

AI For Early Detection of ASD: A VGG16-Based Child Facial Image Classification

Mera Kartika Delimayanti¹, Aftina Zahrayani^{*2}

¹Department of Computer and Informatics Engineering, Politeknik Negeri Jakarta, Indonesia

²Department of Information System, Faculty of Computer Science, Universitas Amikom Purwokerto, Indonesia

e-mail: mera.kartika@tik.pnj.ac.id¹, aftinazahrayani@gmail.com²

Article History	Abstract
Received : 06 January 2026 Revised : 22 February 2026 Accepted : 02 March 2026 Published : 28 March 2026 Keywords: Autism Spectrum Disorder, CNN, VGG16, Facial Image Classification, Deep Learning.	Autism Spectrum Disorder (ASD) is a neurodevelopmental condition whose global prevalence continues to rise, yet conventional diagnostic procedures remain costly, time-consuming, and inaccessible in many regions. This study develops an ASD classification system based on children's facial images using a Convolutional Neural Network (CNN) with VGG16 transfer learning, motivated by clinical evidence of phenotypic facial differences between children with ASD and typically developing children. The model was trained using the Piosenka Autism Image Dataset, comprising 2,936 facial images divided into training (2,536 images), validation (100 images), and testing (300 images) sets with a balanced class distribution. The images underwent preprocessing steps including resizing, pixel normalization, and data augmentation to improve generalization. Experimental results show that the model achieved an overall accuracy of 84%, with a precision of 0.80 and recall of 0.90 for the ASD class, indicating relatively balanced performance across both classes. Evaluation using a confusion matrix and ROC-AUC curve further indicates good discriminative capability. The trained model was deployed as an interactive, Streamlit-based web application that allows users to upload facial images and receive real-time predictions. These findings suggest that the proposed VGG16-based approach has potential as a supportive, non-invasive tool for early ASD screening.

^{*}Corresponding author : aftinazahrayani@gmail.com

1. INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by impairments in social communication and the presence of repetitive behavioral patterns [1], which typically emerge in early childhood. Recent data indicate that the prevalence of ASD continues to increase globally. In the United States, approximately 1 in 31 children aged 8 years (3.2%) were identified as having ASD based on estimates from the CDC's Autism and Developmental Disabilities Monitoring (ADDM) Network in 2022 [1]. Meanwhile, the World Health Organization (WHO) estimates that the global prevalence is approximately 1 in 100 children, although this figure varies across regions due to differences in access to diagnostic services [2].

Early detection of ASD is critically important, as early intervention has been proven to yield positive outcomes in cognitive abilities, daily living skills, and motor development [3]. Studies indicate that early intervention can improve children's intellectual abilities and reduce autistic symptoms even several years after the initial treatment has ended [4]. Toddlers who receive intensive individualized intervention at 18 months of age demonstrate greater improvements in receptive and expressive language, social communication, and daily living skills compared to those who begin intervention at 27 months [5].

However, conventional ASD diagnostic processes still face significant challenges, particularly in developing countries. Major barriers include financial constraints, lack of government initiatives, inadequately trained healthcare professionals, and centralized healthcare systems [6]. In low-income countries, obtaining an autism diagnosis is difficult due to limited service availability and low awareness [7]. In China, with more than 2 million autistic children, there are fewer than 500 child psychiatrists, and pediatricians are reported to have less general knowledge about autism than parents [8]

With the rapid development of artificial intelligence, machine learning and deep learning approaches have been increasingly applied in healthcare to support the diagnosis of neurological disorders. Convolutional Neural Networks (CNNs) have demonstrated superior performance compared to conventional techniques and, in some cases, even radiologists in tasks such as lesion detection and classification [9]. CNNs offer several advantages in medical applications, particularly their ability to automatically learn relevant features from raw input data such as medical images without requiring manual feature extraction [10].

In the context of ASD detection, several recent studies have utilized children's facial images as a data source due to the presence of phenotypic facial differences between children with ASD and typically developing children. Research employing transfer learning based on VGG16 on a unique dataset of clinically diagnosed children achieved a classification accuracy of 95% and an F1-score of 0.95[11]. A comparative study reported that VGG16 with a batch size of 2 and the Adam optimizer trained for 100 epochs achieved a validation accuracy of 99% and a testing accuracy of 87% [12]. Furthermore, comparisons among various pre-trained models such as ResNet34, ResNet50, AlexNet, MobileNetV2, VGG16, and VGG19 showed that ResNet50 achieved the highest accuracy of 92% [12].

Nevertheless, several challenges remain. Previous studies generally relied on limited and less representative datasets, and only a few have implemented their systems as practical applications accessible to the general public. Therefore, this study aims to develop a deep learning based facial image classification system for children using the VGG16 architecture with a transfer learning approach and to implement it as a web-based application using Streamlit. This system is expected to serve as a supportive solution for early ASD detection that is more objective, efficient, and easily accessible.

2. METHOD

This study adopts a quantitative approach with an experimental research design, which is the de facto standard in machine learning research [4]. The primary objective of this study is to develop and evaluate a deep learning based model for classifying children's facial images. Deep learning is selected due to its ability to automatically extract features without the need for manual feature engineering and its proven superiority in medical image classification tasks.

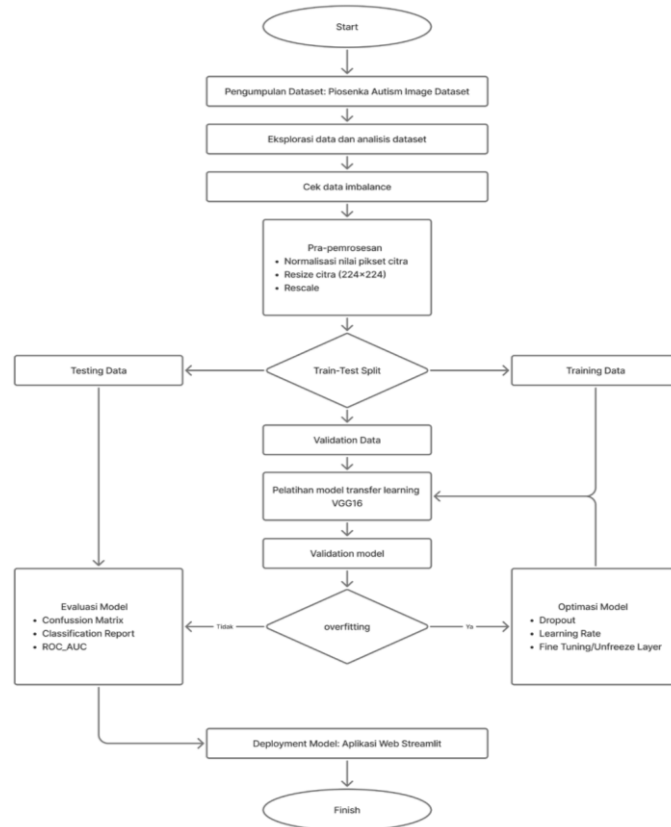


Figure 1 Research Methodology Flow

2.1 Dataset Description

This study uses the Piosenka Autism Image Dataset, obtained from the public repository Zenodo [13]. The dataset contains facial images of autistic and typically developing individuals and is organized into three subsets: training, validation, and testing. All images were collected and anonymized to protect participants' identities [1]. Most of the images were downloaded from autismrelated websites and Facebook pages, with the majority representing children from Europe and the United States [1]. The dataset is classified into two categories: Autism Spectrum Disorder (ASD) and Typical Development (Non-ASD).

The dataset is structured into three separate folders corresponding to the stages of model development:

- Training set: 1,268 images of autistic individuals and 1,268 images of typical individuals (total: 2,536 images) [1]
- Validation set: 50 images of autistic individuals and 50 images of typical individuals (total: 100 images) [1]
- Test set: 150 images of autistic individuals and 150 images of typical individuals (total: 300 images) [1]

Thus, the total number of images used in this study is 2,936 children's facial images. Each image is in JPG format with a resolution of 224×224 pixels and RGB color space [1], which conforms to the input specifications of the VGG16 architecture. The default input size for the VGG16 model is 224×224 pixels [14], which has become a widely adopted standard in deep learning-based image classification research.

a. Class Distribution of the Dataset

The dataset has a class ratio of 1:1 between autistic subjects and typical controls, although the number of male facial images is approximately three times higher than that of females [7]. The balanced distribution between the ASD and Non-ASD categories in each subset (training, validation, and testing) provides several advantages:

- Reduces the risk of model bias toward a particular class

-
- Improves the stability of the neural network learning process
 - Makes evaluation results (accuracy, precision, recall, and F1-score) more representative

Class balance is a critical factor in deep learning-based classification, as imbalanced datasets may cause the model to favor the majority class and degrade its generalization capability.

b. Role of the Training, Validation, and Test Sets

The main purpose of dividing the dataset into a validation set is to prevent overfitting, a condition in which a model performs very well on training samples but fails to generalize and accurately classify previously unseen data [15]. The three subsets serve different roles in model development:

- Training Set

The training set is the largest portion of the dataset and is used to train the machine learning model and optimize its internal parameters [16]. This subset is used to update the neural network weights based on visual patterns in the facial images of autistic and typical children and serves as the primary source of learning for the model.

- Validation Set

The validation set is used during the iterative model development process to tune hyperparameters (such as learning rate or regularization strength) and to assess intermediate model performance in order to detect overfitting [16]. Data in this subset are not used to update the model weights but to evaluate generalization performance and to support the application of early stopping techniques to prevent overfitting.

- Test Set

The test set is a completely independent portion of the data (typically 10–20% of the total dataset) that is held out from both training and validation. It is used after model training is completed to provide an unbiased evaluation of how well the model generalizes to unseen data [16]. This subset is entirely excluded from the learning and validation processes; therefore, the test results reflect the model's performance on truly new data.

Dividing the dataset into training, validation, and testing subsets helps assess how well the machine learning model generalizes to unseen data and prevents overfitting, where a model performs well on training data but fails to generalize to new instances. This strategy aims to avoid data leakage, maintain evaluation objectivity, and obtain a more realistic estimate of model performance under real-world conditions.

2.2 Preprocessing Data

The preprocessing stage is a critical component in the development of image-based deep learning models, as the quality and uniformity of input data strongly influence the model's ability to extract meaningful features and generalize to unseen data. Techniques such as resizing, noise reduction, normalization, contrast enhancement, and feature extraction are essential elements of image preprocessing for classification tasks [17]. Studies have shown that appropriate preprocessing can significantly improve training stability and CNN model accuracy.

a. Image Size Transformation (Resizing)

All images in the dataset were resized to 224×224 pixels to match the input specifications of the VGG16 architecture used in this study. Neural networks require fixed-size inputs; for example, architectures such as VGG16 and ResNet expect images of size 224×224 pixels [12]. Resizing and cropping are therefore crucial to ensure consistent input dimensions in machine learning algorithms, which contributes to improved model performance [17]. A fixed input size must be defined, and all images must be resized accordingly. Without size standardization, differences in image dimensions may result in tensor shape mismatches and disrupt the model training process.

b. Pixel Value Normalization

After resizing, pixel values were normalized from the range 0–255 to 0–1 by dividing each value by 255. Scaling pixel values to a range more suitable for neural networks enables faster and more stable training [18]. Thus, each pixel value in the images was normalized to the interval [0, 1] [19]. Normalization improves model convergence, helps gradients behave more consistently, and often leads to better accuracy [12]. Pixel normalization has been shown to accelerate convergence, reduce the risk of vanishing gradients, and improve the stability of weight updates during CNN training.

c. Data Augmentation

To increase the diversity of the training data and reduce the risk of overfitting, data augmentation techniques were applied to the training subset. Data augmentation is a method for artificially increasing the variability of the training dataset through various transformations, which helps prevent overfitting and enables the model to learn features that are invariant to such transformations [19].

- Augmentation preserves the original class labels while altering the visual representation of the images using several transformations, including:
- Random rotation – rotating images by a certain angle to introduce rotational invariance in the model [19]
- Horizontal flipping – flipping images horizontally (or vertically) to achieve orientation invariance [19]
- Zoom in and zoom out
- Translation (width and height shifting)
- Shear transformation

In most cases, data augmentation has been shown to improve model accuracy, reduce overfitting, and enhance performance on unseen datasets [19].

2.3 Splitting Data

Dataset splitting is a crucial step to ensure that the training, validation, and evaluation processes are conducted objectively and to prevent overfitting [20]. A common practice in image classification research is to divide the dataset into three subsets: a training set, a validation set, and a testing set [1].

- In this study, the dataset was divided as follows:
- Training set (2,536 images) – used to train the model
- Validation set (100 images) – used to monitor model performance during training
- Testing set (300 images) – used for final model performance evaluation

This corresponds to approximately 86% training data, 3.4% validation data, and 10.6% testing data. The choice of these ratios is subjective and involves certain trade-offs: using more training data helps the model find better solutions, while a larger validation set supports more reliable model selection and hyperparameter tuning [5].

The use of three separate subsets aims to:

- Prevent data leakage
- Provide a realistic estimate of model performance
- Maintain the statistical validity of evaluation results

With a structured dataset split and balanced class distribution, the training and evaluation processes are expected to produce a stable, objective classification system with good generalization capability.

2.4 Development Model

The model employed in this study is a Convolutional Neural Network (CNN) based on the VGG16 architecture, implemented using a transfer learning approach. CNNs were chosen due to their ability to automatically and effectively extract spatial features from images and their proven superiority in various medical image classification and facial recognition tasks.

VGG16 is a convolutional neural network proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper “Very Deep Convolutional Networks for Large-Scale Image Recognition” [14]. The VGG16 model achieved a top-5 test accuracy of 92.7% on the ImageNet

dataset, which contains more than 14 million images across 1,000 classes (Autism Parenting Magazine). VGG16 consists of 16 learnable layers, including 13 convolutional layers and 3 fully connected layers [14].

a. Transfer Learning Approach

Transfer learning allows a model to function as a generic visual feature extractor by leveraging previously learned feature maps without training from scratch [21]. In this study, the initial weights of VGG16 pretrained on the ImageNet dataset were used as a feature extractor. The base convolutional layers of VGG16 were frozen to preserve the general visual features learned, while the classification layers at the top were modified to perform binary classification (ASD and Non-ASD).

b. Model Architecture

The implemented model architecture consists of:

- Pretrained VGG16 convolutional base (without the top layer)
- Flatten layer to convert feature maps into a one-dimensional vector
- Dense layer with 128 neurons and ReLU activation function
- Dropout layer with a rate of 0.5 to reduce the risk of overfitting
- Output layer with 2 neurons and softmax activation function

2.5 Training Model

The model training process was conducted after completing the preprocessing and dataset splitting stages. The primary objective of this phase is to adjust the model parameters so that it can optimally learn the visual patterns that distinguish the facial characteristics of children with Autism Spectrum Disorder (ASD) from those of typically developing children.

a. Model Initialization and Freezing of Base Layers

The VGG16 model loaded with pretrained ImageNet weights was used as the feature extraction backbone. All convolutional layers in the base model were frozen to prevent their weights from being updated during the initial training phase. This layer-freezing strategy aims to preserve previously learned general visual features, such as edges, textures, and facial structures [22]. Transfer learning with frozen convolutional layers has been shown to improve training stability and accelerate model convergence, particularly when the target dataset is relatively small [23].

b. Training Parameter Configuration

The model was compiled using parameter settings optimized for the task of children's facial image classification. The optimizer employed was Adam with a learning rate of 0.0001, a batch size of 32, and a maximum training duration of 5 epochs before early stopping was triggered. Adam was selected because it combines the advantages of Adaptive Gradient (AdaGrad) and Root Mean Square Propagation (RMSProp), resulting in faster and more stable convergence even when training on data with varying gradient magnitudes [24].

c. Training and Validation Process

Model training was performed iteratively using the training set, while model performance was evaluated at the end of each epoch using the validation set. This periodic evaluation aims to monitor the model's generalization ability on unseen data during learning and to detect overfitting or underfitting at an early stage [25]. At each epoch, the entire training dataset was processed in small batches, and both loss and accuracy values were computed for the training and validation sets to provide a comprehensive overview of the model's performance [22].

d. Application of Early Stopping

To prevent overfitting, an early stopping technique was applied by continuously monitoring the validation loss. The training process was automatically terminated if no improvement (i.e., decrease) in validation loss was observed for five consecutive epochs [26]. When this condition was met, the best-performing model weights saved earlier were restored and used as the final model.

e. Training Performance Monitoring

During training, two primary metrics were monitored in real time to evaluate the model's learning performance and stability: accuracy and loss, for both the training and validation sets. Accuracy measures the proportion of correct predictions out of the total predictions, while loss quantifies the discrepancy between the model's predictions and the true labels [16]. This monitoring enables researchers to identify learning issues at an early stage and apply corrective interventions when necessary.

2.6 Evaluation

Model evaluation is conducted to measure the system's ability to accurately classify facial images of children with Autism Spectrum Disorder (ASD) and typically developing children. Testing is performed using a test dataset that is not involved in either the training or validation processes to ensure the objectivity of the evaluation results. Several evaluation metrics are employed, including the confusion matrix, accuracy, precision, recall, F1-score, and the Receiver Operating Characteristic (ROC) curve along with the Area Under the Curve (AUC) [27].

a. Confusion Matrix

A confusion matrix is a contingency table that illustrates the comparison between the actual labels and the model's predicted results. This matrix consists of four main components: True Positive (TP), representing ASD samples correctly predicted as ASD; True Negative (TN), representing typical samples correctly predicted as typical; False Positive (FP), indicating typical samples incorrectly predicted as ASD; and False Negative (FN), indicating ASD samples incorrectly predicted as typical [27].

The confusion matrix provides a detailed view of the types of errors made by the model and enables in-depth analysis of classification error patterns.

b. Accuracy

Accuracy indicates the proportion of correct predictions over the total number of test samples and is calculated as:

$$Accuracy = (TP + TN) / (TP + TN + FP + FN)$$

This metric is easy to interpret and provides a general overview of model performance; however, it may be less representative for imbalanced datasets, as it can produce misleading results when one class dominates [27].

c. Precision

Precision measures the model's correctness in predicting the positive class (ASD) and is calculated as:

$$Precision = TP / (TP + FP)$$

A high precision value indicates that most samples predicted as ASD are indeed true ASD cases, thereby minimizing false alarms that could lead to unnecessary anxiety and inefficient use of follow-up evaluation resources [27].

d. Recall (Sensitivity)

Recall measures the model's ability to identify all actual ASD cases and is calculated as:

$$Recall = TP / (TP + FN)$$

This metric is critically important in medical applications, as failing to detect ASD cases (false negatives) may result in delayed early intervention, which is crucial for child development [27].

e. F1-Score

The F1-score is the harmonic mean of precision and recall and is used to balance both metrics:

$$F1 - Score = 2 \times (Precision \times Recall) / (Precision + Recall)$$

The F1-score provides a more stable evaluation in the presence of class imbalance and yields a high value only when both precision and recall are high [27].

f. ROC Curve dan AUC

The ROC curve illustrates the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) at various threshold values, where:

$$FPR = FP / (FP + TN)$$

The Area Under the ROC Curve (AUC) summarizes overall performance with values ranging from 0 to 1. An AUC of 1 indicates perfect classification, 0.5 represents random guessing, and higher AUC values indicate better model performance [27].

In medical contexts, an AUC value greater than 0.9 is considered excellent, 0.8–0.9 good, and 0.7–0.8 fair.

2.7 System Deployment Implementation Using Streamlit

After the deep learning model has been trained and evaluated, the next stage is system deployment so that the model can be directly used by end users. Deployment aims to transform the research model into an interactive application, ensuring that the research outcomes are not merely theoretical but also practical and beneficial in real-world settings. In this study, deployment was implemented using the Streamlit framework, a Python library designed to rapidly and simply build web-based machine learning applications without requiring expertise in frontend web technologies [28].

Streamlit was selected due to several significant advantages for deploying deep learning–based classification systems. It supports direct integration with TensorFlow and Keras models, does not require knowledge of frontend programming languages such as HTML, CSS, or JavaScript, enables rapid and lightweight development, and supports real-time visualization of prediction results. The deployment system architecture consists of three main integrated components. The first component is a Streamlit-based web application user interface, which provides an interface for users to upload children’s facial images. The second component is the deep learning model, namely the fine-tuned VGG16 model stored in .h5 format. The third component is the prediction engine, which is responsible for performing image preprocessing, executing class prediction, and displaying the results [23].

The deployment implementation is carried out through several systematic steps. First, the trained model is saved using the function `model.save("model_autism.h5")`, which enables the system to perform inference without retraining the model [29]. Second, the user interface is built using Streamlit components such as `st.file_uploader()` for image uploading, `st.image()` for displaying images, `st.button()` to initiate prediction, and `st.write()` to present the results. Third, preprocessing during deployment is performed to ensure consistency with the preprocessing stage during training, including resizing the image to 224×224 pixels, normalizing pixel values by dividing by 255, and adding a batch dimension [12]. Fourth, the inference process is conducted by the model predicting class probabilities using the softmax function:

$$P(y = c|x) = \text{softmax}(z_c)$$

The class with the highest probability is selected as the final prediction:

$$\hat{y} = \text{argmax } P(y = c|x)$$

Fifth, the system presents a visualization of the results, including the predicted class, probability scores, and an interpretation of the outcome, in order to enhance the transparency of the AI system for non-technical users [28].

3. RESULTS AND DISCUSSION

The Convolutional Neural Network (CNN) model with a VGG16 architecture was successfully trained using a dataset of children’s facial images that had undergone preprocessing and were split into training, validation, and testing sets. During the training process, the model demonstrated a consistent increase in accuracy and a decrease in loss values for both the training and validation data. This pattern indicates that the model was able to effectively learn the visual characteristics of children’s faces without experiencing significant overfitting. These performance trends can be observed in the accuracy and loss graphs presented in Figure 2.

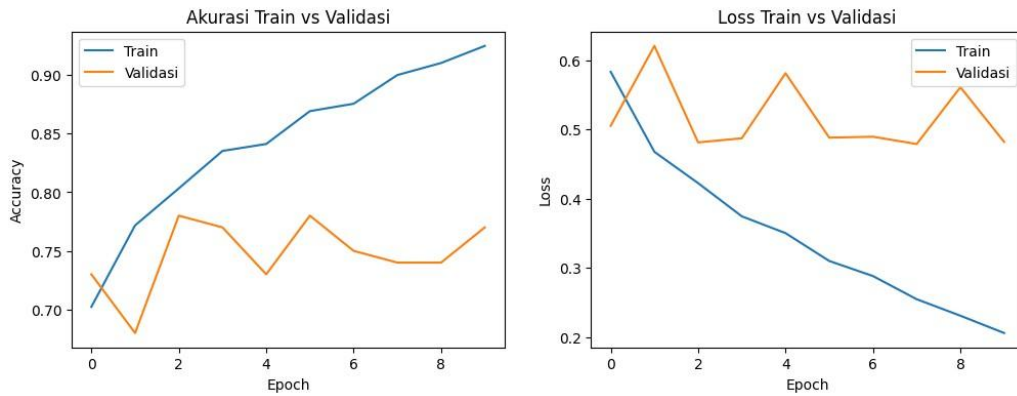


Figure 2 Training Data Accuracy and Loss Values

The model's performance was evaluated using a test dataset consisting of 300 children's facial images, with a balanced class distribution between children with Autism Spectrum Disorder (ASD) and typically developing children. The evaluation results indicate that the model achieved an accuracy of 84%, suggesting that the majority of images were correctly classified. Furthermore, the precision, recall, and F1-score values for each class demonstrate relatively balanced performance, particularly the high recall for the ASD class. This indicates that the model has a strong capability in detecting children with ASD, which is crucial in the context of early detection. It should be noted that the 84% accuracy obtained in this study is lower than the 87–99% accuracy reported by several related works described in the Introduction. This gap is plausibly attributable to differences in dataset size and diversity, since the Piosenka Autism Image Dataset used here is comprised largely of images collected from public websites and social media rather than clinically diagnosed samples used in some prior studies. In addition, this study only fine-tuned the top classification layers while keeping the pretrained VGG16 convolutional base entirely frozen, whereas some higher-performing studies applied deeper fine-tuning of the convolutional layers or used ensembles of multiple pretrained architectures. A summary of the model's performance evaluation results is presented in Table 1.

Table 1 Model Performance Summary

Class	Precision	Recall	F1-Score	Support
Autism	0.80	0.90	0.85	150
Typical	0.89	0.77	0.83	150
Accuracy			0.84	300

Further analysis was conducted using a confusion matrix to identify the types of classification errors that occurred. Based on the confusion matrix shown in Figure 3, most of the data from both classes were correctly predicted. However, a small number of misclassifications in the form of false positives and false negatives were still observed. False negative errors, in which the faces of children with ASD were predicted as typically developing children, are of particular concern as they may lead to delays in the early identification process. These errors are presumed to be influenced by similarities in facial characteristics, variations in facial expressions, and the quality of the images used.

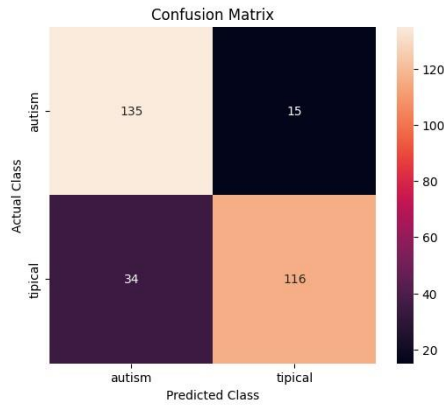


Figure 3 Confusion Matrix

To evaluate the model's ability to comprehensively distinguish between the two classes across various threshold values, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) metric were employed. The ROC curve shown in Figure 4 demonstrates that the model has strong discriminative capability, as indicated by a high AUC value and a curve that deviates substantially from the diagonal line. These results suggest that the model not only performs well at a single decision threshold but also remains stable in classification across different operating conditions.

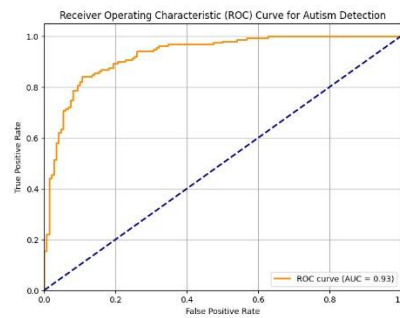


Figure 4 ROC AUC Curve

Overall, the results of this study indicate that the application of a Convolutional Neural Network (CNN) with a VGG16 architecture and a transfer learning approach is effective for classifying the facial images of children with autism. Nevertheless, this study has several limitations, particularly related to the size and diversity of the dataset used, as well as the reliance on static facial images without considering behavioral aspects or other multimodal data. These limitations present opportunities for further development, such as expanding the dataset, fine-tuning the convolutional layers, and integrating multimodal data to enhance the accuracy and reliability of the system. A summary of the limitations and future development directions is presented in Table 2.

Table 2 Summary of the Limitations

Class	Precision
Dataset	Limited dataset size and diversity
Input	Relies solely on static facial images
Model	No further fine-tuning has been performed
Evaluation	No cross-validation, confidence intervals, or significance testing; results are based on a single train/test split
Future Development	Integration of multimodal data, XAI methods, and a larger dataset

4. CONCLUSION

This study successfully developed an Autism Spectrum Disorder (ASD) classification system based on children's facial images using a deep learning approach with a Convolutional Neural Network (CNN) architecture and a VGG16 transfer learning model. The model was trained using the Piosenka Autism Image Dataset, which was divided into training, validation, and testing sets to ensure an objective learning and evaluation process. The experimental results demonstrate that the model achieved an accuracy of 84%, with relatively balanced precision, recall, and F1-score values across both classes, autism and typically developing children, indicating that the model exhibits stable and reliable classification performance. Model performance evaluation using a confusion matrix and ROC AUC curve shows that the system is capable of distinguishing between the two classes with adequate sensitivity and specificity. The high AUC value indicates that the model has strong discriminative ability in classifying facial images of children with ASD and typically developing children. Furthermore, the application of the early stopping technique proved effective in preventing overfitting during the training process, enabling the model to maintain optimal performance on previously unseen data. In addition to model evaluation, this study also implemented the trained system into a Streamlit-based application, allowing the model to be used interactively and practically by users. This implementation demonstrates that the proposed model is not only academically valid but also has potential to be further developed as a technology-based decision support tool for early ASD detection. Nevertheless, this study still has several limitations, including the use of a unimodal approach that relies solely on facial images, the absence of further fine-tuning on the convolutional layers of the VGG16 model, the lack of Explainable Artificial Intelligence (XAI) methods to enhance model interpretability, and the absence of statistical validation, such as cross-validation, confidence intervals, or significance testing, since the reported results are based on only a single train/test split. Therefore, future research is recommended to integrate multimodal data, apply advanced fine-tuning strategies, incorporate explainable AI approaches, and adopt more rigorous statistical validation procedures to improve the reliability, transparency, and accuracy of ASD detection systems in the future.

BIBLIOGRAPHY

- [1] U.S. Centers For Disease Control and Prevention, "Data and Statistics on Autism Spectrum Disorder," [cdc.gov](https://www.cdc.gov/autism/dataresearch/index.html). Accessed: Dec. 01, 2025. [Online]. Available: <https://www.cdc.gov/autism/dataresearch/index.html>
- [2] "Statistics Spotlight: How Many People Have Autism in 2024," [advancedautism.com](https://www.advancedautism.com/post/statistics-spotlight-how-many-people-have-autism-in-2024). Accessed: Nov. 28, 2025. [Online]. Available: <https://www.advancedautism.com/post/statistics-spotlight-how-many-people-have-autism-in-2024>
- [3] L. Zwaigenbaum *et al.*, "Early Intervention for Children With Autism Spectrum Disorder Under 3 Years of Age: Recommendations for Practice and Research," *Pediatrics*, vol. 136, no. Supplement_1, pp. S60–S81, Oct. 2015, doi: 10.1542/peds.2014-3667E.
- [4] K. A. Shaw *et al.*, "Prevalence and Early Identification of Autism Spectrum Disorder Among Children Aged 4 and 8 Years — Autism and Developmental Disabilities Monitoring Network, 16 Sites, United States, 2022," *MMWR. Surveillance Summaries*, vol. 74, no. 2, pp. 1–22, Apr. 2025, doi: 10.15585/mmwr.ss7402a1.
- [5] World Health Organization (WHO), "Autism," [who.int](https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders). Accessed: Nov. 28, 2025. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders>
- [6] M. Pervin, H. U. Ahmed, and Y. Hagmayer, "Effectiveness of interventions for children and adolescents with autism spectrum disorder in high-income vs. lower middle-income countries: An overview of systematic reviews and research papers from LMIC," *Front. Psychiatry*, vol. 13, Aug. 2022, doi: 10.3389/fpsy.2022.834783.
- [7] G. Dawson *et al.*, "Randomized, controlled trial of an intervention for toddlers with autism: The early start Denver model," *Pediatrics*, vol. 125, no. 1, Jan. 2010, doi: 10.1542/peds.2009-0958.
- [8] J.-L. Wu and J. Pan, "The scarcity of child psychiatrists in China," *Lancet Psychiatry*, vol. 6, no. 4, pp. 286–287, Apr. 2019, doi: 10.1016/S2215-0366(19)30099-9.
- [9] A. Esteva *et al.*, "A guide to deep learning in healthcare," *Nat. Med.*, vol. 25, no. 1, pp. 24–29, Jan. 2019, doi: 10.1038/s41591-018-0316-z.

- [10] I. Nafisah *et al.*, “Deep learning-based feature selection for detection of autism spectrum disorder,” *Front. Artif. Intell.*, vol. 8, 2025, doi: 10.3389/frai.2025.1594372.
- [11] 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 Sci.*, vol. 11, no. 11, p. 1446, Oct. 2021, doi: 10.3390/brainsci11111446.
- [12] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.
- [13] R. C. Contreras, “Piosenka autism image dataset,” *Zenodo*, Feb. 2025, Accessed: Jan. 31, 2026. [Online]. Available: <https://zenodo.org/records/14890016>
- [14] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Apr. 2015, [Online]. Available: <http://arxiv.org/abs/1409.1556>.
- [15] S. O. Akinola and O. J. Oyabugbe, “Accuracies and Training Times of Data Mining Classification Algorithms: An Empirical Comparative Study,” *Journal of Software Engineering and Applications*, vol. 08, no. 09, pp. 470–477, 2015, doi: 10.4236/jsea.2015.89045.
- [16] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [17] K. K. Mujeeb Rahman and M. M. Subashini, “Identification of Autism in Children Using Static Facial Features and Deep Neural Networks,” *Brain Sci.*, vol. 12, no. 1, p. 94, Jan. 2022, doi: 10.3390/brainsci12010094.
- [18] I. Nafisah *et al.*, “Deep learning-based feature selection for detection of autism spectrum disorder,” *Front. Artif. Intell.*, vol. 8, 2025, doi: 10.3389/frai.2025.1594372.
- [19] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J. Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.
- [20] I. Ahmad, J. Rashid, M. Faheem, A. Akram, N. A. Khan, and R. ul Amin, “Autism spectrum disorder detection using facial images: A performance comparison of pretrained convolutional neural networks,” *Healthc. Technol. Lett.*, vol. 11, no. 4, pp. 227–239, Aug. 2024, doi: 10.1049/htl2.12073.
- [21] T. Farhat, S. Akram, H. S. AlSagri, Z. Ali, A. Ahmad, and A. Jaffar, “Facial Image-Based Autism Detection: A Comparative Study of Deep Neural Network Classifiers,” *Computers, Materials & Continua*, vol. 78, no. 1, pp. 105–126, 2024, doi: 10.32604/cmc.2023.045022.
- [22] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *2017 International Conference on Engineering and Technology (ICET)*, IEEE, Aug. 2017, pp. 1–6. doi: 10.1109/ICEngTechnol.2017.8308186.
- [23] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?,” Nov. 2014, [Online]. Available: <http://arxiv.org/abs/1411.1792>
- [24] G. Carnevale, F. Farina, I. Notarnicola, and G. Notarstefano, “GTAdam: Gradient Tracking with Adaptive Momentum for Distributed Online Optimization,” *IEEE Trans. Control Netw. Syst.*, vol. 10, no. 3, pp. 1436–1448, Sep. 2023, doi: 10.1109/TCNS.2022.3232519.
- [25] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*. 2021.
- [26] N. Haque, T. Islam, and M. Erfan, “An exploration of machine learning approaches for early Autism Spectrum Disorder detection,” Jun. 01, 2025, *Elsevier Inc.* doi: 10.1016/j.health.2024.100379.
- [27] D. M. W. Powers and Ailab, “EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION.”
- [28] Aurélien Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd Edition. 2019. Accessed: Jan. 31, 2026. [Online]. Available: <https://www.oreilly.com/library/view/hands-onmachine-learning/9781492032632/>
- [29] A. Nur holifah and I. Tahyudin, “Implementasi Teknologi Deep Learning untuk Diagnostik Stroke Otak Berbasis CNN-LSTM-FNN,” *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 3, pp. 538–599, Mar. 2025, doi: 10.52436/1.jpti.538.