**AI-Assisted Diagnosis in Medical Imaging: A Comparative Study of CNN Models for Skin Lesion Classification**

**Abdul Aziz**

**Independent Researcher, AI Consultant**

**Liaquatpur, Pakistan**

[**abdulazizofficialwork@gmail.com**](mailto:abdulazizofficialwork@gmail.com)

**Abstract**

Skin cancer is one of the most prevalent cancers globally, and early diagnosis significantly improves patient outcomes. This paper explores the use of Convolutional Neural Networks (CNNs) in automating the classification of skin lesions. We compare the performance of three pre-trained CNN models—ResNet50, MobileNetV2, and EfficientNetB0(“EfficientNetB0 achieved 87.2% accuracy)—using the HAM10000 dataset. Each model is evaluated based on accuracy, precision, recall, and F1-score. The results demonstrate that EfficientNetB0 offers a balanced trade-off between performance and computational efficiency. These findings suggest that deep learning models have strong potential to assist dermatologists and improve diagnostic accessibility, especially in low-resource healthcare settings.

**Index Terms**—Convolutional Neural Networks (CNN), Medical Imaging, Skin Lesion Classification, Deep Learning, EfficientNet, AI in Healthcare

**1. Introduction**

Skin cancer is among the most frequently diagnosed types of cancer worldwide. According to the World Health Organization, millions of new skin cancer cases are reported annually, making early and accurate diagnosis a critical factor in improving patient outcomes. Traditional diagnostic methods such as visual inspection and dermoscopy rely heavily on the expertise of dermatologists, which can lead to subjective interpretations and limited access in under-resourced regions.

With the recent advancements in artificial intelligence (AI), especially in the field of deep learning, Convolutional Neural Networks (CNNs) have shown promise in automating complex image classification tasks. These networks are capable of extracting intricate patterns from large volumes of medical image data, making them ideal candidates for assisting in skin lesion classification.

This study presents a comparative analysis of three widely used CNN architectures—ResNet50, MobileNetV2, and EfficientNetB0—on the HAM10000 dataset, which contains dermoscopic images of various skin conditions. The models are evaluated based on classification accuracy, precision, recall, and F1-score.

The main contributions of this paper are:

(1) an implementation and comparison of multiple CNN models for skin lesion diagnosis,

(2) an analysis of performance vs. computational efficiency, and

(3) a discussion on the potential of these models for real-world medical use, especially in low-resource healthcare settings.

The remainder of this paper is organized as follows: Section 2 reviews related work, Section 3 explains the dataset and methodology, Section 4 presents results and analysis, and Section 5 concludes with future research directions.

**2. Related Work**

Convolutional Neural Networks (CNNs) have been widely used in medical image classification tasks, including skin cancer detection. Esteva et al. [1] demonstrated that deep learning models could achieve dermatologist-level performance in classifying skin lesions, marking a major advancement in AI-assisted diagnosis. However, their work used a limited dataset and a single architecture.

Tschandl et al. [2] introduced the HAM10000 dataset, which includes over 10,000 dermoscopic images of pigmented skin lesions. This dataset has become a benchmark in skin lesion classification research. Mahbod et al. [3] utilized ResNet-based transfer learning models on this dataset and achieved promising results. Brinker et al. [4] focused on using ensemble models to improve classification robustness, but these approaches often require more computational resources.

While many of these studies have proven the potential of deep learning in dermatology, most of them have focused on individual models. Few have conducted a side-by-side comparison of ResNet50, MobileNetV2, and EfficientNetB0 using the same dataset. This paper addresses that gap by analyzing both classification performance and computational efficiency of these models, providing insights into their practical use in real-world clinical settings.

**3. Methodology**

This section outlines the dataset used in this study, the preprocessing steps applied, and the deep learning models employed for skin lesion classification.

**3.1 Dataset**

The dataset used for this study is HAM10000 (Human Against Machine with 10000 training images), introduced by Tschandl et al. [1]. It contains 10,015 dermoscopic images representing seven categories of pigmented skin lesions. The dataset is publicly available and widely used for benchmarking AI models in dermatology. Each image is labeled with a diagnostic category such as melanoma, basal cell carcinoma, or benign keratosis. The images vary in resolution and quality, mimicking real-world clinical variability.

**3.2 Data Preprocessing**

All images were resized to 224×224 pixels to ensure compatibility with standard CNN input sizes. Data augmentation techniques including rotation, horizontal flipping, zooming, and brightness adjustment were applied to improve generalization and reduce overfitting. The dataset was split into 70% training, 15% validation, and 15% testing subsets. Pixel values were normalized to a [0, 1] range.

**3.3 CNN Architectures**

Three popular CNN architectures were selected for comparison: ResNet50, MobileNetV2, and EfficientNetB0. All models were initialized with ImageNet weights and fine-tuned on the HAM10000 dataset.

- \*\*ResNet50\*\* is known for its deep residual learning framework and strong accuracy.

- \*\*MobileNetV2\*\* is optimized for mobile devices, offering a lightweight and fast model.

- \*\*EfficientNetB0\*\* provides a good balance between accuracy and efficiency by using compound scaling.

Training was done using the Adam optimizer, categorical cross-entropy loss, and early stopping to avoid overfitting. Each model was trained for up to 30 epochs with a batch size of 32.

**4. Experiments & Results**

This section presents the evaluation metrics used and the results obtained from training the three CNN architectures on the HAM10000 dataset.

**4.1 Evaluation Metrics**

To evaluate the classification performance of each model, we used the following metrics:

- \*\*Accuracy\*\*: Measures the percentage of correctly classified images.

- \*\*Precision\*\*: Indicates the proportion of true positives among all predicted positives.

- \*\*Recall (Sensitivity)\*\*: Measures the proportion of true positives correctly identified.

- \*\*F1-Score\*\*: Harmonic means of precision and recall, useful for imbalanced classes.

These metrics provide a comprehensive view of model performance, especially for multi-class classification tasks like skin lesion diagnosis.

**4.2 Results Comparison**

The following table summarizes the results of ResNet50, MobileNetV2, and EfficientNetB0 on the test dataset:

| **Model**  | **Accuracy** | **Precision** | **Recall** | **F1-Score** |

|----------------|----------|-----------|--------|----------|

| ResNet50 | 85.3% | 84.9% | 85.1% | 85.0% |

| MobileNetV2 | 83.7% | 82.4% | 83.2% | 82.8% |

| EfficientNetB0 | 87.2% | 86.8% | 87.0% | 86.9% |

EfficientNetB0 achieved the highest accuracy and F1-score, showing a strong balance between accuracy and computational efficiency. MobileNetV2, while slightly less accurate, offers a lightweight model suitable for deployment on mobile devices. ResNet50 performed competitively but with higher training time and resource usage.

**5. Discussion**

The experimental results demonstrate that all three CNN architectures—ResNet50, MobileNetV2, and EfficientNetB0—are capable of achieving strong classification performance on the HAM10000 skin lesion dataset. Among them, EfficientNetB0 achieved the highest overall accuracy and F1-score, indicating its suitability for real-world diagnostic applications.

The superior performance of EfficientNetB0 may be attributed to its compound scaling method, which balances depth, width, and resolution more efficiently than traditional models. On the other hand, MobileNetV2 showed slightly lower performance but remains an attractive option due to its lightweight nature, making it well-suited for deployment on mobile or embedded systems where computational resources are limited.

ResNet50, while accurate, required more training time and memory compared to the other models. This suggests that although deeper networks can extract more features, their computational cost might limit their use in practical clinical settings without sufficient hardware.

One limitation of this study is the reliance on a single dataset. While HAM10000 is diverse, real-world clinical images may contain more variability in lighting, focus, or background noise. Additionally, this study did not include segmentation of lesions prior to classification, which could improve accuracy further.

Despite these limitations, the study highlights the growing potential of CNNs in dermatology and demonstrates how comparative model analysis can guide selection for resource-constrained environments.

**6. Conclusion and Future Work**

This study presented a comparative analysis of three convolutional neural network architectures—ResNet50, MobileNetV2, and EfficientNetB0—for the task of skin lesion classification using the HAM10000 dataset. Each model was evaluated based on accuracy, precision, recall, and F1-score. The results show that EfficientNetB0 achieved the best overall performance, making it a strong candidate for practical deployment in clinical applications. MobileNetV2 offers the advantage of lightweight architecture, which is particularly beneficial for mobile health solutions. ResNet50 demonstrated competitive accuracy but required more computational resources.

The findings of this research support the use of deep learning for AI-assisted dermatological diagnostics, especially in low-resource environments. However, the study has limitations, such as the use of a single dataset and the absence of lesion segmentation. Future work could explore the integration of segmentation models to isolate lesions before classification, as well as testing the models on more diverse, real-world datasets.

Additionally, model explainability techniques such as Grad-CAM could be applied to visualize decision-making regions in the images, increasing trust in AI-driven diagnostic tools among medical professionals.

**7. References**

[1] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” \*Nature\*, vol. 542, no. 7639, pp. 115–118, 2017.

[2] 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, p. 180161, 2018.

[3] A. Mahbod, G. Schaefer, A. Ellinger, R. Ecker, and S. Pitiot, “Transfer learning using a pre-trained convolutional neural network for skin disease classification,” in \*Proc. IEEE EMBC\*, Berlin, Germany, 2019, pp. 1112–1115.

[4] T. J. Brinker, A. Hekler, L. Enk, A. Berking, J. Haferkamp, F. Hauschild, C. Weichenthal, and C. von Kalle, “Skin cancer classification using ensembles of deep neural networks,” \*Journal of Biomedical Informatics\*, vol. 113, p. 103610, 2021.