Journal of Soft Computing Exploration
Homepage: htttps://shmpublisher.com/index.php/joscex
p-ISSN: 2746-7686
e-ISSN: 2746-0991

.

Benchmarking deep transfer learning for imbalanced skin cancer classification: Integrating focal loss, explainable AI, and web deployment

Yazid Aufar[1], Muhammad Daffa Abiyyu Rahman[2], M. Fadli Ridhani[3]
[1]Informatics Engineering, Politeknik Hasnur, Indonesia
[2]Electrical Engineering, Universitas Lambung Mangkurat, Indonesia
[3]Multimedia Engineering Technology, Politeknik Hasnur, Indonesia

Article Info

Article history:

Received March 14, 2026
Revised March 18, 2026
Accepted March 22, 2026

Keywords:

Deep transfer learning
Explainable AI
Focal loss
Skin cancer classification
Web deployment

ABSTRACT

Non-melanoma skin cancer (NMSC) classification faces challenges like severe data imbalance and the "black-box" nature of AI, limiting clinical trust. This study benchmarks four pre-trained convolutional models (ConvNeXt-Tiny, EfficientNetV2-S, DenseNet121, MobileNetV3-Large) for the imbalanced multi-class classification of Squamous Cell Carcinoma, Actinic Keratosis, and benign Nevus. Images were preprocessed using morphological hair removal and inpainting. The methodology integrated a 5-fold Stratified Group-KFold cross-validation, Focal Loss to address class imbalance, and Grad-CAM for Explainable AI (XAI) transparency. Results showed ConvNeXt-Tiny achieved the highest and most stable performance with a Balanced Accuracy of 76.98% (± 0.31 standard deviation) and a Macro F1-Score of 0.7513, significantly outperforming the other architectures. Grad-CAM confirmed the model's precise focus on pathological lesion borders. Ultimately, the optimal model was deployed as a real-time Streamlit web application, establishing a robust and practical clinical decision-support system.

Corresponding Author:

Yazid Aufar,
Department of Informatics Engineering,
Politeknik Hasnur, Kalimantan Selatan, Indonesia
Email: yazid.aufar.ya@gmail.com
https://doi.org/10.52465/joscex.v7i1.20

1. INTRODUCTION

Non-melanoma skin cancer (NMSC), primarily comprising Squamous Cell Carcinoma (SCC) and its pre-malignant precursor, Actinic Keratosis (AK), represents a growing global health burden with a steadily increasing incidence rate [1]–[4]. While NMSC generally exhibits lower mortality rates compared to malignant melanoma, SCC possesses a significant risk of metastasis and local tissue invasion if not diagnosed and managed in its early stages [5], [6]. Traditionally, clinical diagnosis depends on dermatologists visually inspecting dermoscopic images. This manual process requires a considerable amount of time, and its accuracy relies heavily on the individual doctor's clinical experience, making it highly subjective [7]–[9]. In regions with limited access to specialized dermatological care, diagnostic delays can lead to suboptimal patient prognoses

and increased healthcare costs. Consequently, developing an automated, highly accurate, and reliable Computer-Aided Diagnosis (CAD) system is a critical objective within medical informatics [10]–[12].

In recent years, deep transfer learning has become a standard approach for automated skin lesion classification. Instead of training models from scratch, researchers use networks previously trained on large datasets like ImageNet to achieve high diagnostic accuracy more efficiently [13], [14]. Several studies have established baseline methods for skin cancer detection using traditional or lightweight Convolutional Neural Networks (CNNs). For instance, Srinivasu et al. [15] utilized MobileNetV2 integrated with long short-term memory (LSTM) networks for skin disease classification. While computationally efficient, lightweight models often struggle with highly imbalanced multi-class datasets. Similarly, De et al. [16] proposed a CNN-DenseNet model to extract features from histopathological images. Other researchers, such as Khattar and Kaur [10] and Mahmoud and Soliman [11], relied on foundational architectures like ResNet and classic DenseNet to build their automated diagnostic systems.

While these previous studies highlight significant progress in using traditional CNN baselines, a critical gap analysis reveals major limitations. First, real-world dermoscopic datasets, such as the widely used ISIC 2019 archive [17], suffer from extreme class imbalance where benign cases drastically outnumber malignancies. Previous research predominantly relied on standard Cross-Entropy loss functions, which create an algorithmic bias toward the majority class and result in high false-negative rates for critical minority classes like SCC [18], [19]. Second, the traditional architectures used in prior studies [11], [15], [16] rely on standard small convolutional filters that often fail to capture the complex, unstructured textural heterogeneity of NMSC lesions. Finally, the "black-box" nature of deep neural networks creates a substantial "trust deficit" among healthcare professionals [18], [20]. Many existing frameworks lack Explainable Artificial Intelligence (XAI) or are not deployed into practical clinical tools [21]–[23].

To address these limitations, this research proposes a superior framework using modernized architectures. Recently, researchers introduced ConvNeXt, which updates standard CNNs by adopting structural ideas from Vision Transformers [24], [25]. By utilizing larger 7x7 filters, ConvNeXt captures global image patterns more effectively than older baselines, offering a distinct advantage in analyzing complex skin lesions. Furthermore, to eliminate the bias found in earlier studies, we replace the standard loss function with Focal Loss [26], which dynamically penalizes the model for misclassifying the hard-to-detect SCC and AK minority classes.

Therefore, the primary objective of this study is to perform a comprehensive benchmarking of modern architectures (ConvNeXt-Tiny, EfficientNetV2-S) against traditional baselines (DenseNet121, MobileNetV3-Large) for the imbalanced multi-class classification of NMSC. The specific contributions of this research are three-fold: (1) integrating Focal Loss to robustly mitigate extreme class imbalance without artificially altering the dataset; (2) ensuring algorithmic transparency through Gradient-weighted Class Activation Mapping (Grad-CAM) to provide visual evidence that aligns with pathological criteria; and (3) deploying the optimal model into a real-time, Streamlit-based web application to establish its practical feasibility as an interpretable clinical decision-support system.

2. METHOD

This section outlines the methodology for developing a web-based system for automated skin lesion classification, integrating deep learning models with Explainable AI (XAI) techniques. The methodology is structured into the following key subsections: Data Acquisition and Preprocessing, Model Architectures and Training Methodology, Model Evaluation, Explainable AI Techniques, and Deployment of Web-based System for Clinical Testing. The steps of this research will be explained through the flowchart in Figure 1.

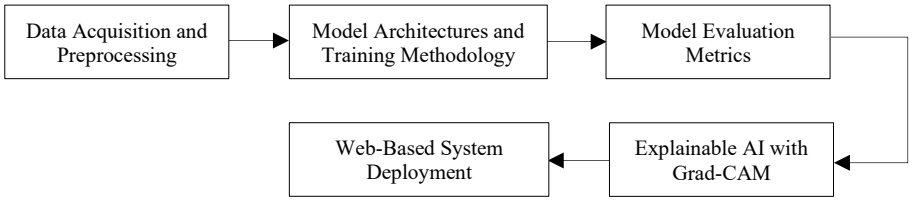

Figure 1. Flow of research

Data Acquisition and Preprocessing

The primary dataset for this research was sourced from the ISIC (International Skin Imaging Collaboration) 2019 challenge archive [17]. This study established a multi-class classification task focusing on three specific diagnostic categories: Squamous Cell Carcinoma (SCC), Actinic Keratosis (AK), and benign Nevus (NV). A defining characteristic of this dataset is its severe class imbalance, where benign NV cases represent approximately 89.6% of the total images, significantly outnumbering the malignant SCC and pre-malignant AK instances. This distribution reflects the high-skew nature of real-world clinical data, where benign lesions are far more prevalent than malignancies.

To ensure that the deep learning models extract features from actual pathological structures rather than extraneous noise, a digital hair removal pipeline was implemented. This preprocessing involved morphological Black-Hat filtering to isolate dark, hair-like artifacts, followed by the Telea inpainting algorithm to reconstruct the underlying skin texture based on neighboring pixels. The effectiveness of this pipeline across the SCC, AK, and NV classes is visually demonstrated in Figure 2. After artifact removal, all images were resized to a uniform dimension of 224×224 pixels and normalized using ImageNet-1k statistics ($mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]$) to ensure compatibility with pre-trained transfer learning architectures.

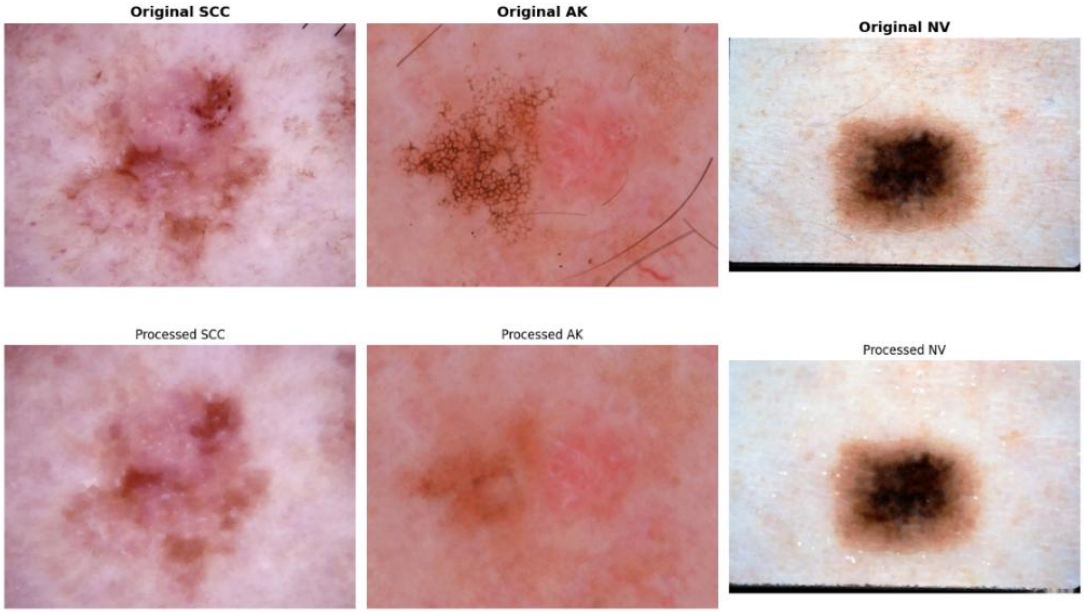

Figure 2. Comparison of original and hair-removed images across NMSC classes

To maintain clinical validity, a group-aware data splitting strategy was strictly enforced using the lesion_id metadata. This prevents "data leakage," where different images of the same lesion might otherwise appear in both training and evaluation subsets, leading to biased results. Approximately 10% of unique lesions were reserved as a hold-out test set. The remaining 90% of the pool was utilized for a 5-fold Stratified Group-KFold cross-validation, ensuring that each fold maintains a consistent class proportion while keeping patient data isolated. The detailed distribution of unique lesions and class-specific image counts across all subsets is summarized in Table 1.

Table 1. Detailed dataset distribution by lesion count and diagnostic class

Subset	SCC	AK	NV
Training Set	431	642	9,295
Validation Set	123	163	2,269
Test Set (Hold-out)	74	62	1,311
Overall Total	**628**	**867**	**12,875**

Furthermore, to mitigate the risk of overfitting on the minority classes (SCC and AK), a dynamic data augmentation strategy was applied during training using the Albumentations library. The augmentation pipeline generated various transformations in real-time, including horizontal and vertical flips, random rotations, color jittering for brightness and contrast, and CoarseDropout (cutout) to force the network to learn robust, distributed feature representations [27]. Representative examples of these dynamic transformations are depicted in Figure 3.

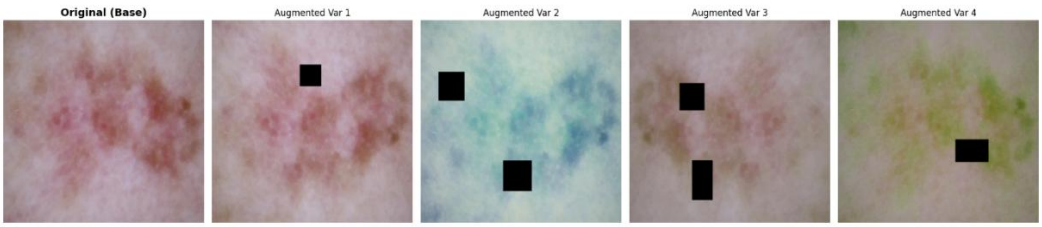

Figure 3. Examples of dynamic data augmentation applied to dermoscopic images

Model Architectures and Training Methodology

This study adopts a deep transfer learning paradigm, utilizing convolutional neural networks (CNNs) pre-trained on the ImageNet-1k dataset. To provide a comprehensive benchmarking analysis, four distinct architectures representing different technological generations were evaluated: EfficientNetV2-S, ConvNeXt-Tiny, DenseNet121, and MobileNetV3-Large.

Feature Extraction Architectures

Each selected model offers unique structural advantages for skin lesion classification:

- ConvNeXt-Tiny: A modernized CNN that adopts macro-designs from Vision Transformers (ViT). It utilizes a large 7×7 kernel and Layer Normalization (LN) to achieve a global receptive field [28]. The transformation block in ConvNeXt is mathematically expressed as shown in Equation (1).

$$\mathcal{F}(X) = Linear_{1\times1}(GELU(Linear_{1\times1}(LN(DWConv_{7\times7}(X))))) + X \quad (1)$$

- EfficientNetV2-S: Optimizes training speed and parameter efficiency through Fused-MBConv layers, which replace depthwise convolutions with standard 3×3 convolutions in the early stages to improve memory access on GPU accelerators [29].
- DenseNet121: Utilizes dense connectivity where each layer receives feature maps from all preceding layers as inputs as shown in Equation (2) [30].

$$x_l = H_l([x_0, x_1, \ldots, x_{l-1}]) \quad (2)$$

- MobileNetV3-Large: A lightweight model designed for efficiency, employing depthwise separable convolutions and Squeeze-and-Excitation (SE) blocks for channel-wise feature recalibration [31].

Addressing Class Imbalance via Focal Loss

As identified in Section 2.1, the dataset exhibits a severe skew toward the benign NV class (~90%). Standard Cross-Entropy loss often fails in such scenarios as it tends to bias the model toward the majority class. To counteract this, all architectures were optimized using Focal Loss (FL) [26]. Focal Loss applies a modulating factor to the cross-entropy loss, down-weighting the loss contribution from "easy" (majority) examples and focusing the gradient updates on "hard-to-classify" (minority) samples such as SCC and AK. The formula is defined as shown in Equation (3).

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} log(p_t) \quad (3)$$

Where p_t is the model's estimated probability for the correct class, α_t is a balancing factor, and γ is the focusing parameter, which was empirically set to 2.0 for all experiments in this study.

Training Protocol and Hyperparameters

The experimental framework was implemented using the PyTorch library. To ensure clinical validity and prevent data leakage, a 5-fold Stratified Group-KFold cross-validation was employed. This strategy ensures that all images belonging to the same $lesion\_id$ are restricted to a single fold, preventing the model from "memorizing" specific patient features during training that might appear in the validation set [32].

All models were trained for 30 epochs per fold using the AdamW optimizer with a weight decay of 0.01. The initial learning rate was set to 1×10^{-4} and was dynamically adjusted using a Cosine Annealing scheduler to ensure stable convergence. Training was conducted on a dual-NVIDIA T4 GPU environment with a batch size of 32. The best-performing model state for each fold was selected based on the highest validation Balanced Accuracy (BAcc) to serve as the final candidate for testing and web deployment.

Model Evaluation Metrics

Due to the severe class imbalance in the ISIC 2019 dataset, standard accuracy is not a reliable metric, as a model could achieve high accuracy simply by predicting the majority class (NV) while ignoring the minority classes (SCC and AK). Therefore, this study evaluates the architectures using robust metrics designed for imbalanced multi-class classification: Balanced Accuracy (BAcc), Macro-Averaged F1-Score, and Area Under the Receiver Operating Characteristic Curve (ROC-AUC) [19], [33].

Balanced Accuracy (BAcc)

Balanced Accuracy computes the arithmetic mean of sensitivity (recall) across all classes. It ensures that the performance on the minority malignant classes carries the same weight as the majority benign class. The formula is defined as shown in Equation (4).

$$\text{Balanced Accuracy} = \frac{1}{N} \sum_{i=1}^{N} \text{Recall}_i \quad (4)$$

where N is the total number of classes (in this case, $N = 3$).

Macro-Averaged F1-Score

The F1-Score represents the harmonic mean of Precision and Recall. In the Macro-Averaged variant, the F1-Score is calculated independently for each class and then averaged, treating all classes equally regardless of their sample size as shown in Equation (5).

$$F1 - Macro = \frac{1}{N} \sum_{i=1}^{N} 2 \times \frac{Precision_i \times Recall_i}{Precision_i \times Recall_i} \quad (5)$$

Standard Deviation and Model Stability

To measure the consistency and stability of the models across the 5-fold cross-validation, the standard deviation (σ) of the Balanced Accuracy was calculated. The standard deviation quantifies the amount of variation or dispersion in the performance metrics across different data folds. A lower standard deviation

indicates higher model stability, meaning the architecture's performance is robust and less sensitive to variations in the training data splits.

Explainable AI with Grad-CAM

A critical barrier to the clinical adoption of Deep Learning is its "black-box" nature. To ensure trust and clinical transparency, Gradient-weighted Class Activation Mapping (Grad-CAM) was integrated into the evaluation pipeline. Grad-CAM leverages the spatial information preserved in the final convolutional layers of the CNN to understand which regions of the dermoscopic image influenced the classification decision the most.

The class-discriminative localization map $L^c_{Grad-CAM}$ is computed by passing the gradients of the target class score y^c with respect to the feature map activations A^k of the last convolutional layer. These gradients are global-average-pooled to obtain the neuron importance weights α^c_k as shown in Equation (6).

$$\alpha^c_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A^k_{i,j}} \tag{6}$$

A weighted combination of the forward activation maps is then followed by a ReLU operation to highlight only the features that have a positive influence on the class of interest as shown in Equation (7).

$$L^c_{Grad-CAM} = \text{ReLU}\left(\sum_k \alpha^c_k A^k\right) \tag{7}$$

In this system, Grad-CAM overlays a heatmap onto the original image, providing dermatologists with visual evidence that the model is actively focusing on the lesion's pathological borders rather than spurious artifacts.

Web-Based System Deployment

To bridge the gap between experimental research and practical clinical utility, the best-performing model from the benchmarking phase was deployed into an interactive web-based application. Developed using the Streamlit framework, the system is designed to provide real-time diagnostic support [34]. The deployment architecture allows users to upload raw dermoscopic images, which are deterministically preprocessed, resized, and passed through the model. The application seamlessly outputs the predicted class probabilities (SCC, AK, or NV) alongside the real-time Grad-CAM visual overlay, acting as a rapid, interpretable second opinion for medical practitioners.

3. RESULTS AND DISCUSSIONS

Result

The experimental results are presented through a multi-faceted evaluation encompassing quantitative benchmarking, algorithmic reliability under extreme class imbalance, and qualitative interpretability through Explainable AI (XAI).

Performance Comparison and Benchmarking Analysis

The primary objective of the benchmarking phase was to identify the most robust architecture for detecting non-melanoma skin cancer (NMSC) within a highly skewed dataset. Performance was averaged over a 5-fold Stratified Group-KFold cross-validation to ensure the generalizability of the results and to prevent any bias resulting from data leakage. As demonstrated in Table 2, the modernized architectures exhibited a significant performance leap compared to traditional CNN baselines.

Table 2. Average 5-fold cross-validation performance metrics

Model Architecture	Balanced Accuracy (BAcc)	Macro F1-Score	Stability (Std. Dev)
ConvNeXt-Tiny	76.98%	0.7513	± 0.31
EfficientNetV2-S	76.28%	0.7219	± 2.05
DenseNet121	55.85%	0.4460	± 4.12
MobileNetV3-Large	47.04%	0.3504	± 5.87

ConvNeXt-Tiny emerged as the superior model, achieving a Balanced Accuracy (BAcc) of 76.98% and a Macro F1-Score of 0.7513. Notably, this model exhibited the highest stability with a standard deviation of only ± 0.31, indicating consistent performance across different data folds. This architectural advantage stems from ConvNeXt's modernized design, which replaces traditional 3×3 convolutions with larger 7×7 kernels, providing a global receptive field similar to Vision Transformers. This allows the network to capture complex, unstructured textural patterns of NMSC lesions.

In contrast, traditional lightweight models like MobileNetV3-Large struggled significantly, with a BAcc of only 47.04%. This failure under severe class imbalance conditions occurs because MobileNetV3 relies heavily on depthwise separable convolutions aimed at minimizing parameter count and computational cost. While computationally efficient, this significantly limits the model's representational capacity. Consequently, the lightweight architecture fails to learn the intricate, heterogeneous features of the minority malignant classes (SCC and AK) and instead overfits to the morphological features of the majority benign class (NV).

Confusion Matrix and Classification Reliability

To further analyze the model's reliability, a Confusion Matrix was generated for the best-performing model, ConvNeXt-Tiny, using the unseen hold-out test set, as shown in Figure 4. Despite the extreme class imbalance where benign NV cases represent approximately 89.6% of the images, the integration of Focal Loss allowed the model to maintain high sensitivity for critical classes.

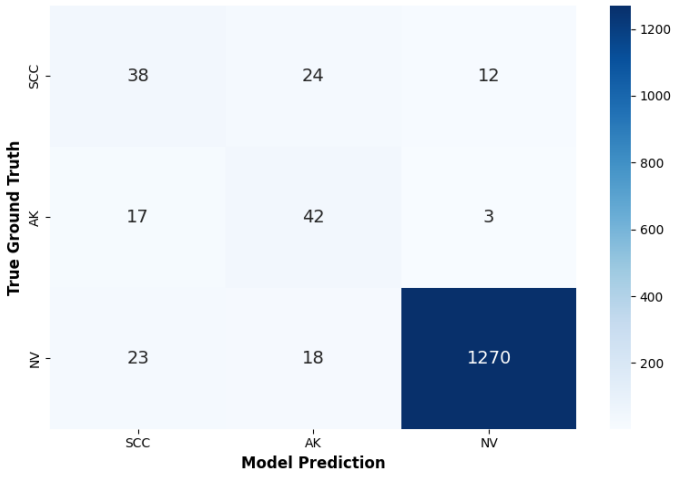

Figure 4. Confusion matrix on test set (convnext_tiny)

The matrix confirms that the model successfully mitigated the majority-class bias. It correctly identified 1,270 out of 1,311 benign NV cases while maintaining high predictive power for the malignant SCC and pre-malignant AK classes. This demonstrates that the algorithmic refinement effectively prioritized the detection of hard-to-classify minority samples.

Spatial Interpretability using Grad-CAM

To move beyond "black-box" predictions and ensure clinical trust, Grad-CAM was utilized to generate spatial activation heatmaps, as depicted in Figure 5.

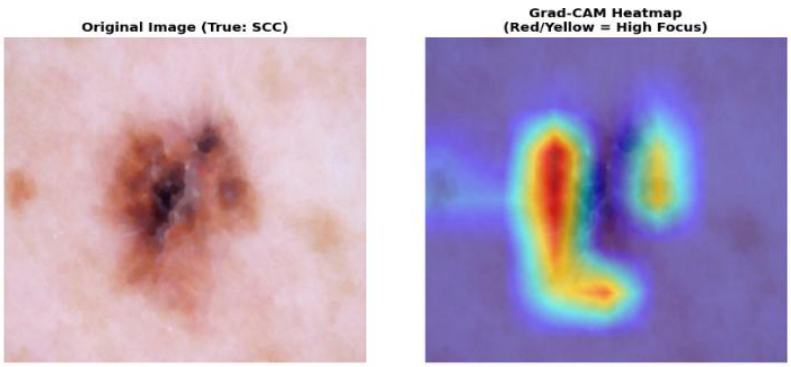

Figure 5. Spatial interpretability using grad-CAM

The heatmaps generated for SCC and AK samples showed a high degree of clinical alignment. The network's attention was strictly localized on the core pathological features, such as irregular pigmentation and scaly textures of the lesions. The blue regions in the heatmap indicate that the model successfully ignored irrelevant background artifacts, such as residual hair or skin folds, which were previously addressed in the preprocessing pipeline.

Practical System Realization

The optimal ConvNeXt-Tiny model was engineered into a functional web-based prototype using the Streamlit framework, illustrated in Figure 6.

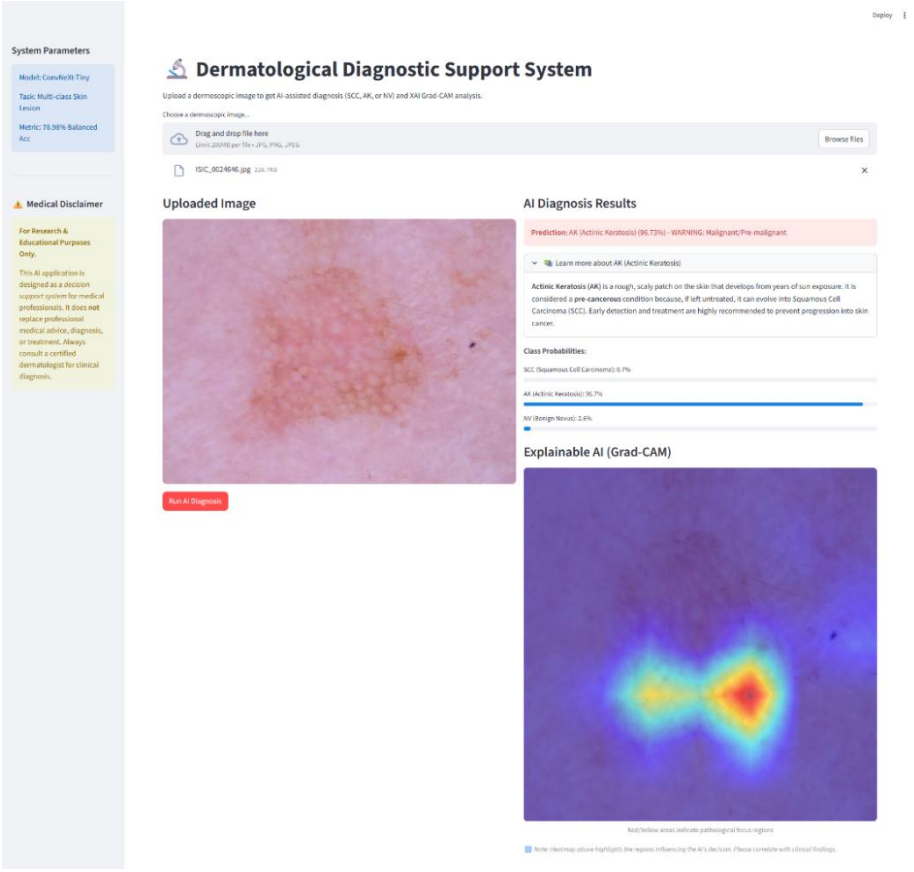

Figure 6. Web application for automated skin lesion classification

.

The application provides a seamless end-to-end pipeline where raw dermoscopic images are preprocessed and analyzed in real-time. The deployment seamlessly outputs predicted class probabilities alongside automated medical descriptions and real-time Grad-CAM overlays, acting as an interpretable second-opinion tool for medical practitioners.

Discussions

Impact of architectural modernization

The findings of this study underscore a significant architectural evolution in medical computer vision. The superior performance of ConvNeXt-Tiny highlights the benefits of bridging the gap between standard convolutions and Vision Transformer (ViT) philosophies. The use of a larger 7x7 kernel provides a wider receptive field, allowing the model to capture more holistic structural information about skin lesions than the localized patterns captured by traditional 3x3 filters.

Effectiveness of focal loss in imbalanced learning

Real-world medical datasets are frequently characterized by severe class imbalance. Standard Cross-Entropy loss functions often produce models that are algorithmically biased toward the majority class. By implementing Focal Loss with a focusing parameter of $\gamma = 2.0$, this study successfully down-weighted the loss contribution from "easy" majority examples and focused gradient updates on "hard-to-classify" minority samples like SCC and AK. This led to a more balanced sensitivity across all categories, which is vital for clinical safety.

Addressing the trust deficit for clinical adoption

The "black-box" nature of deep neural networks creates a substantial "trust deficit" among healthcare professionals. By integrating Explainable AI (XAI) through Grad-CAM, this research provides visual evidence that aligns with established pathological criteria. Furthermore, the transition from experimental benchmarking to a functional web-based prototype establishes the translational feasibility of the research. The resulting system provides an accessible decision-support pipeline that can assist dermatologists, especially in resource-limited settings.

Comparison with Previous State-of-the-Art Studies

To reinforce the success and novelty of this research, Table 3 presents a comparison between our proposed framework and recent state-of-the-art studies focusing on skin lesion classification.

Table 3. Comparison with previous research

Study	Proposed Method / Baseline	Imbalance Handling	Key Performance	Explainability & Deployment
Srinivasu et al. [15]	MobileNetV2 + LSTM	Not explicitly addressed (Standard Cross-Entropy)	Accuracy > 85%	None / No Deployment
De et al. [16]	Hybrid CNN-DenseNet	Not explicitly addressed (Standard Cross-Entropy)	Accuracy 95.7%	None / No Deployment
Mahmoud and Soliman [11]	ANN + SVM	Standard Data Augmentation	Accuracy 94%	None / No Deployment
Proposed Study	**ConvNeXt-Tiny**	**Focal Loss (Dynamic Weighting)**	**BAcc: 76.98% (± 0.31)**	**Grad-CAM & Streamlit Web App**

As illustrated in Table 3, while previous studies by Srinivasu et al. [15], De et al. [16], and Mahmoud and Soliman [11] achieved high accuracy rates, they predominantly operated on balanced datasets or did not explicitly address the performance decay caused by class imbalance. Furthermore, their models remain as 'black-box' systems, which limits clinical adoption. In contrast, our study not only maintains stable performance on highly skewed NMSC data through Focal Loss but also bridges the 'trust deficit' by integrating Grad-CAM for transparency and providing a functional web-based prototype for real-time diagnostic support.

4. CONCLUSION

In direct alignment with the objectives established in the Introduction, this study successfully addressed the limitations of traditional CAD systems by developing, benchmarking, and deploying a modernized deep

transfer learning framework for imbalanced skin cancer classification. As anticipated in our gap analysis, the experimental results confirmed that modernized architectures significantly outperform traditional baselines. ConvNeXt-Tiny was identified as the most effective architecture, achieving a Balanced Accuracy of 76.98% and a Macro F1-Score of 0.7513. Furthermore, the hypothesis regarding algorithmic bias was proven correct; the integration of Focal Loss successfully mitigated the severe data skewness that previously hindered older models. The system also fulfilled the need for clinical transparency by utilizing Grad-CAM to provide visual evidence aligning with pathological criteria. Ultimately, the successful deployment of the optimal model into a Streamlit web application fulfills the primary goal of establishing a practical, interpretable, and real-time clinical decision-support tool. Future research will focus on integrating multimodal data, such as patient metadata, to further enhance the system's diagnostic sensitivity and global generalizability.

REFERENCES

[1] W. Hu, L. Fang, R. Ni, H. Zhang, and G. Pan, "Changing trends in the disease burden of non-melanoma skin cancer globally from 1990 to 2019 and its predicted level in 25 years," *BMC Cancer*, vol. 22, no. 1, p. 836, 2022.

[2] L. Zhou, Y. Zhong, L. Han, Y. Xie, and M. Wan, "Global, regional, and national trends in the burden of melanoma and non-melanoma skin cancer: insights from the global burden of disease study 1990–2021," *Sci. Rep.*, vol. 15, no. 1, p. 5996, 2025.

[3] Y. Pan, B. Tang, Y. Guo, Y. Cai, and Y. Y. Li, "Global burden of non-melanoma skin cancers among older adults: a comprehensive analysis using machine learning approaches," *Sci. Rep.*, vol. 15, no. 1, p. 15266, 2025.

[4] L. Nanz, U. Keim, A. Katalinic, T. Meyer, C. Garbe, and U. Leiter, "Epidemiology of Keratinocyte Skin Cancer with a Focus on Cutaneous Squamous Cell Carcinoma," *Cancers (Basel).*, vol. 16, no. 3, p. 606, 2024.

[5] C. Dessinioti and A. J. Stratigos, "Recent Advances in the Diagnosis and Management of High-Risk Cutaneous Squamous Cell Carcinoma," *Cancers*, vol. 14, no. 14. p. 3556, 2022.

[6] M. Catalano, F. Nozzoli, F. De Logu, R. Nassini, and G. Roviello, "Management Approaches for High-Risk Cutaneous Squamous Cell Carcinoma with Perineural Invasion: An Updated Review," *Curr. Treat. Options Oncol.*, vol. 25, no. 9, pp. 1184–1192, 2024.

[7] K. Liopyris, S. Gregoriou, J. Dias, and A. J. Stratigos, "Artificial Intelligence in Dermatology: Challenges and Perspectives," *Dermatol. Ther. (Heidelb).*, vol. 12, no. 12, pp. 2637–2651, 2022.

[8] N. Alshdaifat *et al.*, "Triple-Stream Transformer Architecture for Multi-Class Skin Cancer Classification in Dermoscopic Images," *J. Posthumanism*, vol. 5, no. 3, pp. 1090–1106, 2025.

[9] S. Nazari and R. Garcia, "Automatic Skin Cancer Detection Using Clinical Images: A Comprehensive Review," *Life*, vol. 13, no. 11, p. 2123, 2023.

[10] S. Khattar and R. Kaur, "Computer assisted diagnosis of skin cancer: A survey and future recommendations," *Comput. Electr. Eng.*, vol. 104, p. 108431, 2022.

[11] N. M. Mahmoud and A. M. Soliman, "Early automated detection system for skin cancer diagnosis using artificial intelligent techniques," *Sci. Rep.*, vol. 14, no. 1, p. 9749, 2024.

[12] R. Kaur, H. Gholamhosseini, and M. Lindén, "Advanced Deep Learning Models for Melanoma Diagnosis in Computer-Aided Skin Cancer Detection," *Sensors*, vol. 25, p. 594, 2025.

[13] J. S M, M. P, C. Aravindan, and R. Appavu, "Classification of skin cancer from dermoscopic images using deep neural network architectures," *Multimed. Tools Appl.*, vol. 82, no. 10, pp. 15763–15778, 2023.

[14] S. Haggenmüller *et al.*, "Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts," *Eur. J. Cancer*, vol. 156, pp. 202–216, 2021.

[15] P. N. Srinivasu, J. G. Sivasai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, "Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM," *Sensors*, vol. 21, no. 8, p. 2852, 2021.

[16] A. De, N. Mishra, and H. T. Chang, "An approach to the dermatological classification of histopathological skin images using a hybridized CNN-DenseNet model," *PeerJ Comput. Sci.*, vol. 10, pp. 1–33, 2024.

[17] Salviohexia, "ISIC 2019 Skin Lesion Images for Classification, Kaggle," *https://www.kaggle.com/datasets/salviohexia/isic-2019-skin-lesion-images-for-classification/data*, 2019. .

[18] V. D. Hoang, X. T. Vo, and K. H. Jo, "Categorical Weighting Domination for Imbalanced Classification with Skin Cancer in Intelligent Healthcare Systems," *IEEE Access*, vol. 11, pp. 105170–105181, 2023.

[19] T. M. Alam *et al.*, "An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset," *Diagnostics*, vol. 12, no. 9, p. 2115, 2022.

[20] S. I. Hussain and E. Toscano, "Enhancing Recognition and Categorization of Skin Lesions with Tailored Deep Convolutional Networks and Robust Data Augmentation Techniques," *Mathematics*, vol. 13, no. 9, May 2025.

[21] R. K. Sheu and M. S. Pardeshi, "A Survey on Medical Explainable AI (XAI): Recent Progress, Explainability Approach, Human Interaction and Scoring System," *Sensors*, vol. 22, no. 20, p. 8068, 2022.

[22] G. Yang, Q. Ye, and J. Xia, "Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond," *Inf. Fusion*, vol. 77, pp. 29–52, 2022.

[23] A. M. Antoniadi *et al.*, "Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review," *Appl. Sci.*, vol. 11, no. 11, p. 5088, 2021.

[24] H. Mo and L. Wei, "SA-ConvNeXt: A Hybrid Approach for Flower Image Classification Using Selective Attention Mechanism," *Mathematics*, vol. 12, no. 14, p. 2151, 2024.

[25] Q. Hou, C.-Z. Lu, M.-M. Cheng, and J. Feng, "Conv2Former: A Simple Transformer-Style ConvNet for Visual Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 8274–8283, 2024.

.

[26] M. Yeung, E. Sala, C. B. Schönlieb, and L. Rundo, "Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Comput. Med. Imaging Graph.*, vol. 95, p. 102026, 2022.

[27] T. R. Mahesh, S. B. Khan, K. K. Mishra, S. Alzahrani, and M. Alojail, "Enhancing Diagnostic Precision in Breast Cancer Classification Through EfficientNetB7 Using Advanced Image Augmentation and Interpretation Techniques," *Int. J. Imaging Syst. Technol.*, vol. 35, no. 1, p. e70000, 2025.

[28] C. Yuan, X. Jiang, and Q. Yang, "Channel-Pruning Convolutional Neural Network with Learnable Kernel Element Position Convolution Utilizing the Symmetric Whittaker–Shannon Interpolation Function," *Symmetry (Basel).*, vol. 17, no. 3, p. 390, 2025.

[29] N. Li, J. Xue, S. Wu, K. Qin, and N. Liu, "Research on Coal and Gangue Recognition Model Based on CAM-Hardswish with EfficientNetV2," *Appl. Sci.*, vol. 13, no. 15, p. 8887, 2023.

[30] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, and Y. Liu, "Dense Convolutional Network and Its Application in Medical Image Analysis," *Biomed Res. Int.*, vol. 2022, no. 1, p. 2384830, 2022.

[31] C. Bi, S. Xu, N. Hu, S. Zhang, Z. Zhu, and H. Yu, "Identification Method of Corn Leaf Disease Based on Improved Mobilenetv3 Model," *Agronomy*, vol. 13, no. 2, p. 300, 2023.

[32] J. White and S. D. Power, "k-Fold Cross-Validation Can Significantly Over-Estimate True Classification Accuracy in Common EEG-Based Passive BCI Experimental Designs: An Empirical Investigation," *Sensors*, vol. 23, no. 13, p. 6077, 2023.

[33] P. Thölke *et al.*, "Class imbalance should not throw you off balance : Choosing the right classifiers and performance metrics for brain decoding with imbalanced data," *Neuroimage*, vol. 277, p. 120253, 2023.

[34] Yazid Aufar, Muhammad Helmy Abdillah, and Jiki Romadoni, "Web-based CNN Application for Arabica Coffee Leaf Disease Prediction in Smart Agriculture," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 7, no. 1, pp. 71–79, Feb. 2023.