Journal of Technology and Informatics (JoTI)
Vol. 8, No. 2, Bulan Oktober Tahun 2026
P-ISSN 2721-4842
E-ISSN 2686-6102

Comparative Evaluation of Deep Learning Architectures for Early Skin Cancer Detection Using Dermatoscopic Images

Aurel Elviolita Putri[1*], Agung Mulyo Widodo[2], Nizirwan Anwar[3], Binastya Anggara Sekti[4]
[1,2,3,4] Department of Computer Science, Universitas Esa Unggul, Jakarta, Indonesia
e-mail: aurelviolita00@students.esaunggul.ac.id[1*], agung.mulyo@esaunggul.ac.id[2], nizirwan.anwar@esaunggul.ac.id[3], anggara@esaunggul.ac.id[4]

Article Information

Article History:
Received : 9 January 2026
Revised : 14 July 2026
Accepted : 30 July 2026
Published : 29 October 2026

***Correspondence:**
aurelviolita00@students.esaunggul.ac.id

Keywords:
Convolutional Neural Network, Dermoscopy, HAM10000, Skin Cancer, Transfer Learning

Copyright ©2026 by Author.
Published by Universitas Dinamika.

This is an open access article under the CC BY-SA license.

10.37802/joti.v8i2.1344

Journal of Technology and Informatics (JoTI)
P-ISSN 2721-4842
E-ISSN 2686-6102
https://e-journals.dinamika.ac.id/index.php/joti

***Abstract*:**
Early identification of skin cancer is essential for improving diagnostic reliability and patient outcomes. Although convolutional neural networks (CNNs) have demonstrated promising performance in dermoscopic image classification, comparative evaluations on imbalanced multiclass datasets remain challenging. This study compares three transfer learning-based CNN architectures—VGG16, ResNet50, and MobileNetV2—for seven-class skin lesion classification using the imbalanced HAM10000 dataset. All models were evaluated using the same dataset partitions, image resolution, and performance metrics, while architecture-specific augmentation and loss-function strategies were retained. In particular, VGG16 used MixUp, ResNet50 used CutMix, VGG16 and ResNet50 employed Focal Loss, and MobileNetV2 used label smoothing. Performance was assessed using accuracy, precision, recall, F1-score, confusion matrices, and ROC–AUC. ResNet50 achieved the best overall performance, with 85.71% accuracy and a weighted F1-score of 0.8502, ahead of VGG16 (82.52%, 0.8172) and MobileNetV2 (79.32%, 0.7739). ResNet50 also showed the highest recall for melanoma and dermatofibroma among the three tested configurations, although minority-class sensitivity remained limited. These results favor ResNet50 under the evaluated training configurations; however, differences in augmentation and loss functions prevent attributing performance gaps solely to network architecture.

INTRODUCTION

Skin cancer is now a significant public health issue because its incidence continues to increase worldwide, adding considerably to the global cancer burden [1]. Long-term exposure to ultraviolet (UV) radiation continues to be a major risk factor for melanoma and non-melanoma skin cancers, particularly in tropical countries such as Indonesia [2]. In many regions, delayed diagnosis is still common because of limited public awareness and unequal access to dermatological services, reducing the opportunity for timely treatment and lowering patient survival rates [3]. Consequently, increasing the reliability of early skin cancer identification has become a key objective in clinical practice.

The conventional diagnostic process primarily relies on visual examination and dermoscopy performed by experienced dermatologists [4], [5]. Although these approaches are widely accepted in clinical settings, diagnostic outcomes may vary depending on the clinician's expertise and experience. Such variability becomes more evident in healthcare facilities with limited specialist availability, increasing the likelihood of delayed or inconsistent diagnoses. These challenges have encouraged the development of computer-assisted diagnostic tools that can offer consistent and impartial support for clinical decisions. Recent developments in artificial intelligence, especially deep learning, have improved automated medical image analysis. Convolutional neural networks (CNNs) can learn discriminative representations directly from dermoscopic images for multiclass skin lesion classification [6], [7]. The architectures evaluated in this study have different design principles: ResNet50 uses residual connections, while MobileNetV2 uses inverted residual blocks and linear bottlenecks [8], [9]. Previous skin lesion classification research has also examined deep CNN architectures and ensemble methods [10], [11].

Comparing convolutional neural network (CNN) architectures for skin lesion classification remains challenging because existing studies employ different datasets, preprocessing strategies, data augmentation techniques, and evaluation protocols [12–15]. These methodological differences limit the comparability and reproducibility of reported results, making it difficult to determine whether performance improvements arise from architectural advantages or differences in experimental design. Previous studies using the HAM10000 dataset have explored various deep learning approaches, including transfer learning with DenseNet169 and ResNet50 [10] and optimized CNNs incorporating data augmentation [16]. Other investigations have proposed specialized CNN architectures, such as DeepSkin CNN [12], and ensemble-based approaches to improve skin lesion classification [11]. Although these studies demonstrate the potential of deep learning for automated dermatological image analysis, differences in model configurations and evaluation strategies highlight the need for more consistent comparative assessments.

Importantly, the substantial class imbalance in the HAM10000 dataset presents an additional challenge, particularly for minority classes such as melanoma (MEL), actinic keratosis (AKIEC), and dermatofibroma (DF). In this context, overall accuracy alone may provide an incomplete representation of model performance because high classification accuracy can coexist with poor recognition of underrepresented lesion categories [16–19]. Therefore, class-wise metrics, including precision, recall, and F1-score, are essential for evaluating the ability of CNN architectures to recognize lesions across different diagnostic categories.

To address these limitations, this study presents a comparative evaluation of VGG16, ResNet50, and MobileNetV2 using the HAM10000 dataset [22] under a consistent experimental framework. Existing studies have investigated transfer learning models, optimized CNN architectures, ensemble-based methods [20], and wavelet-based approaches [21] for skin

lesion classification; however, variations in experimental settings make it difficult to determine whether reported performance differences arise from network architectures or differences in data preparation and training strategies. Accordingly, this study evaluates the three models using a shared dataset partitioning strategy, image dimensions, and evaluation protocol, while acknowledging that certain augmentation and loss-function choices differ across architectures, as described in the Methods section. Consequently, the results are interpreted as a comparison of the evaluated model-training configurations rather than an architecture-only ablation. Particular emphasis is placed on class-wise performance for melanoma (MEL), actinic keratosis (AKIEC), and dermatofibroma (DF) to assess the models' ability to recognize underrepresented lesion classes and provide a more comprehensive understanding of the trade-offs between classification performance and model efficiency.

METHOD

This quantitative experimental study evaluated three transfer learning-based convolutional neural network architectures—VGG16, ResNet50, and MobileNetV2—for multiclass skin lesion classification. Each model was assessed using the same HAM10000 train, validation, and test partitions and the same reporting metrics. The models were not trained with fully identical augmentation and loss functions; therefore, observed performance differences may reflect both architectural properties and training strategies. The workflow consisted of dataset preparation, image preprocessing, augmentation, model training, fine-tuning, evaluation, and cross-model analysis, as illustrated in Figure 1.

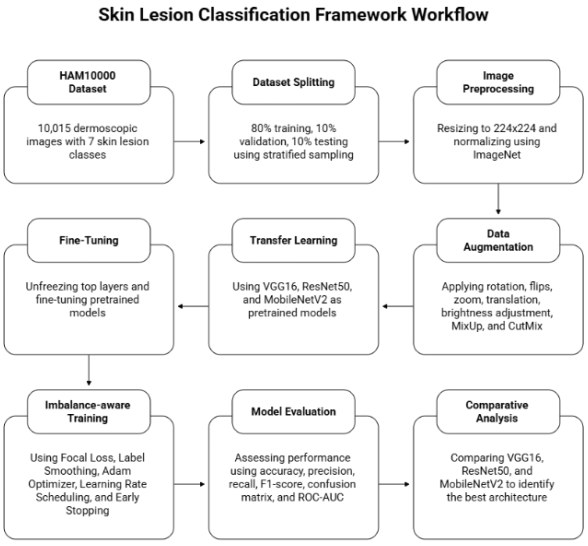

Figure 1. Overall workflow of the proposed methodology.

Figure 1 depicts the complete sequence implemented in this research, covering data preparation, image processing, model development, assessment, and cross-model analysis. The HAM10000 dataset was partitioned into training, validation, and test subsets through stratified sampling so that the initial class proportions were retained. Model performance was evaluated using accuracy, precision, recall, F1-score, confusion matrix, and ROC–AUC.

1. Dataset and Data Splitting

Table 1. Distribution of Skin Lesion Classes in the HAM10000 Dataset

Class	Abbreviation	Number of Images	Percentage
Melanocytic nevi	NV	6,705	66.95
Melanoma	MEL	1,113	11.11

Class	Abbreviation	Number of Images	Percentage
Benign keratosis-like lesions	BKL	1,099	10.97
Basal cell carcinoma	BCC	514	5.13
Actinic keratoses	AKIEC	327	3.27
Vascular lesions	VASC	142	1.42
Dermatofibroma	DF	115	1.15
Total	—	10,015	100.00

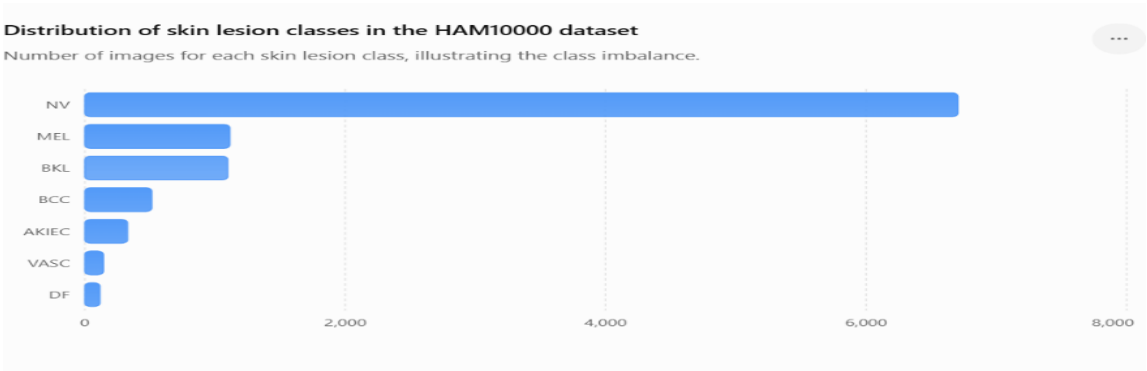

Figure 2. Distribution of skin lesion classes in the HAM10000 dataset.

Table 1 reports the number of images in each lesion category. Melanocytic nevi (NV) dominate the dataset, whereas dermatofibroma (DF) and vascular lesions (VASC) have relatively few samples. This imbalance can lead models to favor majority classes; consequently, class-wise metrics are necessary to assess the recognition of underrepresented categories.

2. Image Preparation and Training Augmentation

Before training, each dermoscopic image was scaled to 224 × 224 pixels, in accordance with the input dimensions required by the selected pretrained CNN models. Pixel intensities were standardized with the ImageNet preprocessing routine assigned to each backbone to maintain compatibility with pretrained weights. To improve generalization during training, the image pipeline included random rotations, horizontal and vertical flips, translations, zoom, and brightness adjustments. The additional augmentation methods were architecture-specific: MixUp was applied to VGG16, whereas CutMix was applied to ResNet50. Thus, the augmentation protocols were not identical across the three models, which should be considered when interpreting comparative performance [23].

3. CNN Architectures and Training

The three evaluated CNN architectures were initialized from ImageNet weights so that previously acquired visual features could be reused. For each architecture, the original classification layer was replaced with a customized classification head composed of Global Average Pooling, dense layers, dropout regularization, and a softmax output layer configured for seven-class prediction [24]. Fine-tuning was performed by making selected higher-level layers of the pretrained networks trainable, allowing the models to adapt features to the skin-lesion classification task. VGG16 and ResNet50 used Focal Loss to address class imbalance, whereas MobileNetV2 employed label smoothing. All models were trained with the Adam optimizer. Because different loss functions and augmentation techniques were used, performance differences cannot be attributed exclusively to network architecture.

4. Evaluation Metrics

Model performance was assessed exclusively on the independent testing dataset to prevent evaluation bias. Several complementary performance metrics were employed, including accuracy, precision, sensitivity, weighted F1, confusion matrices, and ROC–AUC. Accuracy was used to measure overall classification performance, whereas precision, sensitivity, and F1 supplied detailed evidence about prediction quality for each class. Confusion matrices were analyzed to identify common misclassification patterns among lesion categories, while ROC–AUC was used to evaluate the discriminative capability of each architecture across all classes. Collectively, these evaluation metrics provided a comprehensive assessment of model effectiveness, particularly for minority-class recognition under an imbalanced multiclass setting.

RESULTS AND DISCUSSION

This section reports the experimental findings and examines the relative performance of VGG16, ResNet50, and MobileNetV2 within a common experimental setting. Every evaluation used the independent test set to maintain an impartial comparison. Model performance was assessed using accuracy curves, ROC–AUC, confusion matrices, and class-wise evaluation metrics.

1. Overall Classification Performance

The relative behavior of the assessed CNN architectures is presented using the training and validation accuracy curves, ROC–AUC curves, confusion matrices, and classification reports.

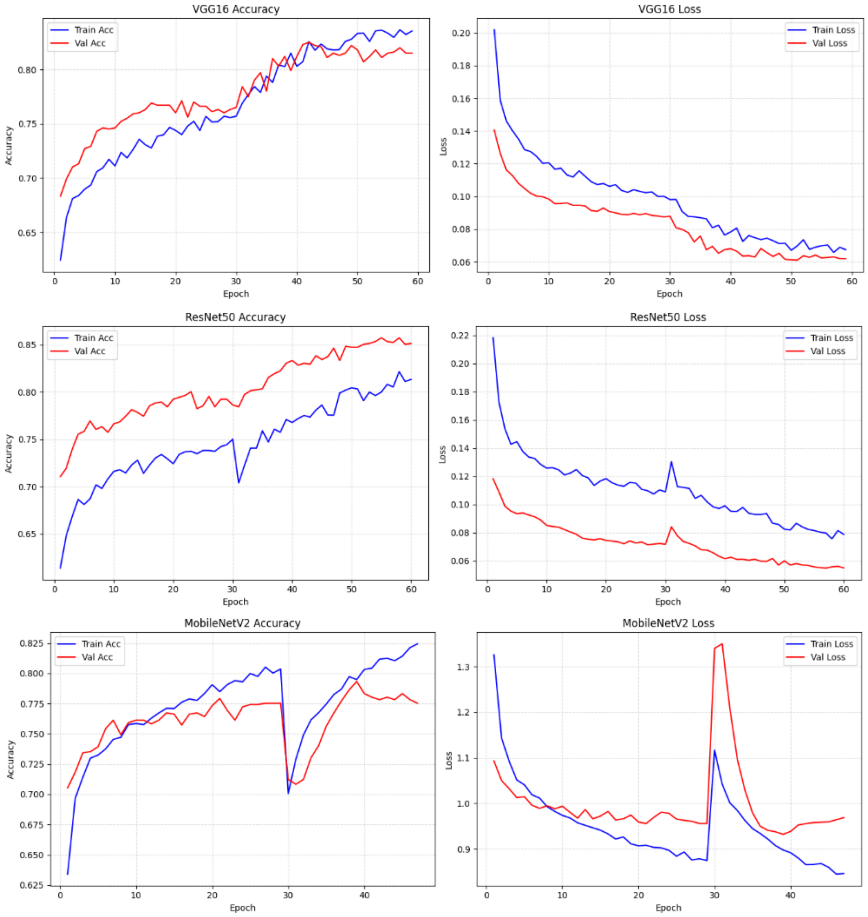

Figure 3. Training and validation accuracy and loss curves of the evaluated CNN architectures.

The learning profiles in Figure 3 show that ResNet50 achieved the most stable convergence during training and validation. Compared with the other architectures, its validation accuracy remained more consistent across epochs, suggesting stronger generalization capability. This behavior can be attributed to the residual learning mechanism, which facilitates gradient propagation and enables deeper feature extraction. Conversely, VGG16 exhibited indications of mild overfitting during the later training stages, while MobileNetV2 showed greater fluctuations in validation performance, implying reduced effectiveness in learning complex multiclass dermoscopic features. These observations are consistent with previous studies highlighting the advantages of residual learning for skin lesion classification [8], [10], [16], [25].

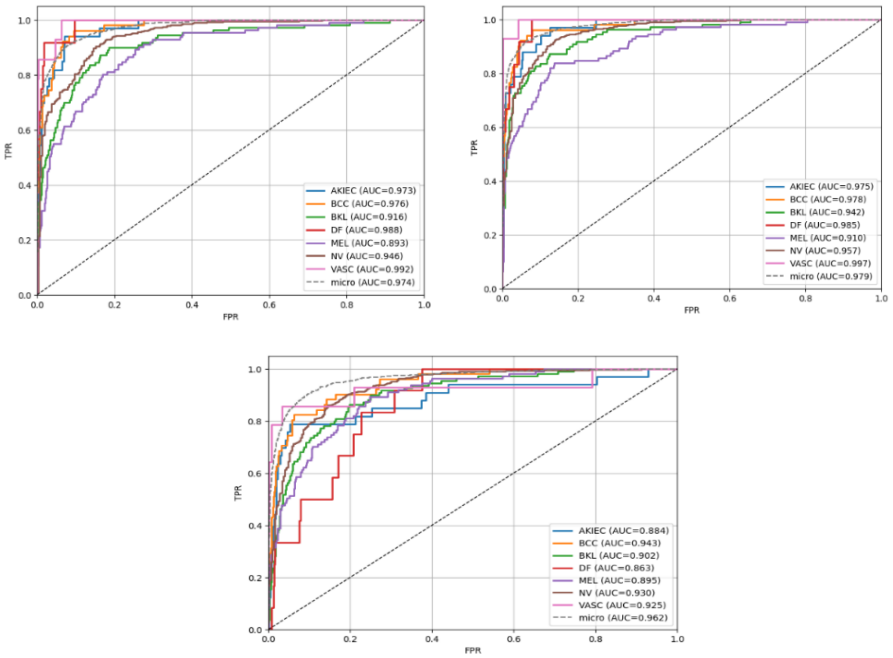

Figure 4. ROC–AUC curves of the evaluated CNN architectures.

The ROC–AUC analysis shown in Figure 4 further confirms the superiority of ResNet50. The model achieved consistently higher discriminative performance across most lesion categories, particularly melanoma (MEL), demonstrating better separation between classes. Although VGG16 produced competitive results for dominant lesion categories, its performance declined when class imbalance became more pronounced. MobileNetV2 achieved comparatively lower ROC–AUC values for minority classes, indicating greater difficulty in distinguishing underrepresented lesions within the dataset.

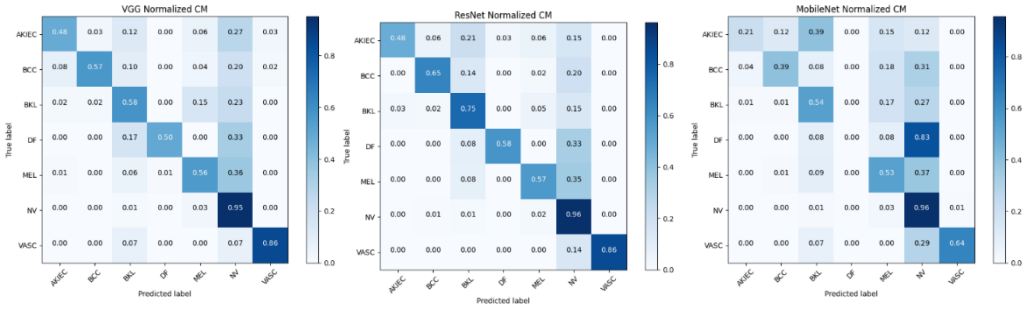

Figure 5. Normalized confusion matrices of the evaluated CNN architectures.

Additional insight is provided by the normalized confusion matrices in Figure 5. ResNet50 generated stronger diagonal patterns, indicating a higher proportion of correctly classified samples across almost all lesion categories. In contrast, VGG16 produced more confusion between melanoma (MEL) and melanocytic nevi (NV), whereas MobileNetV2 struggled to recognize several minority classes, particularly dermatofibroma (DF). Despite the common tendency of all models to misclassify minority lesions as the majority NV class, ResNet50 maintained the most balanced prediction performance when evaluated on the same test set [16], [26].

Table 2. Classification Report of VGG16

	Precision	**Recall**	**F1-Score**	**Support**
AKIEC	0.6957	0.4848	0.5714	33
BCC	0.8529	0.5686	0.6824	51
BKL	0.7033	0.5818	0.6368	110
DF	0.8571	0.5000	0.6316	12
MEL	0.5905	0.5586	0.5741	111
NV	0.8774	0.9507	0.9126	670
VASC	0.8000	0.8571	0.8276	14
Accuracy			**0.8252**	1001
Macro Avg	0.7681	0.6431	0.6909	1001
Weighted avg	0.8179	0.8252	0.8172	1001

Table 2 shows that VGG16 obtained 82.52% overall accuracy, together with a weighted F1 of 0.8172 and a macro F1 of 0.6909. The marked difference between the weighted and macro F1 values suggests that the model remained biased toward majority classes, particularly NV, while exhibiting reduced sensitivity for less represented classes, including AKIEC, DF, and MEL. These results suggest that although VGG16 provides competitive overall classification performance, its robustness decreases under severe class imbalance.

Table 3. Classification Report of ResNet50

	Precision	**Recall**	**F1-Score**	**Support**
AKIEC	0.8000	0.4848	0.6038	33
BCC	0.8049	0.6471	0.7174	51
BKL	0.7217	0.7545	0.7378	110
DF	0.8750	0.5833	0.7000	12
MEL	0.7500	0.5676	0.6462	111
NV	0.8944	0.9612	0.9266	670
VASC	0.9231	0.8571	0.8889	14
Accuracy			**0.8571**	1001
Macro Avg	0.8242	0.6937	0.7458	1001
Weighted avg	0.8519	0.8571	0.8502	1001

Table 3 demonstrates that ResNet50 produced the strongest aggregate result, recording 85.71% accuracy, a weighted F1 of 0.8502, and a macro F1 of 0.7458. Compared with VGG16 and MobileNetV2, ResNet50 produced more balanced precision and recall across both majority and minority classes. The smaller gap between weighted and macro F1-scores indicates improved robustness against class imbalance, confirming the effectiveness of residual learning for multiclass skin lesion classification.

Table 4. Classification Report of MobileNetV2

	Precision	**Recall**	**F1-Score**	**Support**
AKIEC	0.7000	0.2121	0.3256	33
BCC	0.7143	0.3922	0.5063	51
BKL	0.6344	0.5364	0.5813	110
DF	0.0000	0.0000	0.0000	12
MEL	0.5268	0.5315	0.5291	111
NV	0.8591	0.9552	0.9046	670
VASC	0.6923	0.6429	0.6667	14
Accuracy			**0.7932**	1001
Macro Avg	0.5895	0.4672	0.5019	1001
Weighted avg	0.7723	0.7932	0.7739	1001

Table 4 shows that MobileNetV2 recorded 79.32% overall accuracy, with weighted and macro F1 values of 0.7739 and 0.5019. The substantial separation between weighted and macro F1 values suggests substantial bias toward the majority class. Although MobileNetV2 maintained competitive performance for NV, it showed poor recognition of minority classes, including complete failure to identify DF samples. These findings suggest that lightweight architectures may sacrifice discriminative capability when applied to highly imbalanced multiclass dermoscopic datasets.

2. Quantitative Performance Comparison

Table 5. Overall Performance Comparison of CNN Architectures

Model	Accuracy (%)	Precision	Recall	Weighted F1	Macro F1	95% CI
VGG16	82.52	0.8179	0.8252	0.8172	0.6909	80.17–84.87
ResNet50	**85.71**	**0.8519**	**0.8571**	**0.8502**	**0.7458**	**83.54–87.88**
MobileNetV2	79.32	0.7723	0.7932	0.7739	0.5019	76.81–81.83

Table 5 summarizes the measured classification performance. ResNet50 achieved the highest overall accuracy (85.71%) and weighted F1-score (0.8502), followed by VGG16 (82.52% and 0.8172) and MobileNetV2 (79.32% and 0.7739). ResNet50 also attained the highest macro F1-score (0.7458), indicating stronger average class-wise performance in this evaluation. The displayed 95% confidence intervals describe uncertainty in the individual accuracy estimates; they do not, on their own, establish statistically significant differences between models.

3. Minority Class Performance and Clinical Implications

Table 6. Comparison of Minority-Class Recall Across the Evaluated CNN Architectures

Model	MEL	AKIEC	DF
VGG16	55.86	48.48	50.00
ResNet50	56.76	48.48	58.33
MobileNetV2	53.15	21.21	0.000

Table 6 compares the recall performance of the evaluated CNN architectures on clinically important minority classes, including melanoma (MEL), actinic keratosis (AKIEC), and dermatofibroma (DF). ResNet50 achieved the best overall minority-class performance,

particularly for DF and MEL, indicating greater robustness under severe class imbalance. These findings complement the overall performance results by demonstrating the model's ability to classify underrepresented lesions more effectively.

4. Class-wise Error Analysis

All evaluated models performed best on the majority NV class, reflecting the imbalanced distribution of the HAM10000 dataset. However, ResNet50 produced more consistent predictions across minority classes, whereas MobileNetV2 frequently misclassified minority lesions as NV. These findings indicate that residual learning improves feature representation for imbalanced multiclass skin lesion classification.

5. Architectural Comparison, Computational Complexity, and Practical Implications

Table 7. Qualitative Comparison of Computational Characteristics of the Evaluated CNN Architectures

Model	Architecture Characteristics	Computational Complexity	Recommended Deployment
VGG16	Deep CNN with stacked convolutional and fully connected layers	High	High-performance workstations
ResNet50	Residual learning architecture with skip connections	Moderate-High	Clinical decision support systems
MobileNetV2	Lightweight CNN using depthwise separable convolutions	Low	Mobile and edge devices

Table 7 presents the computational characteristics of the evaluated CNN architectures. The results indicate that MobileNetV2 is more suitable for resource-constrained environments because of its lightweight design, whereas VGG16 requires higher computational resources. ResNet50 provides the strongest compromise between prediction quality and computational demand, supporting its applicability for AI-assisted skin lesion classification in clinical settings [11], [25].

CONCLUSIONS AND SUGGESTIONS

This study compared VGG16, ResNet50, and MobileNetV2 for seven-class skin lesion classification using the HAM10000 dataset under a common data-partitioning and evaluation protocol. ResNet50 achieved the best overall performance, with 85.71% accuracy, a weighted F1-score of 0.8502, and a macro F1-score of 0.7458, outperforming VGG16 (82.52% accuracy; weighted F1-score of 0.8172) and MobileNetV2 (79.32% accuracy; weighted F1-score of 0.7739). ResNet50 also demonstrated higher recall for melanoma (MEL) and dermatofibroma (DF), while matching VGG16 in recall for actinic keratosis (AKIEC), highlighting its stronger aggregate performance and recognition of selected lesion classes. Although MobileNetV2 offers a lightweight architecture, its lower classification performance indicates a potential trade-off between model efficiency and predictive capability. These findings identify ResNet50 as a promising candidate for further investigation in automated skin lesion classification, rather than establishing its readiness for clinical deployment. Nevertheless, the use of a single publicly available, highly imbalanced dataset, the absence of external clinical validation, and architecture-specific augmentation and loss-function strategies limit generalizability and prevent the isolated attribution of performance differences to network

architecture alone. Future research should adopt fully matched training protocols and controlled ablation experiments, validate the models on independent clinical datasets, investigate explainable AI and additional imbalance-aware strategies, and benchmark inference time, floating-point operations (FLOPs), and memory consumption to establish their robustness and practical deployment feasibility.

REFERENCES

[1] H. Sung *et al*., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021, doi: 10.3322/caac.21660.

[2] J. Yuan, X. Li, and S. Yu, "Global, Regional, and National Incidence Trend Analysis of Malignant Skin Melanoma Between 1990 and 2019, and Projections Until 2030," *Cancer Control*, vol. 31, pp. 1–11, 2024, doi: 10.1177/10732748241227340.

[3] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer Statistics, 2023," *CA: A Cancer Journal for Clinicians*, vol. 73, no. 1, pp. 17–48, 2023, doi: 10.3322/caac.21763.

[4] H. Bhatt, V. Shah, K. Shah, R. Shah, and M. Shah, "State-of-the-Art Machine Learning Techniques for Melanoma Skin Cancer Detection and Classification: A Comprehensive Review," *Intelligent Systems with Applications*, vol. 19, Art. no. 200290, 2023, doi: 10.1016/j.iswa.2023.200290.

[5] E. Ahishakiye, M. B. van Gijzen, J. Tumwiine, R. Wario, and J. Obungoloch, "A Survey on Deep Learning in Medical Image Reconstruction," *Intelligent Medicine*, vol. 1, no. 3, pp. 118–127, 2021, doi: 10.1016/j.imed.2021.08.004.

[6] A. Esteva, K. Chou, S. Yeung, N. Naik, J. Dean, and R. Socher, "Deep Learning-Enabled Medical Computer Vision," *npj Digital Medicine*, vol. 4, no. 5, 2021, doi: 10.1038/s41746-020-00376-2.

[7] A. Adegun and S. Viriri, "Deep Learning Techniques for Skin Lesion Analysis and Melanoma Cancer Detection: A Survey," *Artificial Intelligence Review*, vol. 53, pp. 6137–6176, 2020, doi: 10.1007/s10462-020-09865-4.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[9] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[10] H. L. Gururaj, N. Manju, A. Nagarjun, V. M. N. Aradhya, and F. Flammini, "DeepSkin: A Deep Learning Approach for Skin Cancer Classification," *IEEE Access*, vol. 11, pp. 50205–50214, 2023, doi: 10.1109/ACCESS.2023.3274848.

[11] P. J., S. G., and R. Deepa, "Medical Image Classification for Skin Cancer Using Deep Learning and Ensemble Methods," in *Proc. 2025 2nd International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF)*, 2025, pp. 1–8, doi: 10.1109/ICECONF65644.2025.11379654.

[12] A. E. Minarno, A. Lusianti, Y. Azhar, and H. Wibowo, "Classification of Skin Cancer Images Using Convolutional Neural Network with ResNet50 Pre-trained Model," *JOIV: International Journal on Informatics Visualization*, vol. 8, no. 4, pp. 2013–2019, 2024.

[13] L. Li, X. Wang, W. Hu, Y. Du, B. Li, and N. N. Xiong, "Deep Learning in Skin Disease Image Recognition: A Review," *IEEE Access*, vol. 8, pp. 208264–208280, 2020, doi: 10.1109/ACCESS.2020.3037258.

[14] A. Mahbod, P. Tschandl, G. Langs, R. Ecker, and I. Ellinger, "The Effects of Skin Lesion Segmentation on the Performance of Dermatoscopic Image Classification," *Computer Methods and Programs in Biomedicine*, vol. 197, Art. no. 105725, 2020, doi: 10.1016/j.cmpb.2020.105725.

[15] C. Hernández-Pérez *et al*., "BCN20000: Dermoscopic Lesions in the Wild," *Scientific Data*, vol. 11, Art. no. 760, 2024, doi: 10.1038/s41597-024-03387-w.

[16] M. Musthafa, M. T. R., V. V., and S. Guluwadi, "Enhanced Skin Cancer Diagnosis Using Optimized CNN Architecture and Checkpoints for Automated Dermatological Lesion Classification," *BMC Medical Imaging*, vol. 24, 2024, doi: 10.1186/s12880-024-01356-8.

[17] M. Z. Khan, M. A. Khan, and N. N. Xiong, "Deep Neural Architectures for Medical Image Semantic Segmentation: Review," *IEEE Access*, vol. 9, pp. 83002–83024, 2021, doi: 10.1109/ACCESS.2021.3086530.

[18] H. S. R. Kumar, C. N. G. Babu, C. P. Vijay, K. Raju, K. L. S. Kumar, K. Prabhavathi, and K. Puttegowda, "Deep Learning-Based Automated Classification of Skin Lesions Using CNN and Computer Vision," *SN Computer Science*, vol. 6, 2025, doi: 10.1007/s42979-025-04375-6.

[19] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer," *Neuroscience Informatics*, vol. 2, no. 1, Art. no. 100034, 2022, doi: 10.1016/j.neuri.2021.100034

[20] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin Lesion Classification Using Ensembles of Multi-Resolution EfficientNets with Meta Data," *MethodsX*, vol. 7, Art. no. 100864, 2020, doi: 10.1016/j.mex.2020.100864.

[21] S. Serte and H. Demirel, "Wavelet-Based Deep Learning for Skin Lesion Classification," *IET Image Processing*, vol. 14, no. 4, pp. 720–726, 2020, doi: 10.1049/iet-ipr.2019.0553.

[22] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 Dataset: A Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions," *Scientific Data*, vol. 5, Art. no. 180161, 2018, doi: 10.1038/sdata.2018.161.

[23] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A Comprehensive Survey of Image Augmentation Techniques for Deep Learning," *Pattern Recognition*, vol. 137, Art. no. 109347, 2023, doi: 10.1016/j.patcog.2023.109347.

[24] M. Hammad, "Skin Cancer Diseases Classification Using Deep Convolutional Neural Network with Transfer Learning Model," *Journal of Physics: Conference Series*, vol. 2128, no. 1, Art. no. 012013, 2021, doi: 10.1088/1742-6596/2128/1/012013.

[25] A. A. Mamun, P. Ray, M. R. U. Nasib, A. Das, J. Uddin, and M. N. Absur, "Optimizing Deep Learning for Skin Cancer Classification: A Computationally Efficient CNN with Minimal Accuracy Trade-Off," in *Proc. 2025 2nd International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM)*, 2025, pp. 1–6, doi: 10.1109/NCIM65934.2025.11159853.

[26] A. M. Sarhan, H. A. Ali, S. Yasser, M. Gobara, A. Kandil, G. Sherif, and E. Moustafa, "Achieving High-Accuracy Skin Cancer Classification with Deep Learning Optimized by Ant Colony Algorithm," *Journal of Electrical Systems and Information Technology*, vol. 12, 2025, doi: 10.1186/s43067-025-00243-8.