

Lightweight deep learning framework for autism spectrum disorder detection from children's facial images

Prafulla Kumari Kannangala Siddaiah¹, Gayathri Koti Madhusudhana²

¹Department of Electronics and Communication Engineering, Faculty of Engineering and Technology,
M. S. Ramaiah University of Applied Sciences, Bengaluru, India

²Department of Electronics and Communication Engineering, School of Engineering, Dayananda Sagar University, Bengaluru, India

Article Info

Article history:

Received Jul 5, 2025

Revised Jul 21, 2026

Accepted Aug 12, 2026

Keywords:

Autism spectrum disorder

Data augmentation

Deep learning

Facial image analysis

Grad-CAM interpretation

Preprocessing

ABSTRACT

Autism spectrum disorder (ASD) refers to a neurodevelopmental disorder that affects children's social interaction, responsiveness, and behavioral patterns. Early identification of ASD is necessary for prompt treatment of children. The existing diagnosis procedures are subjective, take a lot of time, and require interpretation by an expert. In this study, an efficient network baseline 0 (EfficientNetB0)-based deep learning (DL) approach is developed for automatic ASD detection from children's facial images. The model is trained on enhanced training data by using a preprocessing-aware data augmentation strategy. The proposed augmentation method includes image processing operations such as face alignment, Gaussian smoothing, hue, saturation, value (HSV) normalization, contrast-limited adaptive histogram equalization (CLAHE), and horizontal flipping. The experimental analysis demonstrated the effectiveness of the proposed predictive model against existing methods on the same dataset, where the model achieved 93% test accuracy and an area under the curve (AUC) score of 0.97. Additionally, gradient-weighted class activation mapping (Grad-CAM)-based interpretation demonstrated that the model focused on semantically relevant facial regions, which offers visual support and transparency for the classification. The proposed predictive system is lightweight and provides a non-invasive and scalable solution to support ASD screening using facial images.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Prafulla Kumari Kannangala Siddaiah

Department of Electronics and Communication Engineering, Faculty of Engineering and Technology

M. S. Ramaiah University of Applied Sciences

University House, New BEL Rd, M S R Nagar, Mathikere, Bengaluru, Karnataka 560054, India

Email: prafullakumari.ec.et@msruas.ac.in

1. INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that usually manifests during infancy and early childhood, causing difficulties in social interaction, communication abilities, and behavioral development [1]. The common symptoms in ASD include decreased social engagement, communication difficulties, repetitive behaviors, and increased sensory sensitivity [2]. Although the precise cause of ASD has not been fully established, current evidence suggests that both genetic predisposition and environmental influences play important roles in its development [3]. However, early detection of ASD is very important to enhance overall development and improve long-term outcomes for children. But due to the varied nature of symptoms and their similarity with other developmental conditions, early diagnosis of ASD is often delayed or missed altogether in low-resource or rural areas. In practice, the standard methods of ASD

diagnosis include different screening tools such as the childhood autism rating scale (CARS) [4], screening tool for autism in toddlers and young children (STAT) [5], and modified checklist for autism in toddlers (M-CHAT) [6], [7]. These existing approaches are often time-consuming due to the repetitive nature of the diagnostic process, which requires behavioral observations by both professionals and caregiver-reported responses. There are also many other issues such as inconsistent diagnosis, variable reliability, age and cultural limitations, low sensitivity and specificity, and biased analyses due to the subjective nature of reporting [8]–[10]. Artificial intelligence (AI) has demonstrated great success in various medical imaging tasks with the adoption of advanced machine learning (ML) and deep learning (DL) in disease classification and object identification, achieving better accuracy and efficacy of ASD diagnosis [11], [12]. Facial analysis also emerged as an important non-invasive biomarker for ASD screening, as several studies have reported that children with ASD show differences in facial structure, expression patterns, and eye gaze compared to normal children [13]. However, the success of DL approaches requires effective data preprocessing, feature representation, and DL modelling.

In recent state-of-the-art work, researchers have presented many computational schemes using varied techniques. The work carried out by Rashid and Shaker [14] and Alam *et al.* [15] explored transfer-learning approaches considering DL models such as xtreme of inception (Xception), visual geometry group 19 (VGG-19) and residual network 50 layers (ResNet50) on the Kaggle facial image dataset. However, they did not emphasize the data preprocessing strategies they adopted. Similarly, Lu and Perkowski [16] showed strong performance of VGG16-based models on images of clinically diagnosed ASD children. But this work did not focus on addressing inconsistent facial poses identified by other researchers that reduce classifier accuracy. The study of Alam *et al.* [17] used the multitask cascaded convolutional neural networks (MTCNN) model to normalize face orientation before feeding into DL models (ResNet50, Xception, and mobile network (MobileNet)) and reported their outcome with significant improvement in diagnostic accuracy. Toranjsimin *et al.* [18] presented a scheme of cross wavelet transform images using electroencephalogram (EEG) signals, which is passed to a DL model for ASD diagnosis. Apart from considering static facial images, there are other works that have fused dynamic behavioral signals with DL models.

Liu *et al.* [19] have done eye-tracking analysis during face-recognition tasks, using classic ML on gaze-pattern features. Mazumdar *et al.* [20] further combined eye-movement metrics with object-presence features in images and improved classification via random forests and a support-vector-machine classifier. The work by Kumar and Jaiswal [21] introduced a hybrid architecture where visual DL is integrated with a bidirectional long short-term memory (BiLSTM) model to capture temporal emotional dynamics from adults' facial expressions during movie viewing, then extending image-only methods to the time domain. Meanwhile, non-visual neuroimaging modalities provide another orthogonal route. Another work considering medical data for ASD diagnosis is conducted by Sherkatghanad *et al.* [22], where the authors have applied 3D convolutional neural networks (3D CNNs) to resting-state functional magnetic resonance imaging (fMRI) from the autism brain imaging data exchange (ABIDE) dataset and achieved competitive outcomes for ASD classification. The work done by Reza and Zare [23] extracted volumetric and surface-based morphometric features from structural magnetic resonance imaging (MRI) data, then compared different ML classifiers and artificial neural network (ANN), where ANN outperformed with 88.5% accuracy. Tawhid *et al.* [24] considered EEG spectrogram data and applied CNNs on time-frequency representations that improved classification accuracy. Talaat [25] detected facial key points via shape predictors before passing data to CNN classification, thereby validating on internal and external cohorts to ensure generalizability.

Existing studies have highlighted the limitations of traditional ASD diagnostic methods as being subjective, time-consuming, and dependent on expert input. Recent research has explored image-based approaches using transfer learning and DL, multimodal techniques (eye-tracking, EEG, MRI, fMRI, and temporal modelling). However, multimodal approaches often involve high computational complexity and lack scalability for use in low-resource clinical environments. The temporal modelling studies are also associated with issues such as modelling done on domain-mismatched datasets, many of which use facial expression data from unrelated contexts such as movies or emotion benchmarks, which makes them less appropriate for ASD detection. Additionally, facial-image-based models often suffer from low accuracy due to small dataset sizes, and they lack detailed preprocessing steps, which are critical for real-world generalization.

This paper presents a DL-based framework with a comprehensive preprocessing scheme for detecting ASD using children's facial images. Unlike previous works that ignore detailed data preparation steps or depend on domain-mismatched datasets, this study introduces a structured augmentation pipeline that focuses on facial alignment, noise reduction, image enhancement, and standardization to increase the data diversity towards improving the generalization of models on unknown data. The novelty of this work is the inclusion of lightweight methodology using facial images with extensive preprocessing-aware data augmentation, the implementation of efficient DL models, and feature-level model interpretability that makes model decisions more transparent and trustworthy.

2. METHOD

It is significant to realize that the task of automated ASD detection is a multifaceted problem in itself, as recognizing facial expressions from images to identify ASD is challenging as it may share similar high-level features with non-ASD classes, and facial recognition datasets for ASD are small, unbalanced, lacking diversity in terms of ethnicity, age, and lighting, which limits the generalization of the model and increases the difficulty. Therefore, the prime agenda of the reported research work is to design an accurate predictive methodology for ASD recognition using a robust DL approach while adapting to the unique low-level features and high-level characteristics found in face images.

The modelling of the proposed system follows a two-phase implementation approach, viz. i) image preprocessing-aware data augmentation and ii) efficient DL model implementation to address the problem of facial image variations and model overfitting during its feature learning phase. In the first phase, data augmentation is applied to the training set under comprehensive image preprocessing operations to enhance training set diversity. Firstly, the facial image undergoes frontal face alignment operation to address variations in pose. Then, the second operation applies a Gaussian smoothing operation to reduce random noise introduced as intensity variation and compression artefacts. The denoised image is then subjected to color correction using a histogram-based approach, contrast-limited adaptive histogram equalization (CLAHE), and hue, saturation, value (HSV) color space conversion to bring the contrast and brightness tone of the image in the normalized range. Additional data augmentation applied during training is rotational transformation and horizontal reflection to improve the training set variability. In the second phase, a DL model efficient network baseline 0 (EfficientNetB0), is implemented with a transfer learning approach, added with custom layers to learn to learn both fine-grained (low-level features) and abstract representations (high-level-image attributes) and passed to the densely connected layers to classify them. The training of the DL model is carried out using carefully selected hyperparameters for faster convergence and effective generalization on an unseen dataset. In addition, the study also implements gradient-weighted class activation mapping (Grad-CAM) for feature-level model explainability. The schematic outline of the proposed system and implementation workflow is exhibited in Figure 1.

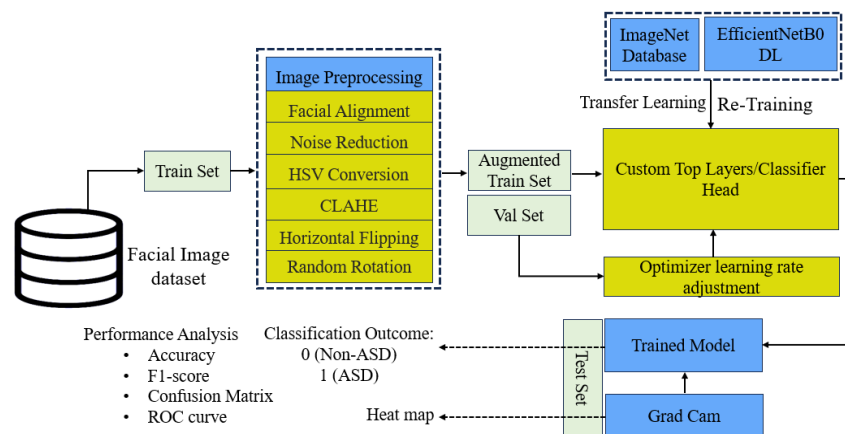


Figure 1. Displays the workflow outline of the proposed methodology

2.1. Dataset

In this research work, the ASD facial image dataset is utilized from Kaggle [26], which contains facial photographs of children aged 2 and 14 years. The dataset consists of 2,480 RGB images in JPEG format and comes with two pre-defined classes (i.e., autistic and non-autistic). This dataset is analyzed and partitioned into three subsets using a stratified sampling strategy, where the train-set includes 2,080 image samples, the val-set comprises 100 samples, and the test-set involves 300 samples. The images exhibit variations in resolution, face orientation, background, and lighting, which require significant preprocessing operations and augmentation for effective model training. Table 1 highlights the dataset distribution.

Table 1. Presents dataset description			
Category	Autistic	Non-autistic	Total
Train	1,040	1,040	2,080
Val (validation)	50	50	100
Testing	150	150	300

2.2. Preprocessing-aware data augmentation

This study introduces a preprocessing-aware data augmentation strategy designed to increase the diversity of the training set D_{train} and improve the generalization of the DL model. Unlike conventional augmentation techniques that apply random transformations, the proposed approach treats preprocessing as a functional component A_θ of the augmentation pipeline such that: $A_\theta = \tau_n \circ \tau_{n-1} \circ \dots \circ \tau_1(x)$, where $\circ$ denotes function composition and each τ_k represents a domain-specific preprocessing operation, namely, face alignment, Gaussian smoothing, HSV normalization, CLAHE, horizontal flipping, and rotation, respectively. The augmented dataset is generated as $D_{Aug} = D_{train} \cup \{(A_\theta(x_i), y_i)\}_{i=1}^N$, where x_i denotes an RGB facial image of size $H \times W$ with 3 channels, and $y_i \in \{0, 1\}$ is the corresponding class label (0: non-ASD, 1: ASD). The obtained D_{Aug} consists of both original and transformed training instances, which enables the DL model f_ϕ parameterized by ϕ to learn from a richer set of facial variations, improving robustness to noise, lighting, and pose variability during inference. The preprocessing operation used in the data augmentation is discussed in Algorithm 1.

Algorithm 1. Preprocessing-aware data augmentation function

Input: D_{train} , σ (for Gaussian smoothing), clip limit c , tile size t , flip probability p_f , rotation range Θ

Output: augmented image $\tilde{x}_i$

Steps:

- 1: Detect bounding box B and landmarks $\{(x_k, y_k)\}_{k=1}^5$ using MTCNN.
 - 2: Let $(x_L, y_L), (x_R, y_R)$ denote left and right eye coordinates.
Compute rotation angle: $\theta = \tan^{-1}(y_R - y_L / x_R - x_L)$
 - 3: Construct affine transform τ_θ and apply rotation to obtain aligned image: $x'_i = \tau_\theta(x_i)$
 - 4: Apply Gaussian filter G_σ to reduce noise: $x''_i = x'_i * G_\sigma$
 - 5: Convert image to HSV: $x''_i \rightarrow \{H, S, V\}$
Apply CLAHE to value channel $V' = CLAHE(C; c, t)$
Merge enhanced V with H, S and convert back to RGB
 - 6: Resize image to 224×224 and normalize pixel values to $[0, 1]$
 - 7: If $\text{uniform}(0, 1) < p_f$:
Apply horizontal flip: $\tilde{x}_i = Flip(\tilde{x}^{enh}_i)$
 - 8: If $\text{uniform}(0, 1) < p_r$:
Apply small random rotation $\theta' \sim u(-\Theta, +\Theta)$
 - 9: Output final augmented image $\tilde{x}_i$
- End

The first step detects facial regions and five key landmarks using the MTCNN [27], which includes proposal network (P-Net), refinement network (R-Net), and output network (O-Net) modules to detect faces and predict the landmarks (step-1). The P-Net serves as the first stage in MTCNN, operating on image patches of size 12×12 pixels. It extracts low-level features via 3×3 convolutional filters to generate initial bounding boxes and predict the presence of a face. The binary classification for face/non-face detection is optimized using the binary cross-entropy (BCE) loss. The output from P-Net is further fine-tuned by the R-Net, which processes 24×24 pixels patches and eliminates incorrect positive predictions and regresses more accurate bounding box coordinates. The network's training incorporates region localization using the L2 loss (Euclidean distance). The final network, O-Net, handles 48×48 pixels inputs and yields the most precise detection and landmark localization. It utilizes deeper feature representations and refines both the classification and localization tasks. Besides classification and bounding box regression, O-Net predicts five facial landmarks corresponding to the eye, nose tip, and mouth corners.

The next step of the algorithm is subjected to facial alignment using the detected eye coordinates (step-2). Here, the angle of rotation θ is calculated considering the coordinates of the left and right eyes, which ensures a correct horizontal eye line and reduction in the pose variance. The algorithm then constructs an affine transformation matrix using the angle θ , and the face is rotated accordingly, which compensates for head tilt and standardizes face orientation (step-3). Further steps are taken towards enhancing image feature representation, where the aligned image is subjected to a Gaussian filter G_σ with kernel size 5×5 and standard deviation $\sigma = 0$ to reduce background noise (step-4). The obtained denoised image is converted from RGB to HSV color space to isolate intensity information, and then CLAHE is applied to the V (value) channel with a clip limit $c = 1$ and tile grid size 8×8 to correct uneven illumination and perform facial texture enhancements (step-5). After enhancement, the HSV image is converted back to RGB format. The enhanced image is then transformed to a 224×224 shape to be compatible with the EfficientNet input layer, followed by pixel normalization in the range $[0, 1]$. Further, geometric augmentations are applied

conditionally to increase diversity with probability $pf = 0.5$ for horizontal flipping and with probability $pr = 0.3$, the image undergoes a small random rotation to introduce further variability (steps 7 and 8). After successful execution of the algorithm, it returns an augmented train set $\tilde{x}_i$.

2.3. Deep learning model

The proposed system adopts the EfficientNetB0 DL model with a pre-trained network approach for the ASD detection task. This model has been selected due to its optimal trade-off between accuracy and computational efficiency, identified based on the preliminary evaluation of several state-of-the-art DL models such as VGG-16, ResNet50, DenseNet121, and Xception, which were evaluated on the raw (un-augmented) dataset. The EfficientNetB0 model was then initialized with pretrained ImageNet weights and configured without its top classification layers. The input layer of the model takes an input image of shape $224 \times 224 \times 3$, and a global max pooling layer is incorporated after the backbone model to convert spatial feature maps into a compact feature representation. A custom classifier head was then added to the frozen EfficientNetB0 base model to enable binary classification, followed by a dense layer with 512 neurons and rectified linear unit (ReLU) activation. To prevent overfitting, the model is regularized using L2 regularization with a weight decay factor of $\lambda = 0.016$, as well as L1-based activation and bias regularization set to $\lambda = 0.006$. To minimize neuron co-adaptation, a dropout layer with a probability of 0.7 is applied during training by randomly disabling neuron activations. Further, a single-neuron dense layer with sigmoid activation is employed as the final classification layer to distinguish ASD from non-ASD cases. Finally, the model is optimized using the Adamax optimizer with a learning rate of 0.001, while BCE is adopted as the loss function, and the training is performed with a batch size of 32 for 30 epochs. To enhance the adaptability of learning, a custom callback mechanism is used, which includes early stopping and learning rate scheduling.

3. RESULTS AND DISCUSSION

The proposed system is implemented using Python, executed within the Anaconda distribution environment on a 64-bit Windows 11 machine with 8 GB RAM. This section reports the obtained experimental outcomes and performance discussion in terms of accuracy, precision, recall, and F1-score. The classification performance is also illustrated with a confusion matrix (CM) visualization and the receiver operating characteristic (ROC) curve to assess the trade-off between sensitivity and specificity. Moreover, model interpretability is performed through Grad-CAM visualizations to show the transparency in the prediction of the facial regions that contributed most to the predictive decision. Finally, a comparative analysis is done to demonstrate the efficacy and strength of the proposed framework against existing approaches using the same dataset.

Figure 2 depicts the training performance of the proposed predictive model, which exhibited training accuracy saturated around 98-99%, which demonstrates an effective feature learning from the augmented train set. The graph trend also shows that the validation accuracy stabilized at 93%, further indicating strong generalization capability on the unseen data. Similarly, Figure 3 shows the corresponding loss curves, where both training and validation losses converge smoothly by epoch 15. It is to be noted that the dataset presented significant challenges due to real-world variability, and without augmentation, the model achieved only 88% validation accuracy. The proposed preprocessing-aware augmentation strategy significantly improved the performance by increasing data diversity.

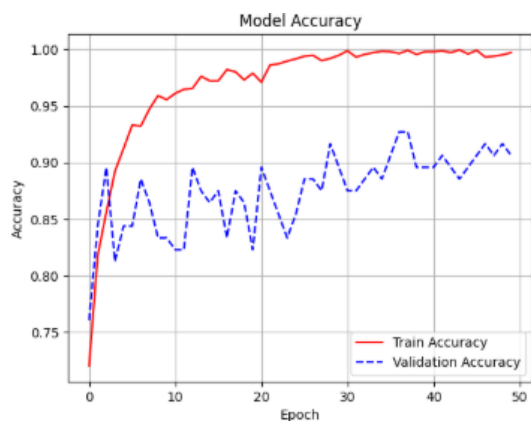


Figure 2. Training accuracy curve

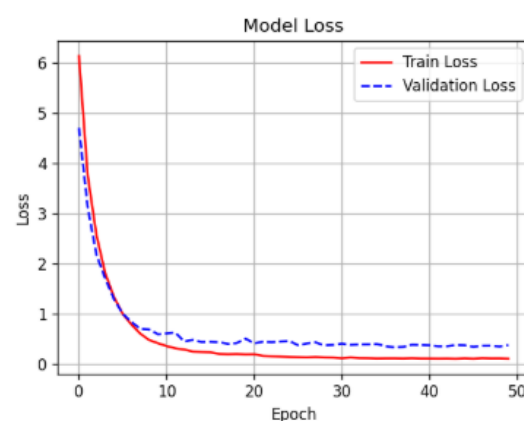


Figure 3. Training loss curve

Figure 4 presents the CM visualization that shows that the model correctly classified 278 out of 300 test samples, thereby achieving 93% test accuracy. It recorded 142 true positives (TP) and 136 true negatives (TN), with only 14 false positives (FP) and 8 false negatives (FN), thereby demonstrating a strong ability to distinguish ASD from non-ASD cases. The next Figure 5 shows the ROC curve with an AUC of 0.970, which indicates excellent discrimination capability of the model with minimal overlap between the two classes. The model showed a high TP rate and low FP rate, which suggests that the model is effective in identifying autistic cases, which reduces the risk of misdiagnosis for initiating the early intervention process.

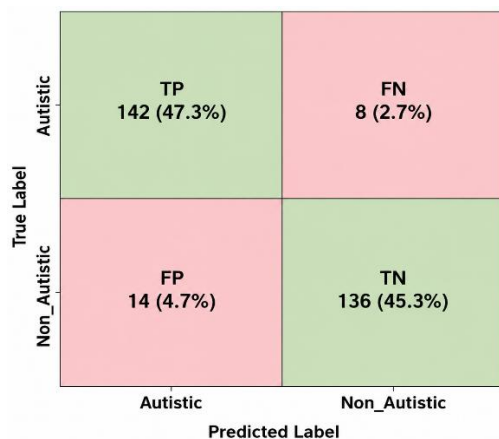


Figure 4. Analysis of CM visualization

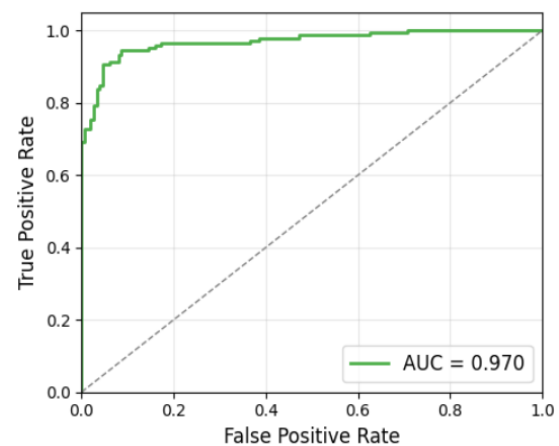


Figure 5. Visualization of ROC curve

The next, Figure 6 shows the class-wise quantified evaluation for the autistic class; the model achieved a precision of 91%, recall of 95%, and an F1-score of 93%, indicating effective identification of ASD cases. Similarly, for the non-autistic case, the model reported 94% precision, 91% recall, and 93% F1-score, which further confirms a balanced performance achieved by the DL model for both categories. Figure 7 illustrates the precision-recall (PR) curve, where the average precision (AP) score reaches to 0.98, which indicates an excellent trade-off achieved by the model between precision and recall.

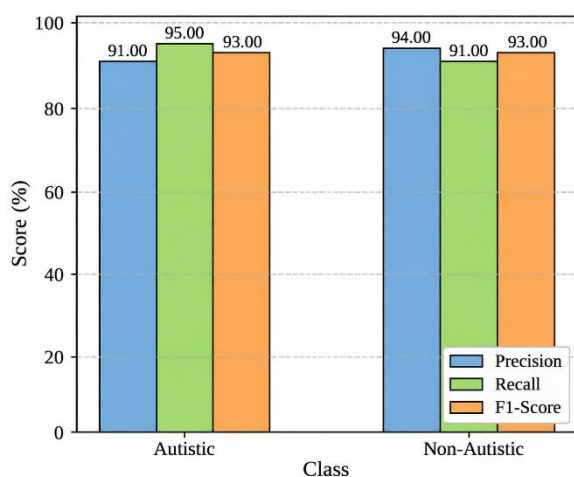


Figure 6. Quantified classification outcome

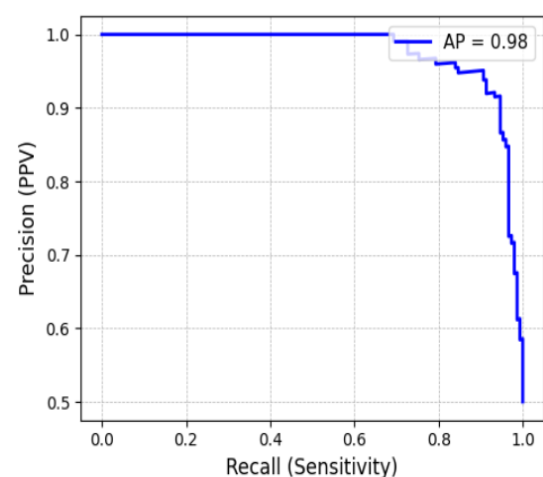


Figure 7. Precision vs. recall

The analysis of Figure 7 reveals a high AP score by the proposed model, which further exhibits strong predictive power even with slight class imbalance issues associated with the input dataset. The PR analysis also complements the ROC curve presented in Figure 5 by exhibiting effective or true ASD cases without producing higher FP. The next, Figure 8 presents feature-level explainability using Grad-CAM [28]

to showcase the model's decision-making rationale. The explanatory visualization shows the original image, the activation heatmap obtained from the final convolutional layer of the predictive model, and the facial attention map obtained by overlaying the heatmap on the input image.

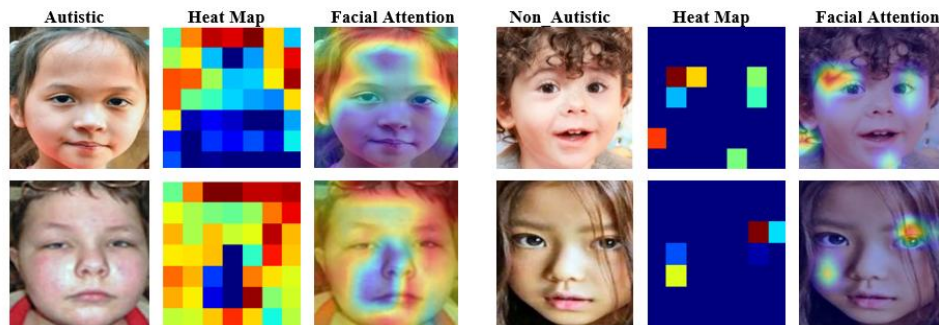


Figure 8. Feature-level explainability using Grad-CAM heatmap for the two random samples from the test set

It can be seen that the model for the autistic case concentrated on the upper facial region around the periocular area and forehead, which indicates that these facial regions provided important features for ASD classification. On the other hand, for the non-autistic case, attention is more distributed across the eyes, mouth, and mid-face regions, that captures typical socio-emotional expression cues. The above insights demonstrate that the model depends on distinguishable facial cues, which validates that the model is leveraging semantically relevant features and supports interpretability in automated ASD screening scenarios. Table 2 presents a comparative evaluation of the proposed EfficientNetB0-based model against several existing ASD classification approaches applied to similar facial image datasets. The results reported in Table 2 were obtained from previously published studies that used the same publicly available facial image datasets. However, differences may exist in dataset partitions, preprocessing methods, and predictive modelling.

Table 2. Comparative analysis

Reference	Model	Accuracy (%)
Kashef [29]	Xception	91.00
Khosla <i>et al.</i> [30]	MobileNet	87.00
Kaur and Gupta [31]	VGG-16	70.00
Ghazal <i>et al.</i> [32]	Autism spectrum disorder deep transfer learning algorithm (ASDDTLA)	87.70
Jahanara and Padmanabhan [33]	VGG-16	84.67
Rahman and Subashini [34]	Xception	90.00
Proposed model	EfficientNetB0	93.00

Based on the analysis from Table 2, it can be seen that the proposed model has achieved the highest accuracy of 93.00%, and outperforms prior methods such as Xception with 91.00% in [29] and 90.00% in [34], MobileNet with 87.00% in [30], and custom models like ASDDTLA with 87.70% accuracy in [32]. The traditional model VGG-16 reported lower accuracies, i.e., 70.00% in [31] and 84.67% in [33], which reflects its limited capacity in learning deep facial representations. The reason behind achieving better performance is the integration of preprocessing-aware data augmentation and the incorporation of regularization techniques in EfficientNetB0, and fine-tuned training configuration further contributed to the model's superior accuracy and generalization on unseen data. The effectiveness of the proposed methodology is further validated by the ablation study carried out in Table 3.

It can be seen that face alignment and illumination normalization resulted in significant improvements by lowering differences in poses and lighting conditions, while geometric augmentation contributed to further improvements. The entire process, incorporating preprocessing as a part of data augmentation, resulted in the highest accuracy and area under the curve (AUC), which proves the importance of considering preprocessing as an inherent part of data augmentation in order to achieve better generalization for the task of ASD faces classification. On the other hand, Table 4 presents the ablation experiment performed on primary training hyperparameters in order to substantiate the choice of model and training configuration that are used in the proposed work. This analysis reveals that a learning rate of 0.001, 30 epochs of training, and a dropout of 0.7 provided a good balance between convergence and generalization.

Table 3. Ablation study on the impact of preprocessing-aware data augmentation

Configuration	Preprocessing	Accuracy (%)	AUC
Baseline model	No preprocessing, raw images	88.0	0.91
+ face alignment	MTCNN-based alignment	89.6	0.93
+ Noise and color normalization	Gaussian smoothing + HSV + CLAHE	91.2	0.95
+ Geometric augmentation	Rotation + horizontal flip	92.1	0.96
Proposed	All preprocessing-aware augmentation steps	93.0	0.97

Table 4. Ablation study on model performance under different hyperparameter settings

Configuration	Learning rate	Epochs	Dropout	Accuracy (%)	AUC
Config-1	0.0001	30	0.7	90.4	0.94
Config-2	0.0005	30	0.7	91.8	0.95
Proposed (final)	0.001	30	0.7	93.0	0.97
Config-3	0.001	20	0.7	91.9	0.95
Config-4	0.001	40	0.7	92.4	0.96
Config-5	0.001	30	0.5	92.1	0.96
Config-6	0.001	30	0.8	91.6	0.95

The experimental outcomes showed that the proposed preprocessing-aware data augmentation strategy improved ASD classification performance compared with the existing baseline models trained on raw images. The reason behind the improved performance is the integration of preprocessing operations and data augmentation, which enabled the model to learn more precise facial features under dynamic conditions such as variations in pose, illumination, and image quality. The use of EfficientNetB0 also supported the classification task with its compound scaling feature and attention mechanism, which provided effective feature extraction with moderate computational complexity. However, there are also some limitations associated with the proposed methodology. The first limitation is the dataset size with limited demographic diversity; there is the possibility that domain bias may appear and influence the generalization capability of the model when introduced to a complex dataset or scenario. The absence of statistical significance testing or confidence intervals is another aspect that can be addressed in future work to further support model reliability. In future, the proposed framework will be extended towards a more optimized approach and will also be validated with medical images for ASD detection.

The proposed framework is intended to act as a decision support system in the clinical context rather than replace medical experts. It can be adopted for decision-making at initial ASD screening to assist clinicians in identifying children who may require comprehensive clinical assessment. The adoption of the proposed model will be helpful in optimizing clinical resources, reducing the workload in hospitals, and supporting earlier referral for diagnostic evaluation. It should also be noted that the final diagnosis will always be performed by qualified healthcare professionals with standardized diagnostic procedures.

4. CONCLUSION

This paper presented a DL-based ASD detection framework using facial color photographs of children. A preprocessing-aware data augmentation strategy was developed to improve training dataset quality for efficient model training. The framework employed EfficientNetB0 for precise feature encoding, ASD specific pattern learning and classification due to its effective feature extraction capability and lower computational complexity. The proposed model is evaluated on raw and un-augmented test set samples and showed an accuracy of 93% with a 0.97 AUC score. The comparative assessment further exhibited the efficacy of the proposed predictive model over different existing methods. The proposed framework is designed to serve as a non-invasive ASD screening tool and assist clinicians during the initial assessment process. Future research work will focus on validating the proposed predictive model on larger facial datasets and medical imaging data by incorporating more optimized approaches. The proposed methodology will be also extended towards incorporating temporal behavioral information and eye-tracking signals, using a dataset of neuroimaging and speech to further improve ASD detection performance.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Prafulla Kumari	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Kannangala Siddaiah														
Gayathri Koti		✓				✓		✓	✓	✓	✓	✓		
Madhusudhana														

C : Conceptualization	I : Investigation	Vi : Visualization
M : Methodology	R : Resources	Su : Supervision
So : Software	D : Data Curation	P : Project administration
Va : Validation	O : Writing - Original Draft	Fu : Funding acquisition
Fo : Formal analysis	E : Writing - Review & Editing	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [PKKS], upon reasonable request.

REFERENCES

[1] E. G. Özgür and S. Balci, "The evolution of autism: a bibliometric analysis," *Acta Medica Nicomedia*, vol. 5, no. 1, pp. 1–7, 2022, doi: 10.53446/actamednicomedia.1059535.

[2] D. J. -Beaulieu and M. L. Sulkowski, "The diagnostic and statistical manual of mental disorders: fifth edition (DSM-5) model of impairment," in *Assessing Impairment: From Theory to Practice*, New York, United States: Springer, 2016, pp. 167–189, doi: 10.1007/978-1-4899-7996-4_8.

[3] Y. Zhang *et al.*, "Biological implications of genetic variations in autism spectrum disorders from genomics studies," *Bioscience Reports*, vol. 41, no. 7, Jul. 2021, doi: 10.1042/BSR20210593.

[4] K. L. Chen, C. H. Lin, T. Y. Yu, C. Y. Huang, and Y. D. Chen, "Differences between the childhood autism rating scale and the social responsiveness scale in assessing symptoms of children with autistic spectrum disorder," *Journal of Autism and Developmental Disorders*, vol. 48, no. 9, pp. 3191–3198, 2018, doi: 10.1007/s10803-018-3585-y.

[5] C. C. Wu, C. H. Chiang, Y. M. Hou, C. L. Chu, and J. H. Liu, "Utility of the Taiwan version of the screening tool for autism in two-year-olds to detect autism in children aged three years," *Journal of Intellectual and Developmental Disability*, vol. 44, no. 3, pp. 337–345, 2019, doi: 10.3109/13668250.2017.1413078.

[6] D. L. Robins and D. Fein, "Modified checklist for autism in toddlers, revised, with follow-up (M-CHAT-R/F), also M-CHAT," in *Encyclopedia of Clinical Neuropsychology*, Cham, Switzerland: Springer, 2018, pp. 1–5, doi: 10.1007/978-3-319-56782-2_1569-2.

[7] Y. G. Dai *et al.*, "Comparison of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F) positive predictive value by race," *Journal of Autism and Developmental Disorders*, vol. 51, no. 3, pp. 855–867, 2021, doi: 10.1007/s10803-020-04428-0.

[8] Y. Huang, S. R. C. Arnold, K. R. Foley, and J. N. Trollor, "Diagnosis of autism in adulthood: a scoping review," *Autism*, vol. 24, no. 6, pp. 1311–1327, 2020, doi: 10.1177/1362361320903128.

[9] G. Leader *et al.*, "Age of autism spectrum disorder diagnosis and comorbidity in children and adolescents with autism spectrum disorder," *Developmental Neuropsychology*, vol. 25, no. 1, pp. 29–37, 2022, doi: 10.1080/17518423.2021.1917717.

[10] S. L. Hyman, S. E. Levy, S. M. Myers, "Identification, evaluation, and management of children with autism spectrum disorder," *Pediatrics*, vol. 145, no. 1, 2020, doi: 10.1542/PEDS.2019-3447.

[11] S. M. M. Hasan, M. P. Uddin, M. Al Mamun, M. I. Sharif, A. Ulhaq, and G. Krishnamoorthy, "A machine learning framework for early-stage detection of autism spectrum disorders," *IEEE Access*, vol. 11, no. 2, pp. 15038–15057, 2023, doi: 10.1109/ACCESS.2022.3232490.

[12] M. Melinda, N. A. C. Andryani, Y. Nurdin, V. Khariyunnisa, Y. Yulita, and I. K. A. Enriko, "Deep learning performance analysis for facial expression based autism spectrum disorder identification," *Radioelectronic and Computer Systems*, vol. 2024, no. 2, pp. 30–40, 2024, doi: 10.32620/reks.2024.2.03.

[13] J. V. -Ortega *et al.*, "Early detection and intervention in autism: emerging behavioral and biological markers," in *Handbook of the Biology and Pathology of Mental Disorders*, Cham, Switzerland: Springer, 2024, pp. 1–26, doi: 10.1007/978-3-031-32035-4_100-1.

[14] A. Rashid and S. Shaker, "Autism spectrum disorder detection using face features based on deep neural network," *Wasit Journal of Computer and Mathematics Science*, vol. 2, no. 1, pp. 74–83, 2023, doi: 10.31185/wjcm.100.

[15] M. S. Alam, M. M. Rashid, R. Roy, A. R. Faizabadi, K. D. Gupta, and M. M. Ahsan, "Empirical study of autism spectrum disorder diagnosis using facial images by improved transfer learning approach," *Bioengineering*, vol. 9, no. 11, 2022, doi: 10.3390/bioengineering9110710.

[16] A. Lu and M. Perkowski, "Deep learning approach for screening autism spectrum disorder in children with facial images and analysis of ethnoracial factors in model development and application," *Brain Sciences*, vol. 11, no. 11, 2021, doi: 10.3390/brainsci11111446.

[17] M. S. Alam, M. M. Rashid, A. R. Faizabadi, and H. F. M. Zaki, "Power of alignment: exploring the effect of face alignment on asd diagnosis using facial images," *IJUM Engineering Journal*, vol. 25, no. 1, pp. 317–327, 2024, doi: 10.31436/ijumej.v25i1.2838.

[18] A. Toranjimin, S. Zahedirad, and M. H. Moattar, "Robust low complexity framework for early diagnosis of autism spectrum disorder based on cross wavelet transform and deep transfer learning," *SN Computer Science*, vol. 5, no. 2, 2024, doi: 10.1007/s42979-023-02564-9.

- [19] W. Liu, M. Li, and L. Yi, "Identifying children with autism spectrum disorder based on their face processing abnormality: a machine learning framework," *Autism Research*, vol. 9, no. 8, pp. 888–898, 2016, doi: 10.1002/aur.1615.
- [20] P. Mazumdar, G. Arru, and F. Battisti, "Early detection of children with autism spectrum disorder based on visual exploration of images," *Signal Processing: Image Communication*, vol. 94, 2021, doi: 10.1016/j.image.2021.116184.
- [21] A. Kumar and U. C. Jaiswal, "Predicting autism spectrum disorder in adults through facial image analysis: a multi-CNN with BiLSTM model," *SN Computer Science*, vol. 6, no. 3, 2025, doi: 10.1007/s42979-025-03783-y.
- [22] Z. Sherkatghanad *et al.*, "Automated detection of autism spectrum disorder using a convolutional neural network," *Frontiers in Neuroscience*, vol. 13, 2020, doi: 10.3389/fnins.2019.01325.
- [23] Z. K. K.-Reza and H. Zare, "Automatic detection of autism spectrum disorder (ASD) in children using structural magnetic resonance imaging with machine vision system," *Middle East Current Psychiatry*, vol. 29, no. 1, 2022, doi: 10.1186/s43045-022-00220-1.
- [24] M. N. A. Tawhid, S. Siuly, H. Wang, F. Whittaker, K. Wang, and Y. Zhang, "A spectrogram image based intelligent technique for automatic detection of autism spectrum disorder from EEG," *PLoS ONE*, vol. 16, 2021, doi: 10.1371/journal.pone.0253094.
- [25] F. M. Talaat, "Real-time facial emotion recognition system among children with autism based on deep learning and IoT," *Neural Computing and Applications*, vol. 35, no. 17, pp. 12717–12728, 2023, doi: 10.1007/s00521-023-08372-9.
- [26] P. Das, "Autistic children facial data set," *kaggle.com*. 2020. Accessed: Oct. 09, 2025. [Online]. Available: <https://www.kaggle.com/datasets/prayashdas/autistic-children-facial-image-dataset>
- [27] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016, doi: 10.1109/LSP.2016.2603342.
- [28] A. P. T. Minh, "Overview of class activation maps for visualization explainability," 2023, *arXiv: 2309.14304*.
- [29] R. Kashef, "ECNN: enhanced convolutional neural network for efficient diagnosis of autism spectrum disorder," *Cognitive Systems Research*, vol. 71, pp. 41–49, Jan. 2022, doi: 10.1016/j.cogsys.2021.10.002.
- [30] Y. Khosla, P. Ramachandra, and N. Chaitra, "Detection of autistic individuals using facial images and deep learning," in *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, Dec. 2021, pp. 1–5, doi: 10.1109/CSITSS54238.2021.9683205.
- [31] N. Kaur and G. Gupta, "Refurbished and improvised model using convolution network for autism disorder detection in facial images," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 2, pp. 883–889, 2023, doi: 10.11591/ijeecs.v29.i2.pp883-889.
- [32] T. M. Ghazal, S. Munir, S. Abbas, A. Athar, H. Alrababah, and M. A. Khan, "Early detection of autism in children using transfer learning," *Intelligent Automation & Soft Computing*, vol. 36, no. 1, pp. 11–22, 2023, doi: 10.32604/iasc.2023.030125.
- [33] S. Jahanara and S. Padmanabhan, "Detecting autism from facial image," *International Journal of Advance Research, Ideas and Innovations in Technology*, vol. 7, no. 2, pp. 219–225, 2021, doi: 10.13140/RG.2.2.35268.35202.
- [34] K. K. M. Rahman and M. M. Subashini, "Identification of autism in children using static facial features and deep neural networks," *Brain Sciences*, vol. 12, no. 1, 2022, doi: 10.3390/brainsci12010094.

BIOGRAPHIES OF AUTHORS



Prafulla Kumari Kannangala Siddaiah     is currently working as assistant professor of Electronics and Communication Engineering at M. S. Ramaiah University of Applied Sciences. She obtained a B.E. degree from Kuvempu University and an M.Tech. degree from Visveswaraya Technical University. She has more than 22 years of academic higher education experience. Her research interests are in AI/ML for medical diagnosis, signal and image processing. Students guided by her have won prizes in project-based competitions. She can be contacted at email: prafullakumari.ec.et@msruas.ac.in.



Dr. Gayathri Koti Madhusudhana     is an associate professor in the Department of Electronics and Communication Engineering, School of Engineering, Dayananda Sagar University, Bengaluru. She has obtained her bachelor's degree in Medical Electronics from VTU, Bangalore in the year 2009. She has obtained her master's degree in VLSI and embedded systems from VTU, Karnataka in the year 2011. Awarded Ph.D. in 2018 from JAIN (Deemed-to-be University). She is having 9 years of teaching experience and 4 years of industry experience. She has published 20 research papers in both international and national journals and presented papers in conferences. She has guided UG and PG students for their academic projects. Her areas of interest are VLSI, wireless communication, satellite communication and navigation systems. She was working as Principal Investigator for ISRO funded projects. She can be contacted at email: gayathri-ece@dsu.edu.in.