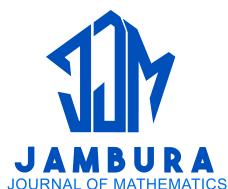


Performance Comparison of VGG16, MobileNetV2, and InceptionV3 Convolutional Neural Networks in Classifying Facial Dermatological Conditions

Fadilah Karamun Nisaa Nadiyah et al.



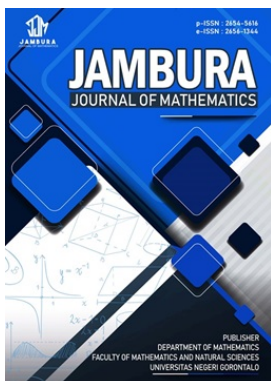
Volume 7, Issue 2, Pages 144–150, August 2025

Received 29 June 2025, Revised 5 August 2025, Accepted 8 August 2025, Published 14 August 2025

To Cite this Article : F. K. N. Nadiyah, N. N. Alifah, S. Nurdiati, E. Khatizah, and M. K. Najib, "Performance Comparison of VGG16, MobileNetV2, and InceptionV3 Convolutional Neural Networks in Classifying Facial Dermatological Conditions", *Jambura J. Math*, vol. 7, no. 2, pp. 144–150, 2025, <https://doi.org/10.37905/jjom.v7i2.33082>

© 2025 by author(s)

JOURNAL INFO • JAMBURA JOURNAL OF MATHEMATICS

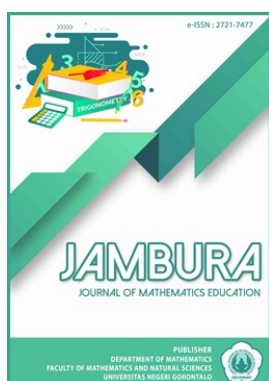


	Homepage	:	http://ejurnal.ung.ac.id/index.php/jjom/index
	Journal Abbreviation	:	Jambura J. Math.
	Frequency	:	Biannual (February and August)
	Publication Language	:	English (preferable), Indonesia
	DOI	:	https://doi.org/10.37905/jjom
	Online ISSN	:	2656-1344
	Editor-in-Chief	:	Hasan S. Panigoro
	Publisher	:	Department of Mathematics, Universitas Negeri Gorontalo
	Country	:	Indonesia
	OAI Address	:	http://ejurnal.ung.ac.id/index.php/jjom/oai
	Google Scholar ID	:	iWLjgaUAAAAJ
	Email	:	info.jjom@ung.ac.id

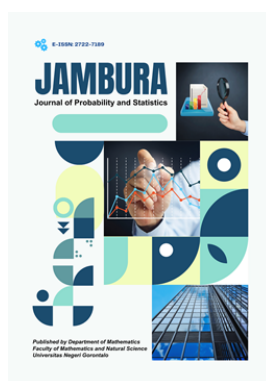
JAMBURA JOURNAL • FIND OUR OTHER JOURNALS



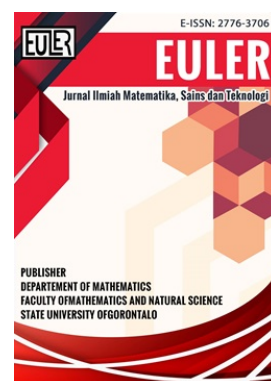
Jambura Journal of Biomathematics



Jambura Journal of Mathematics Education



Jambura Journal of Probability and Statistics



EULER : Jurnal Ilmiah Matematika, Sains, dan Teknologi



Performance Comparison of VGG16, MobileNetV2, and InceptionV3 Convolutional Neural Networks in Classifying Facial Dermatological Conditions

Fadilah Karamun Nisaa Nadiyah¹, Nayla Nur Alifah¹, Sri Nurdiati¹, Elis Khatizah¹, Mohamad Khoirun Najib^{1,*}

¹School of Data Science, Mathematics, and Informatics, IPB University, Bogor 16680, Indonesia

ARTICLE HISTORY

Received 29 June 2025

Revised 5 August 2025

Accepted 8 August 2025

Published 14 August 2025

KEYWORDS

Classification

CNN

MobileNetV2

VGG16

InceptionV3

ABSTRACT. This study investigates the performance of three convolutional neural network (CNN) architectures (VGG16, MobileNetV2 and InceptionV3) in classifying two common facial dermatological conditions: acne and dark spots. A dataset of 235 facial skin images was augmented, then used to train and evaluate each model using standard classification metrics such as accuracy, precision, recall, and F1-score. The results demonstrate that MobileNetV2 achieved the highest classification accuracy of 93.13% while maintaining a relatively low computational cost. The model exhibited perfect precision (1.00) for the acne class and a high recall of 0.99 for the dark spots class, indicating its strong capability in accurately and sensitively identifying both lesion types. All three models demonstrated acceptable classification performance for both acne and dark spots classes, as evidenced by their precision, recall, and F1-scores exceeding 70%. This indicates that each model was capable of capturing relevant discriminative features of both lesion types.



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. *Editorial of JJoM:* Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B. J. Habibie, Bone Bolango 96554, Indonesia.

1. Introduction

Facial skin problems, such as acne and dark spots, are two common conditions that can be identified through visual images and serve as relevant targets for automated classification [1]. These two conditions exhibit distinct visual characteristics, including color, texture, and distribution on the skin surface, which digital cameras can capture. Given these differences, image-based classification approaches offer an effective solution for fast and accurate identification [2].

With the advancement of Artificial Intelligence (AI) technology, deep learning methods, particularly CNNs, have been widely applied in various fields, including medical image classification and facial skin analysis [3, 4]. CNNs are capable of recognizing patterns and features in images, making them suitable for the automated identification and classification of facial skin problems. Several CNN architectures have been developed to improve efficiency and accuracy in image processing [5].

VGG16, MobileNetV2, and InceptionV3 are three popular CNN architectures that vary in terms of depth, computational efficiency, and model complexity [6–8]. VGG16, introduced by Simonyan and Zisserman, is known for its deep and uniform architecture using sequential small 3×3 convolutional filters, which offers high classification accuracy but requires substantial computational resources and training time [9]. MobileNetV2 is a lightweight and optimized model for mobile and embedded devices, utilizing inverted residuals and linear bottlenecks to sig-

nificantly reduce the number of parameters and memory usage without compromising accuracy [10, 11]. In comparison, InceptionV3, an improved version of GoogLeNet, enhances computational efficiency and accuracy by combining multiple convolution filter sizes in a single module, allowing the network to extract features at various scales simultaneously [12].

Based on these considerations, this study aims to classify facial skin problems, specifically acne and dark spots, using CNN methods. The research compares the performance of three different CNN architectures: VGG16, MobileNetV2, and InceptionV3. This study is expected to contribute to the fields of skincare and dermatological health, particularly in developing AI-based systems that can help with early diagnosis and provide more accurate skincare recommendations [13, 14].

The research questions addressed in this study are as follows: (1) How can deep learning technology be utilized to assist in the classification of facial skin problems? (2) How do different deep learning architectures compare in their performance for classifying facial skin conditions? (3) Which deep learning model demonstrates the best performance based on classification evaluation metrics? To answer these questions, the objectives of this study are: (1) to implement the CNN method for classifying facial skin issues, particularly acne and dark spots; (2) to analyze and compare the performance of three CNN architectures: VGG16, MobileNetV2, and InceptionV3 based on evaluation metrics such as accuracy, precision, recall, and F1-score; and (3) to determine the most optimal CNN architecture for facial skin classification

*Corresponding Author.

based on model testing and evaluation results. This study is expected to provide several benefits, including expanding knowledge and serving as a reference for future researchers interested in deep learning applications in medical imaging and dermatological issues, contributing to the development of pattern recognition techniques for facial skin imagery using various CNN architectures, and offering a foundation for the development of AI-based applications in the fields of beauty and dermatology.

2. Methods

2.1. Data source and type

The dataset used in this study consists of digital images representing two types of facial skin conditions: acne and dark spots. In detail, the dataset comprises a total of 235 images, including 140 labeled as acne and 95 labeled as dark spots. The data were obtained as secondary data from an external source, namely the Kaggle platform, which provides a variety of open-access datasets for research and development purposes. All images underwent manual review to ensure their relevance and visual clarity for classification tasks. The class imbalance present in the dataset was also considered during preprocessing and evaluation stages to minimize classification bias. To address the class imbalance, data augmentation was applied in a balanced manner to enhance minority class representation. Although augmentation techniques were applied uniformly, care was taken to generate a comparable number of samples per class in the training set. Furthermore, evaluation metrics sensitive to imbalance such as Precision, Recall, and F1-score reported alongside Accuracy to provide a more comprehensive assessment of model performance.

2.2. Research procedure

This research was conducted through several main stages. Initially, planning and preparation of the data for classification were carried out. The data were collected from the external source Kaggle.

To improve data variation and model generalization, data augmentation techniques were performed including random rotations (± 20 degrees), horizontal flipping, width and height shifts (up to 10% of the image size), shearing transformations (up to 0.2 radians), and brightness adjustment within the range of 70% to 130% of the original intensity. These transformations were implemented using the ImageDataGenerator module in TensorFlow and applied uniformly to all training images to synthetically increase dataset diversity and mitigate overfitting. An example of the augmented data shown in Figure 1. The dataset was then divided into two subsets: training data and testing data, which were used for model training and evaluation respectively.

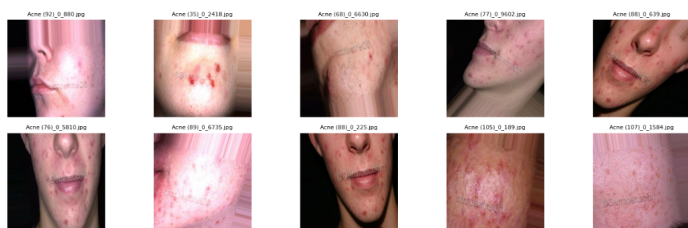


Figure 1. Example of augmentation

Table 1. Layer and hyperparameters VGG16

Layer	Hyperparameter
Conv2D	64 filters, 3×3 filter size, ReLU, same padding, input shape (224, 224, 3)
Conv2D	64 filters, 3×3 filter size, ReLU, same padding
MaxPool2D	2×2 pool size, 2 strides
Conv2D	128 filters, 3×3 filter size, ReLU, same padding
Conv2D	128 filters, 3×3 filter size, ReLU, same padding
MaxPool2D	2×2 pool size, 2 strides
Conv2D	256 filters, 3×3 filter size, ReLU, same padding
Conv2D	256 filters, 3×3 filter size, ReLU, same padding
Conv2D	256 filters, 3×3 filter size, ReLU, same padding
MaxPool2D	2×2 pool size, 2 strides
Conv2D	512 filters, 3×3 filter size, ReLU, same padding
Conv2D	512 filters, 3×3 filter size, ReLU, same padding
Conv2D	512 filters, 3×3 filter size, ReLU, same padding
MaxPool2D	2×2 pool size, 2 strides
Flatten	
Dense	4096 Units, ReLU activation
Dropout	0.5 Neurons
Dense	4096 Units, ReLU activation
Dropout	0.5 Neurons
Dense	1 Units, Sigmoid activation

VGG16 is a CNN architecture developed by Simonyan and Zisserman, known for its deep and well-structured design using repeated 3×3 convolutional layers that allow for increased depth without significantly raising the number of parameters. The architecture proved highly successful in large-scale image classification tasks such as ImageNet and has inspired various compression strategies [9]. In this research, VGG16 contributes as a baseline network capable of capturing detailed spatial features, despite its limitations in size and training time. Its integration supports achieving high classification accuracy with consistent generalization in facial skin condition analysis.

Table 1 outlines the layer configuration of the VGG16 architecture used in this study. VGG16, introduced by Simonyan and Zisserman, employs a consistent design pattern using repeated 3×3 convolutional layers and 2×2 max-pooling layers [15]. This architecture increases network depth without significantly expanding the number of parameters [9]. The model is composed of five convolutional blocks followed by pooling, then three fully connected layers with dropout to prevent overfitting. VGG16 was selected for its strong ability to extract spatial features, making it suitable for classifying dermatological images such as acne and dark spots.

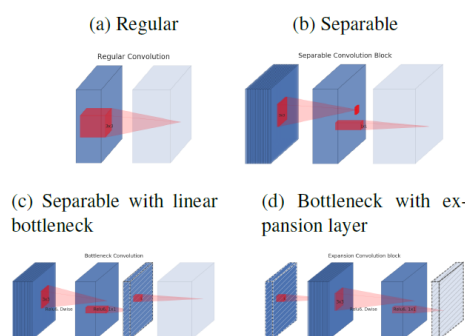


Figure 2. Evolution of separable convolution blocks [10]

Table 2. Layer and hyperparameters InceptionV3

Layer	Hyperparameter
Conv2D	32 filters, 3 × 3 filter size, 2 strides, ReLU activation, valid padding, input shape (224, 224, 3)
Conv2D	32 filters, 3 × 3 filter size, ReLU activation, valid padding
Conv2D	64 filters, 3 × 3 filter size, ReLU activation, same padding
MaxPool2D	3 × 3 pool size, 2 strides
Conv2D	80 filters, 1 × 1 filter size, ReLU activation, valid padding
Conv2D	192 filters, 3 × 3 filter size, ReLU activation, valid padding
MaxPool2D	3 × 3 pool size, 2 strides
Inception block 1	
Branch 1 × 1 = Conv2D	64 filters, 1 × 1 filter size, ReLU activation, same padding
Branch 5 × 5 = Conv2D	48 filters, 1 × 1 filter size, ReLU activation, same padding
Branch 5 × 5 = Conv2D	64 filters, 5 × 5 filter size, ReLU activation, same padding
Branch 3 × 3db1 = Conv2D	64 filters, 1 × 1 filter size, ReLU activation, same padding
Branch 3 × 3db1 = Conv2D	96 filters, 3 × 3 filter size, ReLU activation, same padding
Branch 3 × 3db1 = Conv2D	96 filters, 3 × 3 filter size, ReLU activation, same padding
Branch pool = MaxPool2D	3 × 3 pool size, 1 strides, same padding
Branch pool = Conv2D	32 filters, 1 × 1 filter size, ReLU activation, same padding
GlobalAveragePool2D	
Dense	1024 Units, ReLU activation
Dropout	0.2 Neurons
Dense	1 Units, Sigmoid activation

MobileNetV2 is a CNN architecture optimized for high efficiency on mobile devices, introducing the concepts of inverted residuals and linear bottlenecks, where shortcut connections link narrow layers and non-linear activations are removed in critical points to preserve essential information. The model is lightweight in terms of parameters and computational operations, making it ideal for resource-constrained environments [10]. In this study, MobileNetV2 is applied as a representation of efficient network design, capable of performing real-time classification of facial skin conditions, thereby supporting AI-based solutions deployable on portable or embedded systems.

The architecture of the bottleneck residual block used in MobileNetV2. The design incorporates depthwise separable convolutions and a linear bottleneck structure, where non-linear activation functions are excluded from the bottleneck output to preserve essential information. This configuration enables efficient feature extraction while minimizing computational cost and parameter usage, making it suitable for mobile and embedded applications [10].

InceptionV3 is an advanced evolution of GoogLeNet, featuring a more complex structure that combines multiple convolutional filter sizes within a single module to capture features at various scales [16]. The model incorporates several optimization techniques, including factorization, asymmetric convolutions, and batch normalization, enhancing both computational efficiency and classification performance [17]. In this research, InceptionV3 serves as a balanced architecture combining depth, accuracy, and efficiency, making it suitable for precise facial skin condition classification without significantly compromising computational performance. All three architectures were implemented in this study to classify two facial skin conditions—acne and dark spots—to determine which architecture performs most optimally.

Table 2 presents the detailed architecture of InceptionV3, an advanced evolution of GoogLeNet. InceptionV3 introduces multi-scale processing through the Inception module, which ap-

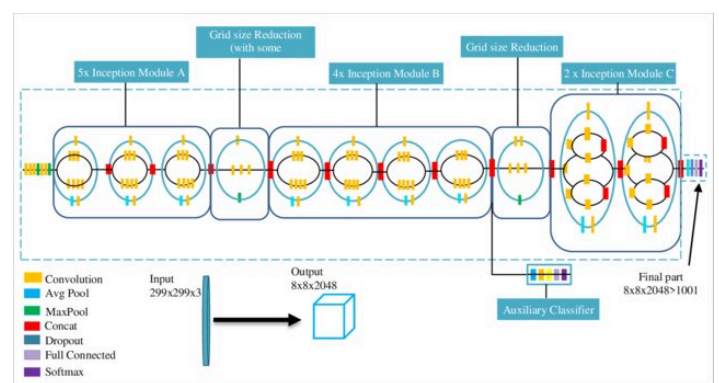


Figure 3. InceptionV3 architecture [18]

plies multiple convolution filter sizes (1 × 1, 3 × 3, and 5 × 5) in parallel within each block [17]. This enables the model to capture features at various scales in a single pass. Additionally, optimization techniques such as factorization and batch normalization are incorporated to enhance computational efficiency and training stability. In this study, InceptionV3 was used to evaluate the performance of a complex yet efficient architecture in classifying facial skin conditions [17].

Table 3. Hyperparameter CNN

Hyperparameter	Value
Optimizer	Adam
Loss function	Binary cross entropy
Epoch	30
Batch size	32
Learning rate	0.0001
Callbacks	Early stopping with patience 5

Model training was performed with various hyperparameter configurations, as described in Table 3. Adam optimizer was selected due to its popularity and advantages, particularly

its usability, computational efficiency and effectiveness in handling large data sets and numerous parameters. The binary cross-entropy loss function was utilized since the classification task involved two categories: acne and dark spots. Training occurred for a maximum of 30 epochs; however, an early stopping mechanism with a patience of 5 epochs was implemented to reduce the potential for overfitting. A model checkpoint was also created to retain only the model with the best performance. Various batch sizes (5, 10, 20, and 40) were first tested, with a batch size of 32 yielding the most advantageous results. The learning rate was set to 0.0001 to ensure fine-grained weight updates during training. Although smaller values slow down convergence, this rate provided sufficient stability and prevented overshooting, especially when fine-tuning deep pre-trained networks on a small dataset. In the MobileNetV2 architecture, we applied regularization techniques with the objective of mitigating overfitting and improving generalization capability model's on unseen validation data. Regularization serves to prevent the model from excessively fitting the training data, including noise or idiosyncratic patterns, thereby promoting more robust performance across diverse and unseen inputs [19].

To evaluate the performance of the classification models, four commonly used evaluation metrics were employed, namely: Accuracy, Precision, Recall, and F1-Score. These metrics are essential for assessing the model's ability to correctly classify facial acne conditions, especially in the presence of class imbalance. The definitions and formulas of each metric are provided below [20]:

- **Accuracy** measures the proportion of total correct predictions:

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

- **Precision** indicates the proportion of true positives among all predicted positives:

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

- **Recall** (also known as Sensitivity) reflects the proportion of actual positives that were correctly identified:

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

- **F1-Score** is the harmonic mean of Precision and Recall:

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

3. Results and Discussion

All CNN architectures were trained and tested on Google Collab and Python 3, with Tesla T4 GPU as hardware accelerator. The accuracy of the CNN architectures is shown in Table 4. The effectiveness of the three CNN models was assessed using accuracy, precision, recall, and F1-score, as detailed in Table 4–Table 5.

The VGG16 model showed a quick yet variable increase in training accuracy. Training was stopped at epoch 18 based on early stopping criteria to avoid overfitting show in Figure 4. However, the validation performance showed significant instability, indicating the model's limited generalization capability on

Table 4. Comparison accuracy and loss model

Model	Accuracy (%)	Loss
VGG16	80.90	0.3961
MobileNetV2	93.13	0.4536
InceptionV3	79.18	0.4250

Table 5. Metric report

	Precision	Recall	F1-Score	Support
VGG16				
Acne	0.88	0.79	0.83	276
Dark spots	0.73	0.84	0.78	190
MobileNetV2				
Acne	1.00	0.89	0.94	276
Dark spots	0.86	0.99	0.92	190
InceptionV3				
Acne	0.82	0.83	0.82	276
Dark spots	0.75	0.74	0.74	190

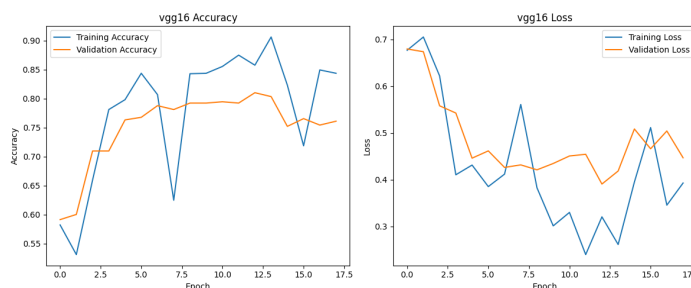


Figure 4. Accuracy and loss VGG16



Figure 5. Example prediction VGG16

unseen data. This was further supported by the divergence between training and validation loss curves, where the validation loss increased while training loss continued to decrease, highlighting overfitting tendencies. In Table The performance metrics reveal that the model achieved higher precision and F1-score for the acne class, while the darkspot class yielded a higher recall value shown in Table 5. This suggests that while the model was more accurate and consistent in predicting acne, it was relatively more sensitive in detecting darkspot lesions, albeit with a trade-off in prediction precision. This is further evidenced by the confusion matrix in Figure 10, where the model correctly classified 218 out of 276 acne images and 159 out of 190 darkspot images. This is further supported by the prediction examples in Figure 5, where the model shows greater accuracy in classifying acne images compared to darkspot. Correct predictions are more frequent for acne, while darkspot cases are often misclassi-

fied, aligning with the higher precision and F1-score for acne and higher recall for darkspot.

In contrast, MobileNetV2 exhibited the most robust and stable performance among the evaluated models shown in Figure 6. The training and validation accuracy curves demonstrated consistent convergence with minimal fluctuations, indicating effective learning and strong generalization capability. Although slight fluctuations in training loss were observed, the validation loss remained stable throughout the training process, suggesting resilience to overfitting. The final test accuracy in Table 4 reached 93.13%, the highest among all models tested, accompanied by highest validation loss (0.4536). These results are reinforced by superior precision, recall, and F1-score values across both classes, confirming the model’s balanced performance in both sensitivity and specificity shown in Table 5. The confusion matrix in Figure 10 further highlights the model’s predictive precision, with 244 out of 276 acne images and 190 out of 190 darkspot images correctly classified. This show also in Figure 7, where all example images are accurate with confidence almost 1.00 for both classes.

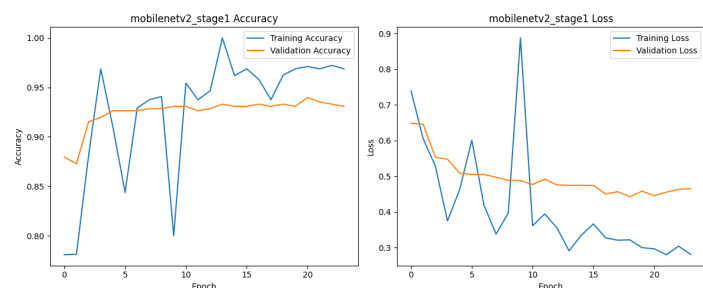


Figure 6. Accuracy and loss MobileNetV2



Figure 7. Example prediction MobileNetV2

In the InceptionV3 model, the training accuracy showed a rapid increase during the initial epochs and stopped at epoch 10, followed by a relatively stable trend with low fluctuations, suggesting that the model continuously adapted to the training data. Both training and validation loss curves remained close and stable, indicating good convergence behavior and generalization capacity. Architecture successfully minimized training errors, as shown in Figure 8. According to Table 5 it attained more balanced metrics for acne and for dark spots. These results are in alignment with confusion matrix in Figure 10, showing that the model correctly classified 228 out of 276 acne images and 141 out of 190 darkspot images. This pattern is consistent with the classification metrics, where higher precision, recall, and F1-scores were observed for the acne class compared to darkspot, demonstrating that the model was more confident and accurate in iden-

tifying acne than darkspot. It show in Figure 9, where the model accurately classified the images with different confidence level, although some misclassifications were present.

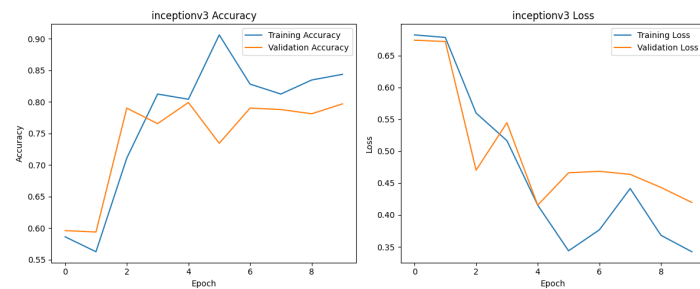


Figure 8. Accuracy and loss InceptionV3



Figure 9. Example prediction InceptionV3

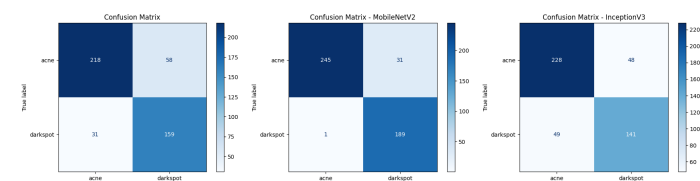


Figure 10. Confusion matrix architectures CNN

From a computational efficiency perspective, the three models present distinct characteristics in terms of parameter size and resource requirements shown in Table 6. VGG16 had the highest computational cost, consisting of approximately 134 million parameters and a model size of 527 MB. This extensive architecture demands substantial memory and processing power, rendering it less suitable for deployment in real-time or edge-based applications, particularly on mobile or embedded devices [21].

Table 6. Comparison parameters VGG16, MobileNetV2 and InceptionV3

Architectures	Total Parameters	Size
VGG16	134,264,641	512.18 MB
MobileNetV2	2.427.713	9.26 MB
InceptionV3	691,905	2.64 MB

MobileNetV2 was designed with computational efficiency as a primary objective. With only 2.4 million parameters and a compact size of 9.26 MB, it achieves significantly reduced memory usage without compromising classification performance.

This efficiency is made possible through the use of depthwise separable convolutions and inverted residual connections, which drastically reduce the number of operations required for inference [22]. MobileNetV2 not only achieved the highest accuracy in this study but also demonstrated superior scalability for deployment in resource-constrained environments.

InceptionV3, presents an intermediate profile with respect to both performance and complexity. Although it contains fewer parameters than VGG16, and a notably smaller model size (2.64 MB), its performance was inconsistent in this study. Despite its theoretical strength in capturing multi-scale features via inception modules [23], the model struggled to generalize well on this specific dataset. This suggests that architectural complexity alone does not guarantee superior performance, particularly when dealing with specialized medical image data.

These findings underscore the importance of balancing model complexity and computational cost, especially for medical imaging applications where real-time processing, storage limitations, and deployment on portable devices are critical considerations [24, 25]. In this regard, MobileNetV2 emerges as the most favorable model, offering a practical trade-off between accuracy and efficiency.

4. Conclusion

This study has evaluated and compared the performance of three CNN architectures—VGG16, MobileNetV2, and InceptionV3—for the classification of skin lesion images, specifically acne and dark spots. Among the trained models, MobileNetV2 outperformed the others in terms of accuracy, stability, and computational efficiency, achieving a classification accuracy of 93.13% with precision, recall, f-1 score highestst for acne class an dark spots class. That ability to generalize well across both types of lesion was evident in the confusion matrix and performance metrics.

VGG16, despite its deeper architecture, showed moderate accuracy and higher computational cost, making it less practical for real-time applications. InceptionV3, although lightweight, showed less consistent performance and struggled with generalization in validation data. These findings highlight the critical importance of selecting models that strike a balance between accuracy and resource efficiency, particularly for mobile or embedded medical diagnostic tools.

Future work may explore the integration of attention mechanisms, ensemble methods, or fine-tuned transfer learning strategies to further enhance classification performance in more complex and imbalanced datasets.

Author Contributions. F. K. N. Nadiyah: Data curation, software, investigation, writing—original draft preparation. N. N. Alifah: Data curation, software, investigation, writing—original draft preparation. S. Nurdianti: Methodology, supervision, writing—review and editing. E. Khatizah: Formal analysis, validation, funding acquisition, writing—review and editing. M. K. Najib: Conceptualization, supervision, project administration, visualization, writing—review and editing, funding acquisition. All authors discussed the results, contributed to the writing of the manuscript, and approved the final version.

Acknowledgement. The authors would like to thank the support of the School of Data Science, Mathematics, and Informatics, IPB University, for

providing the research environment.

Funding. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest. The authors declare that there is no conflict of interest regarding the publication of this paper.

Data availability. The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

- [1] Y. S. Bekmirzaev S, Oh S, “Rethnet: Object-by-object learning for detecting facial skin problems,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0.
- [2] C. Kim, W. Lee, Y. Hong, Y. Han, and M. Jang, “Skin lesion classification using hybrid convolutional neural network with edge, color, and texture information,” *Applied Sciences*, vol. 13, no. 9, 2023, doi: [10.3390/app13095497](https://doi.org/10.3390/app13095497).
- [3] A. Mahalakshmi, C. Padmavathy, N. Priya, G. Priyanka, and B. Archana, “Skin cancer detection using deep learning,” in *2024 10th International Conference on Electrical Energy Systems (ICEES)*. IEEE, August 2024, pp. 1–4.
- [4] W. Chen, C. Chin, T. Tsai, and M. Chin, “Facial skin image classification system using convolutional neural networks deep learning algorithm,” in *2018 9th International Conference on Awareness Science and Technology (iCAST)*. IEEE, 2018, pp. 51–55, doi: [10.1109/ICAWST.2018.8517246](https://doi.org/10.1109/ICAWST.2018.8517246).
- [5] —, “Facial skin image classification system using convolutional neural networks deep learning algorithm,” in *2018 9th International Conference on Awareness Science and Technology (iCAST)*. IEEE, 2018, pp. 51–55, doi: [10.1109/ICAWST.2018.8517246](https://doi.org/10.1109/ICAWST.2018.8517246).
- [6] P. Kaur and R. Kumar, “Performance analysis of convolutional neural network architectures over wireless capsule endoscopy dataset,” *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 1, 2024, doi: [10.11591/eei.v13i1.5858](https://doi.org/10.11591/eei.v13i1.5858).
- [7] R. Klangbunrueang, W. Chansanam, P. Pookduang, and T. Lunrasri, “Ai-powered lung cancer detection: Assessing vgg16 and cnn architectures for ct scan image classification,” *Informatics*, vol. 12, no. 1, p. 18, 2025, doi: [10.3390/informatics12010018](https://doi.org/10.3390/informatics12010018).
- [8] J. Zheng and M. Du, “Study on tomato disease classification based on leaf image recognition based on deep learning technology,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 4, 2023, doi: [10.14569/IJACSA.2023.0140463](https://doi.org/10.14569/IJACSA.2023.0140463).
- [9] H. Qassim, A. Verma, and D. Feinzimer, “Compressed residual-vgg16 cnn model for big data places image recognition,” in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, January 2018, pp. 169–175.
- [10] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4510–4520.
- [11] S. Noor, M. Saleem, N. Ismat, H. Inam, A. Siddiqui, and H. Minhas, “Eye on the future: Deep learning with mobilenetv2 for early eye disease diagnosis,” in *2024 Global Conference on Wireless and Optical Technologies (GCWOT)*. IEEE, 2024, pp. 1–7, doi: [10.1109/GCWOT63882.2024.10805707](https://doi.org/10.1109/GCWOT63882.2024.10805707).
- [12] W. L. H. L. S. W. Z. X. D. M. Z. S. W. M. . M. H. Dai, M., “Pepper leaf disease recognition based on enhanced lightweight convolutional neural networks,” *Frontiers in Plant Science*, vol. 14, 2023, doi: [10.3389/fpls.2023.1230886](https://doi.org/10.3389/fpls.2023.1230886).
- [13] S. Tantry, T. Lal, and S. Rangaswamy, “Skin disease classification using deep learning,” *National Academy Science Letters*, 2024, doi: [10.1007/s40009-024-01523-z](https://doi.org/10.1007/s40009-024-01523-z).
- [14] S. Kalal, G. Ubale, O. Kamble, A. Kashyap, S. Mhamane, and S. Pande, “Evaluating the performance of pre-trained cnns for multi-class skin disease detection,” in *2024 Fourth International Conference on Multimedia Processing, Communication & Information Technology (MPCIT)*. IEEE, 2024, pp. 303–308, doi: [10.1109/MPCIT62449.2024.10892670](https://doi.org/10.1109/MPCIT62449.2024.10892670).
- [15] G. Jablonski, “Vgg based unsupervised anomaly detection in multivariate time series,” in *Proceedings of the International Conference on Computational Science (ICCS)*. Springer, 2020, pp. 1287–1296, doi: [10.1007/978-3-030-50936-1_107](https://doi.org/10.1007/978-3-030-50936-1_107).
- [16] V. Semwal, S. Singh, S. Challa, and N. Dua, “Inception inspired cnn-gru hybrid network for human activity recognition,” *Multimedia Tools and Applications*, vol. 82, pp. 5369–5403, 2022, doi: [10.1007/s11042-021-11885-x](https://doi.org/10.1007/s11042-021-11885-x).

- [17] C. S. Anumol, "Advancements in cnn architectures for computer vision: a comprehensive review," in *2023 Annual International Conference on Emerging Research Areas: International Conference on Intelligent Systems (AICERA/ICIS)*. IEEE, November 2023, pp. 1–7.
- [18] R. Jain, P. Nagrath, G. Kataria, V. S. Kaushik, and D. J. Hemanth, "Pneumonia detection in chest x-ray images using convolutional neural networks and transfer learning," *Measurement*, vol. 165, p. 108046, 2020, doi: [10.1016/j.measurement.2020.108046](https://doi.org/10.1016/j.measurement.2020.108046).
- [19] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, [Online]. Available: <http://www.deeplearningbook.org>.
- [20] J. Tohka and M. van Gils, "Evaluation of machine learning algorithms for health and wellness applications: A tutorial," *Computers in Biology and Medicine*, vol. 132, p. 104324, 2021, doi: [10.1016/j.combiomed.2021.104324](https://doi.org/10.1016/j.combiomed.2021.104324).
- [21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, [Online]. Available: <https://arxiv.org/abs/1409.1556>.
- [22] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, [Online]. Available: <https://arxiv.org/abs/1704.04861>.
- [23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015, [Online]. Available: <https://arxiv.org/abs/1512.00567>.
- [24] S. Rajaraman, S. K. Antani, M. Poostchi, K. Silamut, M. A. Hossain, R. J. Maude, S. Jaeger, and G. R. Thoma, "Pre-trained convolutional neural networks as feature extractors toward improved malaria parasite detection in thin blood smear images," *PeerJ*, vol. 6, p. e4568, Apr. 2018, doi: [10.7717/peerj.4568](https://doi.org/10.7717/peerj.4568).
- [25] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," 2018, [Online]. Available: <https://arxiv.org/abs/1807.11164>.