers
- Flatten layer

(ii) Scores input branch:

- Dense layers

(iii) Concatenation layer to combine outputs from both branches.

(iv) Dense layers for joint processing

(v) Output layer with sigmoid activation for binary classification (autistic or non-autistic)

Details of each of the layers:

(i) Image Input Layer:

Data from images is fed into this layer. Its proportions usually match the height, width, and number of channels of the input images.

(ii) Layers of a Convolutional Neural Network (CNN):

A sequence of convolutional layers processes the input image to extract features from it. A collection of filters is applied to the input by each convolutional layer to capture various patterns and features. To extract the most significant characteristics and reduce spatial dimensions, max-pooling layers are frequently inserted between convolutional layers.

(iii) Flattening Layer:

To transform the multidimensional feature maps into a one-dimensional vector, a flattening layer is used after the convolutional layers. This gets ready the features that were taken out of the pictures for additional processing.

(iv) Scores Input Layer:

The model receives scores from ASD evaluations as input in addition to visual data. Numerous metrics about social effects, repeated behaviors, and the general severity of ASD symptoms may be included in these ratings.

(v) Concatenation Layer:

Along the feature dimension, the features that the CNN layers derived from the picture data and the ASD evaluation scores are concatenated. In doing so, the data from both modalities is combined into a single feature representation, which enables the model to express the correlation between evaluation scores and image attributes.

(v) Dense Layers:

To further process the combined feature representation, one or more dense layers may be utilized after the concatenation layer. The concatenated characteristics undergo nonlinear modifications by these deep layers, which enable the model to discover intricate patterns and connections in the data. The complexity and representational capability of the model are determined by the number of units in thick layers. Before the output layer in the algorithm, dense layers with 64 units are used for feature extraction and categorization.

Lastly, the features that have been processed are sent through an output layer, which is usually made up of a single neuron with a sigmoid activation function. The final output, which is the model's prediction for the binary classification task (such as autistic or non-autistic), is produced by this neuron.

During training, the Adam optimizer combines the benefits of momentum and RMSprop techniques to adjust a neural network's parameters. Using estimations of the first and second moments of the gradients, it determines the adaptive learning rates for every parameter.

Binary cross-entropy loss is chosen as the loss function for binary classification tasks, as seen in the model compilation step. This loss function is suitable for classifying ASD and non-ASD cases.

The Multimodal Concatenation Model as depicted in Fig. 2 makes use of the advantages of each modality to enhance classification performance and offer more reliable predictions for diagnosing autism spectrum disorder by merging data from both picture data and ASD evaluation scores.

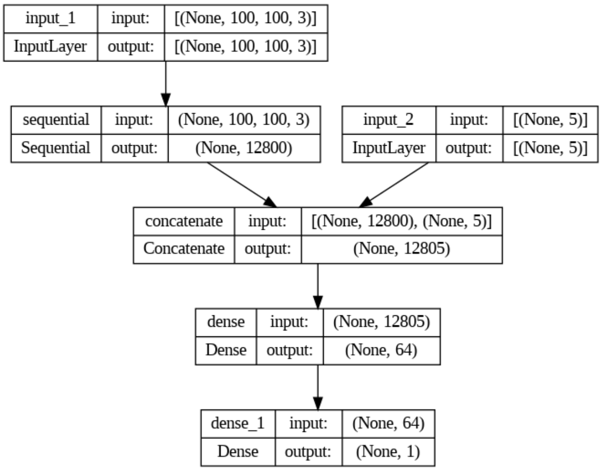
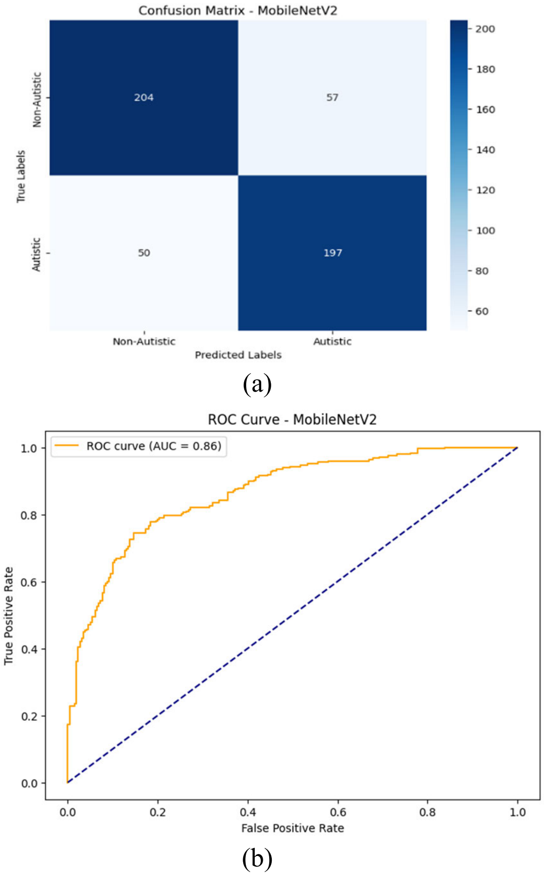

input_1 | input: | [(None, 100, 100, 3)]
InputLayer | output: | [(None, 100, 100, 3)]
sequential | input: | (None, 100, 100, 3)
Sequential | output: | (None, 12800)
input_2 | input: | [(None, 5)]
InputLayer | output: | [(None, 5)]
concatenate | input: | [(None, 12800), (None, 5)]
Concatenate | output: | (None, 12805)
dense | input: | (None, 12805)
Dense | output: | (None, 64)
dense_1 | input: | (None, 64)
Dense | output: | (None, 1)

FIGURE 2. **Novel multimodal concatenation model architecture.**

V. RESULTS AND DISCUSSION

A. EVALUATION METRICS

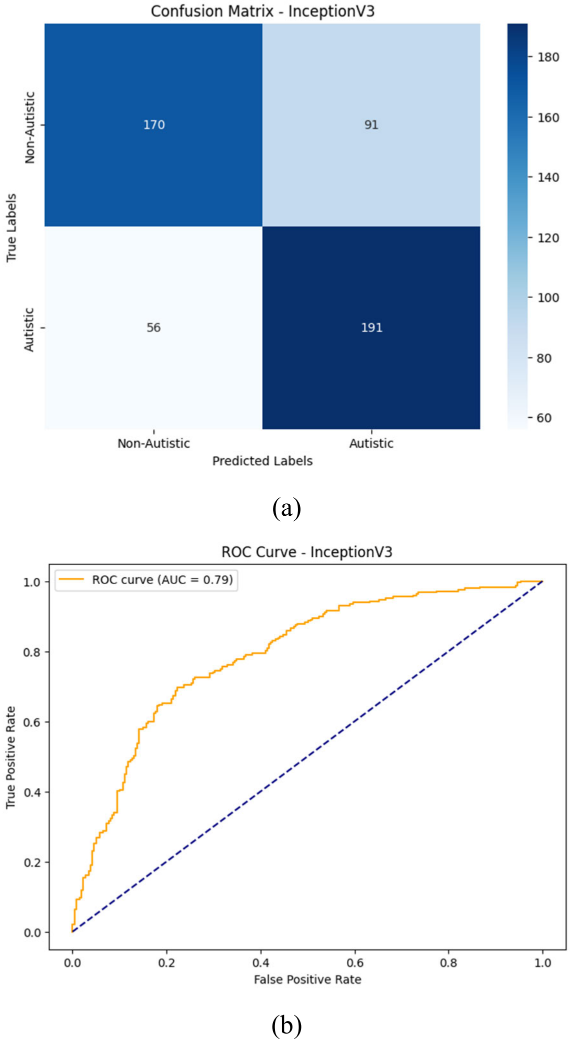
The performance of the proposed models is evaluated using several measures, such as the area under the ROC curve (AUC). These metrics provide insight into how successfully the models classify cases with and without ASD. They account for both true positive and false positive rates. The percentage of correctly categorized cases relative to all instances is known as accuracy. Precision quantifies the accuracy of positive predictions by computing the ratio of true positive instances to the sum of true positives and false positives. Recall, sometimes called sensitivity, is a measure of the model's ability to identify positive occurrences. It is computed as the ratio of true positive instances to the sum of true positives and false negatives. A fair assessment of a classifier's performance is provided by the F1 score, which is the harmonic mean of accuracy and recall. The ROC and AUC curve ultimately represent the chance that the model will correctly categorize a chosen positive instance greater than a chosen negative instance. As a whole, these measures provide a nuanced understanding of the models' effectiveness in distinguishing between cases with ASD and those without, which is essential for assessing their practicality.



(a)

(b)

FIGURE 3. **(a) MobileNetV2 - Confusion Matrix, (b)MobileNetV2 - ROC curve.**



(a)

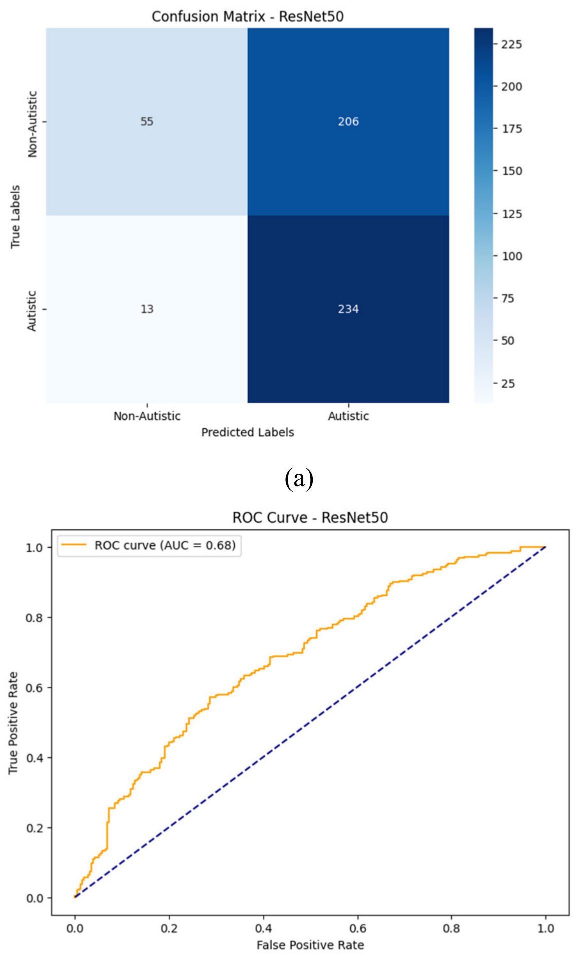
(b)

FIGURE 4. **(a) InceptionV3 - Confusion Matrix, (b) InceptionV3 -ROC curve.**

B. TRANSFER LEARNING MODEL PERFORMANCE

With 78.94% accuracy, MobileNetV2 as shown in Fig. 3 took the lead among the transfer learning models, followed by InceptionV3 (71.06%) as depicted in Fig. 4, and ResNet50 (56.19%) as presented in Fig. 5. ResNet50 demonstrated a comparatively high recall of 94.74%, indicating its efficacy in accurately detecting true positive cases, despite its lower accuracy. Its lesser precision in comparison to the other models, however, also suggested a larger false positive rate. Across a range of parameters, InceptionV3 performed more consistently, showing similar precision, recall, and F1 score.

ResNet50:



(a)

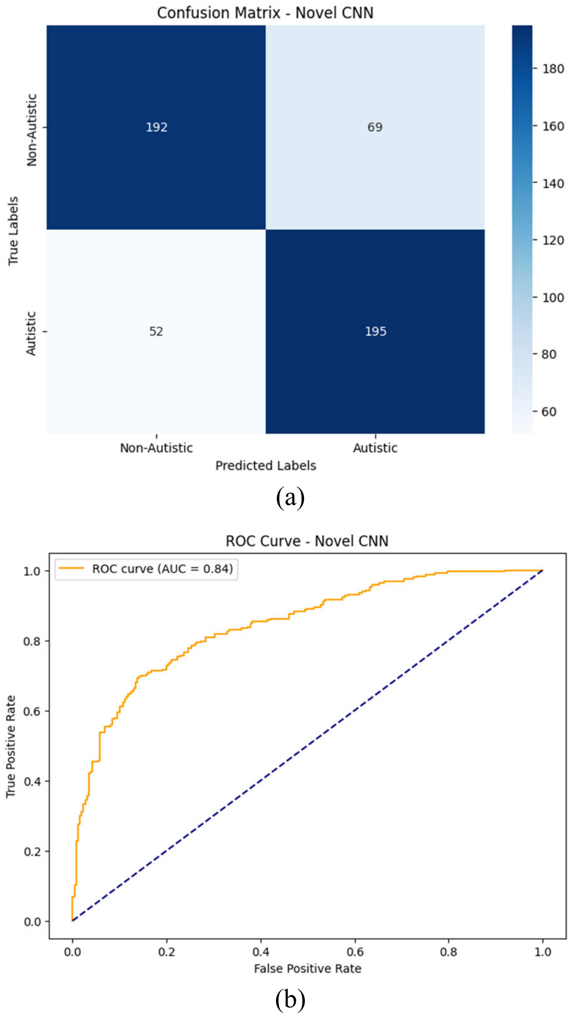
(b)

FIGURE 5. (a) ResNet50-Confusion Matrix, (b). ResNet50 ROC curve.

C. NOVEL CNN AND MULTIMODAL CONCATENATION MODEL PERFORMANCE

The novel CNN model, on the other hand, produced an accuracy of 76.18% and balanced recall and precision of 73.86% and 78.95%, respectively. This model presents a viable substitute by utilizing a special architecture designed for ASD detection applications.

Furthermore, with an astounding accuracy of 97.05%, the multimodal concatenation model substantially surpassed all other models. This model produced a strong classification performance by utilizing both the image data and the ASD evaluation scores. Its capacity to distinguish between ASD and non-ASD patients is demonstrated by its excellent precision, recall, and F1 score, which makes it a viable method for the real world. The graphical representation is depicted in Fig. 6 for the novel CNN model, Fig. 7, Fig. 8, and Fig.9 for the multimodal concatenation model.



(a)

(b)

FIGURE 6. (a) Novel CNN - Confusion Matrix, Novel CNN - (b) ROC curve.

Multimodal Concatenation Model:

VI. INFERENCE

Based on the transfer learning models' performance study, MobileNetV2 outperforms the others with an accuracy of 78.94%, InceptionV3 (71.06%), and ResNet50 (56.19%). ResNet50 exhibits a noteworthy recall of 94.74%, suggesting its usefulness in properly recognizing true positive cases, despite its lower accuracy. Its reduced accuracy, meanwhile, points to a larger false positive rate. On the other hand,

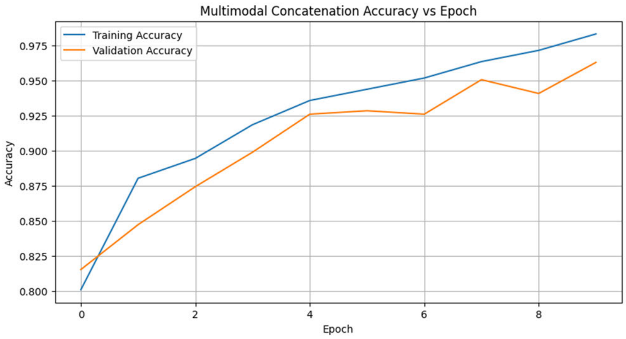


FIGURE 7. Multimodal concatenation accuracy vs epoch plot.

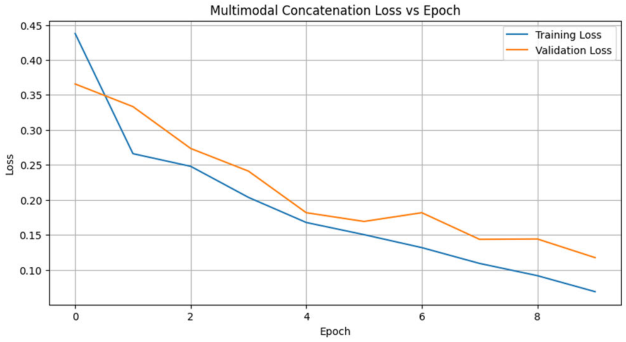
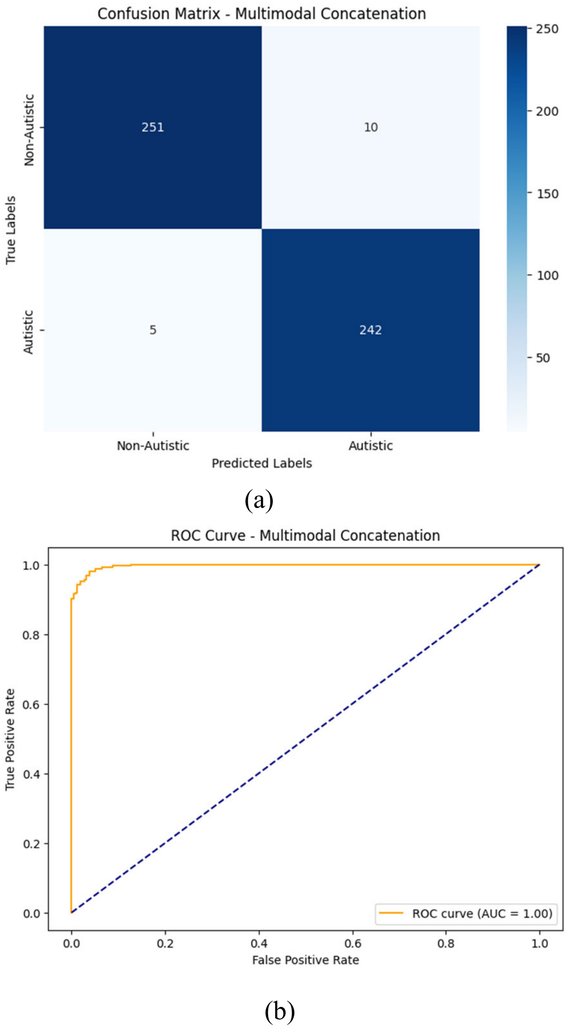


FIGURE 8. Multimodal concatenation loss vs epoch plot.

InceptionV3 performs more consistently across a range of criteria. A promising substitute, the innovative CNN model designed for ASD detection achieves an accuracy of 76.18% with balanced recall and precision. With an accuracy of 97.05%, the multimodal concatenation model remarkably beats all other models, demonstrating its strong classification performance by utilizing both image data and ASD evaluation ratings. The consolidated performance of these models is presented in Table 2 and Table 3. The higher performance of the multimodal concatenation model is attributed to its capacity to synergistically incorporate information from both picture data and scores from the ASD exam. The model's discriminative power is improved by incorporating characteristics from other modalities, which provide a more thorough grasp of the underlying data distribution. This method allows for more accurate and reliable ASD detection since it captures complex interactions between several data kinds, in contrast to single-modal models. Additionally, by utilizing numerous modalities, the model can more effectively adjust to changing data distributions, guaranteeing consistent performance across various circumstances and datasets. As a result, a potential method for detecting ASD that offers improved reliability in clinical settings is the multimodal concatenation model.

These results underscore the importance of exploring diverse model architectures and data modalities for ASD detection. While pre-trained features are leveraged by transfer learning models to give a strong foundation, unique architectures like the innovative CNN offer customized solutions. Moreover, multimodal fusion [29] incorporates complementary data from several sources to improve classification accuracy. Overall, the findings highlight the potential of deep learning approaches for enhancing ASD diagnosis and prognosis.



(a)

(b)

FIGURE 9. (a) Multimodal Concatenation Model - Confusion Matrix, (b). Multimodal Concatenation Model.

TABLE 2. Evaluation metrics.

Model	Precision	Recall	F1-Score	AUC
ResNet50	0.53	0.94	0.68	0.68
InceptionV3	0.67	0.77	0.72	0.78
MobileNetV2	0.77	0.79	0.78	0.86
Novel CNN	0.73	0.78	0.76	0.84
Multimodal Concatenation Model	0.96	0.97	0.96	0.99

TABLE 3. Evaluation metrics (Accuracy).

Model	Accuracy
ResNet50	0.56
InceptionV3	0.71
MobileNetV2	0.78
Novel CNN	0.76
Multimodal Concatenation Model	0.97

VII. CONCLUSION

In conclusion, this study shows that novel CNN architecture, multimodal concatenation techniques, and transfer learning models are all useful for diagnosing autism spectrum disorder (ASD). Through extensive experimentation, it is demonstrated that the multimodal concatenation model outperforms the other models, achieving remarkable accuracy and reliability in ASD classification. This model incorporates images with ADOS scores from ASD evaluations. Subsequent research endeavors may concentrate on optimizing the design of the multimodal model, investigating supplementary data modalities, and augmenting interpretability using advanced methodologies such as attention mechanisms. Further research into the suggested models' adaptability to a variety of datasets and demographics would be helpful for practical implementation. Furthermore, including feedback mechanisms and real-time monitoring in the models should improve their utility in clinical settings. Alongside this a web interface that enables teachers and parents to upload the student's image and test scores and having the model predict the level of autism in the child can be implemented in future works. Overall, this study highlights and lays a solid foundation for advancing ASD detection methodologies and underscores the potential of multimodal approaches in enhancing diagnostic precision and clinical results.

ACKNOWLEDGEMENT

The authors would like to take this opportunity to thank the management of Vellore Institute of Technology for providing the APC and encouragement to carry out this work.

REFERENCES

[1] A. H. Al-Qadri, W. Zhao, M. Li, M. H. Al-Khresheh, and A. Boudouaia, "The prevalence of the academic learning difficulties: An observation tool," *Heliyon*, vol. 7, no. 10, Oct. 2021, Art. no. e08164.

[2] A. M. Hayes, E. Dombrowski, A. H. Shefcyk, and J. Bulat, *Learning Disabilities Screening and Evaluation Guide for Low-and Middle-Income Countries*. RTI Press, 2018.

[3] V. Vilasini, B. B. Rekha, V. Sandeep, and V. C. Venkatesh, "Deep learning techniques to detect learning disabilities among children using handwriting," in *Proc. 3rd Int. Conf. Intell. Comput. Instrum. Control Technol. (ICICICT)*, Aug. 2022, pp. 1710–1717.

[4] N. P. G. Seshadri, S. Agrawal, B. K. Singh, B. Geethanjali, V. Mahesh, and R. B. Pachori, "EEG based classification of children with learning disabilities using shallow and deep neural network," *Biomed. Signal Process. Control*, vol. 82, Apr. 2023, Art. no. 104553.

[5] J. M. David and K. Balakrishnan, "Machine learning approach for prediction of learning disabilities in school-age children," *Int. J. Comput. Appl.*, vol. 9, no. 11, pp. 7–14, Nov. 2010.

[6] C. Song, Z.-Q. Jiang, L.-F. Hu, W.-H. Li, X.-L. Liu, Y.-Y. Wang, W.-Y. Jin, and Z.-W. Zhu, "A machine learning-based diagnostic model for children with autism spectrum disorders complicated with intellectual disability," *Frontiers Psychiatry*, vol. 13, Sep. 2022, Art. no. 993077.

[7] S. Santhiya and C. S. KanimozhiSelvi, "A study on dyslexia detection using machine learning techniques for checklist, questionnaire and online game based datasets," *Appl. Comput. Eng.*, vol. 5, no. 1, pp. 837–842, May 2023.

[8] M. V. Chakraborty, "A survey paper on learning disability prediction using machine learning," *Int. J. Inf. Comput. Sci.*, vol. 6, no. 5, pp. 481–485, 2019.

[9] S. Kaisar, "Developmental dyslexia detection using machine learning techniques : A survey," *ICT Exp.*, vol. 6, no. 3, pp. 181–184, Sep. 2020.

[10] T. S. Poornappriya and R. Gopinath, "Application of machine learning techniques for improving learning disabilities," *Int. J. Elect. Eng. Technol. (IJEET)*, vol. 11, no. 10, pp. 392–402, 2020.

[11] A. Loizou and Y. Laouris, "Developing prognosis tools to identify learning difficulties in children using machine learning technologies," *Cognit. Comput.*, vol. 3, no. 3, pp. 490–500, Sep. 2011.

[12] P. Dhamal and S. Mehrotra, "Deep learning approach for prediction of learning disability," in *Proc. 5th Int. Conf. Intell. Syst., Metaheuristics Swarm Intell.*, Apr. 2021, pp. 77–83.

[13] M. M. Hanumanthappa and A. Sangamithra, "Intelligent predicting learning disabilities in school going children using fuzzy logic K mean clustering in machine learning," *Int. J. Recent Technol. Eng. (IJRTE)*, vol. 8, no. 4, pp. 1694–1698, Nov. 2019.

[14] R. U. Khan, J. L. A. Cheng, and O. Y. Bee, "Machine learning and dyslexia: Diagnostic and classification system (DCS) for kids with learning disabilities," *Int. J. Eng. Technol.*, vol. 7, nos. 3–18, pp. 97–100, 2018.

[15] T. Akter, M. S. Satu, M. I. Khan, M. H. Ali, S. Uddin, P. Lió, J. M. W. Quinn, and M. A. Moni, "Machine learning-based models for early stage detection of autism spectrum disorders," *IEEE Access*, vol. 7, pp. 166509–166527, 2019.

[16] S. Parmar and C. Paunwala, "Early detection of dyslexia based on EEG with novel predictor extraction and selection," *Discover Artif. Intell.*, vol. 3, no. 1, p. 33, Oct. 2023.

[17] N. Ahire, R. N. Awale, and A. Wagh, "Development of EEG-based identification of learning disability using machine learning algorithms," in *Data Modelling and Analytics for the Internet of Medical Things*. Boca Raton, FL, USA: CRC Press, 2023, pp. 141–152.

[18] M. A. Mooney, C. Neighbor, S. Karalunas, N. F. Dieckmann, M. Nikolas, E. Nousen, J. Tipsord, X. Song, and J. T. Nigg, "Prediction of attention-deficit/hyperactivity disorder diagnosis using brief, low-cost clinical measures: A competitive model evaluation," *Clin. Psychol. Sci.*, vol. 11, no. 3, pp. 458–475, May 2023.

[19] M. Garcia-Argibay, Y. Zhang-James, S. Cortese, P. Lichtenstein, H. Larsson, and S. V. Faraone, "Predicting childhood and adolescent attention-deficit/hyperactivity disorder onset: A nationwide deep learning approach," *Mol. Psychiatry*, vol. 28, no. 3, pp. 1232–1239, Mar. 2023.

[20] M. Maniruzzaman, J. Shin, and M. A. M. Hasan, "Predicting children with ADHD using behavioral activity: A machine learning analysis," *Appl. Sci.*, vol. 12, no. 5, p. 2737, Mar. 2022.

[21] I. Bhatti, M. Tariq, Y. Hayat, A. Tariq, and S. Rasool, "A multimodal affect recognition adaptive learning system for individuals with intellectual disabilities," *Eur. J. Sci., Innovation Technol.*, vol. 3, no. 6, pp. 346–355, 2023.

[22] R. Y.-Y. Chan, C. M. V. Wong, and Y. N. Yum, "Predicting behavior change in students with special education needs using multimodal learning analytics," *IEEE Access*, vol. 11, pp. 63238–63251, 2023.

[23] J. Han, G. Jiang, G. Ouyang, and X. Li, "A multimodal approach for identifying autism spectrum disorders in children," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2003–2011, 2022.

[24] J. Chen, M. Liao, G. Wang, and C. Chen, "An intelligent multimodal framework for identifying children with autism spectrum disorder," *Int. J. Appl. Math. Comput. Sci.*, vol. 30, no. 3, pp. 435–448, 2020.

[25] F.-L. Zhu, S.-H. Wang, W.-B. Liu, H.-L. Zhu, M. Li, and X.-B. Zou, "A multimodal machine learning system in early screening for toddlers with autism spectrum disorders based on the response to name," *Frontiers Psychiatry*, vol. 14, Jan. 2023, Art. no. 1039293.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[27] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[29] D. Johnston, H. Egermann, and G. Kearney, "Measuring the behavioral response to spatial audio within a multi-modal virtual reality environment in children with autism spectrum disorder," *Appl. Sci.*, vol. 9, no. 15, p. 3152, Aug. 2019.

[30] C. Gamaethige, U. Gunathilake, D. Jayasena, H. Manike, P. Samarasinghe, and T. Yatanwala, "'SenseA'—Autism early signs and pre-aggressive detector through image processing," in *Proc. Asia Model. Symp. (AMS)*, Dec. 2017, pp. 125–130.

SUSEELA SELLAMUTHU received the B.E. degree in CSE and the M.E. degree (Hons.) in software engineering, in 2002 and 2009, respectively, and the Ph.D. degree from the National Institute of Technology, Tiruchirappalli, in 2021. She is currently an Assistant Professor with the School of Computer Science Engineering, Vellore Institute of Technology, Chennai Campus. She has specialized in the domain of wireless multimedia sensor networks. She has around 19 years of teaching experience in the field of computing science and engineering. She has published more than thirty research papers, five articles, and two book chapters in reputed journals and has written two books. She has acted as a reviewer and has six patents granted in the domain of IoT and governance packages. Her research interest spans wireless networks, cyber-physical systems, computational intelligent algorithms, artificial intelligence, the Internet of Things, machine learning, and deep learning. She is a member of CSI and ISTE.

SHARON ROSE is currently pursuing the Bachelor of Technology degree in computer science with specialization in artificial intelligence and machine learning with VIT Chennai. She expresses a keen interest in pursuing challenging tasks to apply and enhance her skills in software development. She has a good foundation in artificial intelligence and machine learning and is well-equipped with knowledge in the domains of deep learning and natural language processing, which encompass her main field of interest in research.