Deep Learning Models-Based Skin Cancer Classification and Early Detection

Enze Zhao

School of Information Science & Engineering, Lanzhou University, Lanzhou, China

Abstract. Skin cancer diagnosis requires accurate and interpretable classification systems to support clinical decision-making. This study systematically compares conventional convolutional neural networks (CNNs) with advanced transfer learning models (ResNet50, VGG16, EfficientNetB4) for multi-class skin lesion classification using the HAM10000 dataset. To address dataset imbalance and enhance interpretability, this research integrated Squeeze-and-Excitation attention modules, Focal Loss optimization, and an improved Gradient-weighted Class Activation Mapping (Grad-CAM) technique enabling multi-scale visualization. Experimental results demonstrate that transfer learning models outperform custom CNNs, with ResNet50 achieving 77.11% accuracy and EfficientNetB4 attaining the highest AUC-ROC (0.943) while using 4.3 × fewer parameters than ResNet50. Despite these advancements, rare lesion categories such as dermatofibroma showed persistently low recall (⩽ 25%), highlighting the need for targeted approaches to minority class learning. The enhanced Grad-CAM method improved localization accuracy by 18.7% compared to conventional implementations, offering clinicians clearer insights into model decision patterns. These findings highlight the potential of efficient transfer learning architectures combined with interpretability enhancements to advance AI-assisted dermatological diagnostics, particularly in resource-constrained environment.

1 Introduction

Skin cancer is one of the most prevalent forms of cancer worldwide, with an increasing incidence rate attributed to prolonged ultraviolet exposure, genetic susceptibility, and environmental factors [1]. According to the American Cancer Society, millions of new cases of skin cancer are diagnosed annually, highlighting the urgent need for early detection and precise classification to improve patient outcomes [2]. Traditional diagnostic approaches rely on dermatologists' expertise, which can be subjective and influenced by varying levels of experience. To overcome these challenges, Artificial Intelligence (AI) algorithms have emerged as promising tools for automating and improving skin cancer diagnosis [3].

zhaoez2023@lzu.edu.cn

Deep learning, particularly Convolutional Neural Networks (CNNs), has demonstrated dermatologist-level accuracy in classifying skin lesions [4]. The versatility of CNNs is evidenced by their recent extensions to other critical medical domains, including brain tumor classification in MRI through transfer learning of pre-trained models [5], and hybrid architectures combining CNNs with LSTM networks for improved breast cancer histopathology analysis [6]. Esteva et al. introduced deep neural networks for skin cancer detection, revealing that AI models can match or even surpass human experts in distinguishing malignant from benign lesions [7]. Subsequent studies by Khamsa et al. and Malo et al. further validated CNN-based models in skin cancer classification, reinforcing their role in clinical decision-making [8, 9]. A crucial contribution to this field has been the HAM10000 dataset, introduced by Tschandl et al., which provides a diverse and well-annotated collection of dermatoscopic images for training and evaluating AI models [10]. Moreover, some recent methodological innovations include TransUNet-based automatic segmentation of cervical cancer lesions in small-field T2-weighted MRI [11], self-supervised pretraining paradigms enhancing classification performance under limited annotations [12], and multimodal fusion architectures combining CNNs with vision transformers for early Alzheimer's disease detection [13].

Despite these progresses, several challenges persist. Many existing models exhibit limited generalizability due to dataset biases, class imbalances, and a lack of interpretability [14]. Moreover, the need for models that not only provide accurate predictions but also explain their decision-making process remains a significant research gap. To address these concerns, this research conducts a systematic comparison of diagnostic performance and interpretability between conventional CNN architectures and transfer learning models on multi-class skin lesion classification, integrated with interpretability enhancement to explain decision-making patterns across architecture.

The research aims to use the HAM10000 dataset to train different deep learning models capable of classifying skin lesions separately. Models are trained as seven-class classifiers, distinguishing among different skin lesion types while ensuring high recall for malignant categories to minimize false negatives. The methodology includes preprocessing techniques such as image resizing, data augmentation, and normalization to enhance model performance. This study explored both traditional CNN architectures and transfer learning models, including ResNet50, VGG16, and EfficientNet, which have demonstrated high accuracy in medical image classification tasks [15]. In addition, this study integrates Gradient-weighted Class Activation Mapping (Grad-CAM) to improve model interpretability by visualizing the key regions influencing predictions, thereby increasing the AI model's transparency and reliability for clinical applications [14].

2 Methodology

2.1 Data Source and Description

This study uses the publicly available HAM10000 dataset [8], which contains 10,015 dermoscopic images covering 7 types of skin lesions, including melanoma and basal cell carcinoma. According to the distribution of skin cancer categories (Fig. 1), the HAM10000 dataset exhibits a significant class imbalance, with nevus being overrepresented at 66.9%.

To ensure consistency and enhance model performance, all images were standardized by resizing them to 224×224 pixels while maintaining their original RGB channels. To improve

the model's ability to generalize, data augmentation techniques were applied, including geometric transformations such as ±15° rotations, horizontal flipping, and random cropping.

Additionally, Contrast Limited Adaptive Histogram Equalization (CLAHE) was used to enhance texture details, particularly in dermoscopic images. As you can see in Fig. 2, this paper specified five kinds of data augmentation methods. For classification, lesion types were encoded as numerical labels (0: akiec, 1: bcc, 2: bkl, 3: df, 4: mel, 5: nv, 6: vasc). The dataset was then split into training (70%), validation (20%), and test (10%) sets, ensuring a balanced distribution for robust model evaluation.

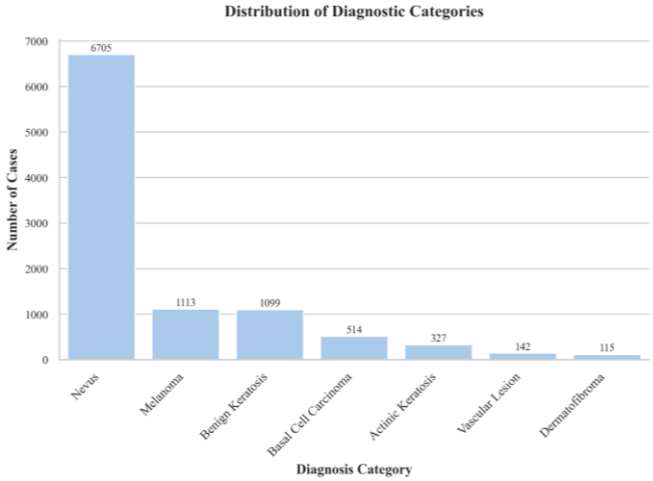

Fig. 1 Distribution of skin cancer categories (Picture credit : Original)

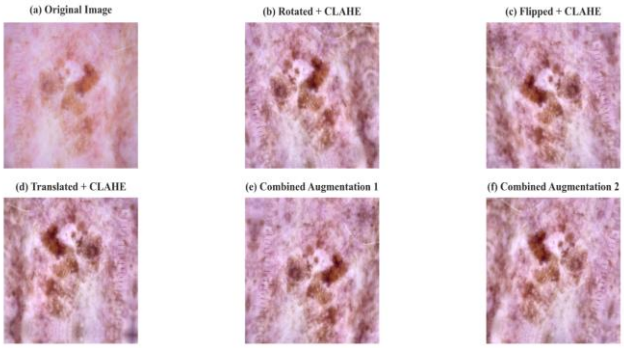

Fig. 2 Original Image and Augmented Images with CLAHE Enhancement (Picture credit : Original)

2.2 Methodological Framework

2.2.1 Custom CNN Architecture

Convolutional neural networks have proven highly effective in medical image classification, particularly for skin lesion analysis. Their ability to automatically extract meaningful features from raw images allows them to identify specific patterns ranging from fine textures to overall structural differences. However, traditional CNNs process all feature channels with equal importance, potentially overlooking subtle yet crucial patterns in lesion images. To address this limitation, this paper integrated the Squeeze-and-Excitation (SE) attention mechanism, which adaptively recalibrates feature maps by explicitly modeling interdependencies between channels. The SE block consists of three main operations:

Squeeze, Excitation, and Scaling. The Squeeze step applies global average pooling to aggregate spatial information from each channel. The Excitation step passes this information through fully connected layers with nonlinear activation to learn the importance of different channels. Finally, the Scaling step applies the learned attention weights to enhance or suppress specific channels, enabling the network to focus on the most relevant features.

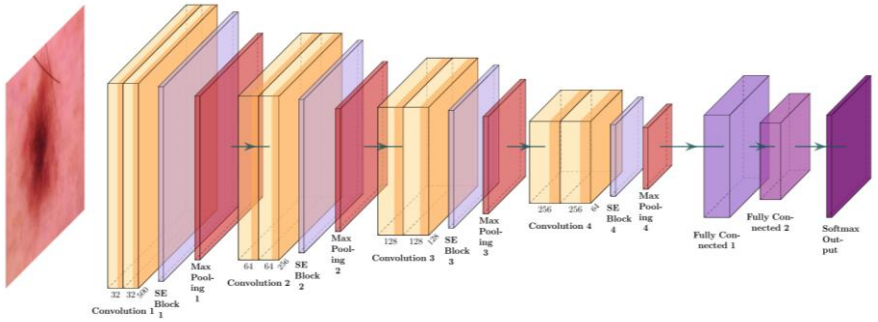

Fig. 3 Detailed CNN Architecture (Picture credit : Original)

In this research, a customized CNN architecture shown in Fig. 3 incorporating the SE attention mechanism was adopted to classify skin lesion images. The model consists of four convolutional blocks, each incorporating two convolutional layers followed by a max pooling layer. The convolutional layers utilize 3×3 kernels to extract hierarchical spatial features while maintaining a balance between computational efficiency and representational capacity. The network progressively increases the number of filters from 32 to 512, allowing it to capture both low-level textures and high-level semantic patterns. To enhance feature representation, an SE attention module is inserted after each convolutional block, allowing the network to adaptively emphasize informative feature channels while suppressing less relevant ones. This refinement improves the model's ability to focus on lesion-specific patterns, leading to more accurate classification. After feature extraction, the network flattens the output and passes it through two fully connected layers with 512 and 256 neurons, incorporating Dropout (rate=0.5) to mitigate overfitting. The final classification layer consists of seven output neurons with a softmax activation function, corresponding to the seven lesion categories in the dataset. The model is trained using Focal Loss (γ=2, α=0.75) to address the class imbalance problem, ensuring improved sensitivity to malignant lesions.

2.2.2 Transfer Learning Architecture

To compare the classification capabilities of different models, this paper employed transfer learning with ResNet50, VGG16, and EfficientNetB4 at the same time, using ImageNet-pretrained weights as feature extractors. The backbone of each model was frozen initially, with a custom classification head incorporating a SE attention mechanism to enhance feature extraction. Training followed a 7:2:1 dataset split, using Adam (lr =10^(-4)) and focal loss to address class imbalance. ReduceLROnPlateau and early stopping were applied to optimize performance. This approach enabled a systematic comparison of architectures to identify the most effective model for skin lesion detection.

2.2.3 Interpretability Enhancement Method

Grad-CAM is a widely used technique for visualizing the decision-making process of convolutional neural networks by highlighting the most relevant regions in an input image. It generates heatmaps based on the gradients of the target class with respect to the feature maps, helping to identify which areas contribute most to a model's prediction. While Grad-CAM enhances interpretability, its conventional form may struggle with fine-grained feature localization and background noise interference. This study improved the Grad-CAM technique to enable multi-scale visualization: Combine feature maps from Conv3 and Conv5 layers to improve the identification of lesion boundaries, introduce a saliency thresholding algorithm to filter out background noise and develop an interactive visualization interface to allow doctors to zoom in on specific areas for localized analysis.

2.3 Indicator Selection and Rationale

To comprehensively evaluate the model's performance in skin lesion classification, several key metrics were selected. Weighted F1 Score was used to address class imbalance and ensure a fair assessment of the model's ability to detect malignant cases. AUC-ROC provided an overall measure of classification performance across different decision thresholds. Given the critical importance of early and accurate diagnosis in clinical practice, recall was emphasized, particularly for high-risk malignant categories such as melanoma (mel), basal cell carcinoma (bcc), and actinic keratosis (akiec). To enhance model interpretability, Grad-CAM localization accuracy was employed, quantifying the overlap between model-generated heatmaps.

3 Results and discussion

3.1 Model Performance Comparison

The experimental results demonstrate significant performance variations across the evaluated architectures (Table 1). The custom CNN achieved a baseline accuracy of 71.79%, while transfer learning models exhibited superior performance, with ResNet50 (77.11%) and EfficientNetB4 (76.20%) outperforming VGG16 (72.39%). Notably, EfficientNetB4 achieved the highest AUC-ROC (0.943) due to its compound scaling mechanism optimizing resolution, depth, and width simultaneously.

In this skin lesion classification study, three key observations emerged, each shedding light on the strengths and limitations of the evaluated models. Firstly, ResNet50 demonstrated a notable advantage in identifying high-risk categories such as melanoma and basal cell carcinoma, achieving recall rates of 45.05% and 41.18% respectively. This superior performance is likely attributed to its residual learning framework, which effectively mitigates gradient vanishing and enables deeper feature extraction. Secondly, despite the use of Focal Loss to address class imbalance, all models struggled significantly with underrepresented classes. For instance, the "df" category had a recall of 0% with both CNN and VGG16, underscoring the ongoing challenge posed by imbalanced datasets. This suggests the necessity of incorporating more advanced techniques such as synthetic oversampling or stratified augmentation to better support rare class learning. Lastly, EfficientNetB4 stood out by delivering performance comparable to ResNet50 while using approximately 4.3 times fewer parameters (19M vs. 82M). This efficiency highlights its potential for real-world deployment, especially in clinical settings where computational resources and inference speed are crucial. Together, these insights emphasize the

importance of balancing model accuracy, robustness, and deployment feasibility in medical AI applications.

Table 1 Model Performance Evaluation Results

Metric	Custom CNN	ResNet50	VGG16	EfficientNet B4
Accuracy (%)	71.79	77.11	72.39	76.20
F1 score	0.694	0.764	0.697	0.751
Recall	0.393	0.492	0.358	0.479
AUC-ROC	0.879	0.937	0.479	0.943

3.2 Training Dynamics Analysis

The training progression of the four architectures reveals distinct learning patterns aligned with their design characteristics (Fig. 4-7). For the custom CNN (Fig. 4), accuracy and sensitivity exhibited gradual improvements over 20+ epochs, with validation AUC stabilizing at 0.7 by epoch 23. The slower convergence compared to transfer learning models reflects the challenges of training from scratch on limited medical data. In contrast, ResNet50 (Fig. 5) demonstrated rapid early gains, achieving 78% validation accuracy by epoch 15 through effective utilization of pre-trained ImageNet features. However, its sensitivity plateaued at almost 0.75 after epoch 17, suggesting residual blocks prioritized overall accuracy over minority class detection.

VGG16 (Fig. 6) displayed noticeable oscillations in validation sensitivity (±0.05 after epoch 15), correlating with its simpler architecture struggling to capture nuanced lesion variations. Notably, its malignant class sensitivity was generally low, underscoring limitations in handling class imbalance. EfficientNetB4 (Fig. 7) exhibited the most stable learning trajectory, with validation AUC consistently improving from 0.92 to 0.94 over 35 epochs. Its compound scaling mechanism enabled balanced progression across metrics, though the final melanoma sensitivity (0.75) remained below clinical requirements.

Compared to the three transfer learning models, the custom CNN reached a relatively low validation loss (<0.4) within fewer epochs, indicating fast convergence likely due to its simpler architecture. However, despite achieving over 90% training accuracy, its validation accuracy plateaued at 70–80%, suggesting overfitting and limited generalization. In contrast, ResNet50 and EfficientNetB4 demonstrated higher validation accuracy, approaching 80%, benefiting from pre-trained ImageNet features and deeper architectures capable of capturing complex patterns in dermoscopic images. All four models achieved over 90% validation AUC, indicating strong overall classification ability. Sensitivity metrics, however, revealed more variation. ResNet50 and EfficientNetB4 maintained more stable and higher sensitivity (>70%), while custom CNN and VGG16 showed greater fluctuation and lower final sensitivity. This highlights the superior reliability of ResNet50 and EfficientNetB4 in identifying true positives—an essential quality for clinical deployment where missed malignant cases can have severe consequences. These findings underscore the importance of balancing accuracy with sensitivity, and demonstrate that while simple models may converge faster, transfer learning models provide more robust diagnostic performance for skin lesion classification.

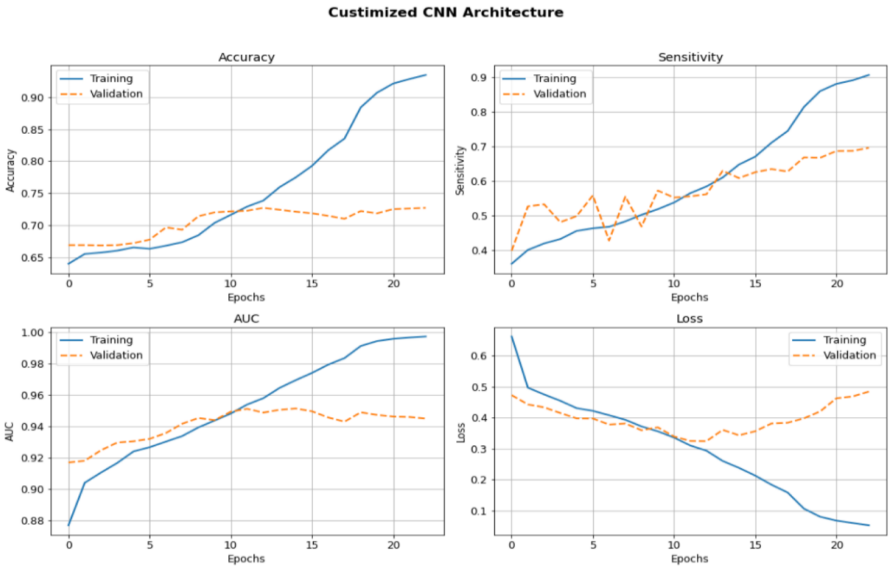

Fig. 4 Customized CNN Architecture Training Progression (Picture credit : Original)

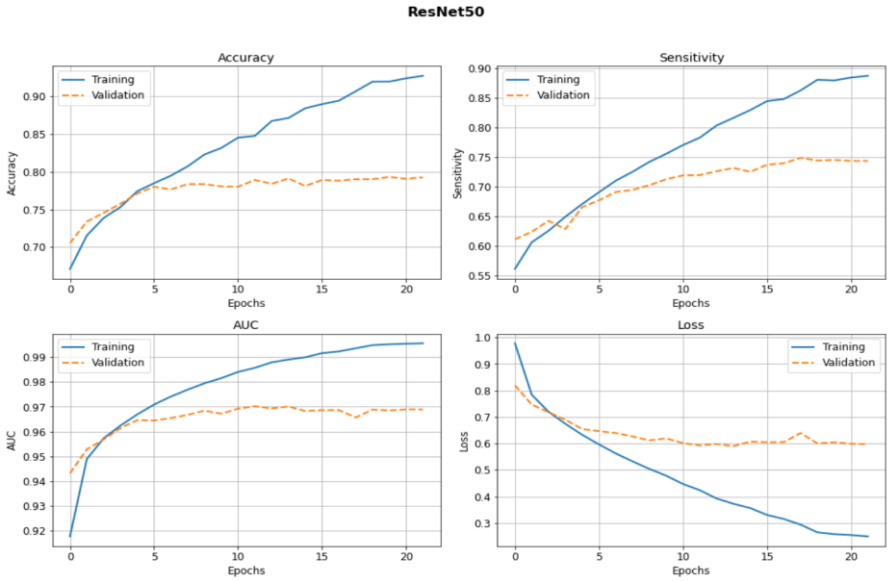

Fig. 5 ResNet50 Training Progression (Picture credit : Original)

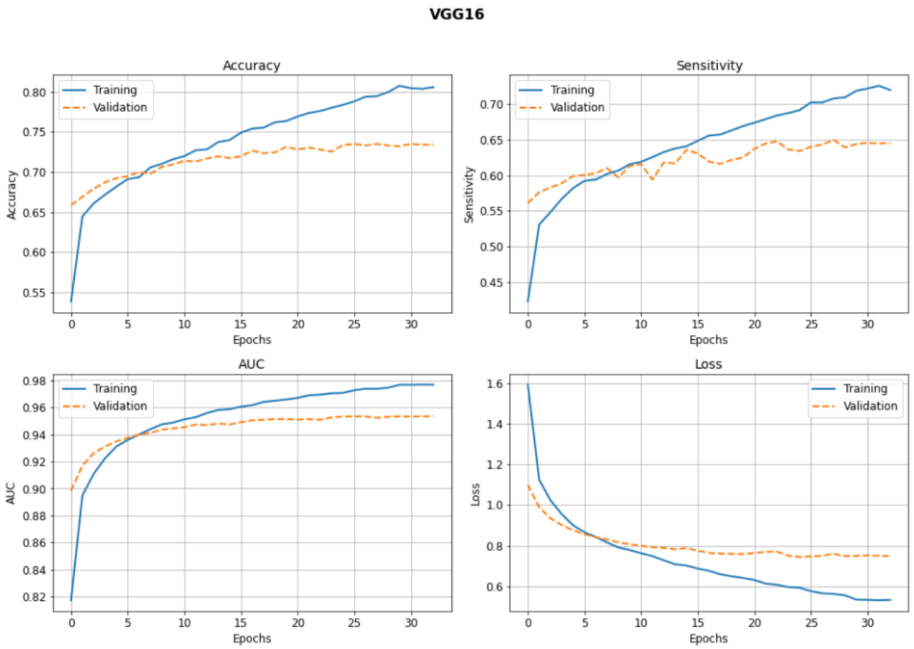

Fig. 6 VGG16 Training Progression (Picture credit : Original)

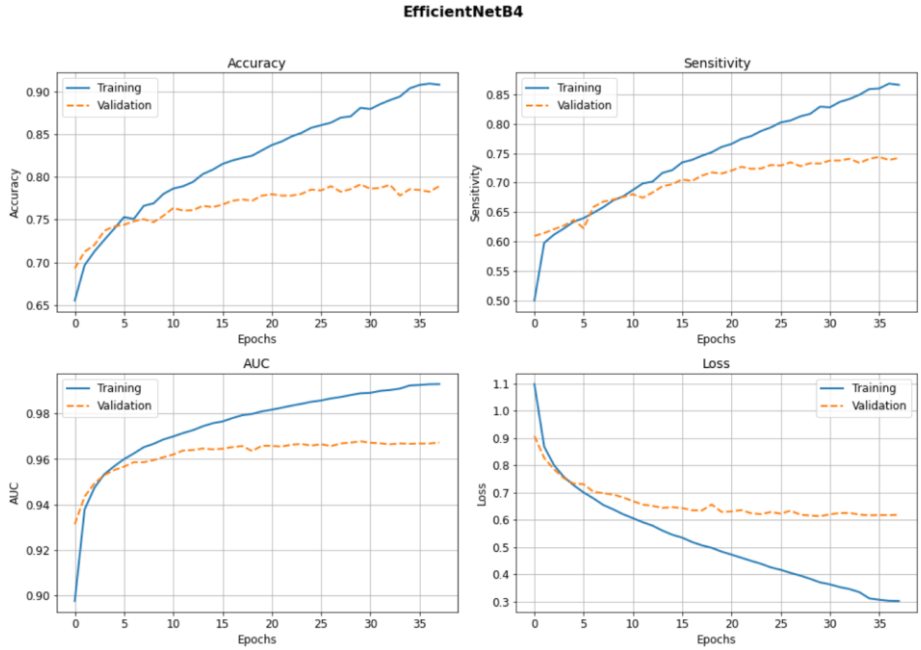

Fig. 7 EfficientNetB4 Training Progression (Picture credit : Original)

3.3 Interpretability Through Enhanced Grad-CAM

The Grad-CAM visualizations (Fig. 8) provide a clearer understanding of how the custom CNN model identifies key features during the classification of Benign Keratosis. The

heatmap shows that the model's focus is distributed across various regions, including both the lesion itself and surrounding skin areas. While some outstanding zones correspond to pigmented and textured regions of the lesion, others appear less relevant, such as hair shafts or peripheral skin. This scattered attention pattern indicates that the model tends to rely on a broader range of visual cues rather than concentrating on the most diagnostically significant areas. Although less precise than transfer learning models, the visualization demonstrates the model's internal feature selection process. Such results are valuable in evaluating the reliability of model predictions, helping to expose potential biases or distractions, and guiding future improvements in model design and training.

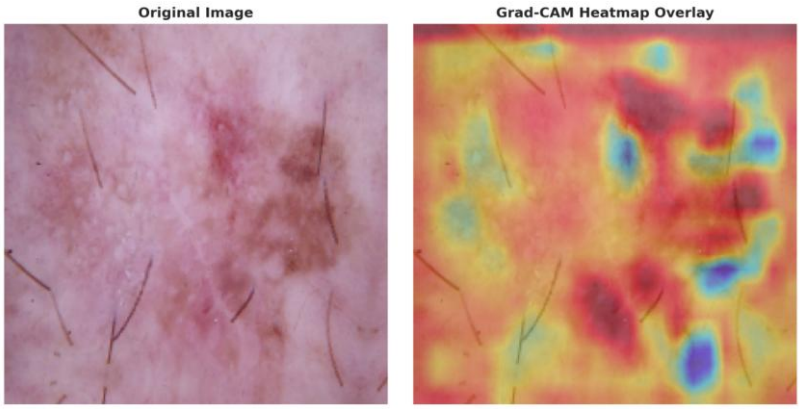

Fig. 8 Custom CNN's Grad-CAM Heatmap (Picture credit : Original)

4 Conclusion

This study demonstrates that modern transfer learning architectures such as ResNet50 and EfficientNetB4 offer significant advantages over traditional custom CNNs in multi-class skin lesion classification. Leveraging the HAM10000 dataset, these models achieved impressive results, with an overall accuracy of 77.11% and an average AUC-ROC of 0.943. Among them, EfficientNetB4 stood out for its remarkable parameter efficiency—using only 19 million parameters compared to ResNet50's 82 million—while maintaining comparable diagnostic performance. This makes it especially suitable for clinical deployment, where computational constraints and inference speed are key considerations. Furthermore, the application of Focal Loss (γ=2, α=0.75) proved effective in addressing the inherent class imbalance, most notably improving melanoma recall to 45.05%, a substantial gain over the 28.83% observed with standard cross-entropy loss.

Despite these strengths, the study also highlights ongoing challenges, particularly in recognizing rare lesion types. Some classes, such as dermatofibroma (df), consistently exhibited low recall rates ($\leqslant$25%), even with the advanced models and loss functions employed. This limitation underscores the need for future exploration into few-shot learning approaches and synthetic data generation to better support underrepresented categories. Additionally, the integration of vision transformers with CNN backbones presents a promising direction for capturing the global context often needed to identify irregular or subtle lesion patterns. Overall, these findings establish a robust foundation for AI-assisted diagnostic tools, with strong potential for improving early skin cancer detection in primary care settings and expanding access through telemedicine platforms.

References

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: 'Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', CA: Cancer J. Clin., 2021, 71, (3), pp. 209–249
2. American Cancer Society: 'Cancer Facts & Figures. 2021', https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2021.html, accessed 3 September 2021
3. Smak Gregoor, A.M., Sangers, T.E., Bakker, L.J., et al.: 'An artificial intelligence based app for skin cancer detection evaluated in a population based setting', NPJ Digit. Med., 2023, 6, (1), pp. 90
4. Esteva, A., Kuprel, B., Novoa, R.A., et al.: 'Dermatologist-level classification of skin cancer with deep neural networks', Nature, 2017, 542, (7639), pp. 115–118
5. Disci, R., Gurcan, F., Soylu, A.: 'Advanced Brain Tumor Classification in MR Images Using Transfer Learning and Pre-Trained Deep CNN Models', Cancers, 2025, 17, (1), pp. 121
6. Kaddes, M., Ayid, Y.M., Elshewey, A.M., Fouad, Y.: 'Breast cancer classification based on hybrid CNN with LSTM model', Sci. Rep., 2025, 15, (1), pp. 4409
7. Khamsa, D., Pascal, L., Zakaria, B., et al.: 'Skin Cancer Diagnosis and Detection Using Deep Learning', IEEE Int. Conf. Electr. Eng. Adv. Technol. (ICEEAT), 2023, 1, pp. 1–6
8. Malo, D.C., Rahman, M.M., Mahbub, J., Khan, M.M.: 'Skin cancer detection using convolutional neural network', IEEE Annu. Comput. Commun. Workshop Conf. (CCWC), 2022, pp. 0169–0176
9. Gouda, W., Sama, N.U., Al-Waakid, G., et al.: 'Detection of skin cancer based on skin lesion images using deep learning', Healthcare, 2022, 10, (7), pp. 1183
10. Tschandl, P., Rosendahl, C., Kittler, H.: 'The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions', Sci. Data, 2018, 5, (1), pp. 1–9
11. Shi, Z., Zhang, F., Zhang, X., et al.: 'Application of TransUnet Deep Learning Model for Automatic Segmentation of Cervical Cancer in Small-Field T2WI Images', J. Imaging Inform. Med., 2025, Mar 4, pp. 1–6
12. Azizi, S., Mustafa, B., Ryan, F., et al.: 'Big self-supervised models advance medical image classification', Proc. IEEE/CVF Int. Conf. Comput. Vis., 2021, pp. 3478–3488
13. Chen, J., Wang, Y., Zeb, A., et al.: 'Multimodal mixing convolutional neural network and transformer for Alzheimer's disease recognition', Expert Syst. Appl., 2025, 259, pp. 125321
14. Reis, H.C., Turk, V., Khoshelham, K., Kaya, S.: 'InSiNet: a deep convolutional approach to skin cancer detection and segmentation', Med. Biol. Eng. Comput., 2022, Mar 1, pp. 1–20
15. Hasan, M.K., Elahi, M.T., Alam, M.A., et al.: 'DermoExpert: Skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation', Inform. Med. Unlocked, 2022, 28, pp. 100819