ADBA
Artificial Intelligence in Applied Sciences
Cutting-Edge Scientific Solutions

e-ISSN: 3108-4060
RESEARCH ARTICLE
Vol.2 / No.1 / 2026 / pp.8-14
https://doi.org/10.69882/adba.ai.2026012

Comparative Evaluation of Convolutional Neural Network Architectures for Automated Skin Cancer Classification: A Study on the ISIC 2018 Dataset

Sifeu Takougang Kingni [*,1] **and Said Baawain** [α,2]

*Department of Mechanical, Petroleum and Gas Engineering, National Advanced School of Mines and Petroleum Industries, University of Maroua, P.O. Box 46, Maroua, Cameroon, αDepartment of Mechanical Engineering, Military Technological College, Muscat, P.O. Box 262, PC 111, Sultanate of Oman.

ABSTRACT The increasing global incidence of skin cancer, particularly lethal malignant melanoma, necessitates the development of robust, automated diagnostic tools to assist dermatologists in identifying subtle pathological markers. In this study, we provide a rigorous comparative evaluation of four state-of-the-art convolutional neural network (CNN) architectures, ResNet-50, DenseNet-169, Inception-v3, and EfficientNet-B0, using the ISIC 2018 (HAM10000) dataset. Our standardized experimental pipeline utilized stratified sampling to address class imbalance, alongside a meticulous preprocessing strategy and data augmentation to ensure model generalization. Quantitative results demonstrate that EfficientNet-B0 outperformed other models, achieving a peak accuracy of 91.84% and a superior F1-Score of 0.8429, despite possessing the most compact parameter footprint of 4.02M. While ResNet-50 exhibited lower diagnostic precision, it offered the fastest inference speed (0.359 ms), highlighting a critical trade-off between accuracy and real-time operational latency. Furthermore, visual validation through Grad-CAM++ confirmed that successful predictions were driven by relevant morphological hallmarks rather than dataset artifacts. Our findings suggest that architectural optimization through compound scaling is more effective than raw model depth for dermatological tasks. Collectively, this work provides a comprehensive framework for selecting deep learning backbones for clinical triage, balancing high-precision diagnostic support with the computational constraints of real-world medical deployment.

KEYWORDS
Convolutional neural networks (CNNs)
Skin cancer classification
EfficientNet-B0
Grad-CAM++
HAM10000 dataset

INTRODUCTION

Skin cancer remains one of the most significant public health challenges globally, with its incidence rates climbing steadily over the past few decades. Among various types, malignant melanoma stands out as the most lethal form, yet it is highly curable if identified in its nascent stages (Leiter *et al.* 2020). However, the diagnostic process is inherently complex; dermatologists must distinguish between a wide array of look-alike lesions, often relying on dermoscopy to visualize sub-surface structures (Siegel *et al.* 2024). Despite the expertise of clinicians, the visual ambiguity of skin lesions, characterized by varying colors, textures, and irregular borders, introduces a level of subjectivity that can lead to diagnostic inconsistency. Consequently, there is an urgent clinical demand for objective, automated screening tools that can support early intervention and improve patient outcomes (Gloster and Neal 2006; Armstrong and Kricker 1995).

The advent of Deep Learning (DL), particularly Convolutional Neural Networks (CNNs), has fundamentally transformed the landscape of medical image analysis (Pacal *et al.* 2024; Cakmak and Pacal 2025b; Cakmak and Maman 2025; Pacal and Cakmak 2025a). These architectures possess the unique ability to automatically learn hierarchical feature representations directly from raw pixel data, capturing intricate patterns that may be imperceptible to the human eye (Cakmak and Zeynalov 2025; Zeynalov *et al.* 2025; Pacal and Cakmak 2025b; Cakmak and Pacal 2025a). In the realm of dermatology, CNNs have shown remarkable potential in automating the classification of skin lesions across diverse diagnostic categories. By leveraging vast repositories of dermatoscopic images, such as the HAM10000 dataset, these models can be trained to recognize the subtle morphological hallmarks of malignancy with a level of precision that occasionally rivals or exceeds that of board-certified specialists (Chaurasia *et al.* 2025; Manju *et al.* 2025).

While literature is replete with various DL approaches, the selection of an optimal architecture remains a non-trivial task. Modern clinical environments require a delicate balance between high diagnostic accuracy and computational efficiency, especially when considering deployment on resource-constrained hardware or real-time diagnostic platforms. Architectures like ResNet have introduced residual learning to overcome the vanishing gradient problem, while DenseNet emphasizes feature reuse through dense connections (Karthik *et al.* 2024; Pacal *et al.* 2025). More recent

Manuscript received: 3 November 2025,
Revised: 28 December 2025,
Accepted: 20 January 2026.

[1]stkingni@gmail.com
[2]saidbaawain14@gmail.com(**Corresponding author**).

innovations, such as EfficientNet, have pushed the boundaries further by scaling depth, width, and resolution simultaneously to maximize performance while minimizing the parameter footprint. However, a comprehensive comparative analysis is necessary to determine how these varying design philosophies perform specifically on multi-class skin lesion datasets (Ozdemir and Pacal 2025).

In this study, we present a rigorous comparative evaluation of four state-of-the-art CNN architectures, ResNet-50, DenseNet-169, Inception-v3, and EfficientNet-B0, to identify the most robust backbone for skin cancer classification. Utilizing the ISIC 2018 (HAM10000) dataset, we implemented a standardized pipeline involving stratified sampling and consistent preprocessing to ensure a fair performance assessment. Beyond traditional accuracy metrics, our analysis delves into the trade-offs between model complexity (Params), computational load (Gflops), and real-world inference speed. By integrating these quantitative results with visual validation through Grad-CAM++, we aim to provide a holistic framework that not only identifies the most accurate model but also offers insights into its clinical reliability and interpretability.

RELATED WORKS

DL frameworks relying on CNNs continue to be refined through advanced optimization and ensemble strategies to enhance diagnostic precision. Farea *et al.* (2024) proposed a hybrid framework that addresses data scarcity by curating a generalized dataset from multiple sources and employing the Artificial Bee Colony (ABC) algorithm to optimize the initial weights of an Xception model. This approach aimed to mitigate the risk of local minima during training, ultimately achieving a high accuracy of 93.04% by effectively fine-tuning the learnable parameters on segmented lesion regions.

Similarly, efforts to maximize feature representation have led to the development of complex ensemble architectures. Akter et al. (2024) (r19 2024) introduced an integrated DL model that fuses the feature outputs of InceptionV3 and DenseNet121 using a weighted sum rule at the score level. Their methodology incorporated extensive data augmentation to resolve class imbalance, resulting in a robust system that achieved a detection accuracy of 92.27% on the ISIC dataset.

Addressing the "black-box" nature of deep learning, explainability has become a central focus alongside classification performance. Attallah (2024) developed "Skin-CAD," an explainable CAD system that aggregates features from multiple CNN layers and employs Principal Component Analysis (PCA) to reduce dimensionality before classification. This system not only classified lesions into seven subtypes with 97.2% accuracy but also integrated LIME (Local Interpretable Model-agnostic Explanations) to provide visual justifications for the model's predictions, thereby enhancing clinical trust.

Furthermore, lightweight CNN architecture remains vital for deployment in resource-constrained environments. Owida *et al.* (2024) designed a custom CNN architecture trained on the HAM10000 dataset, emphasizing the importance of preprocessing techniques such as morphological filtration for hair removal. Their streamlined model achieved a high efficiency of 95.23%, demonstrating that custom-built CNNs can still compete with heavier pre-trained models when data quality is rigorously managed.

The integration of meta-heuristic algorithms for hyperparameter optimization within complex neural architectures has become a prominent research direction to enhance segmentation accuracy. Ali *et al.* (2024) proposed a hybrid framework for dermoscopic image segmentation based on a fully convolutional encoder-decoder network (FCEDN) optimized via the Sparrow Search Algorithm (SpaSA). In this study, the individual wolf method and ensemble ghosting techniques were integrated into the SpaSA to maintain an effective balance between navigation and exploitation during the search process. Their proposed FCEDN-SpaSA architecture achieved high segmentation performance on datasets such as ISBI 2017 and PH2, while the adaptive CNN classification module reached a 91.67% accuracy rate with significantly lower energy, storage space, and memory access compared to conventional incremental learning techniques.

In parallel, researchers have focused on leveraging signal processing techniques to reinforce feature representation in the frequency domain. Claret *et al.* (2024) introduced an innovative approach combining Discrete Wavelet Transformation (DWT) with CNN models for enhanced skin cancer diagnosis. In this methodology, dermoscopic images are decomposed into multiple sub-images characterized by different spatial domains and independent frequencies (LL, LH, HH, HL). By utilizing the Low-Low (LL) features, which retain 50% of the relevant pixels from the original image, the model achieves effective dimensionality reduction and improved computational efficiency. Supported by a softmax activation function, this model achieved a sensitivity of 94% and a specificity of 91% on the HAM10000 dataset, significantly outperforming traditional artificial neural networks (ANN) and multilayer perceptron methods.

To address the challenges of high annotation costs and significant class imbalance in medical datasets, the synergy between active learning (AL) and optimization algorithms has emerged as a critical strategy. Mandal *et al.* (2024) proposed an efficient framework that integrates AL with Particle Swarm Optimization (PSO) to selectively identify the most informative unlabeled instances for expert annotation. This method utilizes PSO to enhance the selection process within the AL framework, ensuring that the model prioritizes training on the most uncertain and challenging samples. Experimental results using the EfficientNetV2M architecture demonstrated that the AL-PSO approach, specifically through the "Least Confidence" strategy, achieved a classification accuracy of 89.44% while requiring only 40% of the labeled training data. This approach offers a robust, cost-effective solution for clinical settings where labeled data is scarce.

MATERIALS AND METHODS

Dataset and Data Preprocessing

The foundation of this study is the HAM10000 ("Human Against Machine with 10,000 training images") dataset, which was prominently featured in the ISIC 2018 challenge (r25 2025). This benchmark repository is comprised of 10,011 high-quality dermatoscopic images collected from multiple clinical sites, representing a broad demographic and a wide variety of acquisition conditions. As showcased in Figure 1, the dataset captures the complex visual morphology of seven distinct diagnostic categories: actinic keratoses and intraepithelial carcinoma (AKIEC), basal cell carcinoma (BCC), benign keratosis-like lesions (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevi (NV), and vascular lesions (VASC). Each category presents unique challenges, such as varying pigment patterns and irregular borders, which the models must learn to navigate for accurate classification.

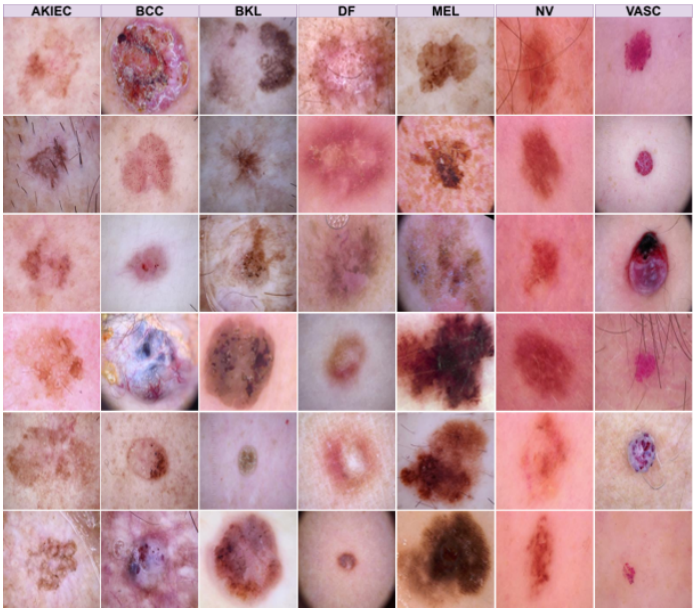

Figure 1 Representative dermoscopic image samples illustrating the morphological diversity of the seven skin lesion classes in the ISIC 2018 dataset

A significant aspect of this dataset is its inherent class imbalance, a characteristic reflective of real-world clinical distributions where benign cases often outnumber malignant ones. The precise numerical distribution of these classes across our experimental subsets is detailed in Table 1. The dataset is heavily skewed toward Melanocytic nevi (NV), which accounts for 6,705 images, whereas classes like Dermatofibroma (DF) are represented by only 115 samples. To address this and ensure the statistical validity of our performance metrics, we utilized a stratified sampling strategy to partition the data. This method meticulously preserved the original class ratios across the training (70%), validation (15%), and test (15%) sets, resulting in 7,005, 1,498, and 1,508 images respectively.

To ensure that the DL architectures could effectively extract meaningful features, we implemented a rigorous preprocessing and standardization pipeline. Raw images were first resized to a uniform resolution of 224×224 pixels to maintain consistency with the input requirements of the pre-trained backbones. Subsequently, pixel values were normalized using the global mean and standard deviation of the ImageNet dataset. This normalization is crucial for medical image analysis as it helps stabilize training dynamics and ensures that the model's attention mechanisms remain focused on lesion-specific pathological markers rather than being distracted by variations in lighting or resolution scale. By standardizing the input data in this manner, we created a level playing field for the comparative evaluation of the different CNN architectures.

Foundational Principles of Convolutional Neural Networks (CNNs)

At the heart of the recent revolution in medical image analysis lies the shift from manual, heuristic-based feature engineering to the automated, data-driven paradigm of CNNs. Unlike traditional computer vision techniques, CNNs are designed to mimic the human visual system by automatically learning hierarchical feature representations directly from raw pixel data. Through a series of specialized layers, primarily convolutional, pooling, and non-linear activation layers, these networks decompose complex dermatological structures into a multi-level abstraction of spatial patterns. Early layers typically capture low-level morphological hallmarks such as edges and color gradients, while deeper layers integrate these into high-level semantic descriptors capable of identifying the subtle diagnostic markers of malignancy. This inherent ability to preserve spatial localities while reducing dimensionality makes CNNs exceptionally robust for classifying diverse skin lesion categories (O'Shea and Nash 2015).

The mathematical rigor of these architectures is further enhanced by advanced structural innovations designed to optimize the learning process. To address the challenges of training deep networks, such as the vanishing gradient problem, modern backbones incorporate specialized design philosophies: ResNet (He *et al.* 2015) utilizes residual learning through skip connections to facilitate the flow of gradients, while DenseNet (Huang *et al.* 2017) promotes feature reuse by connecting every layer to every subsequent layer. More sophisticated frameworks, such as EfficientNet (Tan and Le 2019), employ compound scaling to balance network depth, width, and resolution, thereby maximizing diagnostic precision while maintaining a compact parameter footprint. During the training phase, these models are optimized through the minimization of a Cross-Entropy Loss function, which penalizes discrepancies between predicted and actual diagnostic labels. By employing robust optimizers like AdamW and dynamic learning rate schedulers, the network weights are iteratively refined to settle into an optimal configuration that ensures both high accuracy and clinical reliability.

Data Augmentation Strategy

We opted for a data augmentation strategy based on the default settings of the timm (PyTorch Image Models) library to improve the model's ability to generalize across different clinical scenarios. By leveraging these standard configurations, we introduced a variety of transformations such as random rotations, horizontal flips, and resized cropping, which effectively mimic the natural variability in how medical images are captured and oriented. These techniques ensure that the network does not simply memorize the training data but instead learns to recognize key pathological features regardless of their scale or position within the frame. Relying on the proven defaults of the timm framework allowed us to maintain a rigorous and reproducible training pipeline, providing strong regularization that balances complexity with the need for robust performance on unseen medical datasets (Wang *et al.* 2024; Mumuni *et al.* 2024).

Experimental Design and Training Protocol

To ensure the technical rigor and reproducibility of our experimental framework, we implemented all CNN architectures using the PyTorch library on a high-end workstation equipped with an NVIDIA RTX 5090 GPU (32GB VRAM), which provided the necessary computational power for efficient model convergence. The dataset was partitioned into training (70%, n=7,005), validation (15%, n=1,498), and testing (15%, n=1,508) subsets using a stratified sampling strategy to maintain consistent class proportions across all phases. Prior to training, each image was resized to a uniform 224×224 resolution and normalized according to ImageNet standards, a step crucial for stabilizing the learning process and ensuring the CNN backbones could focus on subtle, lesion-specific morphological features. We optimized the network parameters using the AdamW algorithm, carefully tuning the learning rate and weight decay to maintain a balance between convergence speed and generalization. The final model selection was determined by the best performance on the validation set, and the diagnostic efficacy was rigorously measured on the independent test set using

Table 1 Distribution of the HAM10000 dataset across seven diagnostic categories for training, validation, and testing subsets

Class Name	Total	Train	Val	Test
BKL	1099	769	164	166
DF	115	80	17	18
VASC	142	99	21	22
AKIEC	323	226	48	49
MEL	1113	779	166	168
BCC	514	359	77	78
NV	6705	4693	1005	1007
Grand Total	10011	7005	1498	1508

a comprehensive suite of metrics, including Accuracy, Precision, Recall, and F1-Score, across the seven diagnostic categories.

Performance Evaluation Metrics

To rigorously benchmark the diagnostic efficacy and clinical reliability of the evaluated CNN architectures, we utilized a comprehensive suite of statistical metrics derived from the multi-class confusion matrix. While overall Accuracy (1) provides a global measure of the model's classification success, the inherent class imbalance within the HAM10000 dataset, where certain benign cases significantly outnumber malignant ones, necessitates a more nuanced evaluation. To this end, we employed Precision (2) to quantify the models' predictive exactness and Recall (3), or sensitivity, to ensure the critical detection of malignant lesions that require early intervention. Given the natural trade-off between these two dimensions in dermatological screening, we prioritized the F1-Score (4) as a robust harmonic mean that balances precision and recall. Collectively, these metrics provide a holistic framework for assessing each architecture's ability to generalize across diverse pathological markers while maintaining the high level of precision required for real-world clinical decision support.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

RESULTS

Quantitative Performance and Comparative Analysis

The objective evaluation of the selected CNN architectures was conducted through a multi-dimensional analysis that integrates both statistical classification metrics and computational efficiency parameters. This comprehensive assessment allows for a nuanced understanding of how different design philosophies, ranging from residual learning to compound scaling, respond to the complex morphological variations inherent in dermatoscopic imagery. The following sections detail the empirical findings derived from the independent test set, with a specific focus on the trade-offs between diagnostic precision and the technical constraints of real-time clinical deployment.

As summarized in Table 2, the experimental results reveal a distinct performance hierarchy among the evaluated models. EfficientNet-B0 emerged as the most robust architecture, achieving the highest overall Accuracy of 91.84% and a superior F1-Score of 0.8429. It is particularly noteworthy that EfficientNet-B0 achieved these results with the most compact parameter footprint in the group, utilizing only 4.02M parameters and requiring just 0.734 Gflops. This suggests that its compound scaling method, which simultaneously optimizes network depth, width, and resolution, is highly effective for capturing the intricate textural markers required for skin lesion classification.

In contrast, ResNet-50 demonstrated the lowest classification performance, with an Accuracy of 88.06% and an F1-Score of 0.7752, despite having a significantly larger parameter count of 23.52M. However, the data highlights a critical trade-off regarding operational latency: ResNet-50 recorded the fastest Inference Time (0.359 ms), whereas EfficientNet-B0 exhibited the longest latency at 5.9026 ms. This discrepancy indicates that while EfficientNet-B0 provides the most accurate diagnostic support, ResNet-50 or the mid-performing Inception-v3 (89.52% accuracy) and DenseNet-169 (90.32% accuracy) might be more suitable for high-throughput screening environments or deployment on edge-computing hardware with strict real-time requirements.

A granular look at the classification behavior of the top-performing model is provided by the confusion matrix in Figure 2. The model shows exceptional sensitivity toward NV, correctly classifying 980 instances. However, the matrix also reveals persistent diagnostic challenges; for example, 26 cases of MEL were misidentified as NV, and 13 cases of AKIEC were confused with BKL. These specific error patterns underscore the visual ambiguity between certain malignant conditions and their benign mimics, which remains a primary hurdle in automated dermatological assessment.

Table 2 Quantitative performance comparison of CNN architectures based on accuracy, precision, recall, F1-score, and computational efficiency metrics

Models	Accuracy	Precision	Recall	F1 Score	Params (M)	Gflops	Inference Time (Ms)
ResNet-50 (He *et al.* 2015)	0.8806	0.7972	0.7576	0.7752	23.52	8.2634	0.359
DenseNet-169 (Huang *et al.* 2017)	0.9032	0.8483	0.8245	0.8314	12.5	6.7169	0.7204
EfficientNet-B0 (Tan and Le 2019)	0.9184	0.8905	0.8068	0.8429	4.02	0.734	5.9026
Inception-v3 (Szegedy *et al.* 2016)	0.8952	0.8474	0.8193	0.8302	21.8	5.6719	0.4328

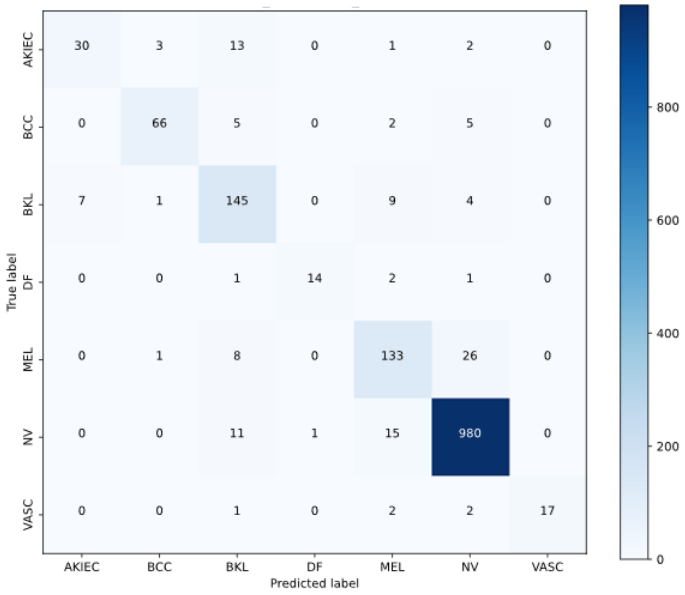

Figure 2 Confusion matrix of the EfficientNet-B0 model illustrating raw classification counts and specific inter-class error patterns across the independent test set

Visual Validation of Model Focus with Grad-CAM++

To ensure that our models were making decisions based on relevant pathological features rather than dataset artifacts, we utilized Grad-CAM++ for visual explanation. As shown in Figure 3, the "Correct Grad-CAM Predictions" row demonstrates that for successfully classified lesions across all seven categories, the model's attention was tightly localized on the lesion's core and its irregular borders. In contrast, the "Misclassified Grad-CAM Predictions" reveal that errors often stemmed from the model focusing on peripheral skin regions or being distracted by clinical artifacts like hair or skin folds, rather than the primary lesion. This visual evidence underscores the necessity of robust preprocessing and the potential of explainability tools to build clinician trust in "black-box" models.

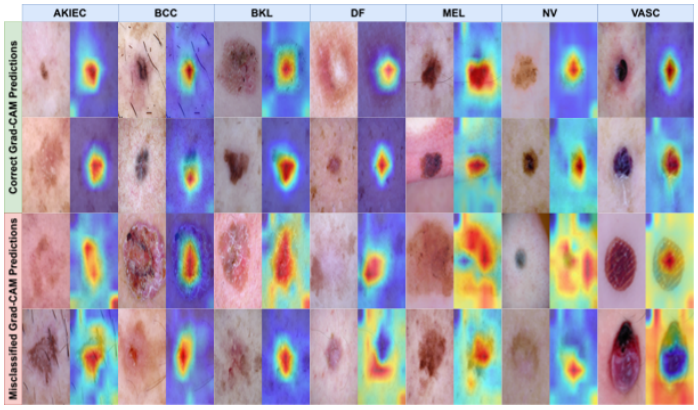

Figure 3 Visual explanation of model interpretability using Grad-CAM++ heatmaps for correct and misclassified skin lesion predictions

DISCUSSION

Interpretation of Key Findings

The results of this study reaffirm that architectural design often outweighs raw model size in medical imaging tasks. The success of EfficientNet-B0 can be attributed to its optimized balancing of network depth, width, and resolution, which proved more effective for the high-frequency textural details of dermatoscopic images than the traditional residual blocks of ResNet-50. Interestingly, the dense connectivity of DenseNet-169 provided a significant boost in recall (0.8245) compared to ResNet-50 (0.7576), suggesting that feature reuse is particularly beneficial for identifying minority classes like DF and VASC. These findings indicate that while accuracy is a valuable metric, the choice of a backbone must be dictated by the specific clinical priority, whether it be the highest possible precision (EfficientNet) or real-time processing speed (ResNet).

Clinical Implications, Limitations, and Future Directions

From a clinical standpoint, the high F1-Scores achieved by our top-performing models suggest they could serve as robust triage tools in primary care settings, potentially reducing the diagnostic

burden on specialists. However, the misclassification of melanoma as benign nevi, as seen in Figure 2, remains a primary limitation. This error is likely exacerbated by the dataset's heavy skew toward the NV class. Future research should focus on integrating cost-sensitive learning or synthetic oversampling (e.g., SMOTE) to specifically penalize "false benign" predictions for malignant cases. Additionally, while our use of Grad-CAM++ provided essential interpretability, moving toward "interpretable-by-design" architectures like Vision Transformers (ViTs) could further enhance the transparency and reliability of automated skin cancer screening in real-world practice.

CONCLUSION

This research underscores that architectural design, specifically the compound scaling of depth, width, and resolution found in EfficientNet-B0, consistently outweighs raw parameter count in complex medical imaging tasks such as skin cancer classification. While EfficientNet-B0 established a new benchmark for accuracy on the HAM10000 dataset, our analysis also revealed an essential operational trade-off: the faster inference of ResNet-50 makes it highly viable for low-latency edge-computing, whereas EfficientNet's superior precision is better suited for high-stakes diagnostic support. The integration of Grad-CAM++ provided a vital layer of interpretability, confirming that our models localized on legitimate pathological markers. Despite these successes, the persistent challenge of distinguishing melanoma from common nevi due to class imbalance remains a hurdle for widespread clinical adoption. Future research should prioritize cost-sensitive learning to penalize "false benign" errors and explore "interpretable-by-design" architectures like Vision Transformers to further enhance transparency and clinician trust in automated screening systems.

Ethical standard

Not applicable.

Availability of data and material

The dataset analyzed for this study is the public dataset, which is available on Kaggle: https://www.kaggle.com/datasets/surajghuwalewala/ham1000-segmentation-and-classification?select=GroundTruth.csv

Conflicts of interest

The authors declare that they have no conflicts of interest.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used artificial intelligence tools in order to improve the readability and language quality of the manuscript. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

LITERATURE CITED

2024 An integrated deep learning model for skin cancer detection using hybrid feature fusion technique.

2025 Skin cancer: Ham10000. Kaggle Dataset.

Ali, R., A. Manikandan, R. Lei, and J. Xu, 2024 A novel spasa based hyper-parameter optimized fcedn with adaptive cnn classification for skin cancer detection. Scientific Reports **14**: 9336.

Armstrong, B. K. and A. Kricker, 1995 Skin cancer. Dermatologic Clinics **13**: 583–594.

Attallah, O., 2024 Skin-cad: Explainable deep learning classification of skin cancer from dermoscopic images by feature selection of dual high-level cnns features and transfer learning. Computers in Biology and Medicine **178**.

Cakmak, Y. and A. Maman, 2025 Deep learning for early diagnosis of lung cancer. Computational Systems and Artificial Intelligence **1**: 20–25.

Cakmak, Y. and I. Pacal, 2025a Comparative analysis of transformer architectures for brain tumor classification. Exploration of Medicine **6**.

Cakmak, Y. and N. Pacal, 2025b Deep learning for automated breast cancer detection in ultrasound: A comparative study of four cnn architectures. Artificial Intelligence in Applied Sciences **1**: 13–19.

Cakmak, Y. and J. Zeynalov, 2025 A comparative analysis of convolutional neural network architectures for breast cancer classification from mammograms. Artificial Intelligence in Applied Sciences **1**: 28–34.

Chaurasia, A. K., P. W. Toohey, H. C. Harris, and A. W. Hewitt, 2025 Multi-resolution vision transformer model for histopathological skin cancer subtype classification using whole slide images. Computers in Biology and Medicine **196**.

Claret, S. P. A., J. P. Dharmian, and A. M. Manokar, 2024 Artificial intelligence-driven enhanced skin cancer diagnosis: leveraging convolutional neural networks with discrete wavelet transformation. Egyptian Journal of Medical Human Genetics **25**: 50.

Farea, E., R. A. A. Saleh, H. AbuAlkebash, A. A. R. Farea, and M. A. Al-antari, 2024 A hybrid deep learning skin cancer prediction framework. Engineering Science and Technology, an International Journal **57**.

Gloster, H. M. and K. Neal, 2006 Skin cancer in skin of color. Journal of the American Academy of Dermatology **55**: 741–760.

He, K., X. Zhang, S. Ren, and J. Sun, 2015 Deep residual learning for image recognition.

Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, 2017 Densely connected convolutional networks.

Karthik, R., R. Menaka, S. Atre, J. Cho, and S. V. Easwaramoorthy, 2024 A hybrid deep learning approach for skin cancer classification using swin transformer and dense group shuffle non-local attention network. IEEE Access **12**: 158040–158051.

Leiter, U., U. Keim, and C. Garbe, 2020 Epidemiology of skin cancer: Update 2019. Advances in Experimental Medicine and Biology **1268**: 123–139.

Mandal, S., S. Ghosh, N. D. Jana, S. Chakraborty, and S. Mallik, 2024 Active learning with particle swarm optimization for enhanced skin cancer classification utilizing deep cnn models. Journal of Imaging Informatics in Medicine **38**: 2472–2489.

Manju, V. N., D. S. Dayana, N. Patwari, K. P. B. Madavi, and K. K. Sowjanya, 2025 Attention-enhanced vision transformer model for precise skin cancer detection. In *Proceedings of the 2025 International Conference on Emerging Technologies in Computing and Communication (ETCC)*.

Mumuni, A., F. Mumuni, and N. K. Gerrar, 2024 A survey of synthetic data augmentation methods in machine vision. Machine Intelligence Research **21**: 831–869.

O'Shea, K. and R. Nash, 2015 An introduction to convolutional neural networks. International Journal of Research in Applied Science and Engineering Technology **10**: 943–947.

Owida, H. A., N. Alshdaifat, A. Almaghthawi, S. Abuowaida, A. Aburomman, *et al.*, 2024 Improved deep learning architecture for skin cancer classification. Indonesian Journal of Electrical Engineering and Computer Science **36**: 501–508.

Ozdemir, B. and I. Pacal, 2025 An innovative deep learning framework for skin cancer detection employing convnextv2 and focal self-attention mechanisms. Results in Engineering **25**.

Pacal, I., M. Alaftekin, and F. D. Zengul, 2024 Enhancing skin cancer diagnosis using swin transformer with hybrid shifted window-based multi-head self-attention and swiglu-based mlp. Journal of Imaging Informatics in Medicine **37**: 3174–3192.

Pacal, I. and Y. Cakmak, 2025a A comparative analysis of u-net-based architectures for robust segmentation of bladder cancer lesions in magnetic resonance imaging. Eurasian Journal of Medicine and Oncology **9**: 268–283.

Pacal, I. and Y. Cakmak, 2025b *Diagnostic Analysis of Various Cancer Types with Artificial Intelligence*. Duvar Yayınları.

Pacal, I., B. Ozdemir, J. Zeynalov, H. Gasimov, and N. Pacal, 2025 A novel cnn-vit-based deep learning model for early skin cancer diagnosis. Biomedical Signal Processing and Control **104**.

Siegel, R. L., A. N. Giaquinto, and A. Jemal, 2024 Cancer statistics, 2024. CA: A Cancer Journal for Clinicians **74**: 12–49.

Szegedy, C., V. Vanhoucke, S. Ioffe, and J. Shlens, 2016 Rethinking the inception architecture for computer vision.

Tan, M. and Q. V. Le, 2019 Efficientnet: Rethinking model scaling for convolutional neural networks.

Wang, Z., P. Wang, K. Liu, P. Wang, Y. Fu, *et al.*, 2024 A comprehensive survey on data augmentation. arXiv preprint .

Zeynalov, J., Y. Cakmak, and I. Pacal, 2025 Automated apple leaf disease classification using deep convolutional neural networks: A comparative study on the plant village dataset. Journal of Computer Science and Digital Technologies **1**: 5–17.

How to cite this article: Kingni, S. T. and Baawain, S. Comparative Evaluation of Convolutional Neural Network Architectures for Automated Skin Cancer Classification: A Study on the ISIC 2018 Dataset. *Artificial Intelligence in Applied Sciences*, 2(1), 8-14, 2026.