Transfer learning-based malnutrition classification using VGG16 and comparative analysis of CNN architectures

Ahmad Fauzi[1], Haerul Yuda Aditiya[1], Maharina[1], Sihabudin[2], Muhammad Ansari Adista[3], Iflan Nauval[3], Natasya Eka Nanda Sonia Puri[1], Candra Zonyfar[1,4]
[1]Faculty of Computer Science, Universitas Buana Perjuangan Karawang, Karawang, Indonesia
[2]Faculty of Economic and Business, Universitas Buana Perjuangan Karawang, Karawang, Indonesia
[3]Faculty of Medicine, Universitas Syiah Kuala, Aceh, Indonesia
[4]Department of Computer Science and Engineering, Sun Moon University, Asan, Republic of Korea

Article Info

Article history:

Received Sep 11, 2025
Revised Jun 16, 2026
Accepted Jul 14, 2026

Keywords:

Child malnutrition
Convolution neural networks
Deep learning
Malnutrition classification
Transfer learning

ABSTRACT

Childhood malnutrition remains a critical public health challenge in developing countries, with Indonesia ranking fifth globally for stunting burden. Traditional anthropometric assessment methods are time-consuming, resource-intensive, and require trained personnel, necessitating efficient computer-based early detection approaches. This study proposes a deep learning-based method for automated nutritional status classification using facial image analysis. We developed and compared multiple transfer learning architectures including visual geometry group 16 (VGG16), densely connected convolutional network 121 (DenseNet121), mobile network version 2 (MobileNetV2), and residual network 50 (ResNet50) for classifying children into three categories: healthy, malnutrition, and stunting. Results demonstrated that VGG16, a simpler architecture trained for only 10 epochs, achieved optimal performance with 91.8% accuracy, and significantly outperforming more complex modern architectures like ResNet50. This finding challenges the conventional assumption that newer, deeper models invariably perform better, and particularly when working with limited medical datasets. The study revealed that longer training durations led to performance degradation due to overfitting, emphasizing the importance of balancing model complexity with dataset characteristics. These findings support the development of practical artificial intelligence (AI)-based malnutrition screening systems suitable for resource-constrained environments, potentially improving early detection capabilities, and public health outcomes in developing regions.

Corresponding Author:

Candra Zonyfar
Department of Computer Science and Engineering, Sun Moon University
Asan 31460, Republic of Korea
Email: candrazonyfar@sunmoon.ac.kr

1. INTRODUCTION

Malnutrition in children under five years of age remains a significant public health challenge in developing countries, including Indonesia [1], [2]. Childhood malnutrition encompasses various conditions such as stunting (impaired linear growth), wasting (acute underweight), underweight, and obesity. Malnutrition in children negatively affects cognitive development and educational achievement in adulthood, directly threatening future human resource potential [3], [4]. United Nations Children's Fund (UNICEF) reports that more than two million children in Indonesia experience nutritional issues such as stunting and wasting [5], with the national stunting rate reaching approximately 37% [6]. This high prevalence places Indonesia as the fifth country globally with the largest stunting burden [7]. The situation is exacerbated by multifactorial determinants, including socioeconomic aspects and environmental conditions such as sanitation and hygiene, which ultimately have significant humanitarian and economic impacts [8]. Therefore, early detection of malnutrition becomes a crucial initial step in addressing childhood malnutrition.

The causal factors of childhood malnutrition are complex and multidimensional. Poverty directly affects limited access to nutritious food for children. Low parental education levels contribute to limited knowledge about children's nutritional requirements. Poor environmental conditions, including inadequate sanitation and limited access to clean water, increase infection risks that can worsen children's nutritional status. Additionally, limited access to healthcare facilities such as community health posts, whether due to distance or transportation cost constraints, and creates barriers to routine child growth monitoring. The impact of malnutrition on children occurs in both short-term and long-term periods. In the short term, malnutrition causes physical growth disorders such as stunting (permanent linear growth failure) and wasting (acute underweight). This condition also weakens children's immune systems, increases susceptibility to infectious diseases, reduces cognitive ability, and motivation for activities. In the long term, malnutrition impacts academic performance decline, economic productivity reduction, and increased risk of non-communicable diseases in adulthood.

Detection and assessment of children's nutritional status are conducted using manual anthropometric measurements. Children under five years undergo measurements of body weight, height, body mass index (BMI), upper arm circumference, and head circumference [4], [7]. However, this approach is susceptible to assessment subjectivity, requires significant time and resources, and must be performed by trained healthcare personnel. Therefore, an efficient computer-based early detection method for malnutrition is needed [9]. Various studies have utilized statistical analysis to measure nutritional status using anthropometric data. These studies explore significant factors related to malnutrition and stunting from large-scale datasets. Statistical analysis of survey data has successfully identified factors influencing stunting, ranging from socioeconomic aspects to sanitation and hygiene conditions [6]-[8]. This analysis is crucial for policy formulation, although the ability to predict nutritional status at the individual level remains limited.

The development of artificial intelligence (AI) has provided breakthroughs in various sectors [10]–[16]. In health and nutrition fields, machine learning (ML) is used to address limitations of traditional approaches [17]. Various studies employ ML to develop malnutrition prediction models in children [18], [19], for early stunting detection [9], [18]–[21], and to predict malnutrition cases at the regional level [22]–[24]. Along with AI advancement, ML limitations can be addressed through deep learning approaches [25], particularly for processing unstructured data such as images [26]. In this context, researchers focus on deep learning approaches [9]. Various deep learning models have been developed to detect malnutrition [27], [28]. Specifically, several studies have demonstrated the effectiveness of facial image analysis for malnutrition detection and stunting classification [29]–[32]. Nevertheless, deep learning research for malnutrition classification still faces challenges, including model performance consistency for varying image datasets.

This study proposes a deep learning approach for child nutritional status classification based on image analysis. We conduct a comprehensive evaluation of various deep learning models using transfer learning, including visual geometry group 16 (VGG16), densely connected convolutional network 121 (DenseNet121), mobile network version 2 (MobileNetV2), and residual network 50 (ResNet50) for nutritional status classification. This research also presents comparative analysis of model architectures to identify the most reliable and effective approach as a foundation for future malnutrition detection system development. This paper consists of several sections. Section 2 presents the method used in this study, including discussion of dataset preprocessing and the proposed model architecture. Section 3 presents results and discussion. Section 4 presents conclusions and recommendations for future research.

2. MATERIAL AND METHOD

The objective of this research was to train and evaluate several deep learning architectures for image classification of children's nutritional status into three categories: healthy, malnutrition, and stunting. The overall research workflow, from data preprocessing to model evaluation, is illustrated in Figure 1. The first stage involves data preprocessing, which aims to standardize the quality and format of images used. Normalization and augmentation processes were applied to address potential bias caused by data imbalance and to enhance the generalization capability of the models. This step is crucial because real-world variations, such as differences in lighting conditions, camera angles, or photo quality, can affect system performance. Through this preprocessing, we expected the models to identify visual patterns more effectively. The subsequent stage encompasses training and evaluation, where several deep learning architectures were directly compared to identify the best-performing model. To ensure objective comparison, we conducted all experiments under uniform conditions, ensuring that the obtained results genuinely reflect the differences in capabilities between models. We performed evaluation using standard metrics including accuracy, precision, recall, and F1-score. Primary focus was given to recall and F1-score metrics to ensure that positive cases (malnutrition and stunting) could be accurately identified without being missed.

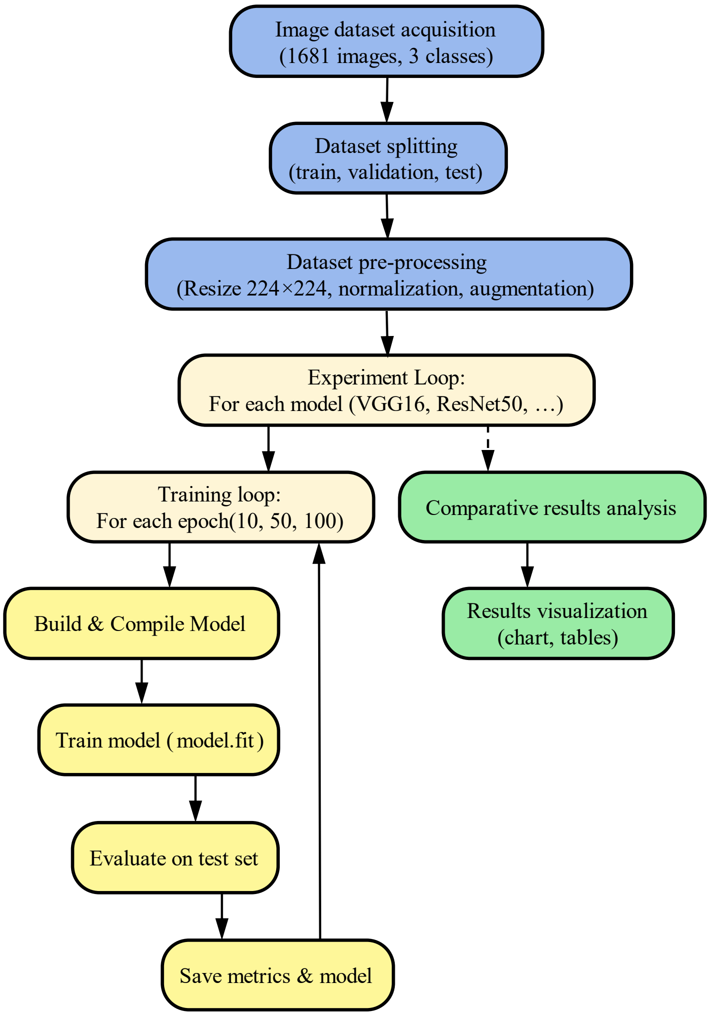


Figure 1. Block diagram of the proposed method

2.1. Data

This study utilized a secondary public dataset named "Deteksi Stunting Dataset" obtained from the Roboflow Universe platform. The dataset contains a total of 1,681 images of children that have been annotated in common objects in context (COCO) JavaScript object notation (JSON) format and classified into three nutritional status classes: 'Healthy', 'Malnutrition', and 'Stunting'. Dataset composition analysis revealed significant class imbalance, where the 'Malnutrition' class contained substantially fewer samples compared to the other two classes. Detailed dataset composition is presented in Table 1, while representative examples from each class are visualized in Figure 2.

To ensure objective model evaluation, we randomly divided the dataset into three parts: training set, validation set, and test set with proportions of 70%, 20%, and 10%, respectively. Table 2 presents the distribution of data samples for healthy and malnutrition classes across training, validation, and test sets.

Table 1. Dataset composition

Class	Number of images	Percentage (%)
Healthy	880	52.3
Malnutrition	46	2.7
Stunting	755	45.0
Total	**1681**	**100**

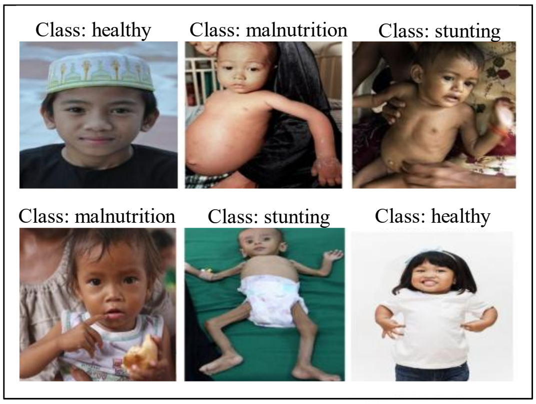


Figure 2. Sample of dataset images from each class

Table 2. Dataset distribution details

Class	Training data	Validation data	Test data	Total
Healthy	619	174	87	880
Malnutrition	24	15	7	46
Stunting	533	145	77	755
Total	**1176**	**334**	**171**	**1681**

2.2. Preprocessing

All images in the dataset underwent a series of preprocessing stages for standardization and augmentation. This stage was crucial for maintaining uniform input quality and enhancing the generalization capability of convolutional neural networks (CNN) models. We initiated this process by resizing all images from their original resolution of 640×640 pixels to 224×224 pixels to match the input dimensions of pre-trained model architectures. Subsequently, we performed pixel value normalization, where the original integer value range was converted to floating-point values in the range using:

$$p_{normalized} = \frac{p_{original}}{p_{max}}$$

This normalization step was essential for stabilizing and accelerating model convergence during training.

Given the limited labeled dataset, we applied data augmentation techniques exclusively to the training set. The augmentation aimed to increase synthetic sample diversity, thereby minimizing overfitting. We designed augmentation parameters to replicate real-world variations, including camera angles, subject positioning, and lighting conditions. Examples of each transformation are presented in Figure 3. The random transformations applied included: rotation of ±20°, horizontal and vertical shifts of ≤20% of image dimensions, shear of ≤20%, zoom of ±20%, and horizontal flipping.

Figure 3. Data augmentation process

2.3. Architecture

This study implemented transfer learning by leveraging pre-trained CNN. This approach proves highly effective for image classification when labeled data is limited, as it enables knowledge transfer from large-scale models trained on ImageNet. Pre-trained models have mastered hierarchical visual features—from edges and corners to textures and complex shapes—allowing this foundation to be efficiently adapted for nutritional status classification. The approach comprised two stages: i) a base architecture consisting of pre-trained models and ii) a custom head classifier positioned on top to generate final predictions. We froze all layers in the base architecture to preserve the features learned from ImageNet during training. Consequently, the optimization process focused on parameters in the relatively simple classifier head.

To evaluate performance, we compared four models: VGG16, ResNet50, DenseNet121, and MobileNetV2. These models not only represent the best achievements of their respective eras but also embody diverse neural network design philosophies, providing comprehensive perspectives on the most suitable architecture for nutritional status classification.

2.3.1. VGG16

Introduced by Simonyan and Zisserman [33] VGG16 is known for its simple yet deep architecture. Its innovation lies in its repeated use of 3×3 convolutional blocks. By stacking several 3×3 layers, VGG16 achieves a receptive field equivalent to larger kernels (5×5 or 7×7) while reducing the number of parameters, and increasing non-linearity. Its uniform structure simplifies understanding and implementation. Although its parameter count is relatively large compared to modern models, VGG16 remains a crucial baseline for evaluating new architectures.

2.3.2. ResNet50

ResNet introduced shortcut connections to overcome the vanishing gradient problem [34]. Instead of learning a direct mapping $H(x)$, each block learns a residual function $F(x) = H(x) - x$, making the final output $y = F(x) + x$. If the identity mapping is already optimal, the network only needs to drive $F(x)$ towards zero, which simplifies optimization and enables the training of deeper networks without performance degradation.

2.3.3. DenseNet121

DenseNet employs dense connectivity, where each layer within a dense block receives and forwards its feature maps to all other layers [35]. This rich flow of information and gradients mitigates the vanishing gradient problem and encourages the direct reuse of foundational features, making DenseNet highly parameter-efficient and more robust on smaller datasets.

2.3.4. MobileNetV2

Designed for computational efficiency on resource-constrained devices, MobileNetV2 utilizes depthwise-separable convolutions and inverted residual blocks [36]. Depthwise-separable convolution splits the process into depthwise and subsequent 1×1 pointwise convolutions, drastically reducing the parameter count. Meanwhile, the inverted residual block follows a "narrow-wide-narrow" pattern, maintaining performance while minimizing memory requirements.

2.3.5. Uniform classifier head

On top of the frozen base architectures, we attached an identical head classifier to ensure a fair comparison.

- Global average pooling 2D (GAP): the GAP layer connected the feature maps to the fully connected layers while reducing dimensionality by averaging each map. This also served as a regularization technique to prevent overfitting [37].
- Dense layer (hidden): a fully connected layer with 1024 neurons used the rectified linear unit (ReLU) activation function, f(x)=max(0,x), to learn non-linear feature combination.
- Dense layer (output): the output layer contains 3 neurons (according to the number of classes) with SoftMax activation:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \; for \; i = 1, \dots, K$$

Algorithm 1. Transfer learning pipeline

```
Require: DATASET_PATH(string), EPOCHS(integer), models_list(dictionary)
Ensure: Trained models and performance comparison results
```

```
                                    ⊳Setup and data preparation
Load COCO dataset
Parse JSON annotations to create DataFrames
Extract class names and count classes
Validate dataset completeness

                                    ⊳Data preprocessing
Configure image size (224×224) and batch size (32)
Create data generatorss with augmentation for training
Apply only rescaling for validation/test sets

function CreateModel(base_model_fn, name, num_classes)
   Load pretrained model (ImageNet weights)
   Freeze base model layers
   Add GlobalAveragePooling2D layer
   Add Dense layer (1024 units, Relu activation)
   Add output layer (num_classes, SoftMax activation)
   Compile with Adam optimizer and categorical crossentropy
   return model
end function

                                     ⊳Training and evaluation pipeline
models←{VGG16, DenseNet121, MobileNetV2, ResNet50}
all_results←{}
for each model in models do
   Create model using transfer learning
   Train model for specified epochs
   Evaluate on test set
   Generate classification report
   Save trained model and results
end for

                                     ⊳Results processing and analysis
Parse model results from saved files
Extract performance metrics (accuracy, precision, recall, F1-score)
Create comparison DataFrame
Identify best performing model based in F1-score

                                     ⊳ Performance visualization
Generate comparison charts for each epoch
Create performance trend analysis
Generate confusion matrices
Export results
```

2.4. Experimental scenario

This experiment was designed for a fair performance comparison among the four different model architectures. The primary variable we investigated was the effect of training duration on model performance at three intervals: 10, 50, and 100 epochs. Each combination of model and target epoch was trained and evaluated under identical conditions. The entire training process was run from scratch for each experiment to ensure the independence of the results. Table 3 summarizes the hyperparameters that were held constant throughout all experiments. By maintaining consistency across these parameters, any performance differences could be directly attributed to the capabilities of the base architecture and the impact of the training duration.

Table 3. Hyperparameter configuration experiment

Parameter	Value/method
Image input size	224×224×3
Base architecture	VGG16, ResNet50, DenseNet121, and MobileNetV2
Optimizer	Adam
Learning rate	0.001
Batch size	32
Loss function	Categorical crossentropy

2.5. Evaluation metrics

To evaluate the predictive performance of each classification model, this study used a series of standard metrics derived from a confusion matrix. The confusion matrix is a table that visualizes model performance by comparing the predicted and actual labels on the test data [38]. The four fundamental components of the confusion matrix in the context of this research (with 'stunting' as the positive class) are:

– True positive (TP): the number of children who are actually stunted and were correctly predicted as stunted.

– True negative (TN): the number of children who are not stunted (e.g., 'healthy') and were correctly predicted as not stunted.
– False positive (FP) (type I error): the number of children who are not stunted but were incorrectly predicted as stunted.
– False negative (FN) (type II error): the number of children who are actually stunted but were incorrectly predicted as not stunted (a missed case).

Based on these components, this study used four primary metrics to assess model performance:

– Accuracy

Measures the proportion of all correct predictions. This metric provides a general overview of performance but can be misleading on imbalanced datasets.

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

– Precision

Measures the reliability of a positive prediction. High precision for the 'stunting' class indicates that if the model identifies a child as stunted, that prediction is highly likely to be correct.

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

– Recall

Measures the model's ability to identify all positive cases. High recall for the 'stunting' class is critical, as it indicates the model can find most of the children who are genuinely stunted with few missed cases.

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

– F1-score

The harmonic mean of precision and recall. This metric is crucial for imbalanced datasets because it requires a balance between both precision and recall. A high F1-score indicates a model that is not only accurate but also comprehensive in its identification.

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

Given the significant class imbalance in this dataset, accuracy alone is insufficient. In medical applications such as stunting detection, reducing false negatives is often considered more critical than reducing false positives. Therefore, the recall and F1-score metrics serve as more relevant evaluation indicators [38]. To report representative scores, this study used precision, recall, and F1-Score based on a weighted average, which calculates the metric for each class according to the proportion of its samples [39].

3. RESULTS AND DISCUSSION

This section presents the results of the comparative experiments. The performance of the four model architectures (VGG16, DenseNet121, MobileNetV2, and ResNet50) was evaluated at three training duration intervals: 10, 50, and 100 epochs.

3.1. Performance comparison between architectures

The evaluation results of the four architectures on the test set are presented in Table 4. This table summarizes the accuracy, precision, recall, and F1-score metrics for each combination of model and training duration. The results show that peak performance was achieved at the 10-epoch training duration. The VGG16 model at 10 epochs reached the highest accuracy of 91.8%, while MobileNetV2 at 10 epochs recorded the highest F1-score of 90.4%. In contrast, ResNet50 consistently demonstrated the lowest performance among the four architectures tested.

Table 4. Evaluation results

Epochs	10				50				100			
Model	Accuracy	F1-score	Precision	Recall	Accuracy	F1-score	Precision	Recall	Accuracy	F1-score	Precision	Recall
DenseNet121	0.8830	0.8713	0.8728	0.8830	0.9064	0.8961	0.8944	0.9064	0.9064	0.8903	0.8750	0.9064
MobileNetV2	0.9123	0.9038	0.8988	0.9123	0.8830	0.8649	0.8488	0.8830	0.8889	0.8769	0.8734	0.8889
ResNet50	0.7602	0.7455	0.7469	0.7602	0.7602	0.7440	0.7287	0.7602	0.6491	0.5914	0.7209	0.6491
VGG16	0.9181	0.8993	0.8821	0.9181	09064	0.8880	0.8723	0.9064	0.8772	0.8625	0.8556	0.8772

This finding reveals an interesting phenomenon: a simpler architecture like VGG16 can outperform more modern and complex architectures such as ResNet50. This indicates that for this specific malnutrition classification dataset, the relevant visual features are likely fundamental, and do not require a highly complex architectural structure. The complexity of ResNet50, which was designed for large-scale datasets, appears to be a disadvantage on this limited dataset, making it more prone to overfitting. The superiority of VGG16 suggests an optimal balance between model complexity and dataset size, where an architecture appropriately scaled to the data's capacity exhibits better generalization. DenseNet121 and MobileNetV2 demonstrated distinct performance characteristics. MobileNetV2, true to its design, proved to be a well-balanced model, achieving the highest F1-score at 10 epochs. This makes it a suitable choice for applications requiring computational efficiency without sacrificing performance. Meanwhile, DenseNet121 exhibited the most consistent learning curve, as its performance did not degrade drastically at higher epochs. This stability is likely due to its design, which is inherently more resistant to overfitting compared to the other models. Figure 4 (in Appendix) illustrates the performance comparison of four models across varying epoch counts. Specifically, Figure 4(a) visualizes the model performance at 10 epochs, where VGG16 achieves the highest scores in term accuracy and recall metrics relative to the other models. Figure 4(b) presents the evaluation results at 50 epochs, highlighting the stable performance of DenseNet121. Finally, Figure 4(c) provides a comparison at 100 epochs, revealing a significant decline in the ResNet50 model's scores compared to its performance at 10 and 50 epochs.

3.2. Analysis of learning trends and overfitting

The learning trend for each architecture as the number of epochs increased is displayed in Figure 5. The graphs show that three models (VGG16, MobileNetV2, and ResNet50) experienced a performance decline after 10 epochs, which indicates overfitting during longer training durations. Conversely, DenseNet121 showed a stable learning curve. The performance degradation in the three models after 10 epochs confirms the onset of overfitting. As the models continued to train, they began to "memorize" specific details of the training data that could not be generalized to new, unseen test data. This phenomenon was amplified by the models' architectural complexity and the limited size of the dataset. This result has significant practical implications: for the task of malnutrition status detection using transfer learning on similar datasets, a shorter training time is not only more computationally efficient but also yields a more accurate model. The stability of DenseNet121 can be attributed to its feature reuse mechanism, which makes it more robust against overfitting.

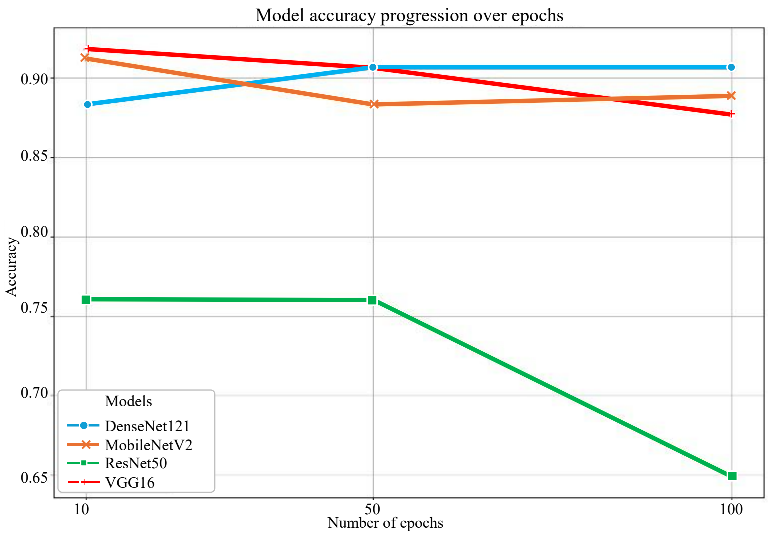


Figure 5. Model accuracy progression over epochs

Achieving optimal performance in a short duration (10 epochs) suggests that in practical applications, the process of retraining the model can be performed quickly and efficiently. This is beneficial when the model needs to be updated with new data, as it does not require substantial computational resources to reach peak performance. In additions a summary of the F1-score performance and consistency for each model is presented in heatmap of the models (Figure 6) and Box Plot chart (Figure 7). The heatmap visually confirms that the highest scores were concentrated at the 10-epoch mark, while the box plot illustrates that the performance distribution of ResNet50 was consistently lower than the other architectures. The heatmap shows that

MobileNetV2 achieved the highest performance at 10 epochs, while DenseNet121 reached its peak at 50 epochs. Furthermore, the confusion matrix analysis reveals that the 'Malnutrition' class is the most difficult to classify due to significant data scarcity, which hinders the model from learning sufficient discriminative features.

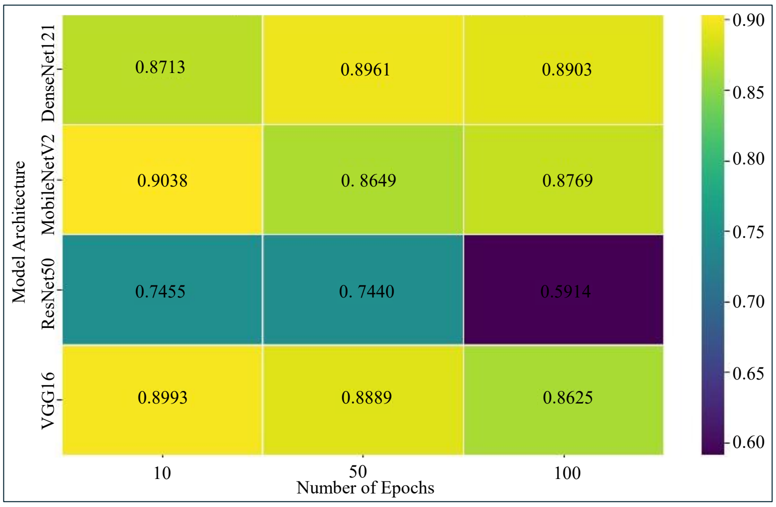


Figure 6. Heatmap of the models (F1-scores per epoch)

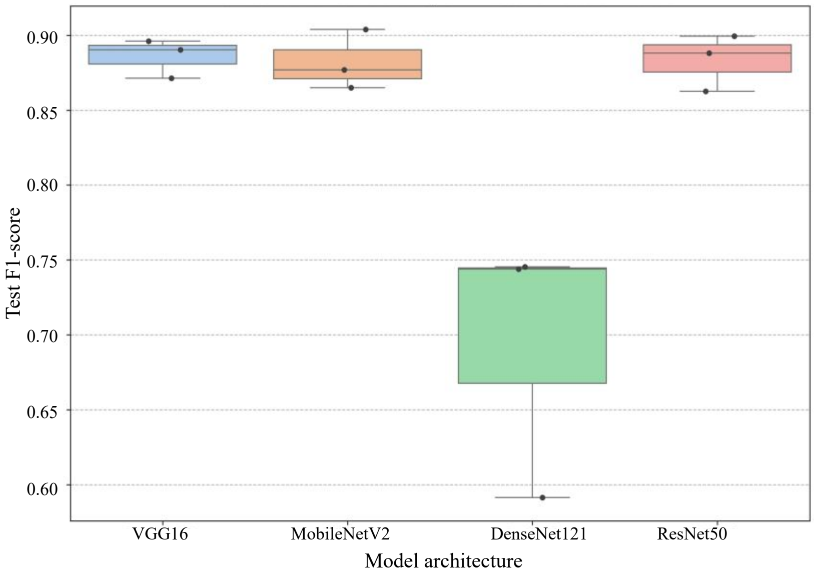


Figure 7. Distribution of F1-scores across the models

3.3. Analysis of the best model

Based on the comparative results, the VGG16 model trained for 10 epochs was selected for in-depth analysis, as it achieved the highest accuracy of 91.8%. The model's ability to discriminate between classes was analyzed using the receiver operating characteristic (ROC) curve, presented in Figure 8. The curve indicates that the model has strong predictive capabilities for the 'Healthy' and 'Stunting' classes, reflected by their high area under the curve (AUC) values. However, its performance on the 'Malnutrition' class proved more challenging, which is consistent with this class's minority status in the dataset.

To understand the model's limitations, we conducted an error analysis by visualizing misclassified test images, as shown in Figure 9. This visualization revealed that the model primarily misclassified images between the 'Stunting' and 'Healthy' classes, indicating a visual feature similarity in certain samples. The model's difficulty in correctly identifying the 'Malnutrition' category likely stems from the significant class imbalance in the dataset. Although the model achieved high overall accuracy, its performance on the minority class remains a primary obstacle. Meanwhile, errors in distinguishing between 'Stunting' and 'Healthy' suggest an inherent limitation of methods that rely solely on visual data. This finding directs future research towards integrating image data with clinical or anthropometric information to develop more comprehensive and reliable predictive models. Additionally, one potential approach to mitigate this issue is the use of generative adversarial networks (GANs) to create synthetic data, which can help balance the dataset and improve minority class recognition.

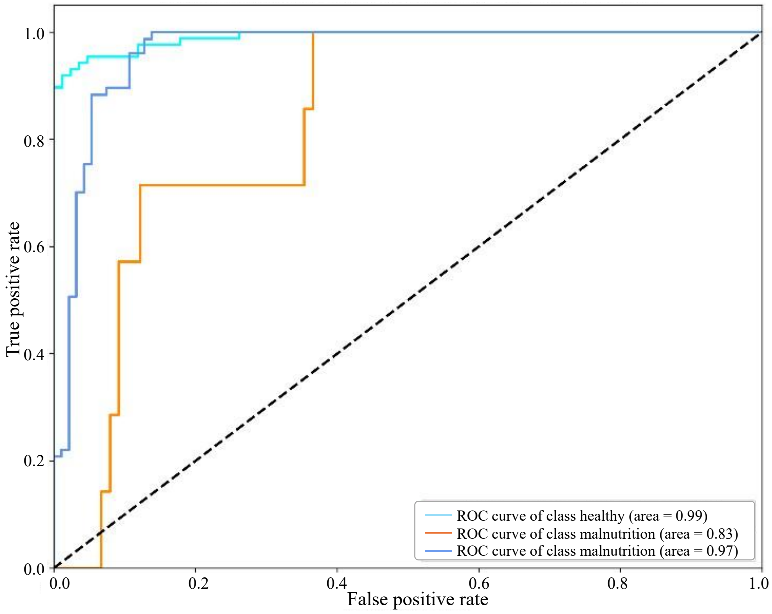


Figure 8. ROC curve of VGG16

Figure 9. Misclassified image by the model

The analysis of the best-performing model revealed two principal limitations: the impact of an imbalanced dataset and the ambiguity of visual features. The difficulty in classifying the 'Malnutrition' class highlights the need for better approaches to handling minority data, such as implementing synthetic data generation or oversampling techniques. Concurrently, misclassifications among the majority classes indicate that a unimodal, image-based approach has its limits. These two limitations form the foundation for future research, which is outlined in the conclusion.

4. CONCLUSION

This study developed a transfer learning-based method and compared the performance of several deep learning architectures to identify the most effective model for classifying malnutrition status. The experimental results showed that a simpler architecture, VGG16, trained for a short duration (10 epochs), and achieved the best performance with an accuracy of 91.8%. This finding is significant when contrasted with more modern and complex architectures like ResNet50, which exhibited the lowest performance and experienced degradation with longer training times. The pattern of longer training leading to reduced performance due to overfitting demonstrates that a brief training duration was most effective for this task. This research provides empirical evidence that challenges the common assumption that newer and deeper models are invariably superior. The results emphasize the need for an optimal balance between model complexity, dataset size, and task-specific characteristics. In the context of medical image classification with limited data, an architecture with an appropriate capacity proved more effective. This indicates that an efficient and accurate malnutrition screening system can be developed without relying on the largest models, a relevant finding for implementation in resource-constrained environments.

The primary constraints are the limited volume and imbalanced composition of the dataset and the high dependency on visual features from images. These limitations present opportunities for future research. For future work, we recommend using larger and more multimodal datasets that combine visual data with supplementary information, such as medical records or anthropometric data. This integration is expected to improve predictive accuracy and reduce bias. Furthermore, exploring advanced regularization techniques, data augmentation, and ensemble methods could potentially enhance model stability and performance.

This research supports the development of AI-based systems in the healthcare sector, particularly for the early detection of malnutrition in children. By leveraging a simple architecture, this study demonstrates that practical technology can be effectively implemented, especially in regions with limited resources. With further research, these findings are expected to contribute to more advanced and sustainable systems, ultimately improving public health services.

FUNDING INFORMATION

This work was supported by a grant from the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia as part of the Regular Fundamental Research program (125/C3/DT.05.00/PL/2025).

AUTHOR CONTRIBUTIONS STATEMENT

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

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ahmad Fauzi	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Haerul Yuda Aditiya		✓	✓	✓		✓	✓	✓	✓	✓	✓			
Maharina	✓		✓	✓	✓	✓	✓	✓		✓	✓	✓		✓
Sihabudin	✓					✓				✓		✓		
Muhammad Ansari Adista	✓				✓	✓				✓				
Iflan Nauval	✓			✓	✓	✓				✓				
Natasya Eka Nanda Sonia Puri	✓					✓				✓				
Candra Zonyfar	✓				✓					✓				✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The authors confirm that the data supporting the findings of this study are available within the article.

REFERENCES

[1] Kementerian Kesehatan RI, "Buku Saku Hasil Survei Status Gizi Indonesia (SSGI) 2022," 2022.
[2] Kementerian Kesehatan RI, "Laporan Hasil Survei Kesehatan Indonesia (SKI) 2023," 2023.
[3] M. Roberts, T. Tolar-Peterson, A. Reynolds, C. Wall, N. Reeder, and G. R. Mendez, "The Effects of Nutritional Interventions on the Cognitive Development of Preschool-Age Children: A Systematic Review," *Nutrients*, vol. 14, no. 3, pp. 1-15, Jan. 2022, doi: 10.3390/nu14030532.
[4] Ramli, K. E. Agho, K. J. Inder, S. J. Bowe, J. Jacobs, and M. J. Dibley, "Prevalence and risk factors for stunting and severe stunting among under-fives in North Maluku province of Indonesia," *BMC Pediatrics*, vol. 9, no. 1, pp. 1-10, Dec. 2009, doi: 10.1186/1471-2431-9-64.
[5] UNICEF, "Nutrition." https://www.unicef.org/indonesia/nutrition. (Accessed Aug. 09, 2025).
[6] T. Beal, A. Tumilowicz, A. Sutrisna, D. Izwardy, and L. M. Neufeld, "A review of child stunting determinants in Indonesia," *Maternal & Child Nutrition*, vol. 14, no. 4, pp. 1-10, Oct. 2018, doi: 10.1111/mcn.12617.
[7] C. R. Titaley, I. Ariawan, D. Hapsari, A. Muasyaroh, and M. J. Dibley, "Determinants of the Stunting of Children Under Two Years Old in Indonesia: A Multilevel Analysis of the 2013 Indonesia Basic Health Survey," *Nutrients*, vol. 11, no. 5, pp. 1-13, May 2019, doi: 10.3390/nu11051106.
[8] H. Torlesse, A. A. Cronin, S. K. Sebayang, and R. Nandy, "Determinants of stunting in Indonesian children: evidence from a cross-sectional survey indicate a prominent role for the water, sanitation and hygiene sector in stunting reduction," *BMC Public Health*, vol. 16, no. 1, p. 669, Dec. 2016, doi: 10.1186/s12889-016-3339-8.
[9] S. M. Janssen, Y. Bouzembrak, and B. Tekinerdogan, "Artificial Intelligence in Malnutrition: A Systematic Literature Review," *Advances in Nutrition*, vol. 15, no. 9, pp. 1-17, Sep. 2024, doi: 10.1016/j.advnut.2024.100264.
[10] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, Apr. 2018, doi: 10.1016/j.compag.2018.02.016.
[11] L. Chen, P. Chen, and Z. Lin, "Artificial Intelligence in Education: A Review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020, doi: 10.1109/ACCESS.2020.2988510.
[12] M. Akhtar and S. Moridpour, "A Review of Traffic Congestion Prediction Using Artificial Intelligence," *Journal of Advanced Transportation*, vol. 2021, pp. 1–18, Jan. 2021, doi: 10.1155/2021/8878011.
[13] W. Ullah and Q. Ali, "Role of artificial intelligence in healthcare settings: a systematic review," *Journal of Medical Artificial Intelligence*, vol. 8, pp. 24–24, Sep. 2025, doi: 10.21037/jmai-24-294.
[14] J. W. Goodell, S. Kumar, W. M. Lim, and D. Pattnaik, "Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis," *Journal of Behavioral and Experimental Finance*, vol. 32, Dec. 2021, doi: 10.1016/j.jbef.2021.100577.
[15] I. K. Nti, A. F. Adekoya, B. A. Weyori, and O. Nyarko-Boateng, "Applications of artificial intelligence in engineering and manufacturing: a systematic review," *Journal of Intelligent Manufacturing*, vol. 33, no. 6, pp. 1581–1601, Aug. 2022, doi: 10.1007/s10845-021-01771-6.
[16] D. Paul, G. Sanap, S. Shenoy, D. Kalyane, K. Kalia, and R. K. Tekade, "Artificial intelligence in drug discovery and development," *Drug Discovery Today*, vol. 26, no. 1, pp. 80–93, Jan. 2021, doi: 10.1016/j.drudis.2020.10.010.
[17] D. Kirk, E. Kok, M. Tufano, B. Tekinerdogan, E. J. M. Feskens, and G. Camps, "Machine Learning in Nutrition Research," *Advances in Nutrition*, vol. 13, no. 6, pp. 2573–2589, Nov. 2022, doi: 10.1093/advances/nmac103.
[18] J. R. Khan, J. H. Tomal, and E. Raheem, "Model and variable selection using machine learning methods with applications to childhood stunting in Bangladesh," *Informatics for Health and Social Care*, vol. 46, no. 4, pp. 425–442, Dec. 2021, doi: 10.1080/17538157.2021.1904938.
[19] F. H. Bitew, C. S. Sparks, and S. H. Nyarko, "Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia," *Public Health Nutrition*, pp. 1–12, Oct. 2021, doi: 10.1017/S1368980021004262.
[20] V. Rani and R. P. S. Rana, "Comprehensive Analytical Study of Existing Malnutrition Prediction Models," in *2025 Third International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, May 2025, pp. 411–416, doi: 10.1109/ICAISS61471.2025.11042152.
[21] S. Khare, S. Kavyashree, D. Gupta, and A. Jyotishi, "Investigation of Nutritional Status of Children based on Machine Learning Techniques using Indian Demographic and Health Survey Data," *Procedia Computer Science*, vol. 115, pp. 338–349, 2017, doi: 10.1016/j.procs.2017.09.087.
[22] E. R. Arumi, S. A. Subrata, and A. Rahmawati, "Implementation of Naïve bayes Method for Predictor Prevalence Level for Malnutrition Toddlers in Magelang City," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 2, pp. 201–207, Mar. 2023, doi: 10.29207/resti.v7i2.4438.
[23] H. Janawisuta, P. H. Gunawan, and Indwiarti, "Early Detection of Stunting in Indonesian Toddlers: A Machine Learning Approach," in *2024 International Conference on Data Science and Its Applications (ICoDSA)*, Jul. 2024, pp. 12–16, doi: 10.1109/ICoDSA62899.2024.10651637.
[24] M. Aryuni *et al.*, "Comparison of Nutritional Status Prediction Models of Children Under 5 Years of Age Using Supervised Machine Learning," in *Proceeding of the 3rd International Conference on Electronics, Biomedical Engineering, and Health Informatics: ICEBEHI 2022*, 2023, pp. 265–277, doi: 10.1007/978-981-99-0248-4_19.
[25] X. Hao, G. Zhang, and S. Ma, "Deep Learning," *International Journal of Semantic Computing*, vol. 10, no. 03, pp. 417–439, Sep. 2016, doi: 10.1142/S1793351X16500045.
[26] D. Jaswal, S. V, and K. P. Soman, "Image Classification Using Convolutional Neural Networks," *International Journal of Scientific and Engineering Research*, vol. 5, no. 6, pp. 1661–1668, Jun. 2014, doi: 10.14299/ijser.2014.06.002.
[27] D. R. E. Rajappan, S. Sanith, and S. Subbulakshmi, "Malnutrition Detection using Deep Learning Models," in *2023 4th IEEE Global Conference for Advancement in Technology (GCAT)*, Oct. 2023, pp. 1–8, doi: 10.1109/GCAT59970.2023.10353327.
[28] A. R. Lakshminarayanan, P. B, R. V, S. Parthasarathy, A. A. A. Khan, and K. J.Sathick, "Malnutrition Detection using Convolutional Neural Network," in *2021 Seventh International Conference on Bio Signals, Images, and Instrumentation (ICBSII)*, Mar. 2021, pp. 1–5, doi: 10.1109/ICBSII51839.2021.9445188.
[29] D. C *et al.*, "Identification of malnutrition and prediction of BMI from facial images using real-time image processing and machine learning," *IET Image Processing*, vol. 16, no. 3, pp. 647–658, Feb. 2022, doi: 10.1049/ipr2.12222.
[30] Yunidar, Roslidar, Yusni, Nasaruddin, and FitriArnia, "CNN Performances for Stunting Face Image Classification," in *2024 International Conference on Electrical Engineering and Computer Science (ICECOS)*, Sep. 2024, pp. 89–94, doi: 10.1109/ICECOS63900.2024.10791288.
[31] E. A. Turjo and M. H. Rahman, "Assessing risk factors for malnutrition among women in Bangladesh and forecasting malnutrition using machine learning approaches," *BMC Nutrition*, vol. 10, no. 1, pp. 1-25, Feb. 2024, doi: 10.1186/s40795-023-00808-8.

[32] B. Rao, M. Rashid, M. G. Hasan, and G. Thunga, "Machine Learning in Predicting Child Malnutrition: A Meta-Analysis of Demographic and Health Surveys Data," *International Journal of Environmental Research and Public Health*, vol. 22, no. 3, pp. 1-15, Mar. 2025, doi: 10.3390/ijerph22030449.

[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint*, 2014, doi: 10.48550/arXiv.1409.1556.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[35] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

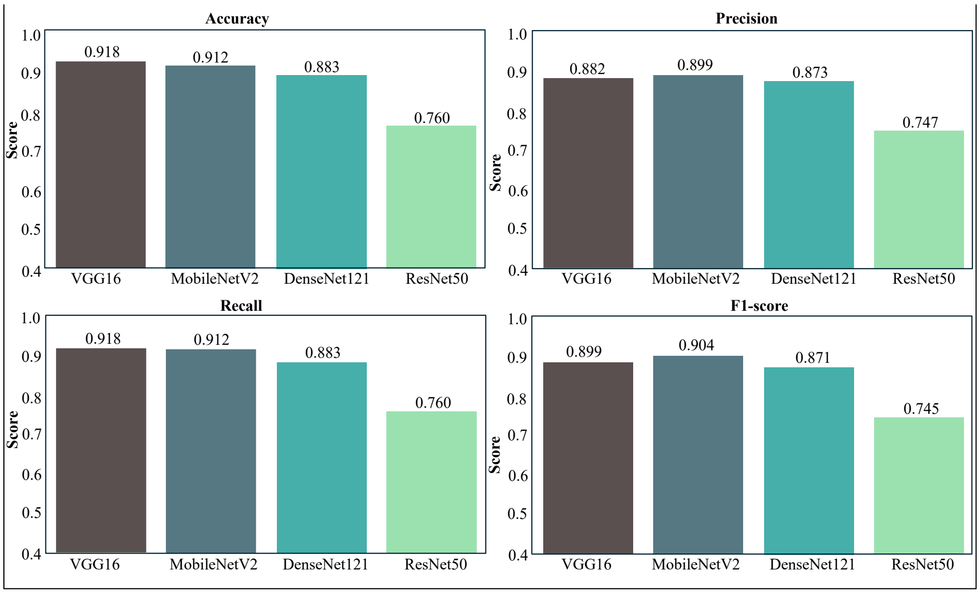
[36] V. Nair and G. E. Hinton, "Rectified linear units improve Restricted Boltzmann machines," in *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, 2010, pp. 807–814, doi: 10.5555/3104322.3104425.

[37] M. Lin, Q. Chen, and S. Yan, "Network In Network," *arXiv preprint*, 2014, doi: 10.48550/arXiv.1312.4400.

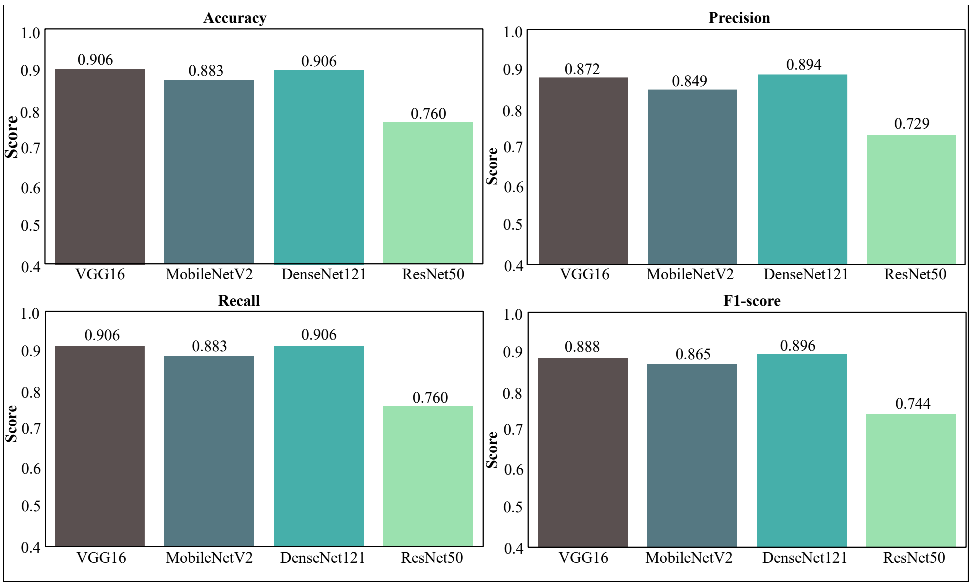
[38] T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLOS ONE*, vol. 10, no. 3, pp. 1-21, Mar. 2015, doi: 10.1371/journal.pone.0118432.

[39] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv preprint*, 2011, doi: 10.48550/arXiv.2010.16061.

APPENDIX

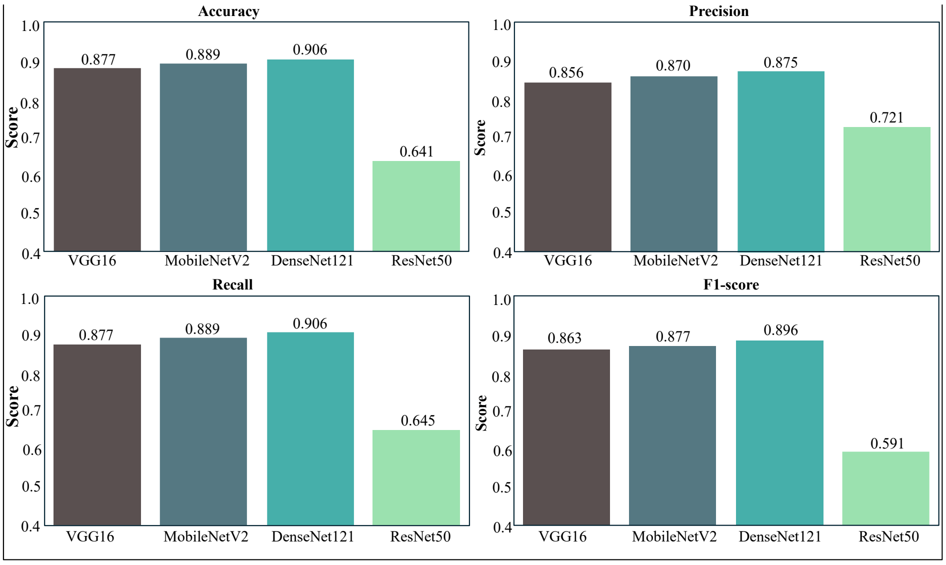


(a)



(b)

Figure 4. Comparison of model performance based on; (a) 10 and (b) 50



(c)

Figure 4. Comparison of model performance based on; (c) 100 epochs *(continued)*

BIOGRAPHIES OF AUTHORS

Ahmad Fauzi holds a bachelor's degree in informatics (S.Kom.) from STMIK ROSMA Karawang Indonesia, a master's degree in informatics (M.Kom.) from STTI Benarif Indonesia Jakarta, and a Doctor of Information Technology from Gunadarma University Jakarta. Currently active as a lecturer in the undergraduate informatics program at the Universitas Buana Perjuangan Karawang. Research interest in artificial intelligence, digital image processing, expert systems, and data science. He can be contacted at email: afauzi@ubpkarawang.ac.id.

Haerul Yuda Aditiya is currently pursuing a Bachelor's degree in Computer Science at Universitas Buana Perjuangan Karawang. His research focuses on computer vision and deep learning, focusing on advanced neural network architectures for image classification. He can be contacted at email: haerulyudaaditiya@gmail.com.

Maharina holds a master information system degree from The School of Informatics and Computer Management LIKMI, Indonesia in 2020. She received her B.Sc. (Computer Science) from Singaperbangsa Karawang University, Indonesia in 2013. She is currently a researcher as well as lecturer at Information System Program in Faculty Computer Science, Universitas Buana Perjuangan Karawang, Indonesia. Her research interest includes machine learning, natural language processing, and kansei engineering. She can be contacted at email: maharina@ubpkarawang.ac.id.

Sihabudin received a Bachelor's degree from Universitas Singaperbangsa Karawang, Indonesia, in 1995, two Master's degrees in Management, first from STIE IPWIJA Jakarta in 2000 and the second from STIM IMMI Jakarta in 2005, and a Doctor from Universitas Pasundan Bandung, Indonesia, in 2019. His research interests are business and financial law, corporate governance, financial technology (FinTech), and micro, small, and medium enterprises (MSME) financing. He can be contacted at email: sihabudin@ubpkarawang.ac.id.

Muhammad Ansari Adista is a prominent neurologist and medical educator affiliated with Syiah Kuala University in Aceh, Indonesia. His research are medical education, neurology, neuro-oncology, and community health. He can be contacted at email: ansariadista@usk.ac.id.

Iflan Nauval is a physician, lecturer, and researcher affiliated with Syiah Kuala University, Aceh. His research interests include public health, obesity, child health, humanitarian studies, and clinical nutrition. He can be contacted at email: iflandr@usk.ac.id.

Natasya Eka Nanda Sonia Puri is undergraduate student in Informatics Engineering Programs, Faculty of Computer Science, Universitas Buana Perjuangan Karawang. She has a keen research interest in data science in healthcare. She can be contacted at email: natasyaenspuri@gmail.com.

Candra Zonyfar earned his Ph.D. from the Department of Computer Science and Electronics Convergence Engineering at Sun Moon University, South Korea. He is currently a postdoctoral researcher at Sun Moon University and a lecturer at Universitas Buana Perjuangan Karawang. His research interests include artificial intelligence, machine learning, and bioinformatics. He can be contacted at email: candrazonyfar@sunmoon.ac.kr.