International Journal of Creative and Open Research in Engineering and Management
ISSN: 3108-1754 (Online)
Volume 02 Issue 10 October-2026 | Impact Factor: 3.5



A Comparative Study of Lightweight and Standard CNN Architectures for Multi-Class Skin Disease Classification on Dermoscopic and Clinical Image Datasets

Manisha Roshini Neethi Koil Pillai
Department of Computer Science
SIES College of Arts, Science and Commerce
Under the guidance of **Dr. Abuzar Ansari**

How to Cite this Article:

Roshini, M. & Pillai, N. K. (2026). A Comparative Study of Lightweight and Standard CNN Architectures for Multi-Class Skin Disease Classification on Dermoscopic and Clinical Image Datasets. International Journal of Creative and Open Research in Engineering and Management, 02(10), 1-9. https://doi.org/10.55041/ijcope.v2i10.150

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.







https://doi.org/10.55041/ijcope.v2i10.150

Abstract

Skin disease classification using deep learning has shown promising results, yet perfor-mance varies significantly across datasets due to differences in image modality, class distri-bution, and data quality. This paper presents a systematic comparison of three widely used convolutional neural network architectures—EfficientNet-B0, MobileNetV2, and ResNet-50—on two distinct skin image collections: the dermoscopic HAM10000 dataset (7 classes) and a large clinical/photographic skin disease dataset (10 classes). Both datasets were care-fully prepared with lesion-level splitting (HAM10000) and near-duplicate removal (second dataset). Models were trained with class-weighted loss, identical hyperparameters, and evaluated over multiple random seeds. On HAM10000, MobileNetV2 achieved the highest mean accuracy of 78.0% and macro-F1 of 0.655. On the second dataset, EfficientNet-B0 obtained the best results with 81.8% accuracy and 0.752 macro-F1, while maintaining high melanoma recall (>0.95). The study highlights the impact of data modality and cleaning on model performance and provides reproducible baselines for future work. This system is intended for research purposes only and is not a clinical diagnostic tool.

Keywords: Skin disease classification, deep learning, EfficientNet, MobileNetV2, ResNet, HAM10000, dermoscopy, clinical images, data cleaning

1 Introduction

Automatic classification of skin lesions and diseases has become an active research area due to the global burden of dermatological conditions and the potential of computer vision to assist non-specialists. Public datasets such as HAM10000 have enabled rapid progress, yet many real-world images are captured with ordinary cameras rather than dermoscopes. Models trained on one modality often degrade when applied to the other.

This work addresses two practical questions:

1. How do standard lightweight and medium-capacity CNNs perform on a well-known der-moscopic dataset versus a larger clinical-style dataset?
2. What is the effect of careful data cleaning (near-duplicate removal) on the clinical dataset?

We train and evaluate EfficientNet-B0, MobileNetV2 and ResNet-50 under identical con-ditions on both collections and report multi-seed statistics, including melanoma recall—a clin-ically important metric.



2 Related Work

Previous studies have extensively benchmarked CNNs on HAM10000, often reporting accura-cies between 75–90% depending on architecture, augmentation and test protocol. Many papers, however, use random image-level splits that can leak information from the same lesion across train and test sets. Clinical photographic datasets are less standardized and frequently contain near-duplicates. Our contribution is a transparent, multi-seed comparison with proper lesion-level splitting and explicit near-duplicate auditing.

3 Datasets and Preprocessing

3.1 HAM10000

- 10,015 dermoscopic images
- 7 classes: actinic keratoses (akiec), basal cell carcinoma (bcc), benign keratosis (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevus (nv), vascular lesions (vasc)
- Strong class imbalance (nv dominates)
- Split by lesion_idusing GroupShuffleSplit (70% / 15% / 15%) to prevent lesion leakage
- Class-weighted Cross-Entropy loss

3.2 Second Dataset (Clinical / Mixed)

- Approximately 27,000 images across 10 classes: eczema, viral warts, melanoma, atopic dermatitis, basal cell carcinoma, melanocytic nevi, benign keratosis-like lesions, psoria-sis/lichen planus, seborrheic keratoses, and fungal infections
- Near-duplicate detection performed via MD5 hashing and visual verification of near-pairs
- Clean train / validation / test splits saved and released

Both datasets used ImageNet normalization and standard augmentations (random flips, ro-tation ±20°, colour jitter) during training.

4 Methods

4.1 Architectures

- EfficientNet-B0
- MobileNetV2
- ResNet-50

All models were initialized with ImageNet pretrained weights. The final classification layer was replaced to match the number of classes (7 or 10).



4.2 Training Setup

- Optimizer: Adam (learning rate 1×10^{-4})
- Loss: Weighted Cross-Entropy
- Mixed-precision training (AMP)
- Batch size 32, input size 224×224
- 5–8 epochs with early checkpointing on validation macro-F1
- Three independent random seeds per model
- Hardware: single consumer GPU (CUDA)

Table 1: HAM10000 Results (mean ±std over 3 seeds)

Model	Accuracy	Macro-F1	Melanoma Recall
MobileNetV2	0.780 ±0.011	0.655 ±0.009	0.604 ±0.033
EfficientNet-B0	0.765 ±0.015	0.635 ±0.015	0.686 ±0.049
ResNet-50	0.722 ±0.010	0.594 ±0.009	0.607 ±0.072

Table 2: Second Dataset Results (mean ±std over 3 seeds)

Model	Accuracy	Macro-F1	Melanoma Recall
EfficientNet-B0	0.818 ±0.004	0.752 ±0.006	0.958 ±0.001
MobileNetV2	0.788 ±0.004	0.712 ±0.010	0.948 ±0.030
ResNet-50	0.778 ±0.018	0.706 ±0.011	0.957 ±0.016

5 Results

On the dermoscopic dataset, the lightweight MobileNetV2 performed best overall. On the larger clinical dataset, EfficientNet-B0 was superior. Melanoma recall was substantially higher on the second dataset, likely due to the larger number of melanoma examples and different visual characteristics.

6 Discussion

The results demonstrate that:

- Architecture ranking is dataset-dependent.
- Proper lesion-level splitting and near-duplicate removal are essential for honest evalua-tion.
- Lightweight models can match or exceed heavier ones when data volume and quality are controlled.
- High melanoma recall on the clinical dataset is encouraging but must be interpreted cau-tiously because of possible residual label noise and domain differences.

Limitations include the relatively short training schedules, single-GPU setting, and the fact that the second dataset is a



heterogeneous Kaggle collection rather than a carefully curated clinical study.

7 Conclusion

We presented a reproducible multi-seed comparison of three popular CNN architectures on two complementary skin image datasets. MobileNetV2 led on HAM10000 while EfficientNet-B0 led on the cleaned clinical dataset. The accompanying code, splits, and seed-level results are publicly available to support further research. This work is strictly for academic and research purposes and is not intended for clinical diagnosis or decision-making.

Ethics Statement

This study used only publicly available datasets. No new patient data were collected. The models are research prototypes and must not be used for medical diagnosis.

Data and Code Availability

All notebooks, cleaned splits, seed results, and near-duplicate verification files are available at:

https://github.com/manisharoshini/Skin-disease-prediction-

AI Assistance Disclosure

Portions of the code and manuscript text were developed with assistance from large language models. The author takes full responsibility for all experimental results, claims, and the final content of the paper.

Acknowledgements

The author thanks Dr. Abuzar Ansari for guidance and SIES College of Arts, Science and Commerce for support.

References

[1] 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.

[2] M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 6105–6114.

[3] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 4510–4520.

[4] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.

[5] A. Esteva *et al.*, “Dermatologist-level classification of skin cancer with deep neural net-works,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[6] N. Codella *et al.*, “Skin lesion analysis toward melanoma detection: A challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), hosted by the Interna-tional Skin Imaging Collaboration (ISIC),” in *Proc. IEEE 15th Int. Symp. Biomed. Imag-ing (ISBI)*, 2018, pp. 168–172.

[7] N. C. F. Codella *et al.*, “Skin lesion analysis toward melanoma detection 2018: A chal-lenge hosted by



the International Skin Imaging Collaboration (ISIC)," arXiv:1902.03368, 2019.

[8] H. A. Haenssle *et al.*, "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Ann. Oncol.*, vol. 29, no. 8, pp. 1836–1842, 2018.

[9] T. J. Brinker *et al.*, "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task," *Eur. J. Cancer*, vol. 113, pp. 47–54, 2019.

[10] N. Gessert *et al.*, "Skin lesion classification using ensembles of multi-resolution Efficient-Nets with meta data," *Methods*, vol. 179, pp. 1–8, 2020.

[11] A. Mahbod *et al.*, "Fusing fine-tuned deep features for skin lesion classification," *Comput. Med. Imaging Graph.*, vol. 71, pp. 19–29, 2019.

[12] C. Yoon *et al.*, "A deep learning-based approach for skin disease classification using Mo-bileNetV2," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, 2021.

[13] M. A. Albahar, "Skin lesion classification using convolutional neural network with novel regularizer," *IEEE Access*, vol. 7, pp. 38306–38313, 2019.

[14] S. S. Chaturvedi, K. Gupta, and P. S. Prasad, "Skin lesion analyser: An efficient deep learning model for multi-class classification of skin lesions using MobileNetV2," in *Proc. Int. Conf. Adv. Comput. Data Sci.*, 2020.

[15] J. Kawahara, A. BenTaieb, and G. Hamarneh, "Deep features to classify skin lesions," in *Proc. IEEE 13th Int. Symp. Biomed. Imaging (ISBI)*, 2016, pp. 1397–1400.

[16] M. E. Celebi, N. Codella, and A. Halpern, "Dermoscopy image analysis: Overview and future directions," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 474–478, 2019.

[17] Z. Mirikharaji *et al.*, "A survey on deep learning for skin lesion segmentation," *Med. Image Anal.*, vol. 88, Art. no. 102863, 2023.

[18] M. K. Hasan *et al.*, "A survey, review, and future trends of skin lesion segmentation and classification," *Comput. Biol. Med.*, vol. 155, Art. no. 106624, 2023.

[19] Y. Li and L. Shen, "Skin lesion analysis towards melanoma detection using deep learning network," *Sensors*, vol. 18, no. 2, Art. no. 556, 2018.

[20] T.-C. Pham *et al.*, "Deep CNN models for classifying and localizing skin lesions," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2020.