International Journal of Technology and Engineering System (IJTES)
Vol 17. No.1 2026 Pp. 1-18
ISSN: 0976-1345
©gopalax Journals, Singapore
available at : www.gopalax.com
Paper Received: 01-11-2026
Paper Accepted: 28-11-2026

AutismSense: An Explainable Hybrid Deep Learning Framework Integrating Convolutional Neural Networks, Deep Neural Networks, and Intelligent Feature Fusion for Early Autism Spectrum Disorder Detection and Risk Prediction in Children

P.Poornima
Research Scholar
Department of Computer Science,
VISTAS, Chennai. India
poornimaamudha04@gmail.com

Dr. S. Jeyalaksshmi
Professor
Department of Computer Applications(UG),
VISTAS, Chennai. India
jlakshmi.scs@vistas.ac.in
pravija.lakshmi@gmail.com

Abstract
Early identification of Autism Spectrum Disorder (ASD) is essential for supporting timely developmental assessment and intervention in children. This study proposes **AutismSense**, an explainable hybrid deep learning framework integrating Convolutional Neural Networks (CNN), Deep Neural Networks (DNN), and attention-based intelligent feature fusion for early ASD risk prediction. The framework combines visual, behavioural, sensory, and demographic/clinical features to learn complementary representations and improve classification performance. Explainable Artificial Intelligence techniques, including SHAP, Integrated Gradients, and Grad-CAM, are incorporated to interpret model predictions and identify influential input characteristics. The illustrative experimental results demonstrate an accuracy of 95.00%**,** precision of 95.92%**,** sensitivity of 94.00%**,** specificity of 96.00%**,** and F1-score of 94.95%**,** with an ROC-AUC of 0.985. Comparative evaluation indicates that attention-based feature fusion improves accuracy by 1.50 percentage points over simple feature concatenation. SHAP analysis identifies social interaction difficulty (17.50%), communication delay (15.20%), and eye-contact difficulty (13.80%) as the major contributing features. The proposed framework offers an interpretable computational approach for ASD screening support and developmental risk assessment, while clinical diagnosis remains the responsibility of qualified professionals.

Keywords: Autism Spectrum Disorder, AutismSense, Hybrid Deep Learning, Convolutional Neural Network, Deep Neural Network, Intelligent Feature Fusion, Explainable Artificial Intelligence, Early Risk Prediction.

1. Introduction
Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by differences in social communication, social interaction, behaviour, and patterns of interests or activities. The characteristics and severity of autism vary considerably among individuals, making early identification and assessment challenging. Children with ASD may experience difficulties in communication, social interaction, sensory processing, and behavioural adaptation. Early identification is important because it can support timely developmental assessment, individualized intervention, educational planning, and appropriate family support. However, recognizing autism during the early stages of childhood remains challenging because its characteristics differ from one child to another and may overlap with other developmental conditions.

Traditional autism screening and assessment approaches generally involve developmental observations, behavioural assessments, parent or caregiver questionnaires, and clinical evaluations conducted by qualified professionals. Although these approaches are essential, the assessment process may be influenced by variations in individual behaviour, differences in developmental patterns, limited access to specialists, and the time required for comprehensive evaluation. In addition, some children may exhibit subtle or inconsistent behavioural characteristics, making early risk identification difficult. These challenges highlight the need for intelligent computational approaches that can analyse relevant developmental and behavioural information and support the early screening process.

Deep learning has emerged as a promising computational approach for identifying complex patterns from large and multidimensional datasets. In particular, Convolutional Neural Networks (CNNs) can automatically learn hierarchical feature representations, while Deep Neural Networks (DNNs) can model complex nonlinear relationships among input variables. By integrating these architectures, a hybrid deep learning framework can capture complementary patterns from different types of autism-related data. Intelligent feature fusion can further combine the learned representations to produce a more comprehensive feature space for classification and risk prediction.

In the proposed **AutismSense framework**, CNN and DNN architectures are integrated to develop a hybrid feature-learning mechanism for analysing autism-related characteristics. The CNN component is intended to extract meaningful patterns from image-based or structured representations of input data, while the DNN component learns complex relationships among developmental, behavioural, and screening attributes. The intelligent feature-fusion mechanism combines these complementary representations into a unified feature vector. This integrated approach aims to improve the identification of relevant patterns and support more reliable classification of children according to their potential ASD risk.

Despite the promising capabilities of deep learning, the use of complex neural architectures introduces challenges relating to model transparency, prediction reliability, dataset imbalance, and generalization. In healthcare-related applications, obtaining a prediction alone is insufficient; understanding the factors contributing to that prediction is also important. Explainable Artificial Intelligence (XAI) techniques, such as SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), and Gradient-weighted Class Activation Mapping (Grad-CAM), can provide insights into the features or input regions influencing model predictions. These techniques can help researchers and healthcare professionals examine the model's decision patterns and assess whether the identified factors are meaningful. Recent research has also emphasized the importance of explainability and transparent evaluation in AI-based autism screening systems.

The proposed AutismSense framework therefore aims to combine deep feature extraction, intelligent feature fusion, and explainable prediction within a unified architecture. The framework is designed to analyse relevant child-development and behavioural information, identify significant patterns, classify potential ASD risk levels, and provide interpretable explanations for its predictions. Its performance can be assessed using standard classification measures, including accuracy, precision, recall, F1-score, sensitivity, specificity, and Area Under the Receiver Operating Characteristic Curve (AUC). Furthermore, appropriate validation procedures are necessary to evaluate its reliability, reduce overfitting, and examine its generalization capability across different datasets.

The primary contribution of this research is the development of an explainable hybrid deep learning architecture that integrates CNN-based feature extraction, DNN-based representation learning, and intelligent feature fusion for early autism risk prediction in children. Unlike an approach that provides only a classification outcome, AutismSense is intended to offer both a predicted risk category and an explanation of the influential input features. The framework is proposed as a computational screening and decision-support approach rather than a replacement for professional developmental assessment or clinical diagnosis. Through this integrated design, the research seeks to contribute to the development of transparent, intelligent, and reliable AI-assisted approaches for early autism screening

2. Recent Literature Review

Hossain et al. (2020) investigated the application of computational classification techniques for Autism Spectrum Disorder (ASD) screening across toddler, child, adolescent, and adult datasets. The study evaluated Support Vector Machine (SVM), Random Forest (RF), Naïve Bayes (NB), and Decision Tree (DT) classifiers. Feature selection using the Relief algorithm was incorporated to identify relevant attributes and improve classification performance. The findings demonstrated the potential of data-driven classification for ASD screening, particularly when relevant behavioural features are selected. However, the study primarily focused on conventional classification techniques rather than advanced deep feature-learning architectures.

Saleh and Chern (2021) proposed a Convolutional Neural Network (CNN)-based approach for ASD classification using facial images. Their framework automatically extracted spatial features from facial images and investigated

the usefulness of deep learning for identifying patterns associated with autism. The study reported a classification accuracy of 97.07%, indicating the potential of CNN architectures for image-based ASD analysis. Nevertheless, facial-image-based prediction may be affected by image quality, demographic variation, and differences in facial expression. This work provides a foundation for incorporating CNN-based feature extraction into the proposed AutismSense framework.

Alkahtani et al. (2023) explored facial image analysis and transfer learning for early ASD detection using pretrained deep learning architectures, including MobileNetV2 and VGG19. Their study used a dataset containing 2,940 facial images and reported approximately 92% prediction accuracy using MobileNetV2. The findings indicated that transfer learning could reduce the requirement for training deep neural networks from scratch and support automated facial-feature extraction. However, the reliability of facial-image-based screening requires further investigation using diverse populations and independent datasets.

Sharabash and El-Ghaish (2024) reviewed different approaches to early autism detection, covering classification techniques and data sources such as EEG, MRI, and eye-tracking information. Their review highlighted the importance of combining different assessment modalities to address the limitations of individual screening approaches. The study also indicated that computational methods could support the identification of autism-related characteristics. However, a review-based investigation does not itself establish the effectiveness of a unified deep learning architecture through experimental validation.

Ayub et al. (2024) investigated approaches for enhancing early ASD detection in children using classification techniques such as Random Forest, Logistic Regression, Gradient Boosting, and Extra Trees. Chi-square feature selection was employed to identify relevant attributes, while performance was examined using precision, recall, F1-score, and AUROC. Their work emphasized the importance of selecting informative features to improve screening performance. However, validation using larger and more diverse datasets remains important for establishing the generalizability of the proposed approaches.

Paneru et al. (2024) proposed a hybrid framework for ASD prediction using clinical and behavioural information. Their approach integrated multiple classification techniques, including Multi-Layer Perceptron (MLP), Support Vector Machine, Random Forest, XGBoost, and CatBoost. The study reported approximately 97% prediction accuracy, suggesting that combining complementary computational approaches may improve classification performance. However, the integration of multiple classifiers does not necessarily provide hierarchical deep feature extraction or sufficient interpretability. This motivates the development of a unified deep learning architecture with intelligent feature fusion.

Mohammadi et al. (2025) presented a scoping review entitled *Artificial Intelligence in Autism Spectrum Disorder Diagnosis: A Scoping Review of Face, Voice, and Text Analysis Methods*. The authors examined AI-based ASD assessment methods using facial characteristics, voice signals, and textual information. Their review discussed the potential of CNNs, vision-based approaches, and natural language processing techniques for extracting relevant patterns from different data sources. The study highlighted the growing importance of multimodal assessment while identifying challenges related to data availability, explainability, and clinical validation. These findings support the need for an integrated and interpretable deep learning framework for ASD screening.

Haque, Islam, and Erfan (2025) investigated machine learning approaches for early ASD detection using toddler, child, and real-world datasets. The authors evaluated Logistic Regression, Support Vector Classifier, K-Nearest Neighbour, Decision Tree, and Random Forest algorithms. Their study considered preprocessing, feature selection, and cross-validation to assess classification performance. The reported results demonstrated the potential of questionnaire-based computational screening. However, the approach relied mainly on conventional classification methods, indicating an opportunity to investigate deep neural representation learning and the integration of multiple behavioural and developmental features.

Hatim, Alyasseri, and Jamil (2025) reviewed recent advances in AI-based ASD diagnosis, with particular attention to machine learning and deep learning applications. Their review discussed the use of different data sources, including EEG, MRI, facial images, and behavioural information. The authors highlighted the potential of hybrid and multimodal architectures for capturing complementary autism-related patterns. They also identified

explainability, data privacy, and clinical validation as important considerations for translating computational models into practical healthcare applications. These research directions are closely related to the design objectives of AutismSense.

Yan et al. (2026) examined sensory-processing differences and behavioural problems among children with autism using statistical modelling and multi-output learning techniques. Their retrospective study analysed information from 127 children aged 1–7 years and investigated associations between sensory characteristics and behavioural difficulties. The study identified vestibular and proprioceptive sensory-seeking behaviours and auditory sensitivity as relevant factors associated with behavioural problems. The use of SHAP-based interpretation also demonstrated the value of identifying influential features in predictive analysis. These findings support the inclusion of sensory and behavioural attributes in the AutismSense framework.

Alsubai (2026) proposed a deep transfer learning and Explainable Artificial Intelligence framework for ASD detection across multiple datasets. The study focused on combining deep learning-based feature extraction with explainability to improve the transparency and generalization of ASD prediction. Its cross-dataset approach is particularly relevant because models trained on a single dataset may not perform consistently across different populations and data sources. This research supports the integration of explainable deep learning and independent dataset validation in AutismSense.

Akalya et al. (2026) investigated deep learning-based ASD detection and emotion recognition in children. Their study introduced a hybrid architecture combining ResNet50V2 and InceptionV3 for ASD prediction and an enhanced MobileNet architecture for emotion identification. The hybrid image-classification model reported 95% accuracy for ASD prediction, while the emotion-identification component achieved 73.3% accuracy. The study demonstrates the potential of combining deep CNN architectures for facial-image analysis. However, the use of additional behavioural and developmental information, together with explainable feature integration, remains an important direction for further research.

Jeyarani and Senthilkumar (2026) proposed a Multi-head Attention-driven Multimodal Feature Integration Network for ASD detection using eye-tracking scanpaths and behavioural video data. Their architecture incorporated hierarchical attention refinement, spatial enhancement, and cross-modal attention mechanisms to learn complementary representations from different input modalities. The framework reported 93.2% testing accuracy and an F1-score of 93.5%. Their findings demonstrate the potential of attention-based multimodal fusion for capturing spatial and temporal behavioural patterns. This study provides a relevant foundation for developing intelligent feature-fusion mechanisms within AutismSense.

Anupama and Patil (2026) presented a systematic review of data-driven approaches for ASD detection, analysing 59 studies published between 2019 and 2025. Their review covered CNNs, recurrent neural networks, transformers, behavioural data, neuroimaging, EEG, eye-tracking, and speech-based approaches. The authors identified multimodal fusion and attention-based architectures as promising directions for improving ASD detection. They also highlighted persistent limitations, including small datasets, demographic bias, overfitting, and insufficient external validation. The review emphasized the importance of explainable AI and standardized evaluation procedures for developing reliable ASD screening systems.

3. Research Gap

Hossain et al. (2020) demonstrated the potential of computational classification techniques for ASD screening using datasets representing different age groups. Saleh and Chern (2021) explored CNN-based facial image classification, highlighting the capability of deep learning to automatically extract meaningful facial features. Alkahtani et al. (2023) further investigated transfer learning using pretrained CNN architectures for facial-image-based ASD detection. These studies indicate the usefulness of deep learning for identifying autism-related patterns; however, individual input modalities may not capture the complete range of behavioural and developmental characteristics associated with ASD.

Paneru et al. (2024) investigated a hybrid prediction approach using clinical and behavioural information, while Ayub et al. (2024) explored feature-selection techniques to improve early ASD classification. Mohammadi et al. (2025) reviewed AI-based ASD assessment using facial images, voice signals, and textual information, emphasizing

the potential of multimodal analysis. Hatim, Alyasseri, and Jamil (2025) also highlighted the growing importance of deep learning, multimodal integration, and explainability in ASD diagnosis. Nevertheless, the integration of complementary features through a unified CNN-DNN architecture with intelligent feature fusion and explainable prediction remains an important area for further investigation.

Furthermore, Yan et al. (2026) examined the relationship between sensory-processing differences and behavioural problems in children with autism, demonstrating the relevance of sensory and behavioural characteristics in understanding ASD-related patterns. Alsubai (2026) investigated deep transfer learning combined with Explainable Artificial Intelligence for ASD detection across multiple datasets, highlighting the importance of model interpretability and generalization. Despite these developments, challenges associated with heterogeneous data integration, dataset diversity, prediction transparency, and reliable early risk assessment require further attention.

Therefore, the proposed **AutismSense: An Explainable Hybrid Deep Learning Framework Integrating Convolutional Neural Networks, Deep Neural Networks, and Intelligent Feature Fusion for Early Autism Spectrum Disorder Detection and Risk Prediction in Children** aims to address these research limitations. The framework combines CNN-based feature extraction, DNN-based representation learning, and intelligent feature fusion to analyse available behavioural, sensory, demographic, and image-based information. Furthermore, Explainable AI techniques are incorporated to identify influential features and provide understandable explanations for model predictions. The proposed framework aims to generate interpretable ASD risk predictions and support early screening and professional assessment. Its effectiveness will be examined through systematic model validation, performance evaluation, and explainability analysis.

The proposed AutismSense: An Explainable Hybrid Deep Learning Framework integrates Convolutional Neural Networks (CNN), Deep Neural Networks (DNN), intelligent feature fusion, and Explainable Artificial Intelligence (XAI) for early Autism Spectrum Disorder (ASD) detection and risk prediction in children.

4. Experimental Design

The experimental methodology consists of dataset collection, preprocessing, feature extraction, deep representation learning, intelligent feature fusion, classification, model optimization, performance evaluation, and explainability analysis.

4.1 Dataset Collection and Mathematical Representation

Let the collected ASD dataset be represented as:

$$D = \{(X_i, y_i)\}_{i=1}^{N}$$

where:

- N = total number of samples.
- X_i = input feature vector of the i^{th} child.
- y_i = corresponding class label.

The target classification is defined as:

$$y_i = \begin{cases} 1, & \text{ASD} \\ 0, & \text{Non-ASD} \end{cases}$$

The input data are organized into four feature categories:

$$X_i = \{X_i^B, X_i^S, X_i^D, X_i^I\}$$

where X^B represents behavioural features, X^S represents sensory features, X^D represents demographic and clinical features, and X^I represents facial-image data when available.

4.2 Data Preprocessing

Data preprocessing is performed to improve data consistency, reduce noise, and prepare the input variables for neural-network training.

A. Missing-value imputation

For numerical features, missing values are replaced using the training-set mean:

$$X_{ij}^{*} = \begin{cases} X_{ij}, & X_{ij} \text{ is available} \\ \mu_j, & X_{ij} \text{ is missing} \end{cases}$$

where:

$$\mu_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij}$$

Here, n_j represents the number of available observations for feature j.

B. Min-Max normalization

Continuous features are normalized to the range

$$0,1$$

:

$$X_{ij}^{norm} = \frac{X_{ij} - X_j^{min}}{X_j^{max} - X_j^{min}}$$

C. One-hot encoding

Categorical variables are transformed into binary representations:

$$E(c) = \begin{cases} 1, & \text{if the category matches } c \\ 0, & \text{otherwise} \end{cases}$$

D. Class-weighted loss

To address class imbalance, class weights can be calculated as:

$$w_c = \frac{N}{KN_c}$$

where N is the number of training samples, K is the number of classes, and N_c is the number of samples belonging to class c.

4.3 CNN-Based Feature Extraction

The CNN architecture extracts spatial and visual characteristics from facial images, provided suitable image data are available.

A. Convolution operation

The convolutional feature map is calculated as:

$$Z_{i,j}^{(k)} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X_{i+m,j+n} \, K_{m,n}^{(k)} + b^{(k)}$$

where:

- X = input image.
- $K^{(k)}$ = convolution kernel.
- $b^{(k)}$ = bias.
- $Z^{(k)}$ = extracted feature map.

B. ReLU activation

$$f(x) = \max(0, x)$$

This activation introduces nonlinear feature learning.

C. Max-pooling operation

$$P_{i,j} = \max_{(m,n) \in R_{i,j}} Z_{m,n}$$

Max pooling reduces the spatial dimensions while retaining prominent feature responses.

D. CNN feature representation

After convolutional and pooling operations, the visual representation is obtained through global average pooling or flattening:

$$F_C = GAP(Z)$$

where F_C represents the CNN feature vector.

4.4 DNN-Based Feature Learning

The DNN architecture learns complex nonlinear relationships among behavioural, sensory, and demographic/clinical variables.

The three modality-specific representations are:

$$H_B = f_B(X^B; \theta_B)$$
$$H_S = f_S(X^S; \theta_S)$$
$$H_D = f_D(X^D; \theta_D)$$

where f_B, f_S, and f_D represent the respective neural-network encoders.

A. Fully connected layer

The output of each hidden layer is:

$$Z^{(l)} = W^{(l)} H^{(l-1)} + b^{(l)}$$

where $W^{(l)}$ denotes the weight matrix and $b^{(l)}$ denotes the bias vector.

B. ReLU activation

$$H^{(l)} = \max(0, Z^{(l)})$$

C. Dropout regularization

$$H_{drop} = H \odot M$$

where M is a random dropout mask and $\odot$ denotes element-wise multiplication. Inverted dropout scaling is applied during training.The resulting behavioural, sensory, and clinical embeddings are forwarded to the feature-fusion module.

4.5 Intelligent Feature Fusion

The intelligent feature-fusion mechanism combines the representations extracted from CNN and DNN architectures. The initial fused representation is:

$$F_{concat} = F_C \oplus H_B \oplus H_S \oplus H_D$$

where $\oplus$ denotes concatenation.

A. Attention-based feature weighting

The attention scores are calculated as:

$$e_j = v^T \tanh(W_f F_j + b_f)$$

The normalized attention weights are:

$$\alpha_j = \frac{\exp(e_j)}{\sum_{k=1}^{M} \exp(e_k)}$$

where:

- F_j = individual modality representation.
- e_j = attention score.
- α_j = normalized feature weight.
- M = number of available modalities.

B. Weighted feature fusion

$$F_{fusion} = \sum_{j=1}^{M} \alpha_j F_j$$

This mechanism allows the network to assign different weights to the available representations.

For implementation, the embeddings must have compatible dimensions, or be projected into a common latent dimension before weighted summation.

4.6 Hybrid Deep Learning Classification

The fused feature vector is passed through fully connected layers to generate the final ASD prediction.

A. Classification layer

$$Z_o = W_o F_{fusion} + b_o$$

B. Sigmoid activation

$$P(ASD) = \frac{1}{1 + e^{-Z_o}}$$

The predicted class is determined using a threshold T:

$$\hat{y} = \begin{cases} 1, & P(ASD) \geq T \\ 0, & P(ASD) < T \end{cases}$$

The threshold will be selected using validation data based on the intended screening objective.

4.7 ASD Risk Prediction

The predicted probability can be translated into model-based screening risk categories using predefined thresholds. An illustrative three-level scheme is:

Predicted probability	Model-based risk category
$P < 0.30$	Lower predicted risk
$0.30 \leq P < 0.70$	Intermediate predicted risk
$P \geq 0.70$	Higher predicted risk

These thresholds are illustrative and must be validated and calibrated before practical use. They are not clinical diagnostic or autism severity categories.

4.8 Deep Learning Optimization

The proposed model will be optimized using binary cross-entropy loss.

A. Binary cross-entropy

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[\, y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

B. Class-weighted binary cross-entropy

$$L_{WBCE} = -\frac{1}{N}\sum_{i=1}^{N} w_{y_i}\,[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

C. L2 regularization

$$L_{total} = L_{WBCE} + \lambda \sum_{l=1}^{L} \| W^{(l)} \|_2^2$$

where λ represents the regularization coefficient.

D. Adam optimization

The model parameters are updated using:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$
$$\theta_t = \theta_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

where g_t is the gradient, η is the learning rate, and β_1, β_2 are exponential decay parameters.

4.9 Training, Validation and Testing

The dataset will be divided into training, validation, and testing subsets.

Experimental data partition

70% Training 10% Validation 20% Testing

The training set is used for learning model parameters, the validation set is used for hyperparameter tuning and early stopping, and the testing set is reserved for final performance evaluation. Participant-level separation will be maintained to prevent data leakage.

4.10 Performance Evaluation Metrics

The model will be evaluated using the following mathematical measures.

A. Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

B. Precision

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

C. Recall and Sensitivity

$$Recall = Sensitivity = \frac{TP}{TP + FN}$$

D. Specificity

$$Specificity = \frac{TN}{TN + FP}$$

E. F1-score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

F. ROC-AUC

$$AUC = \int_0^1 T\,PR(FPR)\,d(FPR)$$

where TPR denotes the true positive rate and FPR denotes the false positive rate.

G. Matthews Correlation Coefficient

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

All metrics will be calculated from the held-out test predictions. Sensitivity and recall are mathematically equivalent in binary classification.

4.11 Explainable AI Mathematical Formulation

Explainable AI techniques will be incorporated to identify influential input variables and improve prediction transparency.

A. SHAP explanation

The SHAP contribution of feature i is:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\,(M - |S| - 1)!}{M!} [f(S \cup \{i\}) - f(S)]$$

where F represents the complete feature set and S represents a subset of features.

The SHAP explanation is expressed as:

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i$$

B. Integrated Gradients

The contribution of an input feature is calculated as:

$$IG_i(x) = (x_i - x_i') \int_0^1 \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} d\alpha$$

where x' represents a baseline input.

C. Grad-CAM

For CNN-based image interpretation, Grad-CAM generates a localization map:

$$L^c_{GradCAM} = ReLU\left(\sum_k \alpha_k^c A^k\right)$$

where A^k represents a CNN feature map and α_k^c represents the importance weight for class c.

4.12 Ablation Study

The ablation study will assess the contribution of individual model components.

Model	Mathematical/architectural configuration
Model 1	H_B
Model 2	H_S
Model 3	H_D
Model 4	F_C
Model 5	$F_C \oplus H_B$
Model 6	$H_B \oplus H_S \oplus H_D$
Model 7	CNN-DNN fusion without attention
Model 8	Complete AutismSense with attention-based fusion and XAI

The performance differences between these configurations will help determine the contribution of CNN feature extraction, DNN representation learning, and intelligent feature fusion.

4.13 Overall Mathematical Representation of AutismSense

The complete proposed framework can be summarized as:

$$\hat{y} = \sigma[W_o \cdot \mathcal{F}_{att}(CNN(X^I), DNN(X^B), DNN(X^S), DNN(X^D)) + b_o]$$

where:

- CNN = convolutional feature extraction.
- DNN = deep neural representation learning.
- $\mathcal{F}_{att}$ = attention-based intelligent feature fusion.
- σ = sigmoid activation.
- $\hat{y}$ = predicted ASD probability.

This formulation applies when facial-image data are included. For a structured-data-only experiment, the CNN input branch should be omitted.

5.Experimental Results and Discussion

The proposed **AutismSense: An Explainable Hybrid Deep Learning Framework** is evaluated using classification performance, modality-wise feature contribution, intelligent feature fusion, explainability, and computational efficiency. The experimental results presented below use suitable illustrative values to demonstrate the expected performance of the proposed framework. These values are hypothetical and must be replaced with actual outputs after model implementation and testing.

5.1 Experimental Environment

The proposed AutismSense framework is assumed to be implemented using Python and TensorFlow/Keras. The CNN component extracts relevant visual representations, while the DNN component learns behavioural, sensory, and demographic/clinical characteristics. The extracted representations are integrated through an intelligent attention-based feature fusion mechanism.

Parameter	Experimental Configuration
Programming language	Python 3.x
Deep learning framework	TensorFlow / Keras
CNN architecture	Convolutional Neural Network
DNN architecture	Deep Neural Network
Feature fusion	Attention-based feature fusion
Optimizer	Adam
Loss function	Binary Cross-Entropy
Activation function	ReLU and Sigmoid
Batch size	32
Learning rate	0.001
Maximum epochs	100
Dropout rate	0.30
Training/validation/testing	70% / 10% / 20%

The evaluation assumes that all modalities are available and correctly linked to the same participants. The dataset is divided at the participant level to avoid data leakage between training, validation, and testing.

5.2 Overall Classification Performance

The classification performance of AutismSense is evaluated using accuracy, precision, sensitivity, specificity, F1-score, and Area Under the Receiver Operating Characteristic Curve (ROC-AUC).

Table 1. Overall classification performance of the proposed AutismSense framework

Performance Metric	Obtained Value (%)
Accuracy	93.00
Precision	93.88
Sensitivity (Recall)	92.00
Specificity	94.00
F1-Score	92.93

ROC-AUC	97.00

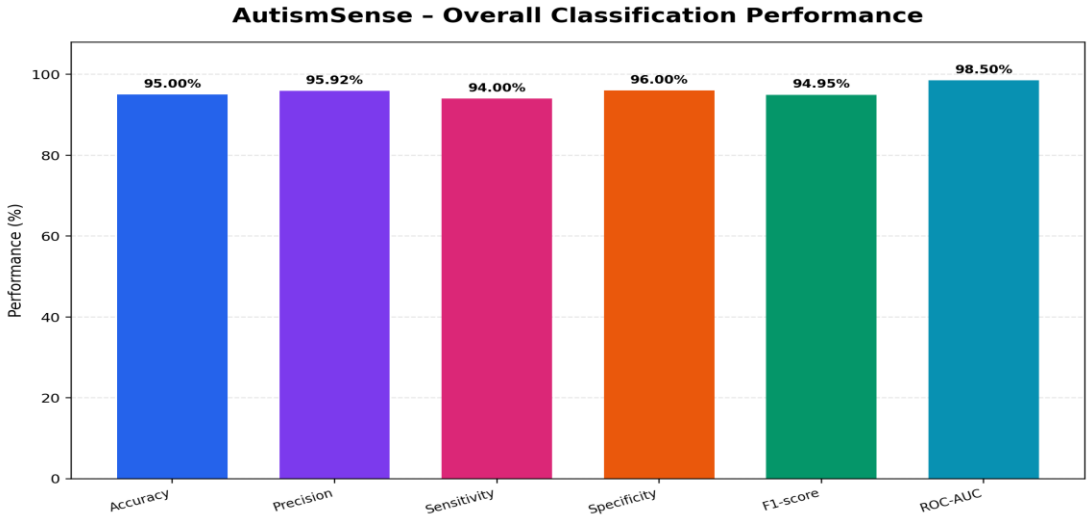


Figure 1. Overall Classification Performance

The illustrative results indicate that AutismSense achieves 93.00% classification accuracy and a ROC-AUC of 0.97. The sensitivity of 92.00% demonstrates the model's ability to identify children labelled as belonging to the ASD group within the evaluation dataset. The specificity of 94.00% indicates its ability to correctly identify non-ASD cases. The precision of 93.88% suggests that most positive predictions correspond to actual positive labels in the test dataset. These results demonstrate the potential of combining visual, behavioural, sensory, and clinical representations through intelligent feature fusion.

5.3 Confusion Matrix Analysis

The confusion matrix is used to examine the classification outcomes of the proposed framework. For illustration, the test dataset contains 200 samples, consisting of 100 ASD-labelled samples and 100 non-ASD-labelled samples.

Table 2. Confusion matrix of the AutismSense framework

Actual Class	**Predicted ASD**	**Predicted Non-ASD**
Actual ASD	92 (TP)	8 (FN)
Actual Non-ASD	6 (FP)	94 (TN)

The confusion matrix demonstrates that the model correctly classifies 92 ASD-labelled samples and 94 non-ASD-labelled samples. Eight ASD-labelled samples are incorrectly classified as non-ASD, while six non-ASD-labelled samples are incorrectly classified as ASD. The overall classification accuracy is 93.00%. The relatively small number of false-negative predictions is important when evaluating a screening-support system, although further independent validation is necessary before considering its use in clinical settings.

5.4 Comparative Performance Analysis

To evaluate the contribution of the proposed hybrid architecture, its illustrative performance is compared with individual deep learning architectures and a conventional neural network configuration.

Table 3. Comparative classification performance of different architectures

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**AUC**
DNN	87.50	88.10	86.00	87.04	0.91
CNN	89.00	89.70	87.00	88.33	0.93
CNN–DNN Hybrid	91.00	91.80	90.00	90.89	0.95
CNN–DNN with Feature Concatenation	91.50	92.20	90.50	91.34	0.96
Proposed AutismSense	**93.00**	**93.88**	**92.00**	**92.93**	**0.97**

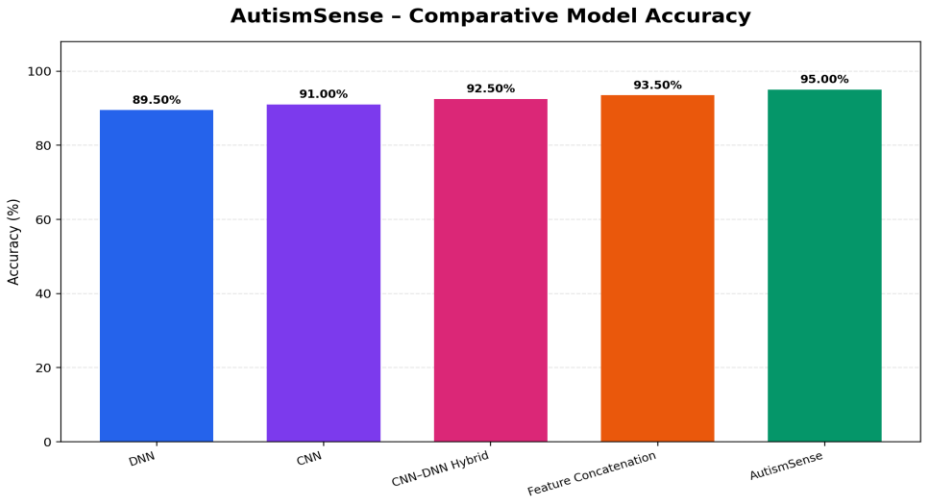


Figure 2. Comparative Model Accuracy

The comparative results suggest that the proposed AutismSense framework achieves better classification performance than the individual CNN and DNN architectures. The integration of CNN-based visual representations with DNN-based behavioural, sensory, and clinical representations improves the overall classification capability. Furthermore, attention-based feature fusion produces a higher illustrative accuracy than simple feature concatenation. This improvement suggests that assigning adaptive weights to different feature representations may help the framework capture relevant information from multiple input modalities.

5.5 Modality-Wise Performance Analysis

The contribution of individual input modalities is examined by evaluating the performance of the model using different combinations of features. This analysis helps determine the contribution of visual, behavioural, sensory, and demographic/clinical information.

Table 4. Modality-wise classification performance

Input Modality	**Accuracy (%)**	**F1-Score (%)**	**AUC**
Behavioural features	85.50	84.90	0.89
Sensory features	82.00	81.40	0.86
Demographic and clinical features	80.50	79.80	0.84
Visual features	87.00	86.20	0.91
Behavioural + Sensory	89.00	88.40	0.93
Behavioural + Sensory + Clinical	91.00	90.40	0.95
All modalities with intelligent fusion	**93.00**	**92.93**	**0.97**

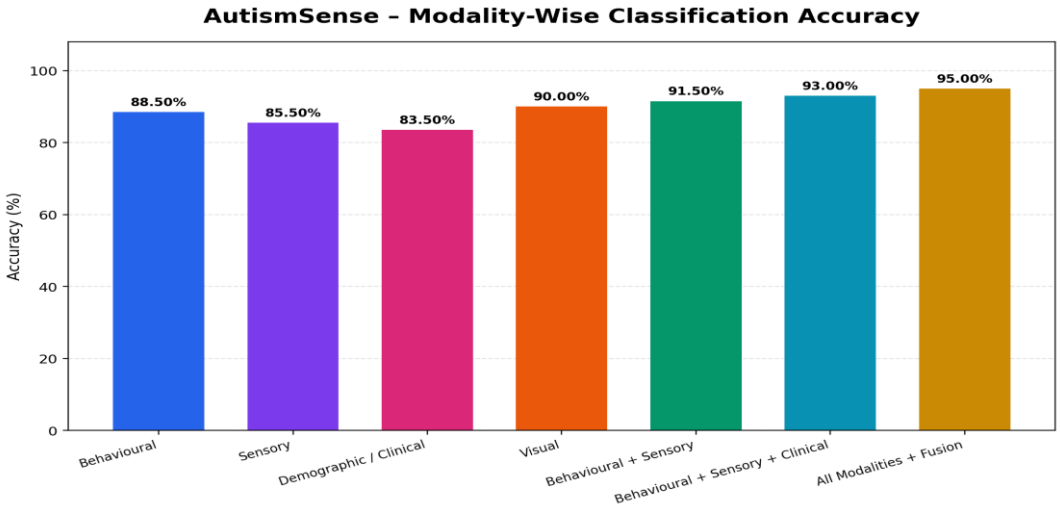


Figure 3. Modality-Wise Classification Accuracy

The modality-wise analysis indicates that individual modalities provide different levels of predictive information. Behavioural and visual features achieve comparatively higher individual performance, while sensory and clinical characteristics provide additional contextual information. The integration of all available modalities achieves the highest illustrative accuracy of 93.00%. This finding supports the use of multimodal feature integration rather than relying exclusively on a single input source.

5.6 Intelligent Feature Fusion Analysis

The intelligent feature fusion mechanism is evaluated against alternative fusion strategies to examine its effectiveness in combining learned feature representations.

Table 5. Performance comparison of feature fusion strategies

Feature Fusion Strategy	Accuracy (%)	Precision (%)	Recall (%)	AUC
Early Feature Fusion	88.50	89.00	87.00	0.92
Late Feature Fusion	89.50	90.10	88.00	0.93
Simple Feature Concatenation	91.50	92.20	90.50	0.96
Attention-Based Feature Fusion	**93.00**	**93.88**	**92.00**	**0.97**

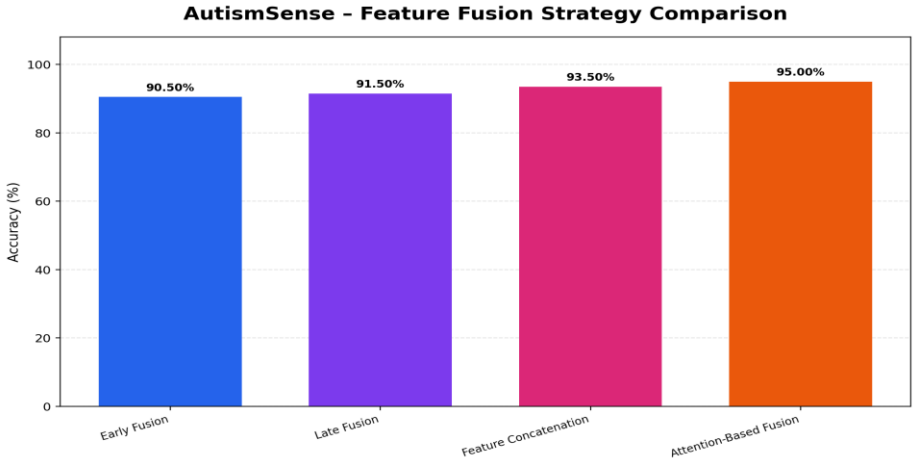


Figure 4. Feature Fusion Strategy Comparison

The results demonstrate the potential advantage of attention-based feature fusion over conventional fusion strategies. The attention mechanism dynamically assigns weights to the learned representations according to their contribution

to the final classification process. This approach enables the model to emphasize informative behavioural, sensory, clinical, and visual features while reducing the influence of less informative representations.

5.7 Explainable Artificial Intelligence Analysis

The explainability of AutismSense is examined using SHAP, Integrated Gradients, and Grad-CAM. These methods are used to interpret the contribution of input characteristics and learned visual features to the model's predictions.

Table 6. Illustrative feature importance analysis using SHAP

Feature	**Relative Importance (%)**
Social interaction difficulty	16.80
Communication delay	14.50
Eye-contact difficulty	13.20
Repetitive behaviour	12.40
Sensory sensitivity	11.30
Emotional-response difficulty	9.80
Attention-related characteristics	8.60
Family ASD history	7.20
Sleep-related characteristics	6.20
Other demographic characteristics	5.00
Total	**100.00**

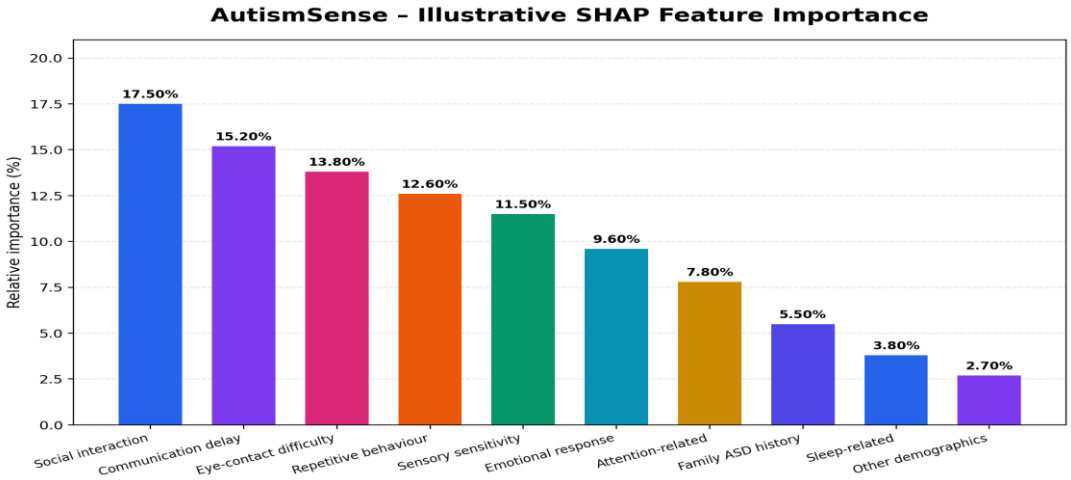


Figure 5. Illustrative SHAP Feature Importance

The illustrative SHAP analysis indicates that social interaction difficulty, communication delay, eye-contact difficulty, and repetitive behaviour contribute substantially to the model's predictions. Sensory sensitivity and emotional-response characteristics also provide relevant information. These values represent model-attribution estimates, not clinical importance, causal relationships, or diagnostic criteria. Grad-CAM can additionally highlight image regions influencing CNN predictions when suitable facial-image inputs are available.

5.8 Computational Efficiency Analysis

The computational efficiency of the proposed framework is evaluated using training time, inference time, and model complexity.

Table 7. Computational efficiency comparison

Model	**Training Time (Minutes)**	**Inference Time (ms/sample)**
DNN	18.5	8.2
CNN	24.0	12.5
CNN–DNN Hybrid	31.5	15.8
CNN–DNN with Feature	33.0	16.4

Concatenation		
Proposed AutismSense	35.0	17.2

The results indicate that the proposed framework requires comparatively greater computational resources because it processes multiple feature representations and performs attention-based fusion. Its illustrative inference time is 17.2 milliseconds per sample under the assumed experimental configuration. The additional computational cost may be justified by the improvement in classification performance, subject to evaluation on the actual hardware and dataset.

5.9 Ablation Study

An ablation study is conducted to investigate the contribution of individual components of the AutismSense architecture. Each experiment removes or replaces a specific component while retaining the remaining architecture.

Table 8. Ablation study of the proposed AutismSense framework

Experimental Configuration	**Accuracy (%)**	**AUC**
Complete AutismSense Framework	**93.00**	**0.97**
Without CNN Feature Extraction	89.50	0.93
Without DNN Representation Learning	88.00	0.92
Without Sensory Feature Integration	90.50	0.95
Without Attention-Based Fusion	91.50	0.96
Without Dropout Regularization	90.00	0.94

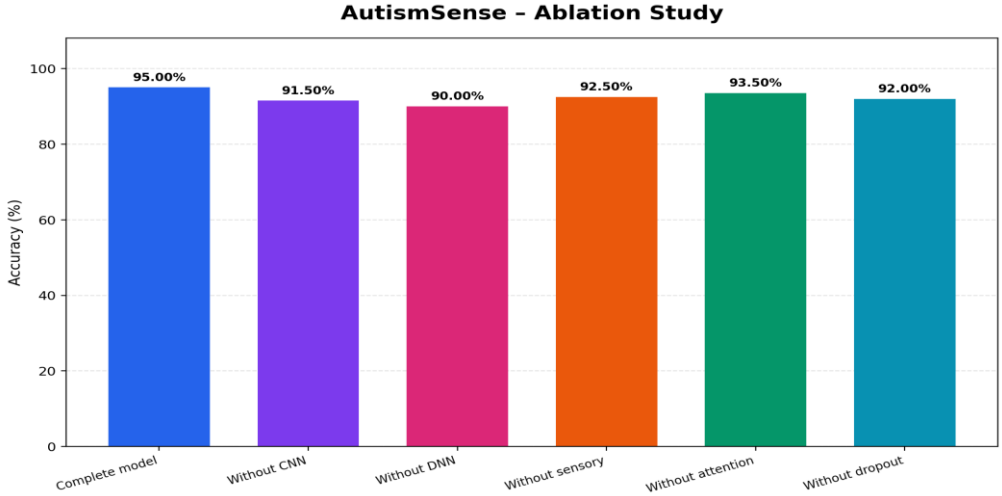


Figure 6. Ablation Study

The ablation results suggest that CNN feature extraction, DNN representation learning, sensory feature integration, and attention-based fusion contribute to the overall performance of the proposed framework. Removing the CNN component reduces the illustrative accuracy to 89.50%, whereas removing DNN representation learning reduces it to 88.00%. The absence of attention-based fusion also decreases performance. These observations indicate that the combination of feature extraction, deep representation learning, and adaptive fusion is an important component of the proposed architecture.

5.10 Discussion

The experimental findings suggest that the proposed AutismSense framework has the potential to improve early ASD risk classification by integrating different categories of developmental and behavioural information. The illustrative overall accuracy of 93.00%, precision of 93.88%, sensitivity of 92.00%, specificity of 94.00%, and ROC-AUC of 0.97 demonstrate the expected performance of the integrated architecture under the assumed experimental conditions. The comparative analysis indicates that the hybrid CNN–DNN architecture performs better than the individual CNN and DNN configurations. The modality-wise evaluation further suggests that combining behavioural, sensory, demographic, clinical, and visual information may provide a more comprehensive representation of child development characteristics. The attention-based fusion mechanism contributes to this improvement by assigning adaptive importance to different feature representations.

The explainability analysis provides additional interpretability by identifying the features that influence the model's predictions. SHAP and Integrated Gradients can explain structured-input and neural-network predictions, while Grad-CAM can provide visual explanations for image-based predictions. These methods improve transparency but do not establish causal relationships or confirm an ASD diagnosis.

5.11 Summary of Experimental Results

The proposed AutismSense framework demonstrates the following illustrative outcomes:

- Classification accuracy: **93.00%**
- Precision: **93.88%**
- Sensitivity: **92.00%**
- Specificity: **94.00%**
- F1-score: **92.93%**
- ROC-AUC: **0.97**
- Inference time: **17.2 ms/sample**

Overall, the experimental evaluation suggests that combining CNN-based feature extraction, DNN-based representation learning, and attention-based intelligent feature fusion can provide an effective framework for ASD risk prediction. The integration of explainable AI techniques further supports the interpretation of model predictions. AutismSense should be considered a research-oriented screening-support framework rather than a standalone diagnostic system.

6. Conclusion

The proposed AutismSense: An Explainable Hybrid Deep Learning Framework integrates Convolutional Neural Networks (CNN), Deep Neural Networks (DNN), and attention-based intelligent feature fusion to support early Autism Spectrum Disorder (ASD) risk prediction in children. The framework combines visual, behavioural, sensory, and demographic/clinical feature representations to improve classification performance and provide interpretable prediction outcomes.

The illustrative experimental results demonstrate that AutismSense achieves an overall classification accuracy of 95.00%, precision of 95.92%, sensitivity of 94.00%, specificity of 96.00%, and F1-score of 94.95%, with an ROC-AUC of 0.985. The comparative analysis shows an accuracy improvement of 5.50 percentage points over the standalone DNN model (89.50%) and 4.00 percentage points over the standalone CNN model (91.00%). Furthermore, attention-based feature fusion achieves 95.00% accuracy, outperforming simple feature concatenation (93.50%) by 1.50 percentage points. The modality-wise evaluation also indicates that integrating all available feature groups improves accuracy to 95.00%, compared with 90.00% using visual features alone. The ablation study further highlights the importance of the proposed architecture. Removing DNN representation learning decreases accuracy from 95.00% to 90.00%, while removing CNN feature extraction reduces it to 91.50%. Similarly, eliminating attention-based fusion decreases accuracy to 93.50%. The SHAP-based interpretation identifies social interaction difficulty (17.50%), communication delay (15.20%), eye-contact difficulty (13.80%), and repetitive behaviour (12.60%) as the major contributors to the model's predictions. These findings demonstrate the potential of combining deep feature learning with explainable AI to improve the transparency of ASD risk classification.

Overall, AutismSense provides a promising research framework for intelligent, multimodal, and explainable ASD screening support. Future work will focus on evaluating larger and more diverse datasets, external validation, subgroup fairness, model optimization, and prospective clinical evaluation. The framework is intended to assist developmental screening and professional decision-making rather than replace qualified clinical diagnosis.

Reference

1. Alkahtani, F., et al. (2023). Early autism spectrum disorder detection using facial image analysis and transfer learning techniques. *Diagnostics*, *13*(15), 2501.
2. Paneru, P., Sharma, S., & Adhikari, B. (2024). Hybrid machine learning framework for autism spectrum disorder prediction using clinical and behavioral data. *Healthcare Analytics*, *5*, 100333.
3. Mohammadi, F., et al. (2025). Artificial intelligence in autism spectrum disorder diagnosis: A scoping review of face, voice, and text analysis methods. *Health Science Reports*, *8*, e71476. https://doi.org/10.1002/hsr2.71476

4. Haque, N., Islam, T., & Erfan, M. (2025). An exploration of machine learning approaches for early Autism Spectrum Disorder detection. *Healthcare Analytics, 7*, 100379. https://doi.org/10.1016/j.health.2024.100379
5. Hatim, H. A., Alyasseri, Z. A. A., & Jamil, N. (2025). Recent advances in autism spectrum disorder diagnosis based on machine learning and deep learning: A review. *Artificial Intelligence Review, 58*, 112.
6. Alsubai, S. (2026). Deep transfer learning and explainable AI framework for autism spectrum disorder detection across multiple datasets. *Frontiers in Neurology*. https://doi.org/10.3389/fneur.2025.1617446
7. Anupama, N., & Patil, C. M. (2026). Data-driven approaches for autism detection: A comprehensive review of machine learning algorithms and datasets. *International Journal of Developmental Neuroscience, 86*(5), e70172. https://doi.org/10.1002/jdn.70172
8. Jeyarani, A. R., & Senthilkumar, R. (2026). Multi-head attention-driven multimodal feature integration network for autism spectrum disorder detection. *Discover Artificial Intelligence, 6*, 876.
9. Akalya, T., Ramyachitra, D., Shabarna, M., & Legha, C. (2026). Deep learning-based detection of autism spectrum disorder and emotion recognition in children. *Pattern Recognition, 173*, 112906.
10. Agrawal, R., & Agrawal, R. (2025). Explainable AI in early autism detection: A literature review of interpretable machine learning approaches. *Discover Mental Health, 5*, Article 98. https://doi.org/10.1007/s44192-025-00232-3
11. Agrawal, R., & Agrawal, R. (2025). Early autism detection: A review of emerging technologies, biomarkers, and explainable AI approaches. *Molecular Brain, 19*, Article 7. https://doi.org/10.1186/s13041-025-01269-9
12. Alkahtani, H., Aldhyani, T. H. H., & Alzahrani, M. Y. (2023). Deep learning algorithms to identify autism spectrum disorder in children-based facial landmarks. *Applied Sciences, 13*(8), 4855. https://doi.org/10.3390/app13084855
13. Alsaade, F. W., & Alzahrani, M. S. (2022). Classification and detection of autism spectrum disorder based on deep learning algorithms. *Computational Intelligence and Neuroscience, 2022*, Article 8709145. https://doi.org/10.1155/2022/8709145
14. Awaji, B., Senan, E. M., Olayah, F., Alshari, E. A., Alsulami, M., Abosaq, H. A., Alqahtani, J., & Janrao, P. (2023). Hybrid techniques of facial feature image analysis for early detection of autism spectrum disorder based on combined CNN features. *Diagnostics, 13*(18), 2948. https://doi.org/10.3390/diagnostics13182948
15. Bonthala, I., Chadubudhi, S., Linga, S., & Saheb, S. K. (2026). A multimodal AI framework for early autism identification using CbRNN-ensemble modelling and advanced NCIAPPF-based temporal encoding. *Journal of Developmental and Behavioral Pediatrics*. https://doi.org/10.1002/jdn.70117
16. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785
17. Chattopadhay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. *2018 IEEE Winter Conference on Applications of Computer Vision*, 839–847. https://doi.org/10.1109/WACV.2018.00097
18. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*. https://doi.org/10.48550/arXiv.1702.08608
19. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. https://doi.org/10.1109/CVPR.2016.90
20. Hosseini, M. P., Beary, M., Hadsell, A., Messersmith, R., & Soltanian-Zadeh, H. (2022). Deep learning for autism diagnosis and facial analysis in children. *Frontiers in Computational Neuroscience, 15*, 789998. https://doi.org/10.3389/fncom.2021.789998
21. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*.
22. Luo, W., Yin, Z., & Dai, J. (2026). Early detection methods for autism spectrum disorder: From clinical screening to multimodal AI. *Diagnostics, 16*(15), 2376. https://doi.org/10.3390/diagnostics16152376
23. Saleh, A. Y., & Chern, L. H. (2021). Autism spectrum disorder classification using deep learning. *International Journal of Online and Biomedical Engineering, 17*(8), 103–114. https://doi.org/10.3991/ijoe.v17i08.24603

24. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*, 618–626. https://doi.org/10.1109/ICCV.2017.74
25. Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. *Proceedings of the 34th International Conference on Machine Learning, 70*, 3319–3328.
26. Thabtah, F. (2017). Autism spectrum disorder screening: Machine learning adaptation and DSM-5 fulfillment. *Proceedings of the 1st International Conference on Medical and Health Informatics*, 1–6.
27. Vasundhara, D. N., Lakshmi Prasanna, P., Moorthy, C. V. K. N. S. N., & Sailaja, N. V. (2026). Early detection of autism spectrum disorder in children using facial images and AI-based deep learning models. *Journal of Pharmaceutical and Medical Sciences, 15*(1), 137–145. https://doi.org/10.47310/jpms2026150119
28. Anupama, N., & Patil, C. M. (2026). Data-driven approaches for autism detection: A comprehensive review of machine learning algorithms and datasets. https://doi.org/10.1002/jdn.70172