ACCSCIENCE PUBLISHING

Journal of Clinical and Translational Research

ORIGINAL ARTICLE

A dual-modal approach for detecting and classifying autism spectrum disorder using behavioral and facial features

M. S. Sankari* and **A. Kannammal**

Department of Electronics and Communication, Faculty of Electronics and Communication Engineering, PSG College of Technology, Coimbatore, Tamil Nadu, India

Abstract

Background: Autism spectrum disorder (ASD) is a neurodevelopmental condition that significantly affects social connection, behavior, and knowledge acquisition. Despite increasing global prevalence, timely diagnosis remains challenging due to heterogeneity in clinical presentation. **Aim:** The aim of this study is to develop a dual-modal framework for early detection of ASD by analyzing behavioral assessment and image data. **Methods:** The proposed framework consists of two independent yet complementary modules. In the behavioral module, questionnaire responses and assessment data were analyzed using an artificial neural network classifier to predict the likelihood of ASD. In the visual module, facial images were analyzed using a DenseNet121-based deep learning model with transfer learning to detect ASD-related traits. Each module independently estimates ASD probability and categorizes severity levels. **Results:** The DenseNet121 model achieved strong performance in image-based ASD detection, with 91.16% (95% confidence interval [CI]: 87.8–94.2) accuracy, 91.2% (95% CI: 87.8–94.2) sensitivity, and 89.8% specificity, including when trained on a relatively small dataset. Independent training of the two modules may improve robustness and reduce modality-specific bias. **Conclusion:** The proposed framework demonstrates potential for enhancing early ASD detection using dual modalities. The findings support the use of deep learning-based approaches to improve detection accuracy. **Relevance for patients:** Early screening of ASD can facilitate timely interventions and personalized care strategies. This method offers a noninvasive, data-driven approach that may support caregivers and healthcare systems in informed decision-making, ultimately benefiting individuals with ASD and their families.

Keywords: Behavioral features; Facial features; Artificial neural network; Transfer learning; Machine learning; Deep learning model; Severity prediction

***Corresponding author:**
M. S. Sankari
(227rlfx01@psgtech.ac.in)

Citation: Sankari MS, Kannammal A. A dual-modal approach for detecting and classifying autism spectrum disorder using behavioral and facial features. *J Clin Transl Res*. 2026;12(1):47-61.
doi: 10.36922/JCTR025400068

Received: September 30, 2025

1st revised: October 27, 2025

2nd revised: December 20, 2025

Accepted: January 15, 2026

Published online: February 5, 2026

1. Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by challenges in language development, cognitive processing, social interaction, and communication,[1] along with reduced empathy and sympathy.[2] Symptoms typically emerge in early childhood, often within the first 3 years of life. A review of data from 2000 through 2014 on 8-year-old children in selected U.S. communities by the

CDC's autism and developmental disabilities monitoring Network showed that ASD prevalence estimates rose more than 150% over this period, from 6.7/1,000 in 2000 to 16.8/1,000 in 2014.[3] Traditionally, diagnostic frameworks (e.g., the international classification of diseases-11[4]) and standardized instruments (e.g., the autism diagnostic observation schedule and the childhood autism rating scale [CARS][5]) are designed for individuals across a wide range of ages and developmental levels, from toddlers to adults. The Social Communication Questionnaire,[6] intended for children aged 4 years and above, consists of 40 items completed by parents under the guidance of a trained clinician. The Autism Spectrum Quotient (AQ)[7] is used to measure autistic traits in the general population and supports the early identification of individuals who may require further assessment. However, the diagnosis remains challenging, especially for marginal case disorders.

The diagnostic statistical manual, Fifth Edition (DSM-5) is a standard classification of mental disorders and serves as a primary diagnostic reference for mental health professionals. As per the DSM-5 criteria,[8,9] ASD is categorized into three functional levels: Level 1, Level 2, and Level 3. Health professionals determine the assigned levels based on the individual's need for outside assistance in their daily routine. Additionally, individuals with ASD may experience issues related to sensory processing difficulties and cognitive challenges. Therefore, health professionals typically diagnose using multiple modalities, such as monitoring facial[10] and behavioral features, assessing motor skills, analyzing eye movements and scan paths, and conducting neuroimaging studies.[11] However, these clinical approaches are often time consuming, costly, and may also be challenging due to communication difficulties experienced by individuals with ASD.

Previous studies have indicated that certain facial features, including a broader upper facial region, increased inter-ocular distance, a pronounced philtrum, and a comparatively shorter mid-face area involving the cheeks and nose, may be associated with ASD. These insights have driven growing interest in facial image-based screening methods. Although convolutional neural networks (CNNs) pre-trained on facial landmarks are increasingly used for ASD identification, current studies present key shortcomings. In particular, limited attention has been paid to the problem of overfitting, which raises concerns about model generalization to diverse populations. Furthermore, most research remains confined to binary classification of ASD versus non-ASD, offering little information about the severity of autistic traits. However, to achieve an acknowledged outcome, relying on a single data modality, such as electroencephalography (EEG),[12] magnetic resonance imaging,[13] facial expressions,[14] or voice data attributes,[15] is insufficient. To the best of our knowledge, most existing studies employing machine learning (ML) or deep learning (DL) techniques have primarily relied on a single modality, and the exploration of ASD severity prediction remains limited. In the current study, by keeping the modalities independent, we aim to ensure that when one model performs sub-optimally on certain samples, the other can still provide reliable classification.

The primary contributions of this research are as follows:

(i) The system employs an artificial neural network (ANN) classifier for analyzing questionnaire responses based on behavioral and demographic data and a fine-tuned DenseNet121 model for image-based data, enabling early detection and classification of ASD.
(ii) The system further predicts severity levels based on both questionnaire responses and image-based features, thereby assisting in suggesting treatment plans.

The remainder of the paper is organized as follows: the next section, Literature Review, presents the foundation upon which our study is built, followed by the Methodology section that describes the dataset employed, the working environment, and the methods used for experimentation, and the Results and Discussion section presents our research findings. Finally, we conclude our paper with the Conclusion section, highlighting the importance of our findings, future directions, and key takeaways.

2. Literature review

In this section, we discuss related research performed on the detection and classification of ASD. Researchers have explored various supervised, unsupervised, and reinforcement learning approaches across diverse data modalities, including behavioral features, neuroimaging, facial expressions, eye tracking, EEG, and voice data. These studies also highlight key challenges, such as data scarcity and the need for interpretability, while emphasizing the potential of ML and DL in advancing ASD research. Moreover, several systematic reviews have examined ML and data mining techniques, as well as feature optimization strategies for ASD prediction.[16-18] Singh *et al.*[19] investigated diagnostic mechanisms for ASD by extensively evaluating various ML models to identify the most significant indicators of ASD in toddlers. Their study initially developed a neural network-based classification model, which was later complemented by experiments using a random forest (RF) classifier.

Hassan and Taher[20] analyzed data from 515 children with autism and applied multiple classification algorithms, reporting high accuracy and F1 scores, with the ANN achieving the best performance. Similarly, Uddin[21] examined different ML classifiers for ASD detection across various age groups, incorporating feature optimization methods. The study used two datasets consisting of 292 and 704 records, respectively, each with 21 attributes. Notably, the RF model achieved 100% accuracy in early ASD diagnosis.

Talukdar *et al.*[22] evaluated ML models on toddler and adolescent datasets using the quantitative checklist for autism in toddlers-10 items and AQ-10 screening tools. The RF classifier achieved the highest accuracy, reporting 93.69% on the toddler and 93.33% on the adolescent dataset, without employing feature selection methods. Bawa *et al.*[23] evaluated various ML algorithms for ASD detection across different age groups. Logistic regression (LR) achieved 94.3% accuracy in children and 99% in adolescents, while support vector machine (SVM) reached 98.5% in adults.

Haque *et al.*[24] evaluated LR, SVM, K-nearest neighbors, decision tree, and RF for predicting ASD in children and toddlers. The results showed that the toddler dataset achieved a mean intersection over union of 100 with SVM and 99.80 with LR. Jahanara and Padmanabhan[25] explored facial feature identification in children using a CNN with a transfer learning approach, using 1,167 samples labeled as autistic or non-autistic. The CNN model, based on the Visual Geometry Group VGG19 architecture, achieved 96% accuracy.

Akter *et al.*[26] proposed a model based on 2,936 facial images using a transfer learning architecture. The evaluation included performance metrics, such as accuracy, sensitivity, false positive rate (FPR), G-mean, F-measure, and false negative (FN) rate. The framework integrated 17 different classifiers, encompassing both ML and DL models. Among them, MobileNet-V1 combined with k-means clustering achieved the highest accuracy of 83%. Alam *et al.*[27] applied deep CNN-based transfer learning approaches using facial images for ASD detection. The modified Xception model achieved 95% accuracy on the test set through empirical evaluation with optimizer selection and hyperparameter tuning.

Hriti *et al.*[28] proposed a classification method that integrates both visual and behavioral data modalities, collected from the same participants, and aligned by generating common attributes and grouped into sub-classes. The ANN and MobileNet models were employed, achieving an accuracy of 97.57%. Rahman and Subashini[29] studied facial photos of autistic children, analyzing 2,936 colored facial images. Various pre-trained CNN models, including MobileNet, Xception, EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2, were used as feature extractors, and a deep neural network was utilized for classification.

Anjum *et al.*[30] used MobileNet, Xception, VGG16, VGG19, and EfficientNet as feature extractors and LR as a binary classifier, achieving 88.33% test accuracy. Our study adopts a dual-modal approach to enhance ASD detection and classification by jointly analyzing image data and the questionnaire responses. The framework employs a fine-tuned DenseNet121 model and an ANN classifier model, ensuring computational efficiency while providing a comprehensive assessment.

3. Methodology

The general architecture of the proposed system, illustrated in Figure 1, leveraged both modalities to capture different aspects of ASD in children using ML techniques. This dual-pronged approach for ASD detection, classification, and severity prediction comprised two key modules: the behavioral module, which utilized questionnaire-based data, and the visual module, which employed image-based data for investigation, experimentation, and analysis.

3.1. Questionnaire data analysis sub-module

The dataset titled "ASD Children Traits"[31] was obtained from the Autism Research Group, University of Arkansas (Computer Science Department), and is publicly available via Kaggle online open-source repository (https://www.kaggle.com/code/mohammedabdelaleem/gathered-asd-datasets). It comprises 1,985 cases of children with 28 behavioral and developmental features. According to repository metadata, the dataset was last updated between 2021 and 2022. The dataset was curated and used for research purposes and contains self-reported or parent/caregiver-reported screening data. Only cases with sufficient behavioral and developmental features were considered to ensure reliable model training and evaluation, comprising several key attributes, including the AQ, assessed through columns A1 to A10, the Social Responsiveness Scale, age (in years), Q-CHAT 10 score, CARS score, speech delay disorder, learning disorder, genetic disorders, depression, global developmental delay/intellectual disability, social/behavioral issues, anxiety disorder, sex, ethnicity, jaundice, and family history of ASD. The CARS score differed from traditional survey-based measures, as it was derived from assessor observations. Each item was rated on a four-point scale from no abnormality to severely abnormal, and the total score was subsequently interpreted to determine the severity of ASD symptoms.

Children who had completed all items of the Q-CHAT 10[32] or AQ-10 behavioral screening questionnaires were included in the training. The AQ-10 scores ranged from 0 to 10; scores exceeding 6 suggested a diagnosis of ASD. Other attributes listed in the record's columns identified the factors that predominantly influenced the diagnosis

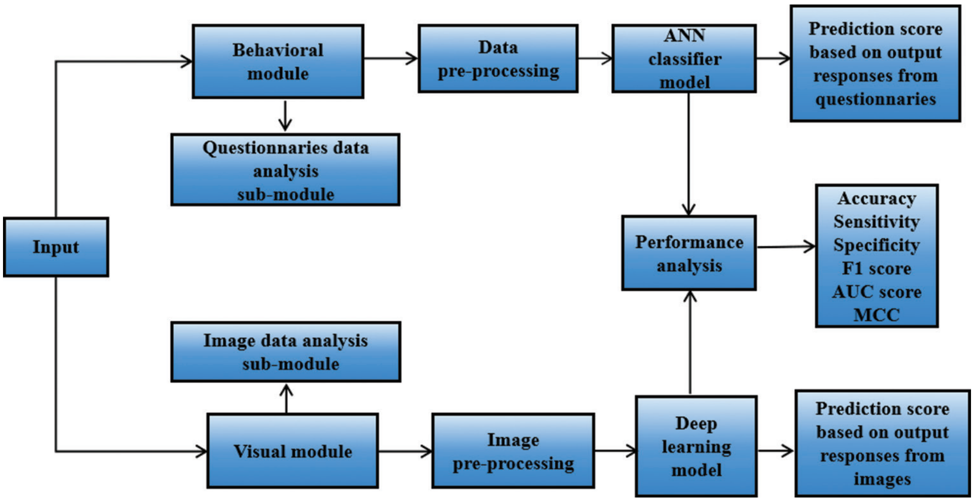

Figure 1. Proposed workflow diagram containing behavioral and visual modules for autism spectrum disorder detection and classification
Abbreviations: ANN: Artificial neural network; AUC: Area under the curve; MCC: Matthews correlation coefficient.

of ASD. The assessment was carried out by the health professional who administered the 10-item questionnaire and scored it on a 5-point scale in the Q-CHAT 10 method. Then, the points for questions 1–9 and question 10 were summed separately. A total score >3 may indicate autism, provided it is confirmed using a comprehensive assessment. During an assessment, responses to ten items (A1–A10) were recorded using five options labeled A, B, C, D, and E. For the questions Q1–Q9, 1 point was assigned if the response fell under columns C, D, or E (Somes, Rarely, or Never). For Q10, 1 point was awarded for responses in columns A, B, or C (Many s a day, A few s a day, or A few s a week), while responses in columns D and E (Rarely or Never) scored 0. A cumulative score >3 indicated potential ASD traits. Overall, the diagnostic framework emphasized two primary domains: social interaction and restricted or repetitive behaviors. Any missing values in the selected features were handled using simple imputation methods (mean or mode, as appropriate). Children with known comorbid neurological disorders unrelated to ASD (if any reported in the dataset) were excluded from the analysis.

Data preparation involved managing missing values, addressing outliers, and handling both categorical and numerical variables. The behavioral module encompassed an ANN classifier to analyze the processed questionnaire data. It predicted the likelihood of ASD by evaluating the behavioral features and responses provided. ANNs shared similarities with the interconnected neural layers of the human brain. The input layer was connected to one or more hidden layers, which in turn were connected to the output layer. During model training, the connection weights and bias terms between neurons were typically initialized with small random values drawn from a uniform or normal distribution to break the symmetry among neurons and enable effective learning through backpropagation. The dataset included a variety of behavioral attributes that were crucial for indicating the early signs of ASD. For example, items such as "Does your child point to share an interest with you?" and "The child looks at you when you call his/her name?" assess social-communication behaviors and interaction deficit among children observed with ASD. Responses to these questions and similar ones were encoded for machine analysis to support further processing and computation, enabling the identification of patterns and decision-making based on the output.

Among the 1,985 records, 1,074 were labeled as ASD and 911 as normal individuals. The participants were predominantly male, comprising 72.9% of the sample, compared to their female counterparts. Participants' ages ranged from 1 to 18 years, with an average of 9.6 years; high participation was observed between 1 and 10 years. The ASD prevalence is based on ethnicity; 16 distinct ethnic groups took part in the test, primarily originating from Europe, Asia, and the Middle East. Among them, white Europeans accounted for the highest number.

Primarily, the dataset revealed missing values in the Qchat_10_Score, social/behavioral issues, and depression columns, which were handled using simple imputation techniques. The mode was calculated and used to replace missing values in the numerical columns. The column "Who_completed_the_test" has six unique values, and the column "Ethnicity" has 16 unique values. These categorical values were converted to numerical columns using one-hot encoding. Ordinal encoding was used for categorical variables with an inherent order, such as gender specification in the dataset. Label encoding was used exclusively for the class variables. Since the dataset attributes were measured on varying scales, normalization was performed using standard scaling techniques (Z-score normalization) to maintain a standard range.

A value was normalized using the standard scalar method as in Equation (1):

$$y = \frac{x - x_{mean}}{x_{std}} \tag{1}$$

Where x indicates the current value of the input x, x_{mean}, x_{std} represent the mean and standard deviation of x, respectively. The proposed system classified outcomes based on probability outputs from both the questionnaire and the visual modules, mapping them to four levels: normal, mild, moderate, and severe. We aim to ensure that the model provides meaningful results and can be used for future ASD diagnosis.

The confidence levels were derived from the prediction probabilities, and the categorization into severity levels was defined, as shown in Table 1. The severity assessment levels were aligned with DSM-5 criteria for social communication deficits, and their characteristics were defined as follows[33]: (i) Normal: A probability score ranging from 0.00 to 0.20, indicating typical developmental responses. There is no evidence of social communication impairment. (ii) Mild: A probability score between 0.21 and 0.50, indicating a mild likelihood of ASD. Some symptoms, such as unusual social responses or atypical attempts to make friends, may be present. Minimal support may be required. (iii) Moderate: A probability score between 0.51 and 0.7, indicating a moderate likelihood of ASD. This level is characterized by noticeable deficits in verbal and non-verbal social communication skills, and substantial support is required. (iv) Severe: Probability score >0.7, indicating a high likelihood of ASD. This stage is associated with severe deficits, including very limited initiation of social interaction, and requires very substantial support.

The ANN is a multi-layer feed-forward network,[34] where signals propagate from the input layer to the output layer during the forward pass. During this phase, the weights and biases were assigned random values to begin training,

Table 1. Classification of autism spectrum disorder severity levels based on prediction probability thresholds, aligned with Diagnostic Statistical Manual-5 criteria for social communication deficits and support requirements

Probability range	Severity level
0.00–0.20	Normal
0.21–0.50	Mild
0.51–0.70	Moderate
>0.70	Severe

and the output was generated as a probability-based prediction. These parameters were iteratively updated using optimization techniques to minimize prediction error during backpropagation. ANN played a central role in analyzing questionnaire data to predict and classify ASD within the proposed system.

The algorithm of the ANN classifier comprised seven steps:

(i) Input: Training dataset: $D = \{(xi, yi)\}_{i=1}^{M}$, where x_i are input features, y_i € {0,1}, and y_i are class labels

(ii) Initialize sequential model:

Add a dense layer with 128 neurons, rectified linear unit (ReLU) activation

Input dimension d, L2 regularization

Add a dense layer with 64 neurons, ReLU activation

L2 regularization

Add dense output layer, sigmoid activation

(iii) Define evaluation metrics:

(True positive [TP], true negative [TN], false positive [FP], FN, Accuracy, Precision, Recall, F1 score, Matthews correlation coefficient [MCC], area under the curve [AUC])

(iv) Configure early stopping:

Monitor accuracy with patience, p, restores best weights when stopping

(v) Compile model: Optimizer = Adam(η), Loss = Binary Crossentropy

(vi) Train model:

Fit model on (X_{train}, Y_{train})

For up to E epochs, with validation split ratio v

Shuffle data each epoch, apply early stopping

(vii) Evaluate model:

Evaluate trained model on (X_{test}, Y_{test})

Output performance metrics and loss

Output: Tested the ANN model and evaluation results.

The model design required tuning several hyperparameters, including network architecture, learning rate, optimizer, batch size, and number of epochs. Among these hyperparameters, batch size played a crucial role, as it directly affects the overall training, the length of each epoch, and the quality of the trained model.

3.2. Image data analysis sub-module

The dataset titled "Autism image data" was obtained from Kaggle online open repository and comprises 2,940 images of children labeled as ASD or non-ASD (https://

data.mendeley.com/datasets/f9dycfvwbt/3).[35] Only facial images of children aged 2–8 years, with sufficient quality, correct face orientation, and distinct facial visibility, were included to support accurate feature extraction and classification. The facial images are available in JPEG format, each with a resolution of 224 × 224 pixels. Approximately 89% of the images represent white children, while the remainder depict Black, East Asian, and other ethnic groups. Duplicated or mislabeled instances were eliminated to maintain dataset integrity. The sample exhibited a male-to-female ratio of approximately 3:1, which aligned with the higher prevalence of ASD reported in males. The distribution between the autistic and non-autistic categories was nearly balanced, ensuring the model was trained on a dataset with minimal class imbalance and supporting fair and reliable classification performance. The image analysis sub-module processed children's facial images to identify key features associated with ASD. The images underwent several pre-processing steps, including cropping and resizing to standardize dimensions for model compatibility while preserving the aspect ratio to avoid distortion. Additional steps, such as normalization and consistent face region detection, were applied to improve uniformity across images. A fine-tuned CNN was employed to analyze the facial images. It was pre-trained on large image datasets, like ImageNet, and was exceptionally adept at recognizing patterns in visual data. It extracted significant facial features, such as a flatter midface, a broader upper face, a shortened philtrum, wider-set eyes, and facial expressions, which were crucial for identifying early signs of ASD. The model predicted the likelihood of ASD by comparing the extracted features against those features typically observed in individuals diagnosed with ASD. The normalization process was used to address inconsistencies and involved scaling pixel values from general ranges, such as (0, 255), to a specific range, such as (0, 1). This specification stabilized and accelerated the training process and ensured uniformity across the dataset. The normalized pixel value of the image is defined as Equation (2).

$$\mathrm{I'}(\mathrm{x},\mathrm{y}) = \frac{\mathrm{I}(\mathrm{x},\mathrm{y}) - \mathrm{mean}(\mathrm{I})}{\mathrm{std\ dev}(\mathrm{I})} \tag{2}$$

Where I (x,y) is the pixel value at position (x, y), and the mean and standard deviation of pixel values are considered. We used a fine-tuned DenseNet121 model, pre-trained on a large image dataset using deep facial features, to enhance performance.

The DenseNet121 architecture was built using dense blocks, in which each layer received inputs from a preceding layer and contributed its own feature maps to subsequent layers.[36] It began with a convolutional layer comprising 64 filters of size 7 × 7 with a stride of 2, followed by a 3 × 3 max-pooling layer with a stride of 2 to reduce spatial dimensions. Non-linearity was introduced through the ReLU activation function. The fully connected head was replaced with a global average pooling layer, which transformed the feature maps into a single feature vector. A softmax activation function was then applied to generate class probabilities. Then, the DenseNet121 model was fine-tuned on a dataset of children's facial images. The early convolutional layers were frozen to preserve low-level feature extraction, while the backbone layers were updated during training to adapt to the classification task. The model's parameters were optimized using the cross-entropy loss function, as shown in Equation (3):

$$L = -\sum_{I=1}^{c} yilog(pi) \tag{3}$$

Where *yi* is the true label, and *pi* is the predicted probability for class *i*.

The model parameters were updated using the root mean squared propagation (RMSprop) optimization algorithm, as defined in Equation (4):

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v_{t+\varepsilon}}}.g_t \tag{4}$$

Where η is the learning rate, and ε is a small constant, v_t is the accumulated moving average of squared gradients at t, and $v_t = \beta v_{t-1} + (1-\beta) g_t{}^2$, β is the decay rate, g_t indicates the gradient at t.

Table 2 illustrates the layer-wise architecture of the DenseNet121 model for prediction and classification. The model began with an input layer (224 × 224 × 3), followed by an initial convolution and pooling stage. The input first passed through an initial convolution layer with a 7 × 7 kernel and a stride of 2, followed by a 3 × 3 max pooling layer to reduce spatial dimensions and capture low-level features. The network then proceeded through four dense blocks, each comprising multiple convolutional layers with 1 × 1 and 3 × 3 kernels, where each layer received inputs from all preceding layers to encourage feature reuse and mitigate the vanishing gradient problem. Between dense blocks, transition layers performed convolutional and pooling operations to reduce spatial resolution while preserving essential feature information. Then, the global average pooling condensed the feature maps into a fixed-length vector. This vector was flattened and fed into fully connected layers with ReLU activations, batch normalization, and regularization to improve

Table 2. Layer-wise architecture of the proposed DenseNet121 model with modifications for autism spectrum disorder image detection

Layer	Output size	Description
Input	224×224×3	RGB image input
Convolution	112×112	7×7 convolution, stride 2
Pooling	56×56	3×3 max pool, stride 2
Dense block (1)	56×56	(1×1 convolution, 3×3 convolution)×6
Transition layer (1)	56×56	1×1 convolution
	28×28	2×2 average pool, stride 2
Dense block (2)	28×28	(1×1 convolution, 3×3 convolution)×12
Transition layer (2)	28×28	1×1 convolution
	14×14	2×2 average pool, stride 2
Dense block (3)	14×14	(1×1 convolution, 3×3 convolution)×24
Transition layer (3)	14×14	1×1 convolution
	7×7	2×2 average pool, stride 2
Dense block (4)	7×7	(1×1 convolution, 3×3 convolution)×16
Global average pooling	1×1 × 1024	Reduces feature maps to a 1D vector
Batch normalization	1024	Normalize activations
Flatten	1024	Converts to a 1D vector
Dense (ReLU)	512	Fully connected layer, L2 regularization
Batch normalization	512	Normalize activations
Dense (ReLU)	512	Fully connected layer, L2 regularization
Dense (Softmax)	Num_classes	Output classification layer

Abbreviations: 1D: One-dimensional; ReLU: Rectified linear unit; RGB: Red, Green, Blue.

generalization. A softmax layer produced class probability scores for classification, enabling the model to assign confidence levels to each predicted class. The DenseNet121 model generated class probability scores, and the class with the highest probability was chosen as the final prediction for each test image.

3.3. ASD severity prediction

The ASD severity prediction was performed using two complementary models applied to independent datasets.[37] The fine-tuned DenseNet121 model processed each input image through convolutional, dense, and pooling layers, generating hierarchical feature representations that were transformed into softmax probabilities. These probabilities were then mapped into predefined severity categories (normal, mild, moderate, and severe). The predicted label, true label, probability score, and severity level were compiled into a result table and exported as a comma-separated values (CSV) file for further analysis. This workflow provided both a quantitative and interpretable assessment of ASD detection and severity estimation, while the model's dense connectivity enhanced accuracy through collective decision-making.

3.4. Experimental results evaluation

The experimental assessment of the proposed framework was performed using a publicly available dataset across different modalities. The findings reveal significant improvements in ASD detection, with greater predictive accuracy compared to existing state-of-the-art methods. The dataset employed in this study comprised two independent sources: children's questionnaire responses and facial image data. While the questionnaire responses capture behavioral attributes, the facial images provide visual features, enabling a dual-modal analysis.

3.5. Performance metrics

The performance evaluation employed several metrics, including accuracy, precision, recall (sensitivity), specificity, F1 score, MCC,[38,39] and AUC. Both the classification report and confusion matrix were analyzed to evaluate the model's predictions. The confusion matrix comprises four outcomes: TP, where an autistic individual is correctly identified; TN, where a non-autistic individual is correctly classified; FP, where a non-autistic person is incorrectly predicted as autistic; and FN, where an autistic individual is mistakenly classified as non-autistic. Quantitative evaluation was performed on the test dataset using the aforementioned metrics, along with receiver operating characteristic (ROC) analysis, as presented in Equations (5-10). Accuracy, in particular, represented the overall correctness of the model's predictions:

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \tag{5}$$

Precision measured the accuracy of positive predictions, primarily in autistic cases.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

Sensitivity measured the model's ability to correctly identify people with ASD symptoms.

$$Recall\,or\,sensitivity = \frac{TP}{TP + FN} \tag{7}$$

The F1 score is the harmonic mean of precision and recall. This metric balanced the importance of both precision and recall. Specificity indicated how effectively the model avoided classifying non-autistic individuals as autistic.

$$Specificity = \frac{TN}{TN + FP} \quad (8)$$

MCC served as a comprehensive metric that evaluated model performance across all classes.

$$MCC = \frac{(\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN})}{\sqrt{(\mathrm{TP} + \mathrm{FP}) \times (\mathrm{TP} + \mathrm{FN}) \times (\mathrm{TN} + \mathrm{FP}) \times (\mathrm{TN} + \mathrm{FN})}} \quad (9)$$

The MCC incorporated all elements of the confusion matrix, providing a more comprehensive and reliable evaluation. The ROC curve was generated by plotting the TP rate (TPR) against the FPR, with TPR on the y-axis and FPR on the x-axis.

$$\text{AUC score} = \frac{TPR + FPR + 1}{2} \quad (10)$$

4. Results and discussion

The DenseNet121 model achieved 91.16% accuracy in detecting ASD, highlighting its capability to identify ASD-related traits from facial image attributes. The ANN classifier achieved 85.1% accuracy in predicting ASD, demonstrating its effectiveness in leveraging behavioral attributes for ASD detection.

4.1. ANN model on questionnaire data

The ANN model was selected for analyzing questionnaire response data for its ability to capture complex nonlinear relationships among behavioral features that other models might overlook. Additionally, the architecture allowed the customization of hidden layers, activation functions, and dropout, offering flexibility to accommodate dataset growth or the inclusion of additional features. The experimentation and analysis were conducted in Google Colab, a cloud-based Jupyter Notebook service that provides access to graphics processing units. For reference, the local system specifications were as follows: Windows 10 operating system, Intel(R) Core(TM) i7-7700 CPU @ 3.6 GHz, 16 GB RAM, and a 64-bit ×64-based processor. Google Colab, an interactive cloud-based platform, was employed to facilitate data analysis. This interactive tool, powered by Python (version 3.12.12, Python Software Foundation, Netherlands) and operated on Microsoft Windows, facilitated data analysis and exploration. It was enhanced with Python libraries, including Numpy, Pandas, Matplotlib, Scikit Learn, Seaborn, and TensorFlow.

The data were processed with appropriate encoding methods, with randomization applied to select data points for training and testing, which were 1,588 and 397, respectively, by following a train and test dataset split ratio of 80:20. The dataset, consisting of 27 input features, was used to perform binary classification of instances into ASD or non-ASD categories. The model was trained for 30 epochs. The sequential application programming interface in Keras was used to define fully connected layers with Dense, specifying input and output dimensions, along with activation functions. ReLU was used for the input and hidden layers, while "sigmoid" was applied to the output layer. The Adam optimizer with a learning rate of 0.0001 was employed, and the binary cross-entropy loss function was used; L2 regularization was considered. Once the model was fine-tuned, its predicted output was compared with the actual target to evaluate performance metrics using the classification report.

The results of the prediction model across epochs for the training and validation sets are shown in Figure 2. Figure 2 illustrates how accuracy changed over epochs for training, validation, and test data, with accuracy scores on the y-axis and the number of training iterations on the x-axis. Similarly, in the loss plot, the y-axis indicates the loss value, and the x-axis denotes the number of training iterations.

The training accuracy (blue) began at approximately 0.55 and steadily increased, reaching above 0.85 by the 30^{th} epoch. Validation accuracy (orange) followed a similar upward trend, starting around 0.65 and improving to about 0.81 by the 30^{th} epoch. The narrow gap between the training and validation curves indicated minimal overfitting, while the flattening of both curves after 20 epochs suggested model convergence. Both training and validation loss decreased consistently across epochs, dropping from about 13.0 to 2.5 by epoch 30. The close alignment of the two loss curves further demonstrated good generalization and stable learning. Overall, the model achieved reasonable accuracy with consistent loss reduction, indicating effective training.

4.2. DenseNet121 model on image data

The DenseNet121 architecture was employed for image data owing to its effectiveness in alleviating the vanishing gradient problem. All experiments were implemented in Python using the Keras framework and executed in the Google Colab environment. Image pre-processing was followed by data augmentation through Keras's "Image Data Generator" to improve model generalization. Data augmentation strategies included random rotations (±5°), horizontal and vertical translations (up to 10% of image dimensions), and horizontal flipping. The dataset was partitioned into training, validation, and test subsets at 75:15:10. Model training was performed for 50 epochs using the RMSprop optimizer with a learning rate of 0.0001 and a batch size of 32. The results of the prediction model, based on accuracy and loss over the number of epochs for the training and validation sets, are shown in Figure 3.

The training accuracy increased rapidly, reaching approximately 99% by epoch 20, while the validation accuracy improved steadily during the first 10 epochs to 85–90%. The gap between training and validation accuracy widened significantly after about 15 epochs, indicating that the model memorized the training data but did not generalize well to unseen validation data. The training loss decreased smoothly, approaching zero by epoch 50. In contrast, the validation loss decreased until epoch 30, then stabilized and fluctuated around 1.0, showing no further improvement. The combination of near-zero training loss and plateaued validation loss indicates slight overfitting, as the model continued to improve on the training data without corresponding gains on the validation data.

The analysis of the ROC curve highlighted how the model balanced the trade-off between the TPR and the FPR. The AUC values presented in Figure 4 provide a comprehensive evaluation of the model's effectiveness. The graphical representation illustrates the performance of the classifiers, where the ANN model achieved an AUC of 0.9152, while the DenseNet121 model attained a higher AUC of 0.96, demonstrating reliable performance with an AUC well above 0.5.

The ROC curve of DenseNet121 shows better classification performance, as reflected in its higher AUC score of 0.96, compared to the ANN model. The model maintained a high TPR while minimizing FP, indicating a strong ability to distinguish between the two classes.

Table 3 presents the performance comparison of the ANN and DenseNet121 models on behavioral questionnaire responses and facial image data. The ANN model achieved a reasonable balance, with high specificity (90.4%) but lower sensitivity (80.5%), leading to more missed ASD cases. In contrast, the DenseNet121 model substantially improved sensitivity (91.2%) while maintaining specificity at 89.8%, resulting in higher overall accuracy (91.16%) and a stronger MCC (0.824). The improvement in MCC from 0.709 to 0.824 indicates that the model effectively distinguished between the ASD and non-ASD categories. All statistical analyses were performed in Python using libraries including

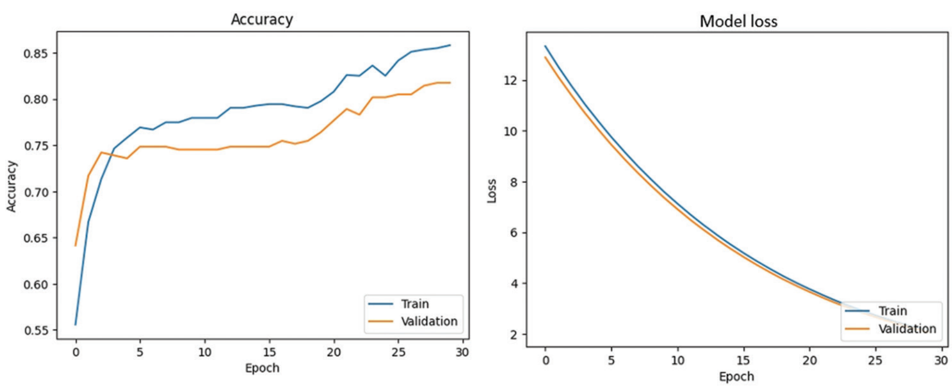

Figure 2. Accuracy and loss curves during artificial neural network training and validation, highlighting learning dynamics and model fit

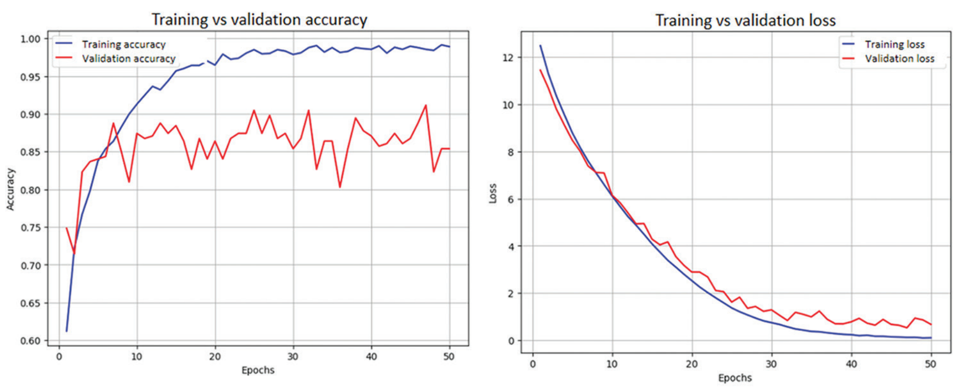

Figure 3. Training and validation accuracy and loss across epochs for the DenseNet121 model, illustrating performance trends

Table 3. Performance comparisons of the artificial neural network and DenseNet121 models on behavioral questionnaire responses and facial image data, showing key evaluation metrics

Dataset/features	Model	Key metrics			
		Precision (%)	Recall (%)	F1 score (%)	MCC
Behavioral questionnaire response	ANN accuracy: 85.1% (95% CI: 82.1–88.7)	90.4 (95% CI: 86.7–93.8)	80.5 (95% CI: 75.7–85.7)	85.4 (95% CI: 81.7–88.8)	0.709 (95% CI: 0.645–0.777)
Facial images	DenseNet121 accuracy: 91.16% (95% CI: 87.8–94.2)	91.1 (95% CI: 87.8–94.3)	91.2 (95% CI: 87.8–94.2)	91.1 (95% CI: 87.7–94.2)	0.824 (95% CI: 0.755–0.885)
Statistical significance	p=0.017 (p<0.05), improvement in accuracy~6% (95% CI: 1.1–10.9)				

Abbreviations: ANN: Artificial neural network; CI: Confidence interval; MCC: Matthews correlation coefficient.

numpy, scipy, and statsmodels. A two-sample Z-test for proportions was conducted to compare the performance of DenseNet121 and ANN. The key assumptions to consider are independence of observations, sufficient sample size, and adequate expected counts. It was used to compare the classification accuracy of two independent models and determine if the observed differences were statistically significant. This test relied on the normal approximation to the binomial distribution, which was valid when the sample sizes were sufficiently large (typically $n > 30$). The diagnostic accuracy of Densenet121 (91.16%, 268/294) was significantly higher than that of the ANN model (85.1%, 338/397). A two-sample Z-test for proportions showed that this improvement was statistically significant (Z = 2.38, p=0.017). Specifically, the effect size based on Cohen's h was small (h = 0.189), and the absolute improvement in diagnostic accuracy was approximately 6.0% (95% confidence interval: 1.1–10.9). As the DenseNet121 model demonstrated a marked reduction in FN, thereby enhancing early detection reliability, it is an essential factor in ASD screening. The higher precision (91.1%) and recall (91.2%) achieved by the DenseNet121 model further support its clinical utility. In addition to the statistically significant improvement, the DenseNet121 model demonstrated clinically meaningful gains, particularly through a 10.7% increase in recall (91.2% vs. 80.5%), indicating a marked reduction in FN. This implies enhanced sensitivity to ASD traits, which is critical for minimizing missed diagnoses and promoting timely clinical assessment.

4.3. Result of ASD severity prediction based on ANN and DenseNet121 models

The predicted label, true label, probability score, and severity category were compiled into a result table and exported as a CSV file for further analysis. This workflow offered a systematic, quantitative, and interpretable assessment of both ASD classification and severity estimation. By incorporating the model's collective decision-making process, the approach enhanced both predictive accuracy and interpretability. In the results, the dataset labels "1" and "0" denote ASD and non-ASD individuals, respectively, following the convention provided by the dataset source.

Table 4. Artificial neural network-based autism spectrum disorder prediction and severity assessment results

Index	True label	ASD probability	Predicted label	Severity level
795	1	0.803926	1	Severe
800	1	0.838478	1	Severe
127	0	0.305518	0	Mild
921	1	0.829160	1	Severe
1572	1	0.508315	1	Moderate
801	1	0.749940	1	Severe
304	1	0.846468	1	Severe
1925	0	0.357174	0	Mild
1135	1	0.519837	1	Moderate
200	1	0.903300	1	Severe

Abbreviation: ASD: Autism spectrum disorder.

The sample predictions in Table 4 show that the model not only classified ASD reasonably well, but also assigned severity level (normal, mild, moderate, and severe) depending on the model's probability scores. The DenseNet121 model utilized dense layers to generate hierarchical feature representations, which were subsequently transformed into softmax probabilities. For example, extremely low probabilities were labeled normal, moderate as mild, and extremely high as severe, and the results were saved to a CSV file. The predicted ASD probabilities were used to assign severity levels, and the results were displayed both in a table and on randomly selected test images. Figure 5 illustrates the ASD severity estimation using the DenseNet121 model for random test images.

Table 5 shows ASD prediction results: the model outputs a probability (ASD probability), predicts a class (Predicted label), and assigns a severity level (normal, mild, severe) based on the probability value. The images show a sample output of an ASD classification model, where children's faces are labeled with predicted severity levels. The model

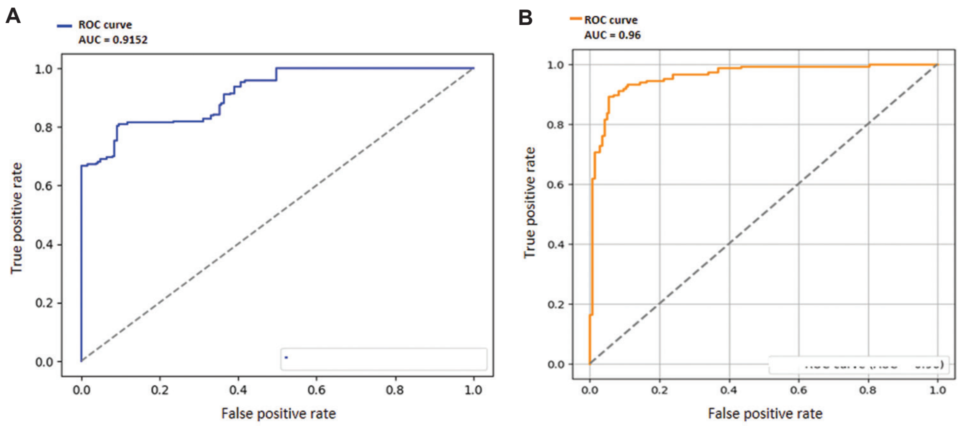

Figure 4. Receiver operating characteristic (ROC) curve with Area under the curve (AUC) illustrating the classification performance of the (A) Artificial neural network and (B) DenseNet121 model

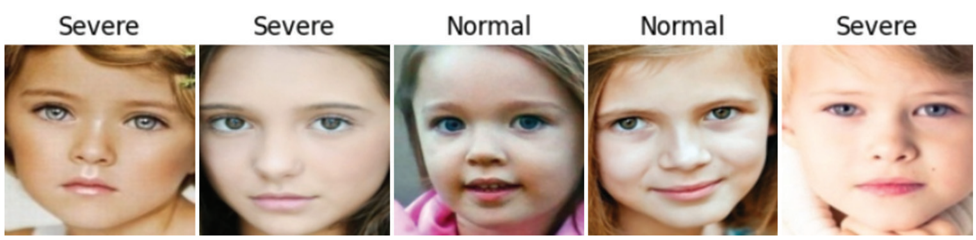

Figure 5. Autism spectrum disorder severity assessed using the DenseNet121 model. Results are presented as test-image predictions illustrated with corresponding confidence levels.

Table 5. DenseNet121-based autism spectrum disorder classification and severity assessment results

Index	True label	ASD probability	Predicted label	Severity level
0	Autistic	0.000230	0	Normal
1	Autistic	0.003804	0	Normal
2	Autistic	0.000059	0	Normal
3	Autistic	0.297353	0	Normal
4	Autistic	0.000020	0	Normal
5	Autistic	0.176527	0	Normal
6	Autistic	0.000212	0	Normal
7	Autistic	0.999946	1	Severe
8	Autistic	0.000032	0	Normal
9	Autistic	0.000438	0	Normal

Abbreviation: ASD: Autism spectrum disorder.

categorizes them into severe (high ASD probability) or normal (low ASD probability). It visually illustrates how the system interprets different faces and assigns severity levels of ASD risk.

Table 6 presents the findings of our approach compared with several existing studies reported in the literature. Previous studies on ASD detection, ML, and DL models have primarily focused on a single modality. Studies based on behavioral features (e.g., Vakadkar *et al.*,[40] Mohanty *et al.*,[42]) achieved accuracy ranging between 81% and 90%, with Mohanty *et al.*[42] reporting perfect specificity (100%). Image-based approaches proposed by Awaji *et al.*,[43] Khan *et al.*,[44] and Nawghare and Prasad[45] have achieved a competitive accuracy of 82–88%. However, these works typically lacked comprehensive reporting of sensitivity and specificity. In contrast, in our proposed dual-pronged method, the behavioral and facial image data processed with ANN and DenseNet121 demonstrated improved performance consistency. DenseNet121 achieves 91.16% accuracy, 91.2% sensitivity, and 89.8% specificity, surpassing most single-modality approaches. This finding highlights the advantage of leveraging complementary data modalities to enhance ASD detection and reduce biases inherent in individual feature spaces.

The current study evaluated behavioral and facial features independently. Each modality provided complementary insights: behavioral traits reflected self-reported or observed tendencies, while facial features offered objective visual cues. Importantly, our severity-based categorization approach extended beyond binary classification. The ANN model, trained on behavioral questionnaire responses, achieved an accuracy of 85.1%, precision of 90.4%, sensitivity of 80.5%, specificity of 90.4%, and an MCC of 0.709. The DenseNet121 model, trained on facial images, demonstrated superior performance with

Table 6. Comparative analysis of various classifier models with existing studies for autism spectrum disorder detection and classification

Reference	Type of data	Methodology	Metric value		
			Accuracy (%)	Sensitivity(%)	Specificity(%)
40	Behavioral attributes	KNN	90.52	NR	NR
		RF	81.52	NR	NR
41	Video data	VGG16	80.90	85.4	NR
42	Behavioral attributes	DNN	85.20	70.4	100.0
43	Facial image data	MobileNet	85.20	85.3	84.7
44	Facial image data	ResNet50	82.00	82.0	NR
		MobileNetV2	85.00	85.0	NR
45	Facial image data	Hybrid-model VGG16, RF	88.34	NR	NR
Proposed method	Behavioral data	ANN	85.10	80.5	90.4
	Facial image data	DenseNet121	91.16	91.2	89.8

Abbreviations: ANN: Artificial neural network; DNN: Deep neural network; KNN: K-nearest neighbors; NR: Not reported; RF: Random forest; VGG: Visual geometry group.

an accuracy of 91.16%, precision of 91.2%, recall of 91.2%, specificity of 89.8%, and an MCC of 0.824. The statistical analysis further confirmed the significance of this improvement, with a p=0.017 (p<0.05) and an approximate 6% increase in accuracy when using facial image features compared to behavioral features. These findings highlight that while both modalities provide valuable insights, facial image analysis contributes more strongly to ASD prediction. The observed increase in accuracy from 85.1% to 91.16% (p=0.017) highlights the improved performance of the facial image-based DenseNet121 model.

Clinically, this improvement reflects a substantial reduction in FN, ensuring that fewer at-risk children are overlooked during early screening and thereby supporting more timely intervention. The increased precision and recall also strengthen confidence in positive detection, enabling pediatricians and psychologists to effectively prioritize children for comprehensive diagnostic evaluation. Importantly, the proposed facial image model is not intended to replace behavioral assessments but to complement them. While behavioral questionnaires capture subjective, parent-reported traits, facial image models analyze objective, phenotype-related indicators. Furthermore, facial image-based systems hold promise for early, accessible screening, with deployment in mobile environments. Such automation may extend ASD screening to rural or resource-limited settings and reduce clinician workload.

We acknowledge potential limitations that may influence the generalization of our findings. The current study was conducted on a small sample drawn from a single dataset, which may not fully represent broader demographic diversity. Factors such as variations in ethnicity, lighting conditions, facial image resolution, and behavioral reporting bias could be confounding variables. The interpretation of the potential dual-pronged approach using behavioral and visual modalities is presented as a prospective direction rather than a definitive conclusion. At present, integrating both modalities may risk bias amplification. However, in our approach, if one modality exhibits bias or reduced sensitivity in certain cases, the other can provide compensatory evidence, thereby enhancing robustness. Furthermore, the use of severity-based thresholds enabled classification into normal, mild, moderate, and severe levels, extending the system beyond binary outcomes.

5. Conclusion

The proposed framework for ASD detection and classification in children demonstrates that DenseNet121 achieved 91.16% accuracy, 91.2% sensitivity, and 89.8% specificity in image-based prediction. While DenseNet121-based deep feature learning demonstrates better performance than other models, issues such as overfitting persist. DenseNet121 mitigates vanishing gradients by reusing features and reducing parameter complexity, thereby improving learning efficiency. By evaluating behavioral and facial modalities independently, this study enhances interpretability and identifies multimodal integration as a promising direction. Importantly, the framework progresses beyond binary ASD detection to include severity stratification, increasing its clinical applicability. However, the current evaluation is limited to a single publicly available dataset. Broader validation using larger, more diverse datasets with subject-independent testing will be essential to strengthen robustness and generalization. Future work should also investigate fusion strategies at the feature or decision level to achieve a unified multimodal prediction. The next step involves interpretable artificial intelligence techniques, such as gradient-weighted class activating mapping or attention roll-outs, to visualize which facial features influence classification. Clinicians can then better trust and understand how model predictions relate to observable ASD markers. Overall, our study provides a strong foundation for hybrid ASD screening systems that support early detection and intervention, demonstrating the potential of ML to deliver accessible and noninvasive clinical decision support.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: All authors
Formal analysis: Sankari M.S
Investigation: All authors
Methodology: Sankari M.S
Writing – original draft: Sankari M.S
Writing – review & editing: All authors

Ethics approval and consent to participate

This study was conducted retrospectively using publicly available human-subject data from the Kaggle open repository. Ethical approval was not required, as confirmed by the license provided with the open-access data.

Consent for publication

Not applicable.

Availability of data

The dataset used in this study is publicly available through the Kaggle open repository:
(i) https://www.kaggle.com/code/mohammedabdelaleem/gathered-asd-datasets.
(ii) https://data.mendeley.com/datasets/f9dycfvwbt/3.

References

1. Lord C, Cook EH Jr., Leventhal BL, Amaral DG. Autism spectrum disorders. *Neuron*. 2000;28(2):355-363.
doi: 10.1016/S0896-6273(00)00115-X
2. Verywell Health Staff. *Do People with Autism Lack Empathy*? Verywell Health; 2011. Available from: https://www.verywellhealth.com/do/people/with-autism-lack-empathy-259887 [Last accessed on 2025 Jul 10].
3. Baio J, Wiggins L, Christensen DL, *et al*. Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *MMWR Surveill Summ*. 2018;67(SS-6):1-23.
doi: 10.15585/mmwr.ss6706a1
4. Sujatha R, Aarthy SL, Chatterjee J, Alaboudi A, Jhanjhi NZ. A machine learning way to classify autism spectrum disorder. *Int J Emerg Technol Learn*. 2021;16(6):182-200.
doi: 10.3991/ijet.v16i06.21575
5. Randall M, Egberts KJ, Samtani A, *et al*. Diagnostic tests for autism spectrum disorder (ASD) in preschool children. *Cochrane Database Syst Rev*. 2018;7(7):CD009044.
doi: 10.1002/14651858.CD009044.pub2
6. Mashudi NA, Ahmad N, Noor NM. Classification of adult autism spectrum disorder using a machine learning approach. *IAES Int J Artif Intell*. 2021;10(3):743-750.
doi: 10.11591/ijai.v10.i3.743
7. Bala M, Ali MH, Satu MS, Hasan KF, Moni MA. Efficient machine learning models for early stage detection of autism spectrum disorder. *Algorithms*. 2022;15(5):166.
doi: 10.3390/a15050166
8. Wiggins LD, Rice CE, Barger B, *et al*. DSM-5 criteria for autism spectrum disorder maximizes diagnostic sensitivity and specificity in preschool children. *Soc Psychiatry Psychiatr Epidemiol*. 2019;54(6):693-701.
doi: 10.1007/s00127-019-01674-9
9. Ramana KS. Machine learning-based novel autism spectrum disorder screening. *Turk J Comput Math Educ*. 2021;12(3):4866-4879.
doi: 10.17762/turcomat.v12i3.4132
10. Is it Autism? *Facial Features that Show Disorder*. *CBS News*; 2011. Available from: https://www.cbsnews.com/pictures/is/it/autism/facial/features/that/show/disorder/9 [Last accessed on 2024 Jul 11].
11. Bhat S, Acharya UR, Adeli H, Bairy GM, Adeli A. Autism: Causative factors, early diagnosis and therapies. *Rev Neurosci*. 2014;25(6):841-850.
doi: 10.1515/revneuro-2014-0033
12. Ardakani HA, Taghizadeh M, Shayegh F. Diagnosis of autism disorder based on deep network trained by augmented EEG signals. *Int J Neural Syst*. 2022;32(11):2250046.
doi: 10.1142/S0129065722500461
13. Khodatars M, Shoeibi A, Sadeghi D, *et al*. Deep learning for neuroimaging-based diagnosis and rehabilitation of

autism spectrum disorder: A review. *Comput Biol Med*. 2021;139:104949.

doi: 10.1016/j.compbiomed.2021.104949

14. Tamilarasi FC, Shanmugam J. Convolutional neural network-based autism classification. In: *Proceedings of the 5th International Conference on Communication and Electronics Systems (ICCES)*. Coimbatore, India: IEEE; 2020. p. 1208-1212.

    doi: 10.1109/ICCES48766.2020.9137926

15. Mohanta A, Mittal VK. Analysis and classification of speech sounds of children with autism spectrum disorder using acoustic features. *Comput Speech Lang*. 2022;72:101287.

    doi: 10.1016/j.csl.2021.101287

16. Ahmed M, Hussain S, Ali F, *et al*. Summarizing recent developments in autism spectrum disorder detection using machine learning and deep learning techniques. *Appl Sci*. 2025;15(14):8056.

    doi: 10.3390/app15148056

17. Poonia A, Kumar S, Raghuwanshi G. Machine learning and deep learning for autism spectrum disorder discovery: A descriptive review. *J Comput Cogn Eng*. 2025;4(3):267-283.

    doi: 10.47852/bonviewJCCE52024757

18. Ding Y, Zhang H, Qiu T. Deep learning approach to predict autism spectrum disorder: A systematic review and meta-analysis. *BMC Psychiatry*. 2024;24(1):739.

    doi: 10.1186/s12888-024-05388-7

19. Singh A, Farooqui Z, Sattler B, Usua U, Helde M. Using Machine Learning Optimization to Predict Autism in Toddlers. In: *Conference: 11th Annual International Conference on Industrial Engineering and Operations Management*. 2021. p. 69206931.

    doi: 10.46254/an11.20211201

20. Hassan MM, Taher SA. Analysis and classification of autism data using machine learning algorithms. *Sci J Univ Zakho*. 2022;10(4):206-212.

    doi: 10.25271/sjuoz.2022.10.4.1036

21. Uddin KMM. A machine learning approach to predict autism spectrum disorder for children and adults using feature optimization. *Netw Biol*. 2023;13(2):37-46.

    doi: 10.17323/1996-7845.2023.2.37

22. Talukdar J, Gogoi D, Singh TP. Comparative assessment of machine learning classifiers for autism spectrum disorder detection. *Healthc Anal*. 2023;3:100178.

    doi: 10.1016/j.health.2023.100178

23. Bawa P, Kadyan V, Mantri A, Vardhan H. Investigating multiclass autism spectrum disorder classification using machine learning techniques. *e-Prime Adv Electr Eng Electron Energy*. 2024;8:100602.

    doi: 10.1016/j.eprime.2024.100602

24. Haque N, Islam T, Erfan M. An exploration of machine learning approaches for early autism spectrum disorder detection. *Healthc Anal*. 2025;7:100379.

    doi: 10.1016/j.health.2025.100379

25. Jahanara S, Padmanabhan S. Detecting autism from facial images. *Int J Adv Res Ideas Innov Technol*. 2021;7(2):219-225.

    doi: 10.13140/RG.2.2.35268.35202

26. Akter T, Ali MH, Khan MI, *et al*. Improved transfer learning-based facial recognition framework for early autism detection. *Brain Sci*. 2021;11(6):734.

    doi: 10.3390/brainsci11060734

27. Alam MS, Rashid MM, Roy R, Faizabadi AR, Gupta KD, Ahsan MM. Empirical study of autism spectrum disorder diagnosis using facial images and transfer learning. *Bioengineering (Basel)*. 2022;9(11):710.

    doi: 10.3390/bioengineering9110710

28. Hriti NS, Shaer K, Momin FMN, Mahmud H, Hasan MK. *Autism Classification using Visual and Behavioral Data*. medRxiv. [Preprint]; 2022.

    doi: 10.1101/2022.10.12.22281002

29. Rahman KK, Subashini MM. Identification of autism in children using static facial features and deep neural networks. *Brain Sci*. 2022;12(1):94.

    doi: 10.3390/brainsci12010094

30. Anjum J, Hia NA, Waziha A, Kalpoma KA. Deep Learning-Based Feature Extraction from Children's Facial Images for Autism Detection. In: *Proceedings of the Cognitive Models and Artificial Intelligence Conference*; 2024. p. 155-159.

    doi: 10.1145/3660853.3660888

31. Bargrizan A. *ASD Final [Dataset]*. Kaggle; 2025. Available from: https://www.kaggle.com/datasets/afarinbargrizan/asd-final [Last accessed on 2025 Jul 16].

32. Allison C, Matthews FE, Ruta L, *et al*. Quantitative checklist for autism in toddlers (Q-CHAT). A population screening study with follow-up: the case for multiple -point screening for autism. *BMJ Paediatr Open*. 2021;5:e000700.

    doi: 10.1136/bmjpo-2020-000700

33. Irmak E. COVID-19 disease severity assessment using CNN model. *IET Image Process*. 2021;15(8):1814-1824.

    doi: 10.1049/ipr2.12175

34. Madhiarasan M, Louzazni M. Artificial neural networks: Architecture, types, and forecasting applications. *J Electr Comput Eng*. 2022;2022:5416722.

    doi: 10.1155/2022/5416722

35. Thillaikarasi R, Kumaresan P. Autism spectrum disorder detection using facial traits: DFDK framework. *Mendeley Data*. 2025;2:V3.

doi: 10.17632/f9dycfvwbt.3

36. Zhou Q, Zhu W, Li F, Yuan M, Zheng L, Liu X. Transfer learning of the resnet-18 and densenet-121 model used to diagnose intracranial hemorrhage in CT scanning. *Curr Pharm Des*. 2022;28(4):287-295.

    doi: 10.2174/1381612828666220208165622

37. *Levels of Autism: ASD Levels of Severity*. Autism Speaks. Available from: https://www.autismspeaks.org/levels-of-autism [Last accessed on 2025 Sep 12].

38. Jacob SG, Sulaiman MMBA, Bennet B. Feature signature discovery for autism detection: An automated machine learning based feature ranking framework. *Comput Intell Neurosci*. 2023;2023:6330002.

    doi: 10.1155/2023/6330002

39. Raj S, Masood S. Analysis and detection of autism spectrum disorder using machine learning techniques. *Procedia Comput Sci*. 2020;167:994-1004.

    doi: 10.1016/j.procs.2020.03.402

40. Vakadkar K, Purkayastha D, Krishnan D. Detection of autism spectrum disorder in children using machine learning techniques. *SN Comput Sci*. 2021;2(5):386.

    doi: 10.1007/s42979-021-00773-8

41. Kojovic N, Natraj S, Mohanty SP, Maillart T, Schaer M. Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children. *Sci Rep*. 2021;11(1):15069.

    doi: 10.1038/s41598-021-94555-0

42. Mohanty AS, Parida P, Patra KC. Identification of autism spectrum disorder using deep neural network. *J Phys Conf Ser*. 2021;1921(1):012006.

    doi: 10.1088/1742-6596/1921/1/012006

43. Awaji B, Senan EM, Olayah F, *et al*. Hybrid techniques of facial feature image analysis for early detection of autism spectrum disorder based on combined CNN features. *Diagnostics (Basel)*. 2023;13(18):2948.

    doi: 10.3390/diagnostics13182948

44. Khan B, Bhatti SM, Akram A. Autism spectrum disorder detection in children using deep learning and facial images. *Bull Bus Econ (BBE)*. 2024;13(1):557-572.

    doi: 10.61506/01.00241

45. Nawghare P, Prasad J. Hybrid CNN and random forest model with late fusion for detection of autism spectrum disorder in toddlers. *MethodsX*. 2025;14:103278.

    doi: 10.1016/j.mex.2025.103278