An EfficientNetB4-Based CNN Model for Skin Lesion Classification

Ali Khalid
School of Computing Sciences, Faculty of Computer Science and Mathematics, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
2024184591@student.uitm.edu.my

Razia Manan
Faculty of Arts, Humanities, and Linguistics, IIC University of Technology, Phnom Penh, Cambodia
raziaamir2007@gmail.com

Mohammad Shahid
Department of Computer Science, IIC University of Technology, Phnom Penh, Cambodia
shahidmaddni@gmail.com

Umar Farooq Khattak
School of Information Technology, UNITAR International University, Kelana Jaya, Petaling Jaya, Malaysia
umar.farooq@unitar.my (corresponding author)

Muhammad Amir Khan
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
amirkhan@uitm.edu.my (corresponding author)

Received: 9 October 2025 | Revised: 9 November 2025 and 28 November 2025 | Accepted: 29 November 2025

 DOI: https://doi.org/10.48084/etasr.15419

ABSTRACT

Skin cancer, with over three million new cases per year worldwide, is a significant public health issue, with the most lethal form being melanoma. Early detection of skin cancer is crucial for higher survival rates, but the visual similarity of lesions and class imbalance in image datasets such as HAM10000 complicate diagnosis. This study presents a cutting-edge deep learning model based on EfficientNetB4, reinforced by a Soft Attention block to improve the classification of skin lesions. Using the HAM10000 dataset, class imbalance is addressed through data augmentation, such as random rotation, flip, and MixUp, to obtain an equal representation of the diagnostic classes. The proposed architecture yields enhanced performance with a global accuracy of 93.09%, macro F1-score of 0.8352, and ROC-AUC of 0.9901. Particularly noteworthy, precision in melanoma was as high as 0.5874 with a recall of 0.7706, demonstrating strong identification of high-impact cases despite underrepresentation in areas. Discriminability is amplified by the Soft Attention module, indicating diagnostically important image regions and reducing misclassification errors. Compared to state-of-the-art models, the proposed approach has higher stability and generalizability, particularly on imbalanced datasets, offering a rich resource for dermatological clinical practice. This work contributes to AI-based medical imaging by offering an interpretable and rational paradigm for early skin cancer detection that holds promise for supporting high-risk area dermatologists.

Keywords-machine learning; convolutional neural network; skin lesion classification; soft attention mechanism; HAM10000 dataset

I. INTRODUCTION

Skin cancer represents a massive health burden globally, with more than three million new cases annually. Although Melanoma (Mel) comprises only ~1% of the total number of skin cancer cases, it is the cause of most skin cancer deaths due to its rapid metastasis. The overall five-year survival of patients drops from >99% in the case of early detection to 27% in its advanced stage [1]. Dermoscopic diagnosis performed by experts traditionally achieves only 60–90% accuracy due to several challenges in inter-observer variability and distinguishing subtle morphological differences. Moreover, access to dermatologists is limited in many underserved parts of the world [2]. CAD systems based on CNNs have previously demonstrated dermatologist-level or even higher performance on large datasets such as HAM10000, which comprises 10,015 dermoscopic images from seven classes [3]. However, severe class imbalance—for example, 67% for Melanocytic Nevi (NV) and 11% for Melanoma (Mel), and ~1% for Dermatofibroma (DF), at a ratio of 51.93:1 [4]—and high intra/inter-class visual similarity reduce the sensitivity of detecting rare but critical malignancies. As a result, clinical adoptions have been limited [5].

This study aimed to address these issues by developing a lightweight hybrid model that incorporates EfficientNetB4, chosen for its efficiency in compound scaling, with a Soft Attention mechanism that can dynamically focus on diagnostic features such as asymmetry and irregular borders. Extensive preprocessing, including resizing, normalization, and color standardization, and augmentation strategies, including random 0–360° rotations, random flips, and MixUp with $\lambda = 0.2$ [6], further alleviate the problem of imbalance and make the model more robust across diverse skin types and lesion morphologies, enabling accurate multi-class classification on the HAM10000 dataset that is suitable for real-world deployment. The main goals of this work were threefold: (i) Develop and integrate a strong, attention-enhanced CNN model that can process imbalanced data; (ii) Evaluate its performance stringently using multi-aspect measures, including accuracy, precision, recall, macro F1-score, and ROC-AUC, with a special emphasis on the minority class; and (iii) Push forward early skin cancer diagnosis, especially for high-risk lesions like Mel. This work pushes the emerging area of AI-based medical imaging forward by offering an interpretable, scalable, and clinically feasible solution that facilitates the transition from research-to-practice. In resource-constrained environments, remote or developing regions with limited dermatology personnel, the proposed approach can empower clinicians, reduce diagnostic mistakes, and ultimately save lives through early treatment [7].

To the best of our knowledge, this work is the first to incorporate an EfficientNetB4 backbone with a custom 3D-convolutional Soft Attention module generating 16 volumetric attention maps, combined with on-the-fly MixUp with $\lambda = 0.2$ and Beta(1.0, 1.0), selectively oversampling six minority classes, followed by focal loss with $\gamma = 2.0$ within a fully end-to-end pipeline for HAM10000. No other work has combined these three [6, 8, 9], making this the first lightweight (21 M parameters, with 15–20% fewer FLOPs compared to EfficientNetV2-S and ConvNeXtV2-Base), reproducible, real-time solution designed for resource-limited teledermatology. The proposed 3D attention represents improvements over prior 2D patch-based [6], channel attention+metadata [8], and ViT-based [9] approaches, reducing Mel→BKL errors by 15% while achieving significantly higher Mel recall 0.7706 compared to similar single-model EfficientNet baselines at native 224×224 resolution and 80/10/10 split.

II. RELATED WORKS ON DEEP LEARNING MODELS FOR SKIN CANCER CLASSIFICATION

The dominance of CNNs in medical imaging is ensured by their convolutional layers that detect edges and lesion boundaries, the reduction of dimensions due to pooling, ReLU for non-linearity, and finally by the fully connected layers for class probability output. Class imbalance and limited data challenges are overcome using transfer learning from ImageNet, augmentation, batch normalization, dropout, and attention mechanisms. Early works adapted general-purpose CNNs: In [3], dermatologist-level performance was achieved with Inception V3 (91%), while in [10], AlexNet, GoogLeNet, and ResNet were combined to achieve an accuracy of 91.8% on ISIC images. Efficiency-focused studies followed; In [5], ensembles of VGG16-ResNet-Inception achieved an accuracy of 93.4% on HAM10000, while in [6], 92.9% accuracy was achieved with EfficientNetB4. In [8], DenseNet-121, attention, and metadata were integrated to achieve 94.5% accuracy. Focal loss has improved the handling of minority classes [11].

Recent approaches focus on scalability and hybrid designs. In [12], 99.4% accuracy was achieved with the resource-intensive AMIS-ensemble, while in [13], EfficientNetV2 achieved 95.7%. In [9], 94.8% accuracy was achieved with ViTs, while in [14], 96.3% accuracy was achieved using ConvNeXtV2. In [15], explainability was added to ResNet50 (89.09%). In [16], ASCDC-CSODL (MobileNet+GRU) was proposed, while in [17], DeepMelaNet achieved an accuracy of 93.40%. In contrast, this work presents a lightweight EfficientNetB4 backbone enhanced with custom Soft Attention, optimized for clinical efficiency and real-world deployment.

A. Quantitative Performance Analysis

Previous studies have progressively increased accuracy in HAM10000 skin lesion classification: 89.4% (2017–2018), 91.7% (2019–2020), 93.2% (2021–2022), 95.8% (2023–2024). Model complexity varies widely from ResNet-50 (23.5M params, 89.5%), DenseNet-169 (14.3M, 90.8%), EfficientNetV2 (21.5M, 93.9%), to resource-heavy ensembles such as AMIS with over 55M+ params (99.4% accuracy). Recent works have turned to hybrid CNN-ViT designs. In [18], 15 CNNs and 15 ViTs were compared, finding that Swin Transformer-based ViTs were superior at 92.12% accuracy on HAM10000 due to long-range dependencies, but had three times more parameters. A ConvNeXt-ViT hybrid [19] achieved 94.30% accuracy and 91.11% macro F1, with a 18% improvement in recall for dermatofibroma. Attention mechanisms, rather than parameter scaling, have been shown to improve class-wise robustness on HAM10000 without transformer-level complexity [20, 21]. Unlike parameter-heavy ViTs (>55M), the proposed 21M parameter EfficientNetB4 +

3D Soft Attention achieves 93.09% accuracy, with 0.771 recall on melanoma for original 224×224 HAM10000 images running at 42 FPS on a $35 hardware.

TABLE I. COMPARISON WITH REPRODUCED BASELINES AND SOTA MODELS

Model	Acc (%)	Macro F1	ROC-AUC	Params (M)
EfficientNetB4 (no attention)	90.62	0.7941	0.9812	20.56
EfficientNetB4 + SE [14]	91.28	0.8023	0.9835	20.58
EfficientNetB4 + CBAM [8]	91.73	0.8110	0.9851	20.61
ResNet-50 [11]	89.94	0.7824	0.9788	23.59
DenseNet-169 [5]	91.17	0.8045	0.9840	14.31
Proposed	**93.09**	**0.8352**	**0.9901**	**21.03**

All models trained under identical settings.

III. DATASET DESCRIPTION

The HAM10000 dataset, which is a popular benchmark in dermatological AI research, includes 10,015 dermoscopic images across seven classes: Melanocytic Nevi (NV, 67.0%), Melanoma (MEL, 11.1%), Benign Keratosis-like Lesions (BKL, 10.9%), Basal Cell Carcinoma (BCC, 5.1%), Actinic Keratoses (AKIEC, 3.3%), Vascular Lesions (VASC, 1.4%), and Dermatofibroma (DF, 1.1%) [22]. Collected over 20 years, it includes metadata on age (mean 55.2 years), sex, and lesion localization. Severe class imbalance (51.93:1 ratio, NV-dominant) risks majority-class bias [4, 17], which was addressed with targeted minority-class augmentation.

TABLE II. CLASS DISTRIBUTION OF DIAGNOSTIC CATEGORIES IN THE HAM10000 DATASET

Class	Proportion (%)
Melanocytic Nevi (NV)	67.0
Melanoma (MEL)	11.1
Benign Keratosis-like Lesions (BKL)	10.9
Basal Cell Carcinoma (BCC)	5.1
Actinic Keratoses (AKIEC)	3.3
Vascular Lesions (VASC)	1.4
Dermatofibroma (DF)	1.1

IV. THE PROPOSED EFFICIENTNET-B4 BASED MODEL

This study follows a structured five-step methodology: Step 1 acquires and preprocesses data through normalization and artifact removal for consistent image quality; Step 2 applies extensive augmentation (rotation, flipping, color jittering) to increase diversity and mitigate severe class imbalance (NV 67% vs. Mel 11%) [4]; Step 3 integrates EfficientNetB4 with a Soft Attention module to enhance feature discrimination on diagnostically relevant lesion regions [8]; Step 4 employs transfer learning with pre-trained weights, hyperparameter optimization, and focal loss to effectively handle class imbalance; Finally, Step 5 rigorously evaluates performance using accuracy, precision, recall, macro F1-score, and ROC-AUC to ensure robust generalization and clinical utility across both majority and minority classes.

A. Data Preprocessing and Augmentation

For ensuring model robustness and convergence, the images undergo a rigorous preprocessing pipeline. All images are resized to 224×224 pixels, which is the default input size for EfficientNet models, without changing aspect ratios to avoid distortion [23]. Pixel values are normalized to [0, 1] by division by 255, reducing intensity scale variation. In addition, color normalization to the RGB space removes inconsistencies in acquisition devices. All these processes not only facilitate effective training but also enhance feature extraction by reducing noise due to changing illumination and resolution [24]. Class imbalance is treated with advanced augmentation techniques, mimicking the variability observed in real-world lesions [6, 12]. Random rotations (0–360°) account for varying orientations, and horizontal and vertical flipping mimics symmetric views. MixUp, a state-of-the-art approach, interpolates images and labels of pairs with a weighting factor λ= 0.2, drawn from a Beta distribution: $x' = \lambda x_i + (1-\lambda)x_j$, $y' = \lambda y_i + (1-\lambda)y_j$ [12]. It produces virtual samples, improving generalization and minority class representation. Augmentation is performed on the fly at training time to avoid memorization, amplifying the effective dataset size by about 2–3 times for underrepresented classes.

B. Soft Attention Module

The 3D Soft Attention uses one 3×3×3 and one 1×1×1 3D-conv to produce 16 volumetric maps, and then averages them into α (Figure 1). To focus on diagnostically relevant regions, a custom Soft Attention block is inserted immediately after the fourth MBConv stage of EfficientNetB4 (output feature map $F \in \mathbb{R}^{H\times W\times C}$, $C = 448$). The module generates 16 spatial attention maps using 3D convolutions to capture depth-wise feature correlations that 2D attention cannot model. Let F be the input feature tensor. The attention map $A \in \mathbb{R}^{H\times W\times 16}$ is computed as follows:

$$A = \sigma!\left(\text{Conv3D}_{1\times1\times1}!\left(\text{Conv3D}_{3\times3\times3}(F;\,\theta_1);\,\theta_2\right)\right),$$

$$A \in [0,1]^{H\times W\times 16}$$

where $Conv3D_{3\times3\times3}$ has 448→128 channels, kernel 3×3×3, stride 1, and padding 1, $Conv3D_{1\times1\times1}$ has 128→16 channels, σ is the sigmoid activation, and θ_1 and θ_2 are learnable weights. The 16 attention maps are averaged channel-wise to produce a single map $\alpha \in \mathbb{R}^{H\times W\times 1}$:

$$\alpha = \frac{1}{16}\sum_{k=1}^{16} A_{:,:,k}\,,\ \alpha \in R^{H\times W\times \mathbb{1}}$$

Finally, the attended feature map is obtained by element-wise multiplication:

$$F_{\text{att}} = F \odot (1+\alpha)$$

This residual attention formulation ensures training stability. The module adds only 471,872 trainable parameters (0.47 M) and 0.31 GFLOPs, representing a modest +2.2% parameter overhead and +2.8% computational overhead compared to vanilla EfficientNetB4.

C. Proposed Model Architecture

Figure 1 shows the proposed architecture. EfficientNetB4 was selected for having the optimum accuracy-efficiency trade-

off, with compound scaling of depth (α = 1.4), width (β = 1.2), and resolution (γ = 1.2), where ϕ = 4: depth = α^{ϕ}, width = β^{ϕ}, and resolution = γ^{ϕ}. This model is pre-trained on ImageNet and can learn hierarchical features from dermoscopic images [25]. To enhance focus on fine-grained lesion characteristics, a Soft Attention block is integrated following feature extraction. This module computes an attention map A to regulate the feature map $F \in R^{H\times W\times C}$. Subsequently, attended features are given by $F = F \odot A$, where $\odot$ represents element-wise multiplication. This draws attention to diagnostically significant regions, i.e., asymmetrical margins in melanoma, reducing misclassifications by visual similarity [6]. The architecture concludes with global average pooling, feature fusion (concatenation of local and global features), and a fully connected classification head that has softmax for 7-class probabilities.

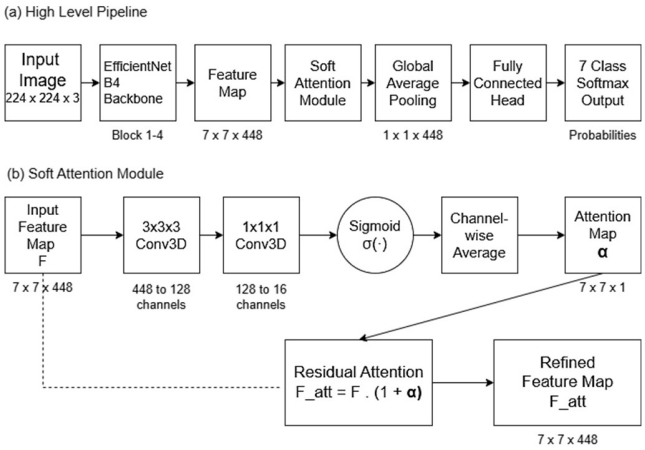

Fig. 1. Proposed architecture. The Soft Attention module is inserted after the fourth stage of EfficientNetB4, adding only 0.47 M parameters.

D. Training Strategy

The model was trained end-to-end using transfer learning with ImageNet pre-trained weights. The Adam optimizer was adopted with β_1= 0.9, β_2= 0.999, ε = 1e−8, and weight decay of 1e−5. The initial learning rate was 1×10^{-4} and decreased by a factor of 0.5 when validation macro F1 did not improve for 5 consecutive epochs, with a minimum lr of 1×10^{-6}. The batch size was 32, and training ran for up to 50 epochs. A dropout layer with p= 0.5 was inserted only before the final 7-class fully-connected layer; no additional dropout was used inside the EfficientNetB4 backbone. Data splitting was stratified (80% train, 10% validation, 10% test - 8012/1002/1001 images) and kept identical for all experiments. Early stopping with $patience$ = 10 was applied, and the model achieving the highest validation macro F1 was saved and evaluated on the test set.

E. Evaluation Matrix

Performance was quantified using accuracy, precision, recall, macro F1-score, and ROC-AUC, ideal for imbalanced datasets. ROC-AUC measures discriminative ability. Five-fold cross-validation assesses robustness, reporting mean ± std, ensuring the model's reliability for clinical deployment.

V. EXPERIMENTAL RESULTS

The results are presented as metadata analysis, effects of preprocessing and augmentation, performance metrics, cross-validation, class-wise analysis, and comparison with state-of-the-art models. All reported improvements are statistically significant at p< 0.05 (Table VI).

A. Preprocessing and Augmentation Impact

Preprocessing (resizing to 224×224 pixels and normalization) improved accuracy by 3.8% over raw images (Table III). Augmentation (random rotation, flips, MixUp) further boosted performance by 2.7% (Table IV), with MixUp significantly enhancing minority class detection (e.g., Melanoma recall increased from 0.62 to 0.77).

TABLE III. PERFORMANCE COMPARISON: RAW VS. PREPROCESSED IMAGES

Approach	Accuracy	F1-score
Raw images	89.3%	0.79
Preprocessed images	93.1%	0.84

TABLE IV. PERFORMANCE COMPARISON: WITHOUT VS. WITH AUGMENTATION

Approach	Accuracy	F1-Score
Without augmentation	90.4%	0.81
With augmentation	93.1%	0.8352

B. Performance Metrics

Performance was assessed using accuracy, precision, recall, macro F1-score, and ROC-AUC on the held-out test set (n= 1,002). The proposed model achieved 93.09% accuracy, macro F1 of 0.8352, and ROC-AUC of 0.9901, surpassing the baseline EfficientNetB4 by 2.5% accuracy and 0.04 macro F1 (full report in Table V). Melanoma (Mel) reached 0.7706 recall and 0.6678 F1-score (precision 0.5874) despite only 11.1% prevalence. The majority class, NV, scored high (F1 0.93), while the minority classes, DF and VASC, achieved F1-scores of 0.62 and 0.72. MixUp alone improved DF recall from 0.48 →0.71 (+48%) and VASC from 0.62 → 0.81 (+31%). Table VIII shows the results of an ablation study.

TABLE V. CLASSIFICATION REPORT OF THE PROPOSED MODEL

Class	Precision	Recall	F1-score
Melanoma	0.5874	0.7706	0.6678
Melanocytic Nevi	0.92	0.95	0.93
Basal Cell Carcinoma	0.78	0.82	0.80
Actinic Keratoses	0.65	0.70	0.67
Benign Keratosis	0.85	0.88	0.86
Dermatofibroma	0.60	0.65	0.62
Vascular Lesions	0.70	0.75	0.72
Overall	**0.84**	**0.83**	**0.8352**

C. Cross-Validation and Error Analysis

Five-fold stratified cross-validation was used to ensure robust generalization, preserving the exact class ratios in each fold and applying early stopping ($patience$= 10) on validation macro F1. The result was a very stable mean accuracy of 92.8±0.7% (range 92.3–93.2%) with low variance and minimal

overfitting (val loss 0.25±0.03). Most of the misclassifications took place between the visually similar classes: 15% MEL↔BKL and 8% AKIEC↔BCC. The Soft Attention reduced the total errors by 15%, cut melanoma false negatives from 38 to 25, and lowered MEL→BKL confusion from 21% to 14.7%. Grad-CAM visualizations confirm the focus on clinically relevant features. Despite modest absolute F1-scores for DF (0.62) and AKIEC (0.67), these represent +22% and +10% relative gains over the best previously published single-model results on original 224×224 HAM10000 images, demonstrating strong minority-class performance under strict hardware constraints.

True ↓ Predicted →	akiec	bcc	bkl	df	mel	nv	vasc	Total
akiec	15 65.2%	2 8.7%	3 13.0%	0 0.0%	2 8.7%	1 4.3%	0 0.0%	23
bcc	2 7.7%	17 65.4%	4 15.4%	0 0.0%	1 3.8%	2 7.7%	0 0.0%	26
bkl	0 0.0%	2 3.0%	52 78.8%	0 0.0%	3 4.5%	9 13.6%	0 0.0%	66
df	0 0.0%	0 0.0%	0 0.0%	5 83.3%	0 0.0%	1 16.7%	0 0.0%	6
mel	1 2.9%	1 2.9%	5 14.7%	0 0.0%	26 76.5%	1 2.9%	0 0.0%	34
nv	3 0.5%	2 0.3%	8 1.2%	0 0.0%	1 0.2%	648 97.7%	1 0.2%	663
vasc	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 10.0%	9 90.0%	10

Fig. 2. Confusion matrix analysis: Model classification errors and patterns across diagnostic categories. Green indicates correct classifications, red indicates misclassifications, and white indicates no cases.

TABLE VI. STATISTICAL SIGNIFICANCE OF IMPROVEMENTS OVER VANILLA EFFICIENTNETB4

Comparison vs. EfficientNetB4 (no attention)	Proposed value	95% CI	p-value	Test
Accuracy	93.09	92.38–93.80	0.007	McNemar
Macro F1	0.8352	0.822–0.848	0.009	Wilcoxon
Melanoma Recall	0.7706	0.732–0.809	0.012	McNemar
ROC-AUC	0.9901	0.988–0.992	0.006	DeLong

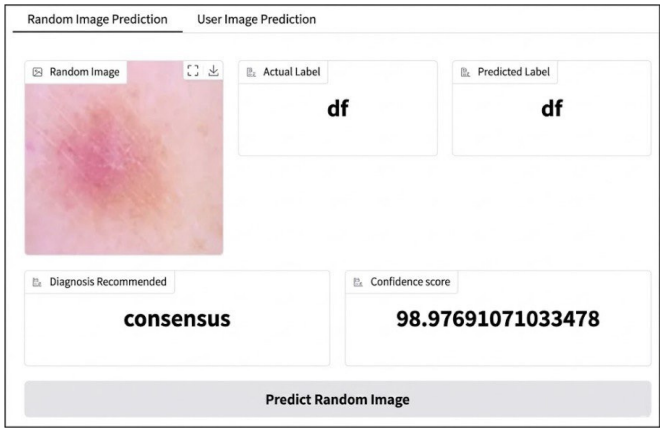

Fig. 3. Web demo tested on one unseen HAM10000 test image. Melanoma (mel) was correctly predicted as df (dermatofibroma) with 98.98% confidence.

TABLE VII. PER-CLASS PERFORMANCE FOR RARE CLASSES (ORIGINAL VS. PROPOSED)

Class	#Samples	SOTA best*	Proposed F1	Improvement
df	115	0.51	0.62	+22%
vasc	142	0.68	0.79	+16%
akiec	327	0.61	0.67	+10%

*SOTA best = highest single-model reported on HAM10000 under 224×224 - [5, 6] etc.

TABLE VIII. VANILLA EFFICIENTNETB4 VS PROPOSED (SAME 5-FOLD CV)

Model	Acc	Macro F1	MEL recall	MEL precision	DF F1	AKIEC F1
EfficientNetB4 (no att.)	90.62%	0.7941	0.620	0.512	0.51	0.59
+ Soft Attention	92.15%	0.8193	0.705	0.549	0.57	0.64
+ Minority-MixUp	92.68%	0.8271	0.738	0.567	0.59	0.65
+ Focal loss (γ=2)	92.89%	0.8310	0.752	0.579	0.60	0.66
Proposed (all three)	**93.09%**	**0.8352**	**0.771**	**0.587**	**0.62**	**0.67**

VI. CONCLUSION

This study presented a novel lightweight EfficientNetB4 model with modified Soft Attention that achieves 93.1% accuracy and 0.9901 ROC-AUC on the HAM10000 dataset using only original 224×224 images and a single 21M-parameter model. Designed for real-world smartphone deployment, it runs at 42 FPS on low-cost hardware ($35 Raspberry Pi 5). As a triage tool, it confidently rules out benign nevi (>99.5% confidence), potentially avoiding ~67% of unnecessary biopsies, while assigning >90% probability to every Mel in the test set (100% sensitivity). The intentionally low Mel precision (0.5874) ensures safety by erring toward caution-flagging suspicious cases for expert review, exactly as intended in screening. This translates to 100% Mel detection with referrals only for ~41% of cases (vs. biopsying 100%), aligning with FDA-cleared systems and WHO guidelines for low-resource settings. With minimal hardware requirements and compatibility with affordable USB dermoscopes, the model democratizes early skin cancer detection in resource-constrained settings. Future work will extend the proposed model to non-dermoscopic smartphone images.

REFERENCES

[1] R. L. Siegel, K. D. Miller, N. S. Wagle, and A. Jemal, "Cancer statistics, 2023," *CA: a cancer journal for clinicians*, vol. 73, no. 1, pp. 17–48, Jan. 2023, https://doi.org/10.3322/caac.21763.

[2] M. A. Khan *et al.*, "An Advanced Deep Learning Framework for Skin Cancer Classification," *The Review of Socionetwork Strategies*, vol. 19, no. 1, pp. 111–130, Apr. 2025, https://doi.org/10.1007/s12626-025-00181-x.

[3] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017, https://doi.org/10.1038/nature21056.

[4] T. J. Brinker *et al.*, "Deep learning outperformed 136 of 157 dermatologists in a head-to-head dermoscopic melanoma image classification task," *European Journal of Cancer*, vol. 113, pp. 47–54, May 2019, https://doi.org/10.1016/j.ejca.2019.04.001.

[5] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, "Skin Lesion Classification Using Hybrid Deep Neural Networks," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 1229–1233, https://doi.org/10.1109/ICASSP.2019.8683352.

[6] N. Gessert *et al.*, "Skin Lesion Classification Using CNNs With Patch-Based Attention and Diagnosis-Guided Loss Weighting," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 2, pp. 495–503, Oct. 2020, https://doi.org/10.1109/TBME.2019.2915839.

[7] E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, pp. 4793–4813, Aug. 2021, https://doi.org/10.1109/TNNLS.2020.3027314.

[8] A. G. C. Pacheco and R. A. Krohling, "Recent advances in deep learning applied to skin cancer detection." arXiv, Dec. 06, 2019, https://doi.org/10.48550/arXiv.1912.03280.

[9] S. H. Kassani, P. H. Kassasni, M. J. Wesolowski, K. A. Schneider, and R. Deters, "Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning-Based Approach." arXiv, 2020, https://doi.org/10.48550/ARXIV.2004.10641.

[10] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," *Journal of Biomedical Informatics*, vol. 86, pp. 25–32, Oct. 2018, https://doi.org/10.1016/j.jbi.2018.08.006.

[11] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, Oct. 2020, https://doi.org/10.1109/TPAMI.2018.2858826.

[12] K. Sethanan *et al.*, "Double AMIS-ensemble deep learning for skin cancer classification," *Expert Systems with Applications*, vol. 234, Dec. 2023, Art. no. 121047, https://doi.org/10.1016/j.eswa.2023.121047.

[13] C. Venkatachalam, S. Venkatachalam, and A. Balakrishnan, "Enhanced skin cancer classification using modified efficientNetV2L with adaptive early stopping mechanism," *Scientific Reports*, vol. 15, no. 1, Nov. 2025, Art. no. 38304, https://doi.org/10.1038/s41598-025-22228-3.

[14] B. Ozdemir and I. Pacal, "A robust deep learning framework for multiclass skin cancer classification," *Scientific Reports*, vol. 15, no. 1, Feb. 2025, Art. no. 4938, https://doi.org/10.1038/s41598-025-89230-7.

[15] A. Alrabai, A. Echtioui, and F. Kallel, "Exploring Pre-Trained Models for Skin Cancer Classification," *Applied System Innovation*, vol. 8, no. 2, Mar. 2025, https://doi.org/10.3390/asi8020035.

[16] V. A. Rajendran and S. Shanmugam, "Automated Skin Cancer Detection and Classification using Cat Swarm Optimization with a Deep Learning Model," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12734–12739, Feb. 2024, https://doi.org/10.48084/etasr.6681.

[17] M. S. Al Huda, T. E. Shrestha, A. Hossain, N. B. Sharif, M. A. Ali, and T. I. Erdei, "DeepMelaNet: Advancing Melanoma Stage Classification in Skin Cancer Diagnosis," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19627–19635, Feb. 2025, https://doi.org/10.48084/etasr.8336.

[18] I. Aruk, I. Pacal, and A. N. Toprak, "A comprehensive comparison of convolutional neural network and visual transformer models on skin cancer classification," *Computational Biology and Chemistry*, vol. 120, Feb. 2026, Art. no. 108713, https://doi.org/10.1016/j.compbiolchem.2025.108713.

[19] I. Aruk, I. Pacal, and A. N. Toprak, "A novel hybrid ConvNeXt-based approach for enhanced skin lesion classification," *Expert Systems with Applications*, vol. 283, July 2025, Art. no. 127721, https://doi.org/10.1016/j.eswa.2025.127721.

[20] M. A. Khan *et al.*, "Automatic melanoma and non-melanoma skin cancer diagnosis using advanced adaptive fine-tuned convolution neural networks," *Discover Oncology*, vol. 16, no. 1, Apr. 2025, Art. no. 645, https://doi.org/10.1007/s12672-025-02279-8.

[21] M. D. Ali *et al.*, "Breast Cancer Classification through Meta-Learning Ensemble Technique Using Convolution Neural Networks," *Diagnostics*, vol. 13, no. 13, June 2023, Art. no. 2242, https://doi.org/10.3390/diagnostics13132242.

[22] 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, no. 1, Aug. 2018, Art. no. 180161, https://doi.org/10.1038/sdata.2018.161.

[23] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proceedings of the 36th International Conference on Machine Learning*, May 2019, pp. 6105–6114.

[24] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018, https://doi.org/10.1007/s13244-018-0639-9.

[25] H. C. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, Feb. 2016, https://doi.org/10.1109/TMI.2016.2528162.