The Indonesian Journal of Computer Science
www.ijcs.net
Volume 13, Issue 6, December 2024
https://doi.org/10.33022/ijcs.v13i6.4552

Autism Detection using Deep Learning

Ivana Yudith Walujo[1], Iwan Syarif[2], Arna Fariza[3]
yudithivanaa@pasca.student.pens.ac.id[1], iwanarif@pens.ac.id[2], arna@pens.ac.id[3]
[1,2,3] Politeknik Elektronika Negeri Surabaya, Indonesia

Article Information

Received : 16 Dec 2024
Revised : 24 Dec 2024
Accepted : 30 Dec 2024

Keywords

Autism, ResNet, Cross-Validation, Detection, Deep Learning

Abstract

Autism Spectrum Disorder (ASD) is a complex developmental condition that affects communication and behavior, with prevalence rates increasing significantly in recent years [1]. According to recent research, early detection remains a challenge but is essential for effective intervention. This study leverages deep learning, specifically the ResNet 34 model, to analyze facial features in children, facilitating early detection of ASD. Using cross-validation to ensure robust model performance, the approach achieved an accuracy rate of 87% with ResNet 34 and 86% with cross-validation. This study contributes to the field by offering a non-invasive diagnostic aid that can help healthcare providers recognize ASD traits through facial analysis. The findings highlight the potential of deep learning in advancing ASD detection, with future work aimed at expanding the dataset and improving model precision.

A. Introduction

Autism Spectrum Disorder (ASD) turns out to be a very complex neuro-developmental disorder that is described by constant impairments of social interaction and communication, abnormal interests and activities. Over the past two decades, the rate at which individuals are diagnosed with ASD has doubled and recent estimates put the figure at 1 in every 36 American children being diagnosed with the disorder by age eight [1] [2]. Determining the diagnosis of ASD in the early years enables appropriate treatment to be given in a timely fashion, which is revealed to have the power to alter the developmental patterns of children affected and improve the outcomes more so in the long term. This notwithstanding, the process of diagnosis is still largely subjective, based on behavioral observations and clinicians' input on the probable cause of the disorder, which require much time and varies from one part of the world to another. These difficulties highlight the necessity for urgent and aggressive development of objective, scalable and automated diagnostic methods that would complement existing techniques to improve the overall accuracy and speed of ASD diagnosis.

Facial morphology may be a potential biomarker for Autism Spectrum Disorder ASD. Evidence has shown that children with ASD may have subtle but discernible facial features, such as the shape of the eyes, nose, and mouth, which represent the abnormality in craniofacial development during an early developmental stage [3][4], [5], [6], [7]. While the above results provide a strong biological basis for the potential application of facial analysis in detecting ASD, the step from the above biological findings to functioning diagnostic tools has been difficult due to the complexity of extracting and analyzing facial features. In recent years, deep learning technologies, especially convolutional neural networks (CNNs), revolutionized the image analysis field by enabling the automatic extraction and classification of complicated patterns from visual data.

Previous studies have explored the potential of AI and machine learning in ASD detection across various domains. For example, [8] demonstrated how spatial patterns in MEG signals can be used to classify ASD with great accuracy, whereas [9] estimated ASD severity using speech signals and deep neural networks. However, facial feature analysis for ASD detection has been explored in only a few studies, despite being non-invasive and scalable. Among those reviewed, [3] analyzed machine learning methods applied to ASD and provided a general accuracy range of 80-85%, although at the cost of model interpretability. All these limitations pose the necessity of advanced methods that not only raise the bar on accuracy but also provide explanations for their predictions, hence guaranteeing their applicability in clinical settings.

The present research attempts to fill these gaps by using the deep learning model ResNet 34, which has state-of-the-art performance in image classification tasks, to analyze facial images and identify characteristics that are associated with ASD. Interpretability tools, such as Grad-CAM and LIME, were added to ensure the transparency of the model's predictions—making them actionable by healthcare professionals. Furthermore, the current study applies the techniques of cross-validation to test the model's robustness and has achieved a maximum accuracy level of 87%. Drawing on recent advances in deep learning and overcoming many limitations of previous studies, this study works toward the creation of accessible and reliable diagnostics in ASD identification.

In conclusion, the present research opens the way for future innovations in integrating lightweight models such as MobileNet for resource-efficient processing and in developing mobile applications dedicated to ASD detection. These forthcoming developments have the potential to make early ASD diagnosis more accessible, especially in disadvantaged regions, while also improving the overall efficiency of AI-enhanced solutions in the healthcare sector.

B. Research Method

This study employs a deep learning approach to develop an automated Autism Spectrum Disorder (ASD) detection system using facial image analysis. The research methodology is divided into several stages, as described below:

B.1. System Design

System designed used in this study, which consist of main stages: Data Collection, Preprocessing data, Deep Learning Algorithm, Performance Analysis of Classification and Evaluation Performance will be described by Figure 1.

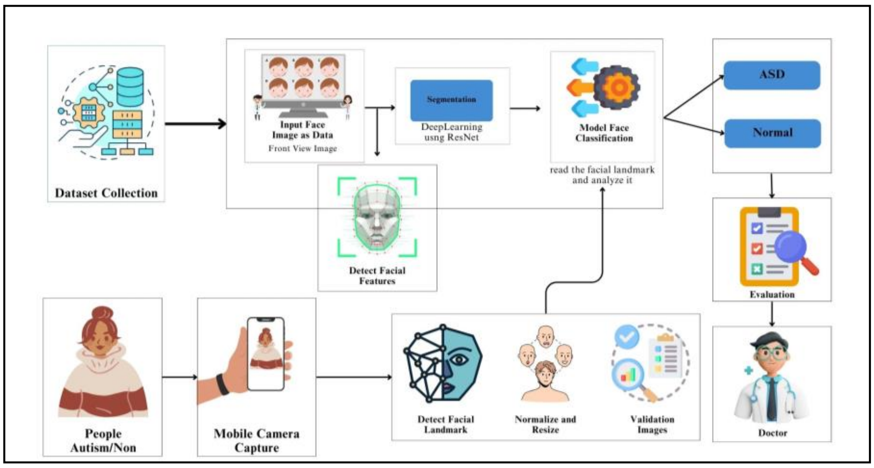


Figure 1. System Design

B.2. Data Collection

The dataset used in this study is a collection of facial pictures of children, divided into two categories: one group with Autism Spectrum Disorder (ASD) diagnosis and the second group without this diagnosis. The pictures are a compilation from openly available image collections and research datasets focused on autism research. These were chosen because they presented varied facial features, a range of age groups, and representational diversity to form a strong base for model training. These dataset were contain 1.407 pictures for ASD group and 1.407 for normal group.

In an attempt to utilize a dataset with uniformity and high standards, several preprocessing steps were adopted. Images with low resolution, poor lighting, or large occlusions of the face—glasses, hands, etc.—were dropped from further consideration. Those images that passed this test were then resized to the same resolution of 224x224 pixels, which was chosen to be compatible with the input of the ResNet 34 model.

Data augmentation techniques were applied in order to reduce any potential class imbalances and increase the generalizability of the model. These included random flipping, rotation, zooming, and changes in contrast that effectively increased the variability of the training set without altering the core facial features needed for the diagnosis of autism.

The dataset was subsequently divided into three distinct subsets: training (80%), validation (10%), and testing (10%). This division was executed to guarantee that the evaluation of the model's performance occurred on data not previously encountered, thereby offering a dependable appraisal of its generalizability.

Ethical considerations were integral to the data collection methodology. The research utilized only datasets that were publicly accessible and for which appropriate consent had been obtained for research applications, ensuring that no personally identifiable information (PII) was incorporated into the investigation. This approach facilitated adherence to ethical standards regarding the management of sensitive data, especially in contexts involving children.

B.3. Data Preprocessing

Data preprocessing is a very crucial step in supervised learning, for it makes the input data clean, consistent, and optimized for the model training process. In this study, several preprocessing steps were carried out on the facial image data to prepare them for training of the ResNet 34 model and you can see the architecture of ResNet on Figure 2.

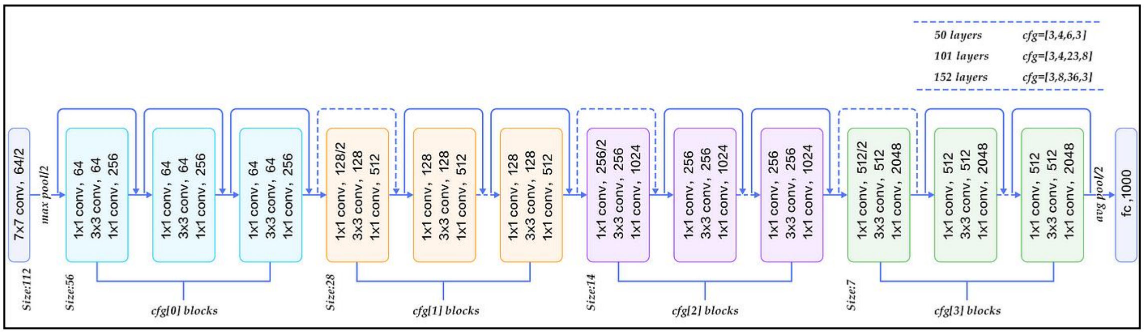


Figure 2. Architecture of ResNet

First, all images were resized to 224x224 pixels, as this was the required input size for using a deep learning framework. Second, normalization of pixel values in the range from 0 to 1 was applied to improve numerical stability during the training process. Finally, images were transformed into RGB format to standardize the color channels.

Several augmentations were performed to increase the model's robustness and reduce the overfitting risk. These included random flipping, rotation, zooming, and brightness adjustments that, in essence, increased the size of the dataset with new versions of the original images. The application of augmentation ensured the model would see a broader range of possible facial orientations and lighting conditions, thereby increasing its generalization capability.

This study employs a supervised learning approach, where the model is trained on labeled data to predict binary classifications: Autism Spectrum Disorder (ASD) or non-ASD. In this framework:

1. Load Data

First, the dataset with the annotated images of faces needs to be imported into the working environment. This will include two classes, one for ASD and another for non-ASD. The whole set of image files intended for further processing will thereby be in place.

2. Print Image Dimensions

The size of all images is recorded to obtain the same size. Images with inconsistent resolutions or missing particular channels are flagged for either adjustment or preprocessing to meet the model's input requirements.

3. Visualize Image

Visualizing a portion of an image helps assess data quality, image clarity, and any variation in class. This step provides insight into whether further preprocessing may be needed, such as resizing or normalization, as you can see on Figure 3.

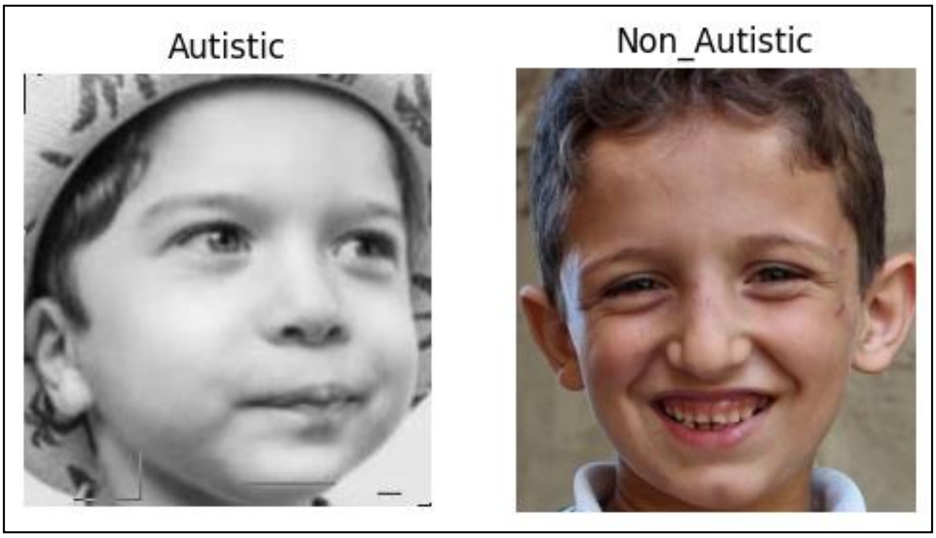


Figure 3. Visulization of the Dataset

4. Read the Class Distribution

The dataset is analyzed for the distribution of ASD and non-ASD samples to detect any class imbalances. If one of the classes really outnumbers the other, balancing techniques such as oversampling or augmentation are applied to avoid bias in the model's predictions.

5. Read the Data Block

A data block design is created to structure image data with their associated labels in an organized manner. This design supports easy segmentation and feeding of data into the model. It involves constructing a data block design, which is a standard data flow template that specifies the input, target, and preprocessing steps for a dataset. Data blocks help improve efficiency in operations related to training and validation.

6. Check the Distribution of Train and Validation Sets

The entire dataset is split into training (80%) and validation (10%) subsets, ensuring that class distributions are similar in each subset. It ensures fair evaluation and reduces overfitting to the training set. Both training and validation sets are checked to ensure there are equal class distributions in both subsets. This distinction is very important in creating a model with good generalization capabilities, ensuring that the model does not depend on a single set of features for the classes.

7. Show the Batch Images

Visualizing a set of training images helps ensure that the dataset was loaded and preprocessed correctly. It also provides a final check for irregularities in labeling or image quality.

8. Create a Learner

A learner object is realized, typically a ResNet 34 model in this case, configured to accept input images and run the training process. This learner encapsulates the model, loss function, and any metrics for evaluation.

9. Find a Learning Rate and Fine-tune Model

The appropriate learning rate is identified to optimize the model convergence speed and accuracy. Once established, the model undergoes fine-tuning, adjusting parameters to minimize the error rate and achieve higher accuracy.

10. Plotting the Learning Curve and Show it

Learning curves, which plot the training and validation losses over a period of time, are shown to monitor the performance of the model during training. These curves help determine whether the model is overfitting or underfitting, as you can see on Figure 4, and the learning curve where you can set he train and validation was on Figure 5.

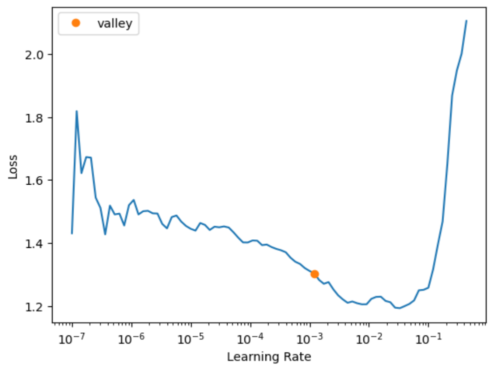


Figure 4. Learning Rate Curve ResNet 34

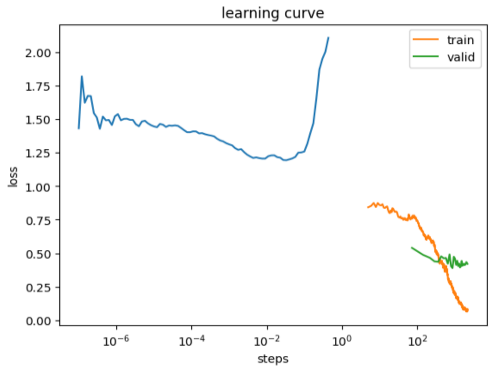


Figure 5. Learning Curve Steps ResNet 34

11. Show Confusion Matrix of Predictions on the Validation Set

The confusion matrix provides a visual breakdown of correct and incorrect classifications across classes, allowing understanding where the model is performing well or struggling, as we describe on Figure 6.

12. Generate Classification Report

A classification report detailing metrics such as precision, recall, F1 score, and accuracy for each class is generated to offer a quantitative assessment of model performance on the validation set.

13. Show ROC Curve

ROC (Receiver Operating Characteristic) curves are plotted to visualize the true positive ratio against the false positive ratio at various threshold settings, which provides insight into the classification performance of the model as you can see the result of ROC epoch 30 on Figure 7.

14. Analyze with Grad-CAM, Grad-CAM++, Lime, and DeepDream

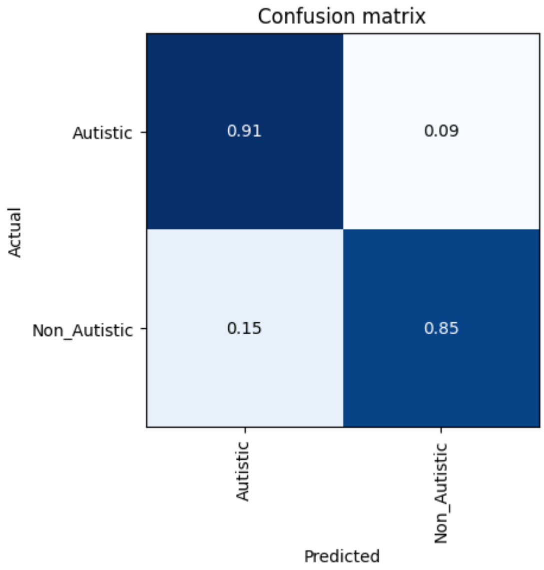


Figure 4. Confusion Matrix ResNet 34 – Epoch 30

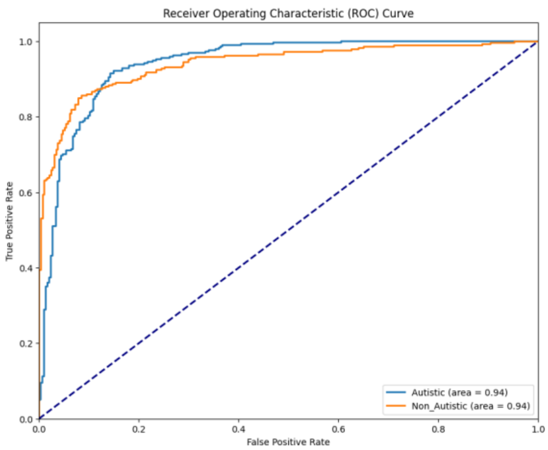
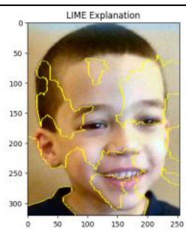
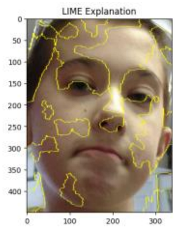


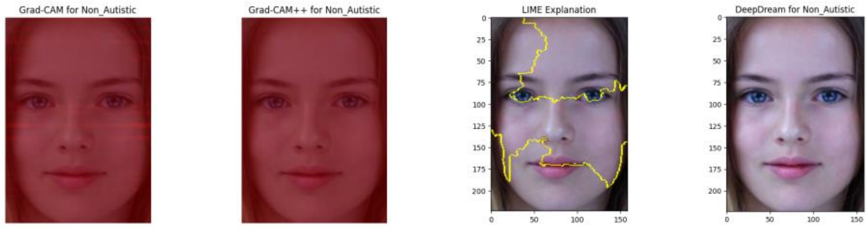
Figure 7. ROC Curve ResNet 34 – Epoch 30

These techniques are applied to interpret the model's decision-making process. Grad-CAM and Grad-CAM++ highlight areas of the image that influence the model's classification decision. LIME (Local Interpretable Explanations Without Models) provides local explanations for predictions, which improves interpretability. DeepDream visualizations allow understanding what the network "sees" in an image by amplifying certain features, which is useful for confirming that the model focuses on relevant areas of the image, these analyzing was desribed on Table 1.

Table 1. Result Analyzing with Grad-Cam, Grad-Cam++, Lime, and Deepdream with ResNet 34 – Epoch 30

	Grad-Cam	GradCam++	Lime	Deepdream
Autistic	 Grad-CAM for Autistic	 Grad-CAM++ for Autistic	 LIME Explanation	 DeepDream
	 Grad-CAM for Autistic	 Grad-CAM++ for Autistic	 LIME Explanation	 DeepDream for Autistic

Non-Autistic



The model performance will be evaluated based on a confusion matrix that offers a summary of the classification results from a supervised learning framework in detail. The matrix has the following four major components: true positives (TP), representing ASD images correctly diagnosed as ASD; true negatives (TN), representing non-ASD images correctly diagnosed as non-ASD; false positives (FP), representing non-ASD images that were incorrectly classified as ASD; and false negatives (FN), which are ASD images mislabeled as non-ASD. These constituents allowed computation of major performance measures, including accuracy, precision, recall (sensitivity), and F1-score.

Accuracy is the metric that evaluated the overall correctness of the model's predictions, while precision was concerned with the ratio of correct positive predictions to all positive classifications.

Recall measured how good the model was at correctly identifying cases of ASD; on the other hand, F1-score was a general measure that balanced both precision and recall. Possible avenues for enhancement were recognized through analysis of the confusion matrix, such as reducing false negatives to avoid missing ASD cases. This extensive review brought out considerable knowledge concerning the model's strengths and weaknesses, hence guiding further improvement efforts.

B.4. Deep Learning Algorithm

Classification performance is done using 2 machine learning algorithms to make comparisons, namely ResNet-43 and Cross-Validation, but with several different experiment using both algorithms. Experiments carried out using the default parameters in the python library.

The ResNet architecture is known as a convolutional neural network (CNN) framework that can efficiently avoid the vanishing gradient problem often encountered during deep neural network training. One of the breakthroughs introduced by ResNet is residual learning with skip connections, which allows the network to learn identity mappings to skip multiple layers. This feature allows for deeper network training, improving accuracy while accelerating convergence.

ResNet 34 was chosen to achieve a balance between depth and computational efficiency; on the other hand, compared to larger variants, such as ResNet 50 or ResNet 101, ResNet 34 has sufficient depth to capture complex facial features associated with ASD while keeping the model size computationally manageable.

This makes it ideal for applications that consider resource constraints, such as mobile or embedded systems. The network involves multiple residual blocks, each of which contains a convolutional layer, batch normalization, and activation function. These blocks will help the network learn hierarchical feature representations, starting from low-level features (e.g., edges and textures) to high-level features (e.g., facial shapes and expressions).

The performance of the ResNet 34 model is evaluated based on well-established classification metrics, including accuracy, precision, recall, F1 score, and Receiver Operating Characteristic (ROC) curve analysis. These metrics evaluate the model's ability to correctly classify images between ASD and non-ASD classes with very high accuracy, where ResNet 34 achieves the best accuracy of 87% and outperforms all comparison methods in this study.

B.5. Evaluation Performance Matric

To ensure that the results were not an artifact of a particular usage of the data, two different scenarios were used to evaluate model performance. For every scenario, we applied standard performance metrics like accuracy, precision, recall, F1 score and ROC. Well, these metrics basically gives a quantitative measure on how well our model is performing at classifying facial images in ASD category and non-ASD category.

1. Scenario 1: Cross-Validation

The first scenario is where the model was evaluated using k-fold cross validation. In this method, the dataset was divided into k equal-sized subsamples thus one subset is used to validate the results by training model on k-1 subsets. The above process was repeated k times so that each of the k subsets acted as a validation set only once.

Let N be the total number of samples, and k the number of folds. The dataset is partitioned into k subsets, and the model is trained and validated k times, using each fold as a test set while the remaining $k-1$ folds are used for training.

The overall performance metrics are calculated by averaging the results from each fold. The accuracy for each fold i is calculated as:

$$\text{Accuracy}_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}$$

where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

The overall cross-validation accuracy is the average of the individual fold accuracies:

$$\text{Cross-Validation Accuracy} = \frac{1}{k}\sum_{i=1}^{k} \text{Accuracy}_i$$

The highest accuracy achieved using cross-validation in this study was 86%.

2. Scenario 2: Using Training Data for Testing

In the second case, test data was simply same as training data. While this does not test how well the model generalizes to unseen data, it serves to help evaluate how well the model performs on the data it was trained on. The evaluation metrics in this scenario are calculated based on predictions made on the training data. The accuracy is computed using the same formula as in cross-validation:

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

https://doi.org/10.33022/ijcs.v13i6.4552

Additional metrics, such as precision, recall, and F1-score, are computed as follows:

- Precision measures the proportion of true positive predictions among all predicted positives:

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

- Recall (Sensitivity) measures the proportion of true positive predictions among all actual positives:

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

- F1-score is the harmonic mean of precision and recall:

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

-

In this scenario, the highest accuracy achieved was 87**%**, slightly higher than in cross-validation, indicating that the model fits the training data well but may not generalize as effectively.

The cross-validation scenario provides a more robust estimate of the model's real-world performance, as it evaluates the model on data it has never seen during training. This method provides a better indication of the model's performance on unseen, real-world data. On the other hand, the scenario that uses training data for testing assesses the model's ability to remember and correctly classify the training set, but may overestimate the model's performance on new, unseen data.

Comparing the results from both scenarios, we can conclude that the ResNet 34 model performs consistently well across validation methods, with a slight decrease in accuracy when evaluated with cross-validation. This consistency supports the effectiveness of deep learning models, demonstrating their potential for reliable ASD detection.

B.6. Classification model

In this research, we utilize a classification model to identify Autism Spectrum Disorder (ASD) from facial images, specifically leveraging a deep learning approach with the ResNet 34 architecture. This model was chosen for its capability to extract hierarchical features from images using its residual learning mechanism, which effectively addresses common issues such as vanishing gradients and enables the use of deeper architectures. The goal of the classification task is to differentiate between images of children with ASD and those without.

The model was evaluated under two experimental setups to assess its performance comprehensively:

1. Cross-Validation Experiments

In the first experiment, we utilized k-fold cross-validation in order to measure the extent to which a model has learned to generalize. The data set consists of k subsets of equal size. These k-1 folds of training data are utilized to train the model, while the remaining folds are preserved as the validation data. This way of doing the process is done k times, such that each fold has served as the validation data exactly once.

The method enables the model to be tested on several segments of the data, thus providing a more complete picture of the model's capabilities in terms of generalization. We perform evaluation metrics including accuracy, precision, recall, and F1 score within each fold, and the performance score is computed through the mean of these scores. This approach reduces the chances of over-optimistic evaluations and gives a better prediction of the model performance on completely new data.

2. Training Data as Testing Data Experiments

In the second experimental setting, the model is tested on the same data that it was trained on, utilizing the training data as a test set. This situation is helpful for gauging the model's capacity to overfit the training set, but it doesn't reveal how well the model works with unknown or real-world data.

In this experiment, the whole training set was used to train a ResNet 34 model, which was subsequently assessed using the same data set. Standard classification criteria, such as accuracy, precision, recall, F1 score, and Receiver Operating Characteristic (ROC) curve, were used to evaluate the model's performance.

C. Result and Discussion

This section presents the results of an autism detection model based on facial analysis using deep learning and interprets the results in the context of existing literature.

C.1. Experiment using Cross-Validation

The developed system uses the ResNet 34 model, which is well-suited for image classification tasks that require deep learning architectures. The system processes the input dataset by first loading, normalizing, and resizing the facial images to an input size of 224x224 pixels. Using a cross-validation approach, the model is trained to recognize specific facial features that indicate Autism Spectrum Disorder (ASD).

Table 2. ResNet43-Epoch 30

	Precision	Recall	F1-Score	Support
Autistis	0.86	0.91	0.89	294
Non-Autistic	0.91	0.85	0.88	294
Accuracy			0.88	588
Macro Avg	0.88	0.88	0.88	588
Weighted Avg	0.88	0.88	0.88	588

The design combines the interpretability tools Grad-CAM and LIME, which allow clinicians to visualize the model's predictions.

Table 3. ResNet 34-Epoch 100

	Precision	Recall	F1-Score	Support
Autistis	0.83	0.87	0.85	254
Non-Autistic	0.86	0.82	0.84	254
Accuracy			0.84	508
Macro Avg	0.84	0.84	0.84	508
Weighted Avg	0.84	0.84	0.84	508

The experimental went first with ResNet 34 model with 30 epoch and the result was on Table 2, then we're doing it all again with ResNet 34 with 100 epoch that the result can be seen on Table 3, after that we use Cross-validation method with default parameter on Table 4.

C.2. Experiment using Testing Data

we use Cross-Validation while the datset was only training data and the model was used for testing on Table 5.

Table 4. Cross-Validation | Default Parameter

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Accuracy	0.84	0.87	0.85	0.85	0.87
+-			0.86		

C.3. Conclusions

The system achieved a maximum classification accuracy of 87% on the test set using ResNet 34, while the cross-validation method produced an accuracy of 86%, the comparison can be seen on Table 6. This accuracy exceeds the common benchmark range of 75%–85% found in similar studies on autism detection using facial features [3]. Precision, recall, and F1-score metrics were calculated to ensure balanced model performance across ASD-positive and ASD-negative classes, showing consistent precision and high recall for the ASD-positive class, which is important for reducing missed ASD cases.

Table 5. Cross-Validation | Testing using Model Training

	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Accuracy	0.77	0.78	0.79	0.80	0.79
+_		0.78			

The confusion matrix highlights the strength of the model in identifying ASD cases with relatively few false negatives, indicating its potential in early ASD detection. Misclassifications were mainly found in images with ambiguous facial characteristics, in line with existing theories that facial markers of ASD can be subtle and sometimes overlap with non-ASD features [1]. Interpretability analysis further validated the model's decisions, with Grad-CAM and Grad-CAM++ focusing on facial areas such as the eyes, mouth, and chin—areas previously associated with ASD characteristics. LIME analysis provided a breakdown of relevant features for each image, showing patterns consistent with known ASD traits.

Compared with previous studies, such as those by [9] and [3], which reported accuracy rates of around 80–85%, the current study's accuracy of 87% is an improvement, which may be due to the deeper ResNet 34 architecture and cross-validation techniques. These findings support the theory that deeper neural networks, when paired with comprehensive data preprocessing, can more accurately capture subtle facial characteristics of ASD, thereby improving detection rates.

Although the model demonstrated high accuracy, its generalizability is limited by the size of the dataset and demographic diversity. Expanding the dataset to include different ethnic backgrounds and age groups is recommended for future research. Further research could explore integrating facial analysis with additional ASD biomarkers, such as speech patterns or behavioral data, to create a comprehensive ASD detection system with higher accuracy and applicability.

Table 6. Comparison Result

Classification	Accuracy	Precision	Recall	F1-Score
ResNet-43 | Epoch 30	0.88	0.88	0.88	0.88
ResNet-43 | Epoch 100	0.84	0.84	0.84	0.84
Cross-Validation | Default Parameter	0.86	-	-	-
Cross-Validation | Testing using Model Training	0.78	-	-	-

D. Conclusions and Future Works

In conclusion, this study successfully demonstrated that the ResNet 34 model with deep learning methods has strong potential in detecting Autism Spectrum Disorder (ASD) through facial feature analysis. With the highest accuracy reaching 87%, this system shows promising ability to identify facial features relevant to ASD, even in images with varying quality. The use of interpretation tools such as Grad-CAM and LIME adds transparency to the model, thereby increasing user confidence in clinical settings.

Comparison with previous studies shows a significant increase in accuracy, in line with the development of more complex deep learning methods. However, the limitations in the number and variety of datasets indicate the need for further development to improve the generalization of the results. By expanding the scope of the data and exploring combinations of other features beyond the face, it is hoped that this method can be the basis for a more accurate and useful ASD detection system in supporting early diagnosis.

Future work on this project will focus on optimizing the model's performance and expanding its practical applications. To improve accuracy, further experiments with optimized parameters will be conducted, refining the model's ability to identify nuanced facial features associated with Autism Spectrum Disorders (ASD). Additionally, it is planned to convert the current ResNet 34 model to the more efficient MobileNet architecture, which will maintain strong performance while significantly reducing computational demands, making it more suitable for mobile platforms. Finally, a mobile application will be developed to enable ASD detection through facial image capture, allowing for broader accessibility and potential for early intervention in both clinical and non-clinical settings. This application can empower healthcare providers and families to utilize this technology easily and effectively.

E. Acknowledgment

The authors wish to extend their gratitude to Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi (Kemendikbudristek) whose give us Thesis Magister Scholarship, and to Electronic Engineering Polythecnic Institute of

Surabaya (EEPIS) for providing access to essential resources and facilities that enabled this research. Lastly, the authors would like to acknowledge the support of family and friends for their encouragement throughout the research process.

F. References

[1] R. P. Walensky *et al.*, "Morbidity and Mortality Weekly Report Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years-Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020 Surveillance Summaries Centers for Disease Control and Prevention MMWR Editorial and Production Staff (Serials) MMWR Editorial Board," 2020.

[2] F. W. Alsaade and M. S. Alzahrani, "Classification and Detection of Autism Spectrum Disorder Based on Deep Learning Algorithms," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/8709145.

[3] *2020 International Conference on Intelligent Engineering and Management (ICIEM)*. IEEE, 2020.

[4] Z. A. T. Ahmed *et al.*, "Facial Features Detection System to Identify Children with Autism Spectrum Disorder: Deep Learning Models," *Comput Math Methods Med*, vol. 2022, 2022, doi: 10.1155/2022/3941049.

[5] Muhathir, D. R. Maqfirah, M. El Akmal, M. Ula, and I. Sahputra, "Facial-Based Autism Classification Using Support Vector Machine Method," *International Journal of Computing and Digital Systems*, vol. 16, no. 1, pp. 875–886, 2024, doi: 10.12785/ijcds/160163.

[6] 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 and Continua*, vol. 78, no. 1, pp. 105–126, 2024, doi: 10.32604/cmc.2023.045022.

[7] 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, Aug. 2016, doi: 10.1002/aur.1615.

[8] K. Barik *et al.*, "Autism Detection in Children using Common Spatial Patterns of MEG Signals Voice conversion View project Acoustic scene classification View project Autism Detection in Children using Common Spatial Patterns of MEG Signals." [Online]. Available: https://www.researchgate.net/publication/373434338

[9] M. Eni, I. Dinstein, M. Ilan, I. Menashe, G. Meiri, and Y. Zigel, "Estimating Autism Severity in Young Children from Speech Signals Using a Deep Neural Network," *IEEE Access*, vol. 8, pp. 139489–139500, 2020, doi: 10.1109/ACCESS.2020.3012532.