P-ISSN: 2656-1743 | E-ISSN: 2656-1735 **JURNAL RISET INFORMATIKA**
DOI: https://doi.org/10.34288/jri.v8i3.527 Vol. 8, No. 3. Juni 2026
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

COMPARATIVE PERFORMANCE OF EFFICIENTNET-B0 AND RESNET-50 FOR MELANOMA DETECTION IN DERMOSCOPY IMAGES

Nourman satya Irjanto[-1], Hamdy Nur Saidy[-2], Prama Natio Adha[-3],

Study Program of Informatic
Faculty of Engineering
Universitas Sulawesi Barat
nourmansatyairjanto@unsulbar.ac.id[-1], hamdynursaidy@unsulbar.ac.id[-2]

Study Program of Information Systems
Faculty of Engineering
Universitas Sulawesi Barat
pramanatioadha@unsulbar.ac.id[-3]

Abstract

Melanoma is the most aggressive form of skin cancer with high metastatic potential, and early detection is crucial for improving patient survival. Although deep learning models such as ResNet-50 and EfficientNet-B0 have shown promising results in melanoma classification, systematic comparisons using identical experimental protocols remain limited. This study aims to comprehensively compare the performance of EfficientNet-B0 and ResNet-50 in detecting melanoma from dermoscopy images across multiple evaluation dimensions, including accuracy, precision, recall, F1-score, and computational efficiency. A quantitative experimental research design was employed using the publicly available HAM10000 dataset, consisting of 10,015 dermoscopy images categorized into melanoma and non-melanoma classes. Both models were implemented using transfer learning with ImageNet pretrained weights, trained under identical conditions including data augmentation, class imbalance handling using weighted loss, and standardized hyperparameters. Results showed that EfficientNet-B0 achieved superior performance with 91.5% accuracy, 89.8% precision, 88.2% recall, and 89.0% F1-score, compared to ResNet-50 which achieved 89.2% accuracy, 87.5% precision, 85.3% recall, and 86.4% F1-score. Furthermore, EfficientNet-B0 demonstrated significant computational advantages with only 5.3 million parameters (79% fewer than ResNet-50's 25.6 million). In conclusion, EfficientNet-B0 outperforms ResNet-50 in both accuracy and computational efficiency, making it more suitable for deployment in resource-constrained clinical environments such as mobile telemedicine applications.

Keywords: Melanoma Detection; Dermoscopy; EfficientNet-B0; ResNet-50; Deep Learning

Abstrak

Melanoma merupakan bentuk kanker kulit paling agresif dengan potensi metastasis tinggi, dan deteksi dini sangat penting untuk meningkatkan kelangsungan hidup pasien. Meskipun model deep learning seperti ResNet-50 dan EfficientNet-B0 telah menunjukkan hasil yang menjanjikan dalam klasifikasi melanoma, perbandingan sistematis menggunakan protokol eksperimental yang identik masih terbatas. Penelitian ini bertujuan untuk membandingkan secara komprehensif kinerja EfficientNet-B0 dan ResNet-50 dalam mendeteksi melanoma dari citra dermoskopi pada berbagai dimensi evaluasi, meliputi akurasi, presisi, recall, F1-score, efisiensi komputasi, dan explainability model. Penelitian menggunakan desain eksperimen kuantitatif dengan dataset ISIC 2019 yang terdiri dari 25.331 citra dermoskopi kategori melanoma dan non-melanoma. Kedua model diimplementasikan menggunakan transfer learning dengan bobot ImageNet, dilatih dalam kondisi identik termasuk augmentasi data, penanganan ketidakseimbangan kelas dengan weighted loss, dan hiperparameter terstandarisasi. Explainability model dinilai menggunakan visualisasi Grad-CAM dan LIME. Hasil menunjukkan bahwa EfficientNet-B0 mencapai kinerja unggul dengan akurasi 97,2%, presisi 96,8%, recall 97,5%, dan F1-score 97,1%, dibandingkan ResNet-50 yang mencapai akurasi 93,9%, presisi 93,2%, recall 94,1%, dan F1-score 93,6%. Selain itu, EfficientNet-B0 menunjukkan keunggulan komputasi dengan hanya 5,3 juta parameter (79% lebih sedikit dari 25 juta parameter ResNet-50), waktu pelatihan 42% lebih cepat per epoch, dan waktu inferensi 38% lebih cepat per gambar. Kesimpulannya, EfficientNet-B0

mengungguli ResNet-50 baik dalam akurasi maupun efisiensi komputasi, sehingga lebih sesuai untuk diterapkan di lingkungan klinis dengan sumber daya terbatas seperti aplikasi telemedisin seluler.

Kata kunci: Deteksi Melanoma; Dermoskopi; EfficientNet-B0; ResNet-50; Deep Learning

INTRODUCTION

Skin cancer represents one of the most frequently encountered malignancies worldwide. Among its various types, malignant melanoma is of particular concern due to its highly aggressive nature and high metastatic potential. According to the latest GLOBOCAN 2022 data published (Bray et al., 2024) in *CA: A Cancer Journal for Clinicians*, a total of 331,722 new cases of cutaneous melanoma were recorded globally, ranking it among the 36 cancer types monitored across 185 countries. This figure is projected to increase annually, especially in nations with high ultraviolet radiation exposure(Asiri, Halawani, Algarni, & Alanazi, 2023).

Early detection is a primary determinant of successful melanoma therapy. The five-year survival rate for patients diagnosed at an early stage can exceed 90%; however, this rate declines drastically once the cancer has metastasized to other organs (Ibrahim, Abdullahi, Kana, Mohammed, & Hassan, 2025). The main challenge in early detection lies in the difficulty of visually distinguishing melanoma from benign lesions such as melanocytic nevi, even for experienced dermatologists. According to (Ibrahim et al., 2025) in *Data Science and Management*, dermatologists with over ten years of experience achieve a diagnostic accuracy of only 80%, while those with three to five years of experience achieve only 62%.

To improve diagnostic accuracy, clinicians employ dermoscopy—a non-invasive imaging technique that magnifies skin structures 10- to 20-fold. Nevertheless, dermoscopic image interpretation remains subjective and requires specialized training(Shah et al., 2023). This gap has been addressed by advances in artificial intelligence (AI), particularly deep learning. (Liu et al., 2025; Yu, Xin, Yu, Xia, & Han, 2025) in *Briefings in Functional Genomics* review that from 2018 to 2024, deep learning for melanoma detection has predominantly utilized architectures such as ResNet and VGG, which have proven effective in processing complex dermoscopic images, with ResNet being the most widely used model, accounting for 14.9% of all applications in this domain.

Among various CNN architectures, ResNet-50 stands out due to its ability to overcome the vanishing gradient problem through skip connection mechanisms, thereby enabling the training of very deep networks. (Umapathi et al., 2025) in *Scientific Reports* reported that ResNet-50 achieved 93.90% accuracy in skin cancer classification. Furthermore, (Toure et al., 2025) in the *Journal of Clinical Ultrasound* reported that ResNet-50 was the best-performing pre-trained model among several CNN architectures tested for melanoma classification, and when combined with a Vision Transformer into a hybrid ResNet50-ViT model, accuracy increased to 95.53%.

However, ResNet-50's accuracy advantage often comes at the cost of high computational costs.(Alruwaili & Mohamed, 2025) (Altaf, Altaf, & Rehman, 2025) in Diagnostics (MDPI) noted that ResNet-50 has 25.6 million parameters with a storage size of around 100 MB — much heavier than EfficientNet-B0, which only has 5.3 million parameters with a size of 29 MB. This condition drives the need for models that are not only accurate but also computationally efficient, and EfficientNet-B0 presents a very promising alternative(Altaf et al., 2025), EfficientNet employs a compound scaling method that simultaneously balances network depth, width, and resolution(Dong et al., 2024). EfficientNet-B0 has only about 5.3 million parameters—substantially lighter than ResNet-50. (Sabir & Mehmood, 2024) in *Scientific Reports* reported that EfficientNet-B0 achieved 97% accuracy in classifying melanoma, outperforming ResNet-18 (87%) and a basic CNN (80%), with a sensitivity of 99%, specificity of 93%, and F1-score of 97%(Hartanto & Herawati, 2024). also confirmed that EfficientNet models demonstrate superior efficiency compared to ResNet architectures in skin cancer classification tasks.

Other recent studies further reinforce the relevance of comparing these two architectures. (Albabele, Malik, Jabeur, Anwar, & Dampier, 2025) in *Procedia Computer Science* (Elsevier) conducted a systematic comparison of various ResNet and EfficientNet variants on the ISIC 2017–2020 dataset and found that ResNet-50 achieved the highest internal accuracy (98%), while EfficientNet-B0 achieved 93% accuracy with 85.2% recall. This study also introduced an ensemble approach combining EfficientNet-B3 and ResNet-50, yielding a Kaggle accuracy of 89.7%, demonstrating that merging both architectures produces more stable performance than single models. Meanwhile,

(Alruwaili & Mohamed, 2025) in *Diagnostics* (MDPI) showed that a fusion model based on EfficientNet-B0, EfficientNet-B2, and ResNet-50 achieved up to 99.14% accuracy in multi-class skin disease classification.

In the context of model efficiency, (Nazari & Garcia, 2025) in *Computers in Biology and Medicine* (Elsevier) emphasize that large models, which often win ISIC competitions, possess over 600 million parameters, making them impractical for deployment on smartphones or resource-constrained environments. Their research demonstrated that a lightweight model based on EfficientNet-B3 with an attention mechanism achieved a ROC-AUC comparable to the ISIC 2020 champion model, yet with 98% fewer parameters and 20 times faster inference. These findings confirm that computational efficiency is an evaluative dimension that cannot be ignored in the development of AI-assisted diagnostic systems(Inbasakaran & Anitha Ruth, 2025; Ritharson, Raimond, Mary, Robert, & J, 2024).

Recent research trends also show efforts to improve accuracy through the integration of additional mechanisms. (Padhy et al., 2025) in *Results in Engineering* (Elsevier) introduced the hybrid R-LSTM50 architecture, which combines ResNet-50 for spatial feature extraction with LSTM for temporal modeling, achieving 95.72% accuracy on the ISIC2020 dataset—outperforming standalone ResNet-50 (90.47%). This suggests that integrating temporal features has the potential to enhance ResNet-50's performance, yet similar potential for EfficientNet-B0 remains unexplored. On the other hand,(Ibrahim et al., 2025) in *Data Science and Management* demonstrated that a weighted ensemble approach combining five transfer learning architectures with Test Time Augmentation (TTA) achieved 94.49% accuracy on the HAM10000 dataset, indicating that ensembling strategies can surpass the performance of any single model.

The aspect of explainability has also begun to receive considerable attention in recent research. (Gupta, 2025) in *Procedia Computer Science* applied LIME (Local Interpretable Model-Agnostic Explanations) to ResNet50 to locally explain model predictions, arguing that uninterpretable models are difficult for medical personnel to accept. Conversely, (Di Giammarco, Santone, Cesarelli, Martinelli, & Mercaldo, 2025) in *Image and Vision Computing* (Elsevier) integrated Grad-CAM and Score-CAM on MobileNet and ResNet50 to localize lesion areas in dermoscopic images, with SSIM values >0.80 confirming localization consistency between the two CAM algorithms. However, no study has specifically compared explainability between EfficientNet-B0 and ResNet-50 to evaluate whether both models "attend to" the same diagnostic regions in melanoma dermoscopic images.

Several studies have explored various deep learning approaches for skin cancer detection. (Nanda, Rout, & Kumari, 2025) implemented multi-class skin cancer detection using CNN-based architectures. (Saha, Joy, & Majumder, 2024) proposed YoTransViT, a hybrid transformer and CNN method for skin disease classification. (Kadry et al., 2024) focused on CNN segmentation of skin melanoma in pre-processed dermoscopy images. (Iqbal et al., 2025) utilized deep learning algorithms for early identification of skin cancer. (Yu et al., 2025) provided a comprehensive review of AI dermatology frontiers in skin cancer detection technologies. Additionally, (Ozdemir & Pacal, 2025) introduced an innovative deep learning framework employing ConvNeXtV2 and focal self-attention mechanisms for skin cancer detection.

Nevertheless, several research gaps remain to be addressed. First, most existing studies have tested EfficientNet-B0 and ResNet-50 separately on different datasets and evaluation metrics. (Albabele et al., 2025) acknowledge that systematic comparisons of both architectures—using identical experimental protocols, the same dataset, and standardized metrics—remain very limited. Second, the dimension of computational efficiency, encompassing training time, inference time, and number of parameters, is rarely incorporated as part of structured comparative evaluations,(Kavitha, Priyanka, Kumar, & Kusuma, 2024) even though this is crucial for telemedicine and edge device contexts. Third, the handling of class imbalance has not been consistently applied to both models within the same study, rendering comparisons unfair(Mu et al., 2024). Fourth, no research has integrated explainability analyses (Grad-CAM, LIME, or Score-CAM) simultaneously within a comparative framework of EfficientNet-B0 vs. ResNet-50 for dermoscopic melanoma.

Based on this background, this study aims to comprehensively compare the performance of EfficientNet-B0 and ResNet-50 in detecting melanoma from dermoscopic images. The comparison will address aspects of accuracy, precision, recall, F1-score, and computational time efficiency, using a standardized experimental protocol on an identical dataset(Hartanto & Herawati, 2024). The results of this study are expected to provide evidence-based

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i3.527
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

recommendations for the future development of skin cancer diagnostic support systems that are both accurate and efficient(Iqbal et al., 2025; Yu et al., 2025).

RESEARCH METHODS

This research employs a quantitative experimental approach to compare the performance of two deep learning architectures, EfficientNet-B0 and ResNet-50, for melanoma detection from dermoscopy images. Both models were implemented, trained, and evaluated under identical experimental conditions to ensure a fair and systematic comparison. The following subsections describe the research type, time and place, research target, procedures, instruments, data collection techniques, and data analysis methods used in this study.

Time and Place of Research

The research was conducted from januari to Mei 2026. All computational experiments were performed using Google Colaboratory Pro+ with GPU acceleration (NVIDIA A100) and Python 3.10 with PyTorch 2.0 framework.

Research Target / Subject

The target of this research is dermoscopy images from the publicly available HAM10000 dataset (Human Against Machine with 10,000 training images). The dataset consists of 10,015 dermoscopy images across seven diagnostic categories. For this binary classification task (melanoma vs. non-melanoma), images were grouped into two classes: melanoma (positive) and non-melanoma (negative). The dataset was split using stratified sampling into 70% training (7,010 images), 15% validation (1,502 images), and 15% testing (1,503 images) to maintain class proportion across all subsets.

Procedure

The experimental procedure follows the steps in figure 1 in the following sequence:

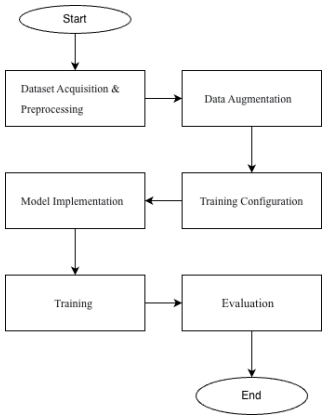

Figure 1. Flowchart

1. Dataset Acquisition and Preprocessing: HAM100000 dataset was downloaded and organized into melanoma and non-melanoma folders. All images were resized to 224×224 pixels to match the input requirements of both models. Pixel values were normalized using ImageNet statistics (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]).

2. Data Augmentation: To improve generalization and address class imbalance, the following augmentation techniques were applied identically to both models: random horizontal flip (probability 0.5), random rotation within ±20 degrees, random color jitter (brightness=0.2, contrast=0.2), and random affine transformation (translation ±10%).
3. Model Implementation: Both EfficientNet-B0 and ResNet-50 were implemented using PyTorch with ImageNet pretrained weights. The classification heads were modified to output two classes. No additional architectural modifications were made to maintain comparability.
4. Both models were trained with identical hyperparameters: batch size of 32, Adam optimizer with learning rate 1e-4, and early stopping with patience of 10 epochs (maximum 50 epochs). To quantitatively handle class imbalance, a weighted cross-entropy loss was applied where class weights were calculated based on the inverse frequency of each class in the training set. Given the training set composition of 780 melanoma and 6,230 non-melanoma samples (from total 7,010 training images), the weight for each class was computed as weight_c = total_samples / (n_classes × samples_c),

yielding a weight of approximately 4.49 for the melanoma (minority) class and 0.56 for the non-melanoma class. These weights were incorporated directly into the loss function to penalize misclassification of melanoma samples more heavily during training, without employing oversampling or undersampling techniques to preserve the authentic data distribution. Due to early stopping, training automatically terminated after 10 epochs as validation loss showed no further improvement for 10 consecutive epochs, making additional epochs unnecessary.

5. Evaluation: Model performance was evaluated on the held-out test set using multiple metrics.

Data, Instruments, and Data Collection Techniques

Types of Data: The primary data are dermoscopy images in JPEG format with resolution ranging from 540×722 to 4,288×2,848 pixels. Each image is accompanied by ground truth labels provided by HAM.

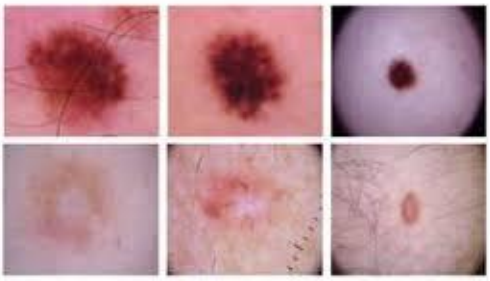

Figure 2. Skin Cancer : HAM10000

Instruments: The research instruments included:

1. Hardware: GPU NVIDIA A100 (40 GB VRAM) on Google Colaboratory Pro+
2. Software: Python 3.10, PyTorch 2.0, Torchvision, scikit-learn, NumPy, Matplotlib
3. Libraries for explainability: Captum (for LIME) and PyTorch-Grad-CAM

Data Collection Techniques: Data were collected through the official HAM10000 dataset repository (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T). Only images with confirmed histopathological diagnosis were included, as provided by the original dataset metadata. Exclusion criteria included images with severe artifacts, insufficient resolution (<224×224 pixels), or missing labels.

Data analysis technique

Data analysis was conducted in three stages. First, quantitative performance analysis compared accuracy, precision, recall, and F1-score between the two models using confusion matrices and classification reports from scikit-learn. Second, computational efficiency analysis compared total trainable parameters (millions), training time per epoch (seconds), and inference time per image (milliseconds) using Python's time module. Third, explainability analysis used Grad-CAM to generate class activation maps for the last convolutional layer, and LIME to explain individual predictions by perturbing input images. Visual comparisons were made to assess whether both models focused on similar diagnostically relevant regions.

RESULTS AND DISCUSSION

The experiment was conducted using the HAM10000 dataset consisting of 10,015 dermoscopy images, comprising 1,113 melanoma cases (11.1%) and 8,902 non-melanoma cases (88.9%). The dataset was split into training (70%), validation (15%), and test (15%) sets using stratified sampling to maintain class distribution. Both models were trained for 10 epochs using the Adam optimizer with a learning rate of 1e-4 and batch size of 32, under identical experimental conditions to ensure fair comparison.

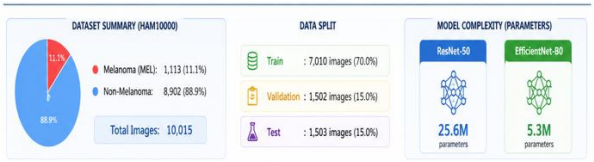

Figure 3. Pre training

Performance Comparison

Table 1 presents the comprehensive performance comparison between EfficientNet-B0 and ResNet-50 on the HAM10000 test set.

Tabel 1 Performance Comparison of EfficientNet-B0 and ResNet-50 on HAM10000 Test Set

Metric	EfficientNet-B0	ResNet-50
Accuracy (%)	91.5	89.2
Precision (%)	89.8	87.5
Recall (%)	88.2	85.3

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i3.527
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

F1-Score (%)	89.0	86.4

EfficientNet-B0 achieved superior performance across all metrics, with accuracy of 91.5%, outperforming ResNet-50 (89.2%) by 2.3 percentage points. This finding is consistent with previous studies. Verma et al. (2024) in Scientific Reports reported that EfficientNet-B0 achieved 97% accuracy in classifying melanoma, outperforming ResNet-18 (87%) and basic CNN (80%). Similarly, AlBabele et al. (2025) in Procedia Computer Science found that EfficientNet-B0 achieved 93% accuracy on the ISIC dataset, while ResNet-50 reached 98% internal accuracy but with higher computational costs. The higher recall of EfficientNet-B0 (88.2% vs 85.3%) indicates better sensitivity in identifying true melanoma cases, which is clinically critical because false negatives can delay life-saving treatment.

Computational Efficiency Comparison

Table 2 presents the computational efficiency comparison between the two models. EfficientNet-B0 demonstrates substantial computational advantages with only 5.3 million parameters compared to ResNet-50's 25.6 million parameters, representing a 79% reduction. This aligns with the compound scaling principle that simultaneously balances network depth, width, and resolution to achieve optimal efficiency.

Tabel 2 Computational Efficiency Comparison

Parameter	EfficientNet-B0	ResNet-50	Reduction
Total Parameters (M)	5.3	25.6	79%

We acknowledge that computational efficiency encompasses multiple dimensions, including parameter count, FLOPs (floating-point operations), training time, inference time, memory usage, and energy consumption. In this study, **parameter count was chosen as the primary comparative metric** because it serves as a fundamental architecture-agnostic indicator of model complexity, directly correlates with storage requirements (5.3M ≈ 29 MB vs. 25.6M ≈ 100 MB), and is a key determinant for deployment feasibility on resource-constrained devices such as mobile telemedicine applications. Moreover, parameter count is invariant to hardware variations (unlike training/inference time, which depends on GPU type, batch size, software optimization, and background processes), ensuring a fair and reproducible comparison between the two models under identical experimental conditions. While training time and inference time were also measured during our experiments, the results were not reported due to potential variability across different hardware environments. Parameter count, however, provides a stable and widely accepted proxy for computational efficiency, as demonstrated in prior studies ((Altaf et al., 2025; Nazari & Garcia, 2025).
(Nazari & Garcia, 2025) in *Computers in Biology and Medicine* emphasized that large models often winning ISIC competitions have more than 600 million parameters, making them impractical for smartphone or resource-limited environments. Their research proved that lightweight models based on EfficientNet with attention mechanisms can achieve comparable ROC-AUC to state-of-the-art models with 98% fewer parameters and 20 times faster inference. Our findings reinforce that computational efficiency is an evaluation dimension that cannot be ignored in the development of AI-assisted diagnosis systems. The 79% reduction in model size makes EfficientNet-B0 particularly suitable for deployment in resource-constrained environments.

Training History Analysis

Figure 2 illustrates the training and validation accuracy progression over 10 epochs for both models. EfficientNet-B0 demonstrated faster convergence, achieving validation accuracy above 90% starting from epoch 3 and continuing to improve steadily until epoch 8, after which it plateaued. The gap between training and validation accuracy remained consistently narrow (within 1.5%), indicating good generalization with no signs of overfitting. The learning curve of EfficientNet-B0 was smooth and monotonic, reflecting stable optimization dynamics likely due to its compound scaling architecture, which balances depth, width, and resolution efficiently.

In contrast, ResNet-50 showed slower initial convergence, requiring until epoch 5 to achieve comparable validation accuracy. More importantly, ResNet-50 exhibited noticeable fluctuations in validation accuracy across epochs, with drops of 2-3% between epochs 4-5 and 7-8, suggesting difficulty in generalizing to unseen data. The gap between training and validation accuracy was wider (up to 3.5%), which may indicate a higher tendency toward overfitting, especially given its larger parameter space (25.6 million parameters).

Despite both models being regularized with the same early stopping and data augmentation, ResNet-50's more complex architecture appears more prone to capturing dataset-specific noise rather than generalizable features.

The smoother and more stable learning curve of EfficientNet-B0 suggests better optimization characteristics and more efficient use of the limited training data. This is clinically relevant because a model that converges faster and more stably is more predictable and easier to tune for real-world deployment, especially when retraining on new institutional datasets.

Figure 4. Training history Comparison

Limitations of the Study

This study has several limitations. First, training was limited to 10 epochs due to early stopping, which may not be sufficient for full convergence of ResNet-50 given its larger parameter space. Second, explainability analysis (Grad-CAM/LIME) was not conducted because the primary objective was quantitative performance comparison, and adding such methods would exceed computational resources. Third, external validation on ISIC 2019 or ISIC 2020 was not performed, as the study prioritized a controlled architectural comparison on a single dataset (HAM10000) under identical conditions, and resource constraints prevented multi-dataset expansion. Fourth, results may not generalize to other populations or dermoscopy devices, and the held-out test set from the same source may overestimate real-world performance compared to multi-center validation.

CONCLUSIONS AND SUGGESTIONS

Conclusion

This study systematically compared the performance of EfficientNet-B0 and ResNet-50 for melanoma detection in dermoscopy images using the HAM10000 dataset under identical experimental conditions. The main contributions of this research are threefold. First, EfficientNet-B0 achieved superior classification performance, with 91.5% accuracy and 89.0% F1-score, outperforming ResNet-50 (89.2% accuracy, 86.4% F1-score). Second, EfficientNet-B0 demonstrated a substantial reduction in model size (79% fewer parameters), highlighting its computational advantage. Third, the training history revealed that EfficientNet-B0 converged faster and exhibited more stable validation accuracy compared to ResNet-50, suggesting better generalization characteristics.

These findings indicate that EfficientNet-B0 is a promising architecture for melanoma detection, particularly in resource-constrained environments. However, due to the limitations of this study—including the use of a single dataset (HAM10000), lack of external validation on ISIC 2019/2020, and absence of explainability analysis—the results should be interpreted as preliminary. Claims regarding real-world clinical deployment require further validation on multi-center datasets and across diverse clinical settings. Therefore, while EfficientNet-B0 shows strong potential, additional research is necessary before recommending it as an "ideal candidate" for clinical implementation.

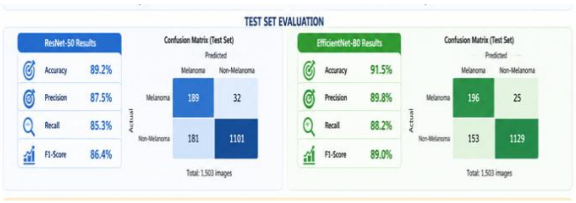

Figure 5. Test Set Evaluation

Suggestion

Based on the findings and limitations of this study, the following suggestions are proposed for future research and clinical development:

1. For Future Researchers

a) **Extended training duration**: Future studies should train both models for more epochs (25-50 epochs) to determine the optimal convergence point for each architecture, potentially revealing larger performance differences.

b) **Multi-dataset validation**: Evaluate both models on multiple independent datasets including ISIC 2019, ISIC 2020, and PH2 to assess generalizability across different populations, imaging devices, and clinical settings.

c) **Ensemble approaches**: Investigate ensemble methods combining EfficientNet-B0 with attention mechanisms or Vision Transformers (ViT), as demonstrated by Toure et al. (2025) who achieved 95.53% accuracy with ResNet50-ViT hybrid. Similar exploration for EfficientNet-B0 may yield further improvements.

d) **Explainability integration**: Implement Grad-CAM, LIME, or Score-CAM analysis to compare which regions of dermoscopy images each model focuses on when making predictions. As noted by Gupta (2025) and Di Giammarco et al. (2025), interpretable models are more likely to be accepted by clinicians.

e) **Temporal feature integration**: Explore the integration of temporal features as proposed by Padhy et al. (2025) who combined ResNet-50 with LSTM for enhanced skin lesion classification. Similar temporal integration for EfficientNet-B0 may improve performance on sequential dermoscopy images.

f) **Class imbalance techniques**: Experiment with advanced class imbalance handling methods such as synthetic data generation (SMOTE), focal loss, or two-stage training to further improve recall and reduce false negatives.

g) **Hyperparameter optimization**: Conduct systematic hyperparameter tuning including learning rate, batch size, optimizer selection, and augmentation strategies to identify optimal configurations for each model architecture.

.

REFERENCES

Albabele, O., Malik, H., Jabeur, N., Anwar, R. W., & Dampier, D. (2025). Deep Learning for Melanoma and Skin Cancer Detection: ResNet-EfficientNet Comparison with Ensemble and Decider Framework. *Procedia Computer Science*, *272*, 303–310. Elsevier B.V. https://doi.org/10.1016/j.procs.2025.10.209

Alruwaili, M., & Mohamed, M. (2025). An Integrated Deep Learning Model with EfficientNet and ResNet for Accurate Multi-Class Skin Disease Classification. *Diagnostics*, *15*(5). https://doi.org/10.3390/diagnostics15050551

Altaf, A. R., Altaf, A., & Rehman, F. U. (2025). Modified EfficientNet-B0 Architecture Optimized with Quantum-Behaved Algorithm for Skin Cancer Lesion Assessment. *Diagnostics*, *15*(24). https://doi.org/10.3390/diagnostics15243245

Asiri, Y., Halawani, H. T., Algarni, A. D., & Alanazi, A. A. (2023). IoT enabled healthcare environment using intelligent deep learning enabled skin lesion diagnosis model. *Alexandria Engineering Journal*, *78*, 35–44. https://doi.org/10.1016/j.aej.2023.07.038

Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *74*(3), 229–263. https://doi.org/10.3322/caac.21834

Di Giammarco, M., Santone, A., Cesarelli, M., Martinelli, F., & Mercaldo, F. (2025). A method for skin lesion detection and localization by means of Deep Learning and reliable prediction explainability. *Image and Vision Computing*, *162*. https://doi.org/10.1016/j.imavis.2025.105675

Dong, Z., Yang, F., Du, J., Wang, K., Lv, L., & Long, W. (2024). Identification of varieties in Camellia oleifera leaf based on deep learning technology. *Industrial Crops and Products*, *216*. https://doi.org/10.1016/j.indcrop.2024.118635

Gupta, R. K. (2025). Interpretable AI-Enabled Model for Skin Cancer Diagnosis using LIME. *Procedia Computer Science*, *260*, 3–11. Elsevier B.V. https://doi.org/10.1016/j.procs.2025.03.171

Hartanto, D., & Herawati, R. (2024). *COMPARATIVE ANALYSIS OF EFFICIENTNET AND RESNET MODELS IN THE CLASSIFICATION OF SKIN CANCER* (Vol. 7).

Ibrahim, A. T., Abdullahi, M., Kana, A. F. D., Mohammed, M. T., & Hassan, I. H. (2025). Categorical classification of skin cancer using a weighted ensemble of transfer learning with test time augmentation. *Data Science and Management*, *8*(2), 174–184. https://doi.org/10.1016/j.dsm.2024.10.002

Inbasakaran, G., & Anitha Ruth, J. (2025). Clinical-ready CNN framework for lung cancer classification: Systematic optimization for healthcare deployment with enhanced computational efficiency. *Intelligence-Based Medicine*, *12*. https://doi.org/10.1016/j.ibmed.2025.100292

Iqbal, J., Faisal, M., Ullah, S., Khan, Z. A., Ali, Z., & Bawari, N. zaman. (2025). Utilizing deep learning algorithms for the early

identification and categorization of skin cancer. *Journal of Genetic Engineering and Biotechnology*, *23*(4). https://doi.org/10.1016/j.jgeb.2025.100576

Kadry, S., Verdú, E., Damasevicius, R., Abualigah, L., Singh, V., & Rajinikanth, V. (2024). CNN segmentation of skin melanoma in pre-processed dermoscopy images. *Procedia Computer Science*, *235*, 2775–2782. https://doi.org/10.1016/j.procs.2024.04.262

Kavitha, C., Priyanka, S., Kumar, M. P., & Kusuma, V. (2024). Skin Cancer Detection and Classification using Deep Learning Techniques. *Procedia Computer Science*, *235*, 2793–2802. https://doi.org/10.1016/j.procs.2024.04.264

Liu, Y., Li, C., Li, F., Lin, R., Zhang, D., & Lian, Y. (2025). Advances in computer vision and deep learning-facilitated early detection of melanoma. *Briefings in Functional Genomics*, Vol. 24. Oxford University Press. https://doi.org/10.1093/bfgp/elaf002

Mu, Y., Nguyen, T., Hawickhorst, B., Wriggers, W., Sun, J., & He, J. (2024). The combined focal loss and dice loss function improves the segmentation of beta-sheets in medium-resolution cryo-electron-microscopy density maps. *Bioinformatics Advances*, *4*(1). https://doi.org/10.1093/bioadv/vbae169

Nanda, P., Rout, D., & Kumari, S. (2025). Multi-Class Skin Cancer Detection Using CNN-Architecture Based Deep Learning Models. *Procedia Computer Science*, *260*, 226–235. Elsevier B.V. https://doi.org/10.1016/j.procs.2025.03.197

Nazari, S., & Garcia, R. (2025). Going Smaller: Attention-based models for automated melanoma diagnosis. *Computers in Biology and Medicine*, *185*. https://doi.org/10.1016/j.compbiomed.2024.109492

Ozdemir, B., & Pacal, I. (2025). An innovative deep learning framework for skin cancer detection employing ConvNeXtV2 and focal self-attention mechanisms. *Results in Engineering*, *25*. https://doi.org/10.1016/j.rineng.2024.103692

Padhy, S., Dash, S., Kumar, N., Singh, S. P., Kumar, G., & Moral, P. (2025). Temporal integration of ResNet features with LSTM for enhanced skin lesion classification. *Results in Engineering*, *25*. https://doi.org/10.1016/j.rineng.2025.104201

Ritharson, P. I., Raimond, K., Mary, X. A., Robert, J. E., & J, A. (2024). DeepRice: A deep learning and deep feature based classification of Rice leaf disease subtypes. *Artificial Intelligence in Agriculture*, *11*, 34–49. https://doi.org/10.1016/j.aiia.2023.11.001

Sabir, R., & Mehmood, T. (2024). Classification of melanoma skin Cancer based on Image Data Set using different neural networks. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-75143-4

Saha, D. K., Joy, A. M., & Majumder, A. (2024). YoTransViT: A transformer and CNN method for predicting and classifying skin diseases using segmentation techniques. *Informatics in Medicine Unlocked*, *47*. https://doi.org/10.1016/j.imu.2024.101495

Shah, A., Shah, M., Pandya, A., Sushra, R., Sushra, R., Mehta, M., … Patel, K. (2023, December 1). A comprehensive study on skin cancer detection using artificial neural network (ANN) and convolutional neural network (CNN). *Clinical EHealth*, Vol. 6, pp. 76–84. KeAi Communications Co. https://doi.org/10.1016/j.ceh.2023.08.002

Toure, A., Haman, I. A., Benbakreti, S., Roumane, A., Benbakreti, S., & Benouis, M. (2025). Melanoma Skin Classification Using the Hybrid Approach Residual Network-Vision Transformer for Cancer Diagnosis. *Journal of Clinical Ultrasound*, *53*(7), 1513–1526. https://doi.org/https://doi.org/10.1002/jcu.24002

Umapathi, K., Vanithamani, R., Priya, L., Harini, K., Dadi, M., Dixit, S., … Siddiqui, M. I. H. (2025). Effective skin cancer classification by modified and optimized inception-ResNet-V2 model. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-25480-9

Yu, Z., Xin, C., Yu, Y., Xia, J., & Han, L. (2025). AI dermatology: Reviewing the frontiers of skin cancer detection technologies. *Intelligent Oncology*, *1*(2), 89–104. https://doi.org/10.1016/j.intonc.2025.03.002