

UDC 004.92

doi: 10.32620/reks.2025.1.11

Junidar JUNIDAR<sup>1</sup>, Melinda MELINDA<sup>2</sup>, Dinda D. DIANNUARI<sup>2</sup>,  
Donata D. ACULA<sup>3</sup>, Zulfan ZAINAL<sup>4</sup>

<sup>1</sup> Department of Informatics, Mathematics and Natural Sciences Faculty,  
Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

<sup>2</sup> Department of Electrical Engineering and Computer, Engineering Faculty,  
Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

<sup>3</sup> Department of Computer Science, College of Information and Computing Sciences,  
University of Santo Tomas, Manila, Philippines

<sup>4</sup> Departement of Computer Engineering, Engineering Faculty, Universitas Serambi Mekkah,  
Banda Aceh 23245, Indonesia

## FACE AUTISTIC CLASSIFICATION BASED ON THERMAL USING IMAGE ENSEMBLE LEARNING OF VGG-19, RESNET50V2, AND EFFICIENTNET

The **subject** of this paper is the detection of Autism Spectrum Disorder (ASD) traits using facial recognition based on thermal images. The **goal** of this study was to evaluate and compare the performance of various Convolutional Neural Network (CNN) architectures in classifying thermal facial images of children with ASD, thereby facilitating the early identification of autistic traits. The **tasks** addressed include preprocessing a dataset of thermal facial images to prepare them for model training; conducting classification using three CNN architectures VGG-19, ResNet50V2, and EfficientNet; and assessing their performance based on accuracy, precision, recall, and F1-score metrics. The **methods** employed involved training these CNN models on a balanced dataset of 4,120 thermal facial images and splitting them into training, validation, and test sets. Each model underwent extensive training to determine its ability to effectively classify autism and non-autism classes. The **results** revealed that ResNet50V2 achieved the highest accuracy of 98.82%, followed by VGG-19 and EfficientNet with accuracies of 96.47% and 96.07%, respectively. ResNet50V2 also demonstrated superior generalizability due to its lower validation loss and higher classification accuracy compared to other architectures. **Conclusion.** The scientific novelty lies in: 1) introducing thermal imaging as an effective tool for detecting ASD traits; 2) demonstrating the superior performance of ResNet50V2 in classifying thermal facial images with high accuracy and generalization; and 3) exploring EfficientNet for the first time in this domain, highlighting its potential for improving autism diagnostic systems. This study contributes to advancing noninvasive methods for ASD detection and paves the way for further applications of deep learning in clinical diagnostics.

**Keywords:** Autism Spectrum Disorder (ASD); Thermal Image; EfficientNet; ResNet50V2; VGG-19.

## 1. Introduction

### 1.1. Motivation for Research

Autism Spectrum Disorder (ASD) is a neurodevelopmental condition in humans characterized by impairments in social interactions, communication (both verbal and non-verbal), and restricted or repetitive behaviors [1, 2]. ASD is a spectrum disorder that manifests in varying severity, with each individual displaying unique symptoms [3]. The latest Centers for Disease Control and Prevention's (CDC) study in 2023 showed that one in 36 children will now be diagnosed with autism. That is up from one in 44 two years ago. Because the report was just released. The data are likely to remain the same through 2024. Across regions such as Asia, Europe, and Africa, the prevalence varies between 0.48% and 3.13%, and the

financial burden of caring for a child with ASD exceeds \$20,000 annually in the United States [4].

Early diagnosis is critical because timely interventions can improve social functioning and overall quality of life [5]. A promising approach for early detection involves facial recognition technologies, particularly thermal imaging, which captures physiological signals in a non-intrusive manner [6]. Thermal imaging was based on the hypothesis that skin temperature changes due to pulsatile blood flow in facial vessels reflect ASD-specific affective states [7].

Machine Learning (ML) algorithms, such as Convolutional Neural Networks (CNNs), have demonstrated exceptional performance in image-processing tasks, including facial recognition [8]. However, research into facial recognition systems tailored for children with autism, particularly those that leverage thermal imaging, remains relatively underexplored.



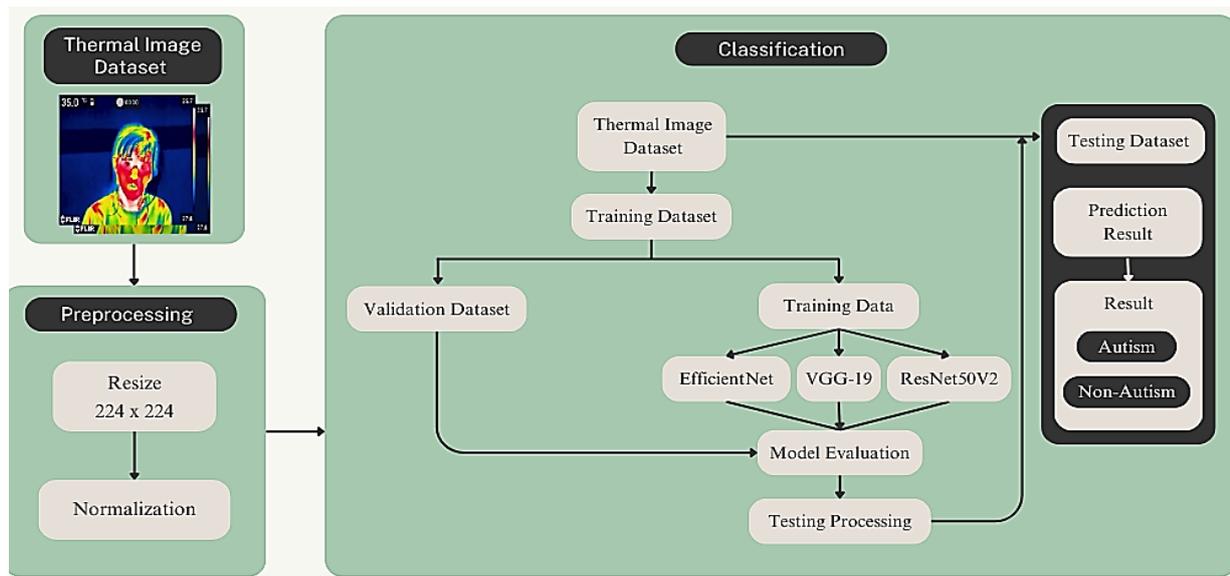


Fig. 1. Research flow in thermal image classification

## 1.2. State of the Art

Previous studies have extensively investigated facial recognition systems using various CNN architectures [9, 10]. For instance, [11] demonstrated a CNN-based facial classification system for children with ASD, achieving an accuracy of 98%. It can be said that this is the closest analog, and it is also based on a thermal imaging dataset. In this study, the dataset used was a previous study [11]. Here, we compare the performance of several architectures, such as ResNet50V2, EfficientNet, and VGG-19, in classifying the faces of autistic children using thermal images. This is because, in the study [11], the authors only tested the dataset using the CNN classification method.

Similarly, [12] employed architectures such as VGG-19, ResNet50V2, and EfficientNet for autism diagnosis, with accuracy rates of 86.5%, 94%, and 85.8%, respectively. The work of [13] used MTCNN for raw data preprocessing and ResNet50V2 for ASD diagnosis based on facial images, achieving an accuracy of 93.97% and an AUC of 96.33%.

Thermal imaging enhances diagnostic accuracy by providing unique physiological data. The integration of advanced CNN architectures, such as VGG-19, ResNet50V2, and EfficientNet, with thermal imaging has demonstrated significant potential for ASD detection [12]. However, EfficientNet remains relatively unexplored in the context of facial recognition for children with autism.

Recent studies have also demonstrated the application of deep learning in various other domains. For instance, in [14], a deep learning method for detecting the nutritional status of children using facial images was proposed, achieving accuracy rates of

99.75% and 100% with AlexNet and ResNet34. In [15], the combination of CNN and LSTM for speech emotion recognition, which addressed gender and style variations, achieved an accuracy of 84.35%. In [16], the application of transfer learning using VGG16, VGG19, and customized CNNs for skin lesion classification, with VGG16 being the most effective for early skin cancer detection, demonstrated the potential for high accuracy in medical applications. In [17], DenseNet201 was used for land cover classification with high-resolution multispectral data, achieving optimal accuracy and highlighting the importance of quality datasets in environmental monitoring.

This study compares the performance of several architectures, such as ResNet50V2, EfficientNet, and VGG-19, in classifying the faces of children with autism using thermal images. Furthermore, this study was proposed because of the differences in facial characteristics between children with ASD and healthy children.

## 1.3. Objective and Approach

This study investigated the effectiveness of the VGG-19, ResNet50V2, and EfficientNet architectures in facial recognition for ASD detection based on thermal images. The performance of these models was evaluated in terms of recognition accuracy, classification precision, and computational efficiency.

The primary contributions of this study are as follows:

1. Enhance the performance of thermal image-based facial recognition systems for children with ASD using advanced CNN architectures.

2. We introduced EfficientNet to the facial recognition domain for children with autism, marking its first application in this context.

3. Demonstrating the superiority of ResNet50V2 over other architectures, such as VGG-19 and EfficientNet.

4. We compared the classification accuracy of autism facial thermal images between the VGG-19, ResNet50V2, and EfficientNet architectures to determine the most effective model.

The structure of this article is organized as follows. Section 2 discusses the case study, including the dataset and its preprocessing steps and the configurations of the employed models. Section 3 presents the experimental results and provides a comprehensive analysis of the model performance, addressing key metrics, such as accuracy and computational efficiency. Section 4 summarizes the conclusions and significant findings, including potential directions for future research.

## 2. Case study

The workflow of this study is shown in Fig. 1, and a description is explained in the next section.

### 2.1. Data Description

The dataset used in this study was a facial thermal image database obtained from a previous study [11]. The dataset was collected from a research project at Universitas Syiah Kuala with My Hope School Banda Aceh, Indonesia. The guardian has approved all the data with proof of a statement of willingness to provide the data or a guidance form. These datasets have two classes: the autism class with 2060 images and the normal class with 2060 images, with a child age range of 3-10 years. The thermal image we mean in the study is a collection of facial data obtained from the study [11]. They used an FLIR E95 thermal camera with an emissivity of 0.98 and resolution of 640x480 pixels. In addition, emissivity is required to read the recorded skin temperature. In addition, the thermal camera is designed to be used in the temperature range of -4 to + 27320F. The thermal camera provides effective identification of the source of hearing and allows for consistent image quality.

### 2.2. Data Preprocessing

Data preprocessing is a vital process in face detection. This process involves a series of steps to prepare facial images before use in model training. In this stage, the image size of the autism and normal datasets was changed from 180x180 to 224x224, corresponding to the size used by the VGG-19, ResNet50V2, and EfficientNet models. Next, the images were normalized

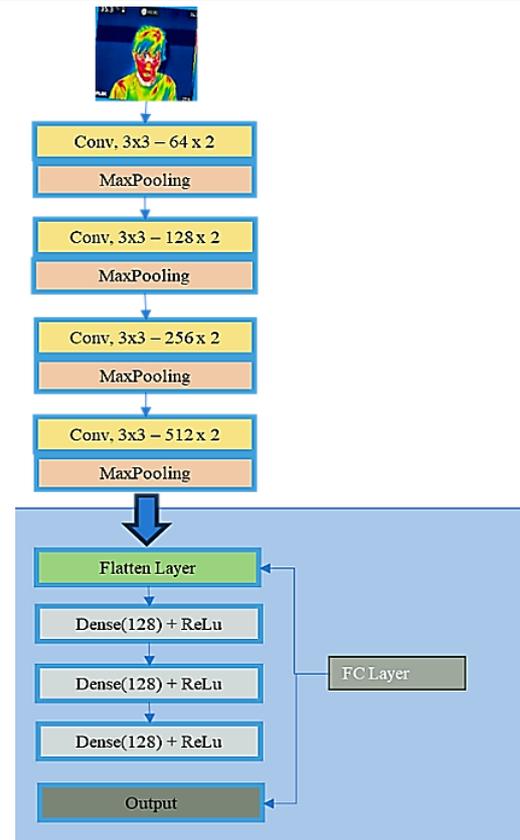


Fig. 2. Illustration architecture of VGG-19

to alter the pixel values in the images to fall within a consistent range, ensuring that the data were ready for the training process.

### 2.3. Image Classification

Image classification in this study involved two main phases: training and testing. During the training phase, a classification model is developed using labeled training data, which is further divided into subsets for training, validation, and testing. In the testing phase, the developed classification model was applied to the unlabeled test data to evaluate its performance.

Before initiating the training process, network parameters were configured to optimize performance and ensure stability. The Adam optimizer was selected due to its efficiency in handling sparse gradients and adapting learning rates, which makes it particularly suitable for deep learning tasks with high-dimensional parameter spaces. To ensure adequate exposure to the dataset while minimizing overfitting risks, the training process was performed over 100 epochs. A batch size of 32 was chosen to strike a balance between computational efficiency and gradient stability. Additionally, a learning rate of 0.0001 was employed to ensure a gradual learning process. This relatively lower learning rate helps maintain model stability and prevents divergence, facilitating steady convergence during training.

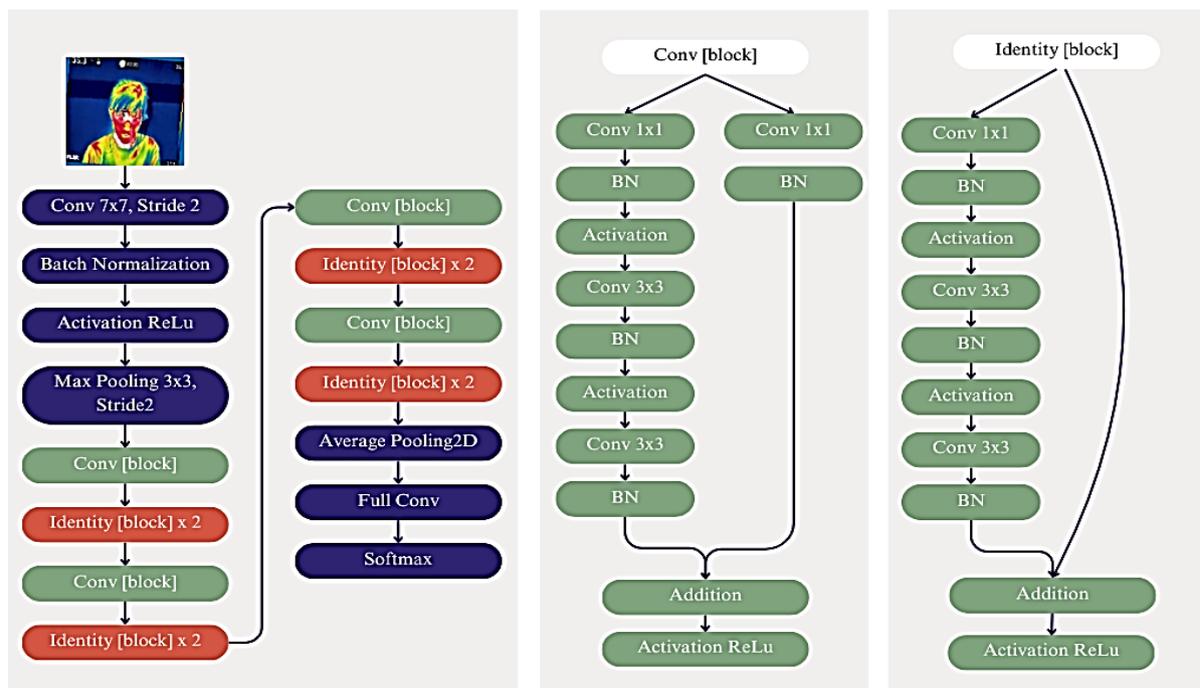


Fig. 3. Illustration architecture of ResNet50V2

## 2.4. Visual Geometry Group 19

As shown in Fig. 2, the architecture of VGG-19 affirms its complexity. This artificial neural network is composed of 19 layers, including 16 convolution layers, 3 fully connected, 5 maxpool, 1 SoftMax, and 1 output layer, all of which are harnessed by CNN techniques [18, 19]. Typically implemented on the ImageNet data set, VGG-19's 3x3 size in each layer allows for deep convolution layers. This feature allows the proposed network to discern even the most subtle details in images [20, 21].

The training process using the VGG-19 architecture begins with inputting the dataset, followed by labeling: "autism" is labeled as "1" and "non-autism" as "0". Next, the data undergoes the convolutional stage, which generates a feature layer with values in the form of matrices that differ from the input data. In this feature layer, the rectified linear unit (ReLU) activation function is applied, which converts negative values to 0 and passes through positive values, resulting in a new matrix. The matrix generated in the previous step is down-sampled through a pooling layer using the max pooling function. Then, the pooling-layer results are flattened, producing a one-dimensional matrix that is connected to the fully connected layer to form several hidden layers with a specified number. After the hidden layers are formed, the process moves to the activation function stage, where softmax activation is used in this study.

## 2.5. ResNet50V2

ResNet50V2 has several phases. This is because this model has 50 convolutional layers (including convolution layers in the residual block) and several fully-connected layers at the end [22, 23]. ResNet's main contribution lies in its residual blocks. This block facilitates the training of deeper networks by addressing the performance degradation problem, which often occurs in deep networks. As shown in Fig. 3, ResNet50V2 also employs an identity block, which is used to keep input dimensions constant when a convolution layer is applied in the residual block [24]. ResNet50V2 no longer exploits post-activation but rather pre-activation, which allows layers to return to the previous layer. This modification reduces the complexity of the network, thereby increasing its efficiency [25, 26].

The classification process using ResNet50V2 begins with batch normalization to allow gradients to flow more easily. Next, features are extracted in the convolutional layers. The process then moves to the max pooling stage, which reduces the spatial dimensions of the feature maps generated by the convolutional layers, thereby increasing translation invariance within the image. Then, the dimensions of the features are further reduced, and the output is streamlined using global average pooling. The final step in classification is to use the fully connected layer, which is a type of dense layer that ensures that each neuron is connected to all neurons in the previous layer.

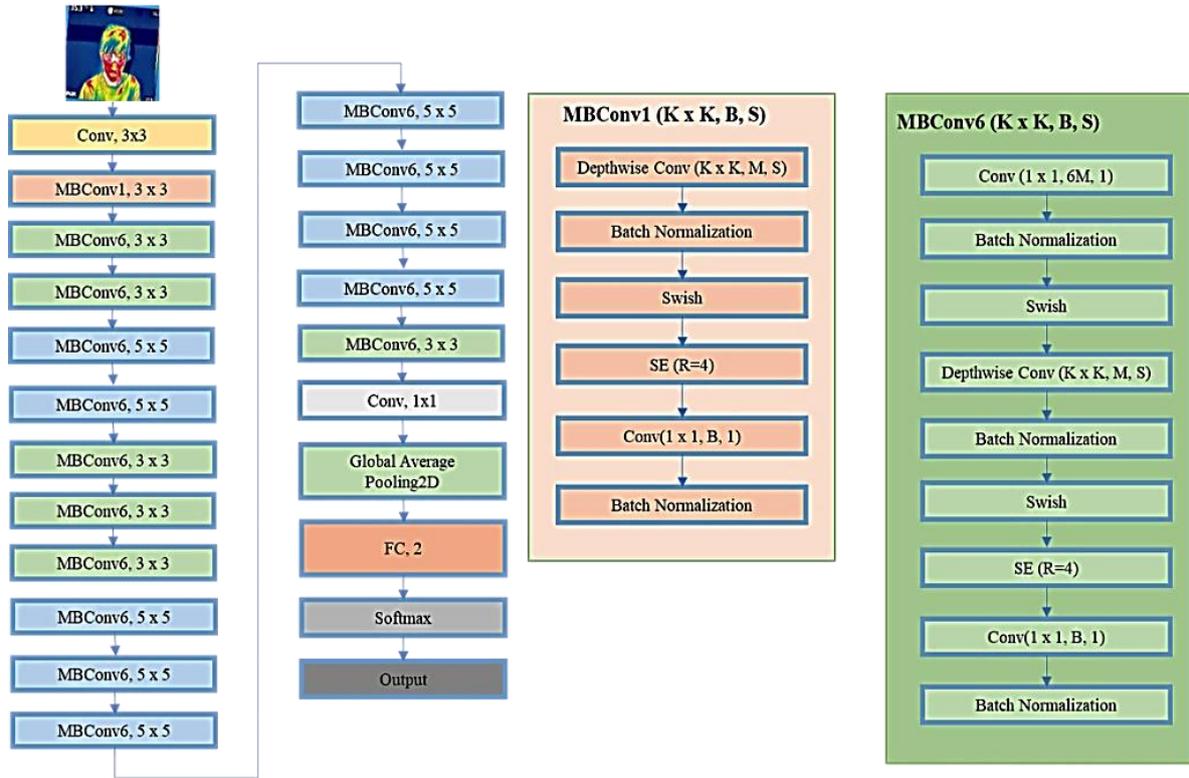


Fig. 4. Illustration architecture of EfficientNet

### 2.6. EfficientNet

EfficientNet is a CNN architectural model that increases accuracy and model precision by reducing parameters and FLOPS (Floating Point Operations Per Second) [27, 28]. This model also uses Squeeze-and-Excitation (SE) optimization to improve model performance. EfficientNet-B0 uses a "compound scaling" approach to balance three critical dimensions of a neural network, namely width, depth, and resolution. This approach allows the model to achieve previously unattainable levels of efficiency without sacrificing accuracy [29, 30]. An illustration of the proposed EfficientNet architecture is shown in Fig. 4.

The classification process using EfficientNet begins by determining the filter size, which is a two-dimensional matrix used to extract information from each pixel of the image. A 3x3 convolution follows this. To optimize computational efficiency, EfficientNet employs depth-wise separable convolution, which consists of two stages: depth-wise and pointwise convolution. Next, global average pooling is applied to the extracted features to reduce feature dimensions and streamline the output. The process then moves to the fully connected layer, which links the extracted features to the class labels. The softmax activation function typically follows this to obtain the probability distribution across classes.

### 2.7. Evaluation

This stage compares the classification performance results of the VGG-19, ResNet50V2, and EfficientNet architectures. Eqs. (1) – (4) can be used to calculate the accuracy, recall, precision, and F1-Score values to determine system performance. The performance results of the model were then compared with those of other studies. Each model requires test parameters as comparison values. The parameters used in this study are described below [31].

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} \tag{1}$$

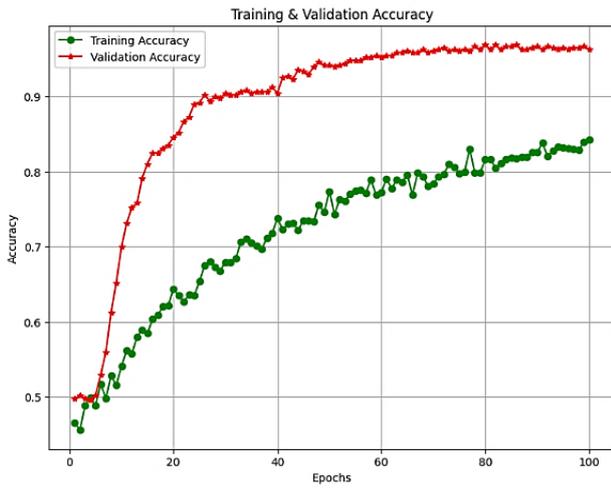
$$\text{Precision} = \frac{(TP)}{(TP + FP)} \tag{2}$$

$$\text{Recall} = \frac{(TP)}{(TP + FN)} \tag{3}$$

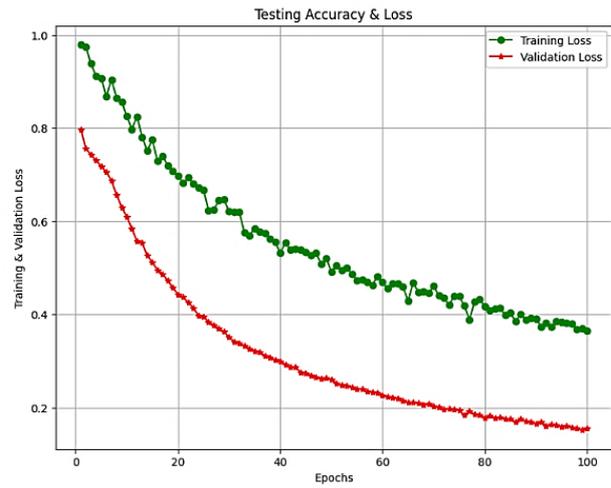
$$\text{F1 - Score} = 2 \times \frac{(\text{Recall} \times \text{precision})}{(\text{Recall} + \text{precision})} \tag{4}$$

The characters in the confusion matrix are explained as follows:

- TP (True Positive): Represents data that are actually positive and correctly predicted as belonging to the positive class;
- TN (True Negative): Represents data that are actually negative and correctly predicted as belonging to the negative class;

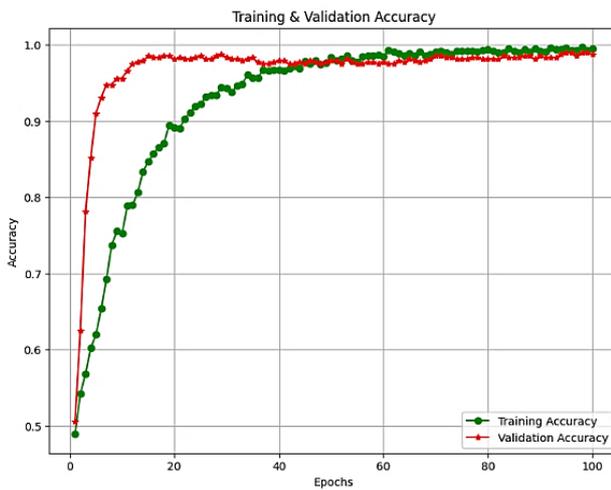


(a)

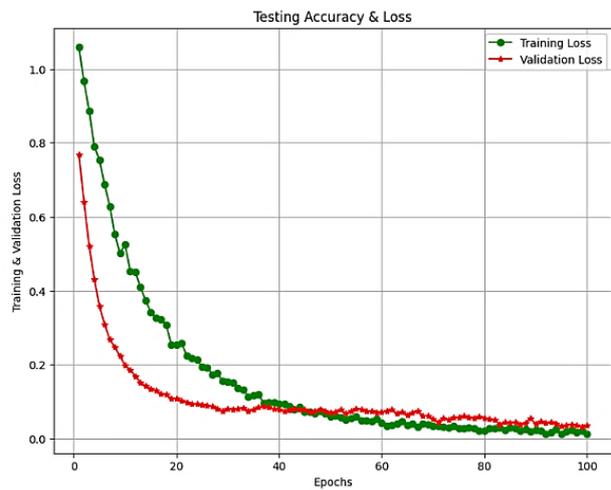


(b)

Fig. 5. Epoch results using VGG-19 architecture: (a) Accuracy vs epoch; (b) Loss vs epoch

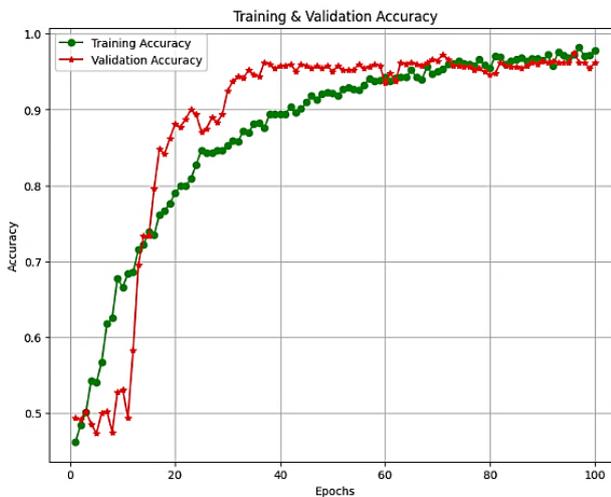


(a)

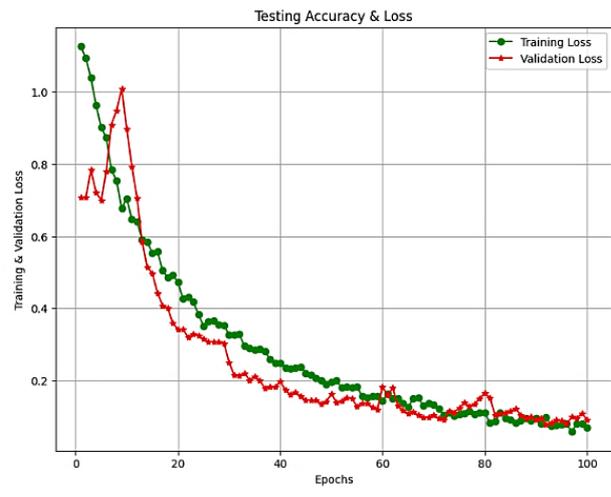


(b)

Fig. 6. Epoch results using ResNet50V2 architecture: (a) Accuracy vs epoch; (b) Loss vs epoch



(a)



(b)

Fig. 7. Epoch results using EfficientNet architecture: (a) Accuracy vs epoch; (b) Loss vs epoch

- FP (False Positive): This class refers to data that are actually negative but incorrectly predicted as belonging to the positive class;
- FN (False Negative): Refers to data that are actually positive but incorrectly predicted as belonging to the negative class.

### 3. Results and Analysis

#### 3.1. Image Thermal Classification

The data were preprocessed in the image classification stage. The training and test data. In the training process, the training data consists of two data sets: training data and validation data. At this stage, 3,610 training data were used, including 3,100 training data and 510 validation data.

#### 3.2. Training Result and Dataset Validation

In this stage, image classification is performed separately using VGG-19, ResNet50V2, and EfficientNet. Figures 5–7 compare the accuracy and loss of each architecture on the training and validation data.

The green curve represents the training accuracy based on Fig. 5(a), which shows the curve of the relationship between epochs and accuracy using the VGG-19 architecture. In contrast, the red curve represents the validation accuracy. Overall, epochs for testing and validation data in each architecture were performed in as many as 100 epochs; based on the explanation in Fig. 5, it is stated that in the 100th epoch for the testing curve, the maximum accuracy level reached 84.23%. Meanwhile, in the validation curve (red), achieving an accuracy of 84.58% only took 20 epochs. This explains why the model works well during the data validation process with the comparison results, as mentioned earlier.

Figure 5(b) shows the variation between training loss (green curve) and validation loss (red curve). At epoch 81, the training loss rate was 40%, and the validation loss was 18%. This indicates that the loss rate obtained in the validation phase is lower compared to the training phase, which affected the resulting accuracy level. On the training loss curve, it can be seen that at epoch 81, the loss level was 40%. In addition, on the validation loss curve with the same number of epochs, the loss level was 18%. Based on this, it can be concluded that the validation phase has a lower loss than the training phase, which in turn influences the resulting accuracy level.

Figure 6(a) shows the variation between training accuracy (green curve) and validation accuracy (red curve). Initially, the training accuracy was 91.14%, and the validation accuracy was 98.12%. However, by epoch

67, the situation changed: training accuracy increased to 99.10%, surpassing validation accuracy, which slightly decreased to 97.71%. This indicates that the model is well-trained (well-fitted), with training performance slightly exceeding validation performance in later epochs.

Figure 6(b) shows a graph illustrating the relationship between epochs and loss using ResNet50V2, where the green curve represents training loss and the red curve represents validation loss. It can be observed that at epoch 15, training loss was relatively high (34.18%, and the validation loss is 13.46%). However, at epoch 92, the training loss was reduced to 1.33%, and the validation loss was reduced to 4.23%. From these data, we conclude that the more epochs were used, the lower the loss level.

Figure 7(a) shows the curve illustrating the relationship between epochs and accuracy, where the green curve represents training accuracy, and the red curve represents validation accuracy. It can be seen that at epoch 11, there is a difference where the training accuracy is higher than the validation accuracy, with the training accuracy being 68.35% and the validation accuracy being 49.38%. This suggests the occurrence of overfitting, where there is a low measurement error but a high testing error; thus, the accuracy is high in measurement, but the performance is low in the test set.

Figure 7(b) shows the curve illustrating the relationship between epochs and loss using the EfficientNet architecture. The green curve represents the loss during the training phase. In contrast, the red curve represents the loss during the validation phase. At epoch 52, the training loss was 17.99%, and the validation loss was 14.22%. At epoch 92, the training loss decreased to 9.68%, and the validation loss decreased to 7.75%. From this, it can be observed that the loss in both the training and validation phases decreased as the number of epoch increased.

Overall, the more epochs a model was trained, the better was the model fit to the training data. However, if a model is drilled for too long, it may start to overfit the training data, which can negatively impact its performance on unseen data. Therefore, it is important to find a balance between the number of epochs and model performance [33].

#### 3.3. Testing Result and Dataset Validation

In the testing phase, the dataset comprised 510 test images (255 of children with autism and 255 normal images).

Figure 8 shows a confusion matrix showing the test results obtained using the VGG-19 model architecture. It can be observed that the model correctly predicted 246

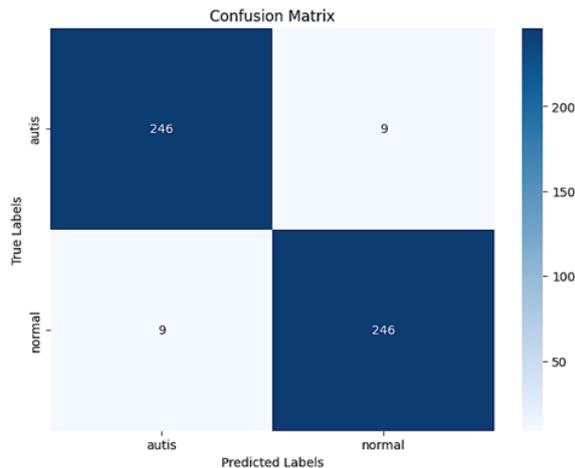


Fig. 8. Testing results using VGG-19

images and made 9 incorrect predictions in the "normal" class. In the "autistic" class, the model also correctly predicted 246 images and made 9 incorrect predictions.

Figure 9 is a confusion matrix showing the test results obtained using the ResNet50V2 model architecture. The model correctly predicted 249 images and made 6 incorrect predictions in the "normal" class. In the "autistic" class, the model correctly predicted 255 images with 0 incorrect predictions. At this stage, the collection of facial thermal image datasets was performed, which were divided into two groups, namely children with autism and normal children, as obtained from the research of Melinda et al. [11] via Google Drive. The dataset was divided into three folders: data validation, training, and testing. The total number of datasets was 4,120 images. In the validation data, there are 510 images; in the training data, as many as 3,100 images; and in the testing data, as many as 510 images.

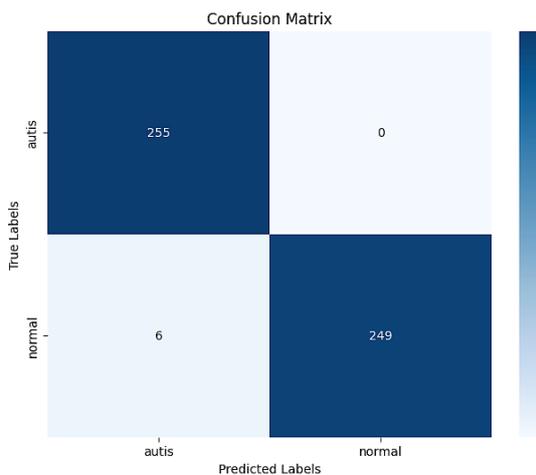


Fig. 9. Testing results using ResNet50V2

Figure 10 presents a confusion matrix showing the test results obtained using the VGG-19 model

architecture. The model correctly predicted 235 images and made 20 incorrect predictions in the "normal" class. In the "autistic" class, the model correctly predicted 255 images with 0 incorrect predictions.

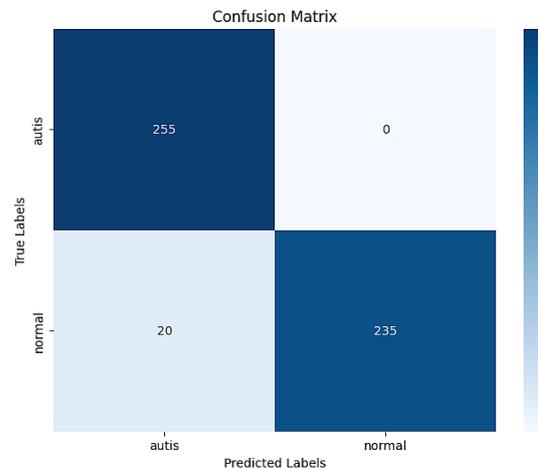


Fig. 10. The testing result using EfficientNet

Based on the test results shown in Figs. 8 – 10, the performance of the VGG-19, ResNet50V2, and EfficientNet architectures was evaluated. Confusion matrices are used to measure the performance of the classification system, and the accuracy, recall, precision, and F1-score are calculated using Eqs. (1) – (4).

Misclassification analysis for Figure 10 reveals that the EfficientNet architecture recorded the highest number of incorrect predictions among the three tested architectures. Specifically, the model made 20 incorrect predictions in the "normal" class, the largest number of errors compared to VGG-19 and ResNet50V2, both of which had fewer errors. The EfficientNet architecture achieves competitive accuracy overall; however, its tendency to misclassify images in the "normal" class underscores potential limitations relative to capturing subtle variations in thermal features specific to this category. This highlights the challenges inherent in feature extraction constraints in EfficientNet, particularly for thermal imaging.

Table 1

The comparison of the accuracy achievements of the three architectures

| CNN's Architecture | Accuracy | Precision | Recall | F1-Score |
|--------------------|----------|-----------|--------|----------|
| VGG-19             | 96.47%   | 96.47%    | 96.47% | 96.47%   |
| ResNet50V2         | 98.82%   | 100%      | 97.7%  | 99.4%    |
| EfficientNet       | 96.07%   | 100%      | 92.72% | 97.9%    |

Table 1 compares the accuracy of the three architectures. The best results were obtained using ResNet50V2, with an accuracy of 98.82%. This accuracy rate is better than that of VGG-19 and EfficientNet, which have accuracy rates of 96.47% and 96.07%, respectively. Thus, we conclude that the ResNet50V2

architecture outperforms VGG-19 and EfficientNet in the facial image classification task of children with autism. This indicates that choosing the appropriate architecture, such as ResNet50V2, can provide better performance in terms of accuracy, precision, recall, and F1-Score compared to the VGG-19 and EfficientNet architectures.

The results of this study are consistent with previous findings that highlighted ResNet's superiority in classification tasks. In this study, ResNet50V2 achieved the highest accuracy for thermal facial image classification of children with autism, outperforming VGG-19 and EfficientNet. Similarly, a study on chest CT images for tuberculosis, pneumonia, and COVID-19 diagnosis found that ResNet was the top performer, achieving 96.6% accuracy and outperforming AlexNet, GoogleNet, and DCNN [32]. These results confirm ResNet's exceptional feature extraction and generalization capabilities, making it a reliable choice for complex datasets in medical diagnostics and autism detection.

Table 2  
Performance comparison of facial recognition methods in autistic children

| Ref.            | Data Type     | Method       | Accuracy |
|-----------------|---------------|--------------|----------|
| [11]            | Thermal Image | CNN          | 98%      |
| [12]            | 2D RGB Image  | VGG-19       | 86.5%    |
|                 |               | ResNet50V2   | 94%      |
|                 |               | EfficientNet | 85.8%    |
| [13]            | 2D RGB Image  | ResNet50V2   | 93.97%   |
| [20]            | 2D RGB Image  | VGG-19       | 98%      |
|                 |               | ShuffleNet   | 88%      |
| Proposed Method | Thermal Image | VGG-19       | 96.47 %  |
|                 |               | ResNet50V2   | 98.82%   |
|                 |               | EfficientNet | 96.07%   |

Table 2 compares the performances of several facial recognition methods that use visual and thermal images in autistic children. Our results demonstrate that the proposed method has a better accuracy value than the previous method, namely 98.4% on the ResNet50V2 architecture.

**Conclusions**

This study successfully evaluated the effectiveness of CNN models for facial recognition in children with autism, specifically, the VGG-19, ResNet50V2, and EfficientNet architectures. The results demonstrate that facial features in children with autism and non-autism exhibit subtle differences at specific points, which are challenging to discern through visual inspection.

However, using CNN-based facial recognition systems, these differences can be effectively identified and classified. Among the models tested, ResNet50V2 achieved the highest accuracy (98.82%, surpassing other architectures). The VGG-19 and EfficientNet models also demonstrated commendable performance, with accuracy levels of 96.47% and 96.07%, respectively, highlighting their reliability for this application.

**In future, we plan to expand** the dataset to include more diverse samples and improve model robustness. In addition, exploring the integration of advanced deep learning techniques and hybrid approaches may further enhance classification accuracy and real-world applicability.

**Contributions of authors:** conceptualization, methodology – Junidar Junidar, Melinda Melinda, Donata D. Acula, Zulfan Zainal, Dinda D. Diannuari; formulation of tasks, analysis – Junidar Junidar, Melinda Melinda, Donata D. Acula; development of model, software, verification – Junidar Junidar, Melinda Melinda, Zulfan Zainal, Dinda D. Diannuari; analysis of results, visualization – Junidar, Junidar, Melinda Melinda, Donata D. Acula, Zulfan Zainal, Dinda D. Diannuari; writing – original draft preparation, writing – review and editing – Junidar, Junidar, Melinda Melinda, Donata D. Acula, Dinda D. Diannuari.

**Conflict of Interest**

The authors declare that they have no conflict of interest related to this research, whether financial, personal, authorship, or otherwise, that could affect the research and its results presented in this paper.

**Financing**

This study was conducted without financial support.

**Data Availability**

The work has associated data in the data repository.

**Use of Artificial Intelligence**

The authors confirm that they did not use artificial intelligence methods while creating the presented work.

**Acknowledgments**

We want to acknowledge this to the Universitas Syiah Kuala, which has supported our study.

All the authors have read and agreed to the published version of the manuscript.

## References

1. Chiarotti, F., & Venerosi, A. Epidemiology of Autism Spectrum Disorders: A Review of Worldwide Prevalence Estimates Since 2014. *Brain Sciences*, 2020, vol. 10, no. 5, article no. 274. DOI: /10.3390/brainsci10050274.
2. Melinda, M., Juwono, F. H., Enriko, I. K. A., Oktiana, M., Mulyani, S., & Saddami, K. Application of Continuous Wavelet Transform and Support Vector Machine for Autism Spectrum Disorder Electroencephalography Signal Classification. *Radioelectronic and Computer Systems*, 2023, vol. 3, no. 107, pp. 73–90. DOI: 10.32620/reks.2023.3.07.
3. Afrin, M. F. N., Freeda, S., Elakia, S., & Kannan, P. AI Based Facial Expression Recognition for Autism Children. *International Journal of Emerging Technology and Innovative Engineering*, 2019, vol. 5, iss. 9, pp. 645–651. Available at: <https://ssrn.com/abstract=3446639>. (accessed 11.11.2024)
4. Lyu, K., Li, J., Chen, M., Li, W., Zhang, W., Hu, M., Zhang, Y., & Feng, X. A Bibliometric Analysis of Autism Spectrum Disorder Signaling Pathways Research in the Past Decade. *Frontiers in Psychiatry*, 2024, vol. 15, article no. 1304916. DOI: 10.3389/fpsy.2024.1304916.
5. Alkahtani, H., Aldhyani, T. H. H., & Alzahrani, M. Y. Deep Learning Algorithms to Identify Autism Spectrum Disorder in Children-Based Facial Landmarks. *Applied Sciences*, 2023, vol. 13, no. 8, 4855 p. DOI: 10.3390/app13084855.
6. Goulart, C., Valadão, C., Delisle-Rodriguez, D., Caldeira, E., & Bastos, T. Emotion Analysis in Children Through Facial Emissivity of Infrared Thermal Imaging. *PLoS ONE*, 2019, vol. 14, no. 3, article no. e0212928. DOI: 10.1371/journal.pone.0212928.
7. Rusli, N., Sidek, S. N., Yusof, H. M., Ishak, N. I., Khalid, M., & Dzulkarnain, A. A. A. Implementation of Wavelet Analysis on Thermal Images for Affective States Recognition of Children with Autism Spectrum Disorder. *IEEE Access*, 2020, vol. 8, pp. 120818–120834. DOI: 10.1109/ACCESS.2020.3006004.
8. Ramasamy, M. P., Krishnasamy, V., & Ramapackiam, S. S. K. Transfer Learning Based Convolutional Neural Network for Classification of Remote Sensing Images. *Advances in Electrical and Computer Engineering*, 2023, vol. 23, iss. 4, pp. 31–40. DOI: 10.4316/AECE.2023.04004.
9. Alamri, J., Harrabi, R., & Ben Chaabane, S. Face Recognition based on Convolution Neural Network and Scale Invariant Feature Transform. *International Journal of Advanced Computer Science and Applications*, 2021, vol. 12, no. 2, pp. 644–654. DOI: 10.14569/IJACSA.2021.0120281.
10. Teoh, K. H., Ismail, R. C., Naziri, S. Z. M., Hussin, R., Isa, M. N. M., & Basir, M. S. S. M. Face Recognition and Identification Using Deep Learning Approach. *Journal of Physics: Conference Series*, 2021, vol. 1755, no. 1, article no. 012006. DOI: 10.1088/1742-6596/1755/1/012006.
11. Melinda, M., Ahmadiar, A., Oktiana, M., Nugraha, M. S. A., Qadrillah, M. A. L., & Yunidar, Y. A Novel Autism Spectrum Disorder Children Dataset Based on Thermal Imaging. *2023 9th International Conference on Computer and Communication Engineering (ICCCCE)*. Kuala Lumpur, Malaysia, IEEE, 2023. DOI: 10.1109/ICCCCE58854.2023.10246072.
12. Alam, M.S., Rashid, M. M., Roy, R., Faizabadi, A. R., Gupta, K. D., & Ahsan, M. M. Empirical Study of Autism Spectrum Disorder Diagnosis Using Facial Images by Improved Transfer Learning Approach. *Bioengineering*, 2022, vol. 9, no. 11, article no. 710. DOI: 10.3390/bioengineering9110710.
13. Alam, M. S., Rashid, M. M., Faizabadi, A. R., & Zaki, H. F. M. Power of Alignment: Exploring the Effect of Face Alignment on ASD Diagnosis Using Facial Images. *IJUM Engineering Journal*, 2024, vol. 25, no. 1, pp. 317–327. DOI: 10.31436/iiu mej.v25i1.2838.
14. Yunidar, Y., Roslidar, R., Oktiana, M., Yusni, Y., Nasaruddin, N., & Arnia, F. Classification of Stunted and Normal Children Using Novel Facial Image Database and Convolutional Neural Network. *Radioelectronic and Computer Systems*, 2024, vol. 109, no. 1, pp. 76–86. DOI: 10.32620/reks.2024.1.07.
15. Hazra, S. K., Erna, R. R., Galib, S. M., Kabir, S., & Adnan, N. Emotion Recognition of Human Speech Using Deep Learning Method and MFCC Features. *Radioelectronic and Computer Systems*, 2022, vol. 104, no. 4, pp. 161–172. DOI: 10.32620/reks.2022.4.13.
16. Oumoulyte, M., Alaoui, A. O., Farhaoui, Y., El Allaoui, A., & Bahri, A. Convolutional Neural Network-Based Skin Cancer Classification with Transfer Learning Models. *Radioelectronic and Computer Systems*, 2023, vol. 108, no. 4, pp. 75–87. DOI: 10.32620/REKS.2023.4.07.
17. Yaloveha, V., Podorozhniak, A., Kuchuk, H., & Garashchuk, N. Performance Comparison of CNNs High-Resolution Multispectral Dataset Applied to Land Cover Classification Problem. *Radioelectronic and Computer Systems*, 2023, no. 2(106), pp. 107–115. DOI: 10.32620/reks.2023.2.09.
18. Alsaade, F. W., & Alzahrani, M. S. Retracted: Classification and Detection of Autism Spectrum Disorder Based on Deep Learning Algorithms. *Computational Intelligence and Neuroscience*, 2022, vol. 2022, article no. 9893015. DOI: <https://doi.org/10.1155/2023/9893015>.
19. Sudha, V., & Ganeshbabu, T. R. A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning. *Computers, Materials and Continua*, 2021, vol. 66, no. 1, pp. 827–842. DOI: 10.32604/cmc.2020.012008.
20. Melinda, M., Oktiana, M., Nurdin, Y., Pujiati, I., Irhamsyah, M., & Basir, N. Performance of ShuffleNet and VGG-19 Architectural Classification Models for Face Recognition in Autistic Children. *International Journal on Advanced Science, Engineering and*

*Information Technology*, 2023, vol. 13, iss. 2, article no. 674. DOI: 10.18517/ijaseit.13.2.18274.

21. Hastari, D., Winanda, S., Pratama, A. R., Nurhaliza, N., & Ginting, E. S. Application of Convolutional Neural Network ResNet-50 V2 on Image Classification of Rice Plant Disease. *Public Research Journal of Engineering, Data Technology and Computer Science*, 2024, vol. 1, iss. 2, pp. 71–77. DOI: 10.57152/predatecs.v1i2.865.

22. Fitriyani, H. I., & Rizkinia, M. Improvement of Xception-ResNet50V2 Concatenation for COVID-19 Detection on Chest X-Ray Images. *2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT)*. Surabaya, Indonesia, IEEE, 2021, pp. 343–347. DOI: 10.1109/EIConCIT50028.2021.9431916.

23. Bayunanda, E., Utami, E., & Ariatmanto, D. Facial Expression Classification Analysis Using Facial Images Based on Resnet-50V2. *2022 6th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*. Yogyakarta, Indonesia, IEEE, 2022, pp. 368–372. DOI: 10.1109/ICITISEE57756.2022.10057622.

24. Sukegawa, S., Tanaka, F., Hara, T., Yoshii, K., Yamashita, K., Nakano, K., Takabatake, K., Kawai, H., Nagatsuka, H., & Furuki, Y. Deep Learning Model for Analyzing the Relationship Between Mandibular Third Molar and Inferior Alveolar Nerve in Panoramic Radiography. *Scientific Reports*, 2022, vol. 12, article no. 16925. DOI: 10.1038/s41598-022-21408-9.

25. Yang, E. H., Amer, H., & Jiang, Y. Compression Helps Deep Learning in Image Classification. *Entropy*, 2021, vol. 23, iss. 7, article no. 881. DOI: 10.3390/e23070881.

26. Wang, L., Shi, C., Lin, S., Qin, P., & Wang, Y. Convolutional Sparse Representation and Local Density Peak Clustering for Medical Image Fusion. *International Journal of Pattern Recognition and Artificial Intelligence*, 2020, vol. 34, no. 7, article no. 2057003. DOI: 10.1142/S0218001420570037.

27. Wang, J., Yang, L., Huo, Z., He, W., & Luo, J. Multi-Label Classification of Fundus Images with EfficientNet. *IEEE*, 2020, vol. 8, pp. 212499–212508. DOI: 10.1109/ACCESS.2020.3040275.

28. Ajmera, P. K., Kabra, S. M., Mall, A., & Lhila, A. AMAizeD: An End to End Pipeline for Automatic Maize Disease Detection. *2023 16th International Conference on Sensing Technology*. Hyderabad, India, IEEE, 2023, pp. 1–6. DOI: 10.1109/ICST59744.2023.10460787.

29. Hridoy, R. H., Akter, F., & Rakshit, A. Computer Vision Based Skin Disorder Recognition Using EfficientNet: A Transfer Learning Approach. *2021 International Conference on Information Technology*. Amman, Jordan, IEEE, 2021, pp. 482–487. DOI: 10.1109/ICIT52682.2021.9491776.

30. Alam, I. N., Kartowisastro, I. H., & Wicaksono, P. Transfer Learning Technique with EfficientNet for Facial Expression Recognition System. *Revue d'Intelligence Artificielle*, 2022, vol. 36, no. 4, pp. 543–552. Available at: DOI: 10.18280/ria.360405.

31. Ahmed, T., & Sabab, N. H. N. Classification and Understanding of Cloud Structures via Satellite Images with EfficientUNet. *SN Computer Science*, 2022, vol. 3, article no. 99. DOI: 10.1007/s42979-021-00981-2.

32. Kaewlek, T., Tanyong, K., Chakkaeo, J., Kladpree, S., Chusin, T., Yabsantia, S., & Udee, N. Classification of Pneumonia, Tuberculosis, and COVID-19 on Computed Tomography Images Using Deep Learning. *Trends in Sciences*, 2023, vol. 20, no. 11, article no. 6974. DOI: 10.48048/tis.2023.6974.

33. Ishida, T., Yamane, I., Sakai, T., Niu, G., & Sugiyama, M. (2020). *Do We Need Zero Training Loss After Achieving Zero Training Error?* <http://arxiv.org/abs/2002.08709>.

Received 18.01.2025, Accepted 17.02.2025

## НАВЧАННЯ В АНСАМБЛІ VGG-19, RESNET50V2 ТА EFFICIENTNET ДЛЯ КЛАСИФІКАЦІЇ АУТИСТИЧНОГО ОБЛИЧЧЯ НА ОСНОВІ ТЕПЛОПОБРАЗУ

Джунідар Джунідар, Мелінда Мелінда, Дінда Діва Діаннуарі,  
Доната Д. Акула, Зульфан Зайнал

Предметом статті є виявлення ознак розладу аутистичного спектру (РАС) за допомогою розпізнавання обличчя на основі теплових зображень. Мета полягає в тому, щоб оцінити та порівняти продуктивність різних архітектур згорткової нейронної мережі (CNN) у класифікації теплових зображень обличчя дітей з РАС, сприяючи ранній і точній ідентифікації аутичних рис. Завдання, які вирішуються, включають попередню обробку набору даних теплових зображень обличчя, щоб підготувати їх для навчання моделі, проведення класифікації з використанням трьох архітектур CNN VGG-19, ResNet50V2 та EfficientNet; і оцінка їх продуктивності на основі показників точності, точності, запам'ятовування та показників F1. Застосовувані методи включають навчання цих моделей CNN на збалансованому наборі даних із 4120 теплових зображень обличчя, поділ їх на набори для навчання, перевірки та тестування. Кожна модель пройшла інтенсивне навчання, щоб визначити її здатність ефективно класифікувати аутизм і не аутизм. Результати показали, що ResNet50V2 досяг найвищої точності 98,82%, за ним йдуть VGG-19 і EfficientNet з точністю 96,47% і 96,07% відповідно. ResNet50V2 також продемонстрував чудове узагальнення завдяки меншій втраті перевірки та вищій точності класифікації

порівняно з іншими архітектурами. Висновок. Наукова новизна полягає в: 1) запровадженні тепловізора як ефективного засобу виявлення ознак РАС; 2) демонстрація чудової продуктивності ResNet50V2 у класифікації теплових зображень обличчя з високою точністю та узагальненням; і 3) дослідження EfficientNet вперше в цій галузі, підкреслюючи його потенціал для вдосконалення систем діагностики аутизму. Це дослідження сприяє вдосконаленню неінвазивних методів виявлення РАС і прокладає шлях для подальшого застосування глибокого навчання в клінічній діагностиці.

**Ключові слова:** розлад аутистичного спектру (РАС); теплові зображення; EfficientNet; ResNet50V2; ВГГ-19.

**Джунідар Джунідар** – Ступінь бакалавра мультимедійних систем, Департамент інформатики, факультет математики та природничих наук, Університет Сія Куала, Банда Ачех 23111, Індонезія.

**Мелінда Мелінда** – Докторський ступінь із обробки мультимедійних сигналів, кафедра електротехніки та комп'ютерної інженерії, інженерний факультет, Університет Сія Куала, Банда-Ачех, 23111, Індонезія.

**Дінда Діва Діаннуарі** – Ступінь бакалавра електротехніки, каф. електротехніки та комп'ютерної інженерії, інженерний факультет, Університет Сія Куала, Банда Ачех, 23111, Індонезія.

**Доната Д. Акула** – Ступінь доктора математичної освіти, Департамент комп'ютерних наук, Коледж інформаційних та обчислювальних наук, Університет Санто-Томас, Маніла, Філіппіни.

**Зульфан Зайнал** – Ступінь бакалавра мультимедійних систем, факультет комп'ютерної інженерії, інженерний факультет, Університет Серамбі Мекка, Банда-Ачех, 23245, Індонезія.

**Junidar Junidar** – Bachelor Degree in Multimedia System, the Department of Informatics, Mathematics and Natural Sciences Faculty, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia,

e-mail: junidar678@usk.ac.id, ORCID: 0009-0009-4273-6669, Scopus AuthorID: 59509214800.

**Melinda Melinda** – Doctoral Degree in Multimedia Signal Processing, the Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia,

e-mail: melinda@usk.ac.id, ORCID: 0000-0001-9082-6639, Scopus AuthorID: 53264423700.

**Dinda D. Diannuari** – Bachelor's Degree in Electrical Engineering, the Department of Electrical and Computer Engineering, Faculty of Engineering, Universitas Syiah Kuala, Banda Aceh, 23111, Indonesia,

e-mail: dindadivadiannuari@gmail.com, ORCID: 0009-0004-3169-2967.

**Donata D. Acula** – Doctoral Degree in Mathematics Education, the Department of Computer Science, College of Information and Computing Sciences, University of Santo Tomas, Manila, Philippines,

e-mail: ddacula@ust.edu.ph, ORCID: 0000-0002-5198-4631, Scopus AuthorID: 57195516506.

**Zulfan Zainal** – Bachelor's Degree in Multimedia Systems, the Department of Computer Engineering, Faculty of Engineering, Universitas Serambi Mekkah, Banda Aceh, 23245, Indonesia,

e-mail: zulfanzainal@serambimekkah.ac.id, ORCID: 0000-0003-3674-5373, Scopus AuthorID: 57209460488.