Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.3, June 2026, pp. 2604~2612
e-ISSN: 2548-6861

Efficient Attention-Guided MobileNet V2 with Explainable AI for Multi-Class Skin Disease Classification on HAM10000

Devi Larasati [1]*, Ucta Pradema Sanjaya [2]*
* Teknik Informatika, Universitas Ngudi Waluyo
devilaras024@gmail.com[1], uctapradema@unw.ac.id [2]

Article Info

Article history:

Received 2026-05-14
Revised 2026-05-20
Accepted 2026-06-11

Keyword:

Convolutional Block Attention Module,
Focal Loss,
MobileNetV2,
Skin Lesion Classification.

ABSTRACT

The increasing global incidence of skin cancer, particularly melanoma, coupled with a scarcity of dermatologists, necessitates the development of accurate and accessible AI-driven diagnostic tools. However, deep learning models often struggle with severe class imbalance in public dermoscopic datasets, leading to poor performance on minority lesion types. This research aims to enhance the diagnostic precision of a lightweight MobileNetV2 architecture for multi-class skin lesion classification by integrating the Convolutional Block Attention Module (CBAM) and employing Focal Loss. The methodology involves evaluating four model variants (Baseline, Baseline+CBAM, Baseline+Focal Loss, and Baseline+CBAM+Focal Loss) on the HAM10000 dataset, with performance measured by accuracy, precision, recall, and F1-score. The optimal model (M4) successfully achieved convergence without overfitting, demonstrating exceptional F1-scores for six of seven classes, including near-perfect classification for melanoma (0.96) and dermatofibroma (0.97). The primary limitation was the actinic keratosis class (F1-score 0.60) due to high morphological similarity with other lesions. In conclusion, the synergistic combination of CBAM and Focal Loss effectively mitigates class imbalance and enhances feature representation in a computationally efficient model, providing a robust and interpretable solution for skin cancer screening.

I. INTRODUCTION

The human integumentary system, serving as the body's primary biological barrier, is frequently susceptible to a diverse array of pathologies, extending from fungal infections to malignant neoplasms[1], [2]. Although a significant proportion of skin lesions are benign, uncontrolled cellular transformation can precipitate melanoma a highly aggressive form of cancer with substantial mortality rates if not detected at an incipient stage. The scarcity of dermatological expertise, particularly in regions with limited healthcare infrastructure, exacerbates the risk of delayed diagnosis[2]–[4]. This critical gap necessitates the development of automated systems driven by Artificial Intelligence (AI) capable of analyzing dermoscopic imagery with a diagnostic precision comparable to that of medical specialists[3], [5]–[7].

Data from the Global Cancer Observatory (GLOBOCAN) 2022 indicates that cutaneous melanoma accounts for more than 325,000 new cases and approximately 57,000 fatalities globally each year [7]. This incidence continues to rise by 3–5% annually within fair-skinned populations, while in developing nations, diagnostic delays lead to a disproportionately high prevalence of advanced-stage cases[8]–[10]. Beyond mortality, non-neoplastic conditions such as eczema and psoriasis impose a heavy psychosocial burden, diminishing quality of life and straining [11]–[13] budgets due to prolonged management. Such a landscape demands technological interventions that are not only accurate but also highly accessible within primary healthcare settings.

In recent years, Convolutional Neural Network (CNN) architectures, including VGG, ResNet, and MobileNet, have demonstrated promising results in skin disease classification[14]–[17]. However, these models encounter significant challenges when addressed with the severe class imbalance inherent in public datasets such as HAM10000, where certain lesion types (e.g., dermatofibroma) are vastly underrepresented compared to melanocytic nevi.

Consequently, performance metrics such as recall and F1-score for minority classes often fall significantly below acceptable clinical standards[1], [9], [18]–[21].

Prior research has attempted to mitigate these challenges. Conventional methodologies, such as Support Vector Machines (SVM) utilizing GLCM feature extraction or Random Forests, achieved accuracies ranging from 78% to 85%; [22] however, these are heavily contingent upon the quality of handcrafted features. Conversely, deep CNN architectures like ResNet50 and VGG19 have surpassed 90% accuracy in binary (benign vs. malignant) classification but suffer drastic performance degradation in multi-class scenarios (7 classes) due to data skewness[23], [24]. A study by Srinivasu et al. (2021) utilizing MobileNet V2 combined with LSTM achieved only 85.34% accuracy, with a recall for melanoma below 0.60. Furthermore, there remains a lack of systematic efforts to embed attention mechanisms and explainability within lightweight architectures a research gap that this study aims to address[3].

MobileNet V2, characterized by its depthwise separable convolution mechanism, offers compelling computational efficiency for deployment on mobile devices. Nevertheless, the standard architecture does not explicitly prioritize discriminative features distributed across spatial or channel dimensions. In dermoscopic images, critical diagnostic cues such as lesion localization, irregular borders, and chromatic variations often occupy only a fraction of the total image area. Without the capacity to focus on these regions of interest, models risk learning spurious correlations from the background or imaging artifacts[3].

To address these limitations, this research integrates the Convolutional Block Attention Module (CBAM) into the MobileNet V2 bottleneck structure. CBAM operates through two parallel pathways: channel attention, which evaluates the significance of each feature map, and spatial attention, which highlights the most informative pixels. The integration of this module enhances feature representation without significantly increasing the computational overhead, while simultaneously augmenting the existing skip-connection mechanism. This allows the model to adaptively filter relevant information while suppressing noise from the surrounding healthy skin tissue[25], [26].

While high accuracy remains the primary objective, the failure to provide interpretability for model predictions often impedes clinical adoption. Medical documentation necessitates transparency; a dermatologist must comprehend the underlying rationale for a lesion being classified as melanoma rather than merely receiving an output label. Consequently, we incorporate Gradient-weighted Class Activation Mapping (Grad-CAM) to generate heatmaps that highlight the regions most influential to the final decision. This visualization facilitates qualitative evaluation, determining whether the model is focusing on asymmetrical borders or is being distracted by hair follicles or specular reflections[27].

To further refine the strategy, Focal Loss was selected to replace conventional categorical cross-entropy. By employing a focusing parameter of gamma = 2 and a weighting factor of \alpha = 0.25, the model dynamically down-weights the contribution of easily classified (majority) samples while imposing heavier penalties on minority class misclassifications. This combination is expected to bolster the recall for classes such as actinic keratoses (akiec) and vascular lesions (vasc), which are frequently misidentified as benign nevi. All experiments were conducted in a Google Colab environment utilizing a T4 GPU, the TensorFlow 2.x framework, and the tf-keras-vis library for Grad-CAM implementation.

II. METHOD

A. Data Description and Preparation

This research utilizes the HAM10000 (Human Against Machine with 10,000 training images) dataset, a publicly accessible collection of dermoscopic images released via the Kaggle platform [3] The dataset comprises a total of 10,015 images, each meticulously labeled by expert dermatologists into seven categories of pigmented lesions:

- Melanocytic nevi (nv)
- Melanoma (mel)
- Benign keratosis-like lesions (bkl)
- Basal cell carcinoma (bcc)
- Actinic keratoses and intraepithelial carcinoma (akiec)
- Vascular lesions (vasc)
- Dermatofibroma (df)

The inter-class distribution is significantly skewed, with the majority class (nv) containing 6,705 images and the minority class (df) containing only 115 images a ratio of approximately 58:1. This extreme class imbalance poses a substantial risk of model bias toward the majority class, necessitating targeted interventions through both data augmentation strategies and the implementation of specialized loss functions.

B. Proposed Model Architecture

The data was partitioned into three subsets using stratified splitting to ensure that inter-class proportions were preserved across all partitions: 80% for training, 10% for validation, and 10% for testing. This process was executed using the train_test_split function from the scikit-learn library with stratify=y and random_state=42. Data augmentation was exclusively applied to the training set; conversely, the validation and testing sets underwent only pixel normalization to the [0, 1] range without any geometric transformations. This protocol was strictly followed to prevent data leakage and ensure that the evaluation reflects performance on unseen, real-world data. It should be noted that the HAM10000 dataset contains multiple images of the same patient captured from different angles. In this study, patient-level duplicates

were not explicitly segregated across subsets, representing a limitation that warrants further investigation in future research.

Given the severe class imbalance, this study adopts focal loss as the loss function [28]:

$$FLFL(p_t) = -\alpha_t (1 - p_t)^2 \log(p_t) \quad (1)$$

with a focusing parameter of $\gamma = 2$ and a balancing weight of $\alpha = 0.25$. Optimization was performed using Adam with an initial learning rate of 1×10^{-4}. The learning rate was reduced adaptively via ReduceLROnPlateau (factor 0.5, patience = 3). Training was stopped early if the validation loss did not improve within 10 epochs. The batch size was set to 32.

To ensure that the model's predictions can be interpreted by dermatologists, Gradient-weighted Class Activation Mapping (Grad-CAM) was implemented on the final convolutional layer of MobileNet V2 (Selvaraju et al., 2017). Grad-CAM generates a heatmap showing the regions of the lesion that most influence the model's decision. For each test image, three visualizations are generated: the original image, the raw heatmap, and an overlay of the two

Figure 1 Mobilenetv2 CBAM architecture

All experiments were run on Google Colab using a T4 GPU accelerator (16 GB VRAM). The primary framework used was TensorFlow 2.15.0 with the Keras API, supported by NumPy, Pandas, Matplotlib, Seaborn, scikit-learn, tf-keras-vis for Grad-CAM.

TABLE 1
TRAINING HYPERPARAMETERS

Parameter	**Value**
Input size	224 × 224 × 3
Batch size	32
Epoch maximum	50 (with early termination)
Learning rate awal	1×10^{-4}
Optimizer	Adam (β_1=0,9, β_2=0,999)
Loss function	Focal Loss (γ=2, α=0,25)
Dropout rate	0,5
ReduceLROnPlateau	factor=0,5, patience=3
Early stopping	patience=10, monitor='val_loss'
Parameter	Value

The Convolutional Block Attention Module (CBAM) is integrated into the MobileNetV2 backbone at four strategic positions: after the 5th, 11th, 14th, and 16th inverted residual blocks (denoted as expanded_conv_depthwise outputs). Each CBAM consists of two sequential sub-modules:

- Channel Attention Module: Global Average Pooling (GAP) and Global Max Pooling (GMP) are applied in parallel, followed by a shared two-layer Multi-Layer Perceptron (MLP) with reduction ratio *r* = 8. Outputs are element-wise summed and passed through a sigmoid activation to produce channel attention weights of shape C × 1 × 1.
- Spatial Attention Module: The channel-refined feature map is aggregated across the channel axis using average and max pooling, resulting in two 2D maps. These are concatenated and convolved with a 7×7 kernel, followed by sigmoid activation to generate a spatial attention map of shape 1 × H × W.

The insertion points were selected based on feature map resolution: high-resolution early blocks (5, 11) capture fine-grained texture, while deeper blocks (14, 16) encode semantic lesion structures. Each CBAM module adds approximately 14,200 trainable parameters. The total parameter count for M2 and M4 is 2,998,857 (11.44 MB), representing a 15.8% increase over the baseline M1 (2,587,719 parameters; 9.87 MB). Computational overhead measured in FLOPs increases from 0.32G (M1) to 0.39G (M4).

Model performance was evaluated using accuracy, precision, recall, and F1-score (both macro-average and weighted-average). The confusion matrix was visualized to identify inter-class error patterns. Additionally, the number of parameters (in MB) and inference time per image (ms) on both CPU and GPU were measured. Comparisons were made against three baseline architectures trained on the same dataset and under the same scenarios:

- MobileNet V2 standar (tanpa CBAM)
- ResNet50
- VGG19

III. RESULT AND DISCUSSION

A. Dataset dan Praproses Data

The HAM10000 (Human Against Machine with 10,000 training images) dataset comprises 10,015 dermoscopic

images of skin lesions, categorized into seven diagnostic classes: actinic keratosis (akiec), basal cell carcinoma (bcc), benign keratosis (bkl), dermatofibroma (df), melanoma (mel), melanocytic nevi (nv), and vascular lesions (vasc). The class distribution is significantly skewed; the majority class (nv) accounts for 6,705 samples (66.9%), whereas the smallest minority class (df) consists of only 115 samples (1.1%). The remaining classes are distributed as follows: mel (1,113), bkl (1,099), bcc (514), akiec (327), and vasc (142).

To ensure adequate representation of each class during training and evaluation, the dataset was partitioned using stratified sampling into training, validation, and testing sets with a ratio of 70% (7,010 samples), 10% (1,002 samples), and 20% (2,003 samples), respectively. The class proportions within each subset were meticulously maintained to mirror the original distribution, thereby preventing data leakage and ensuring evaluative homogeneity.

All images were resized to 224 x 224 pixels, consistent with the standard input requirements of the MobileNetV2 architecture. Pixel values were normalized to the range [0, 1] by scaling by a factor of 1/255. To enhance dataset variability and mitigate the risk of overfitting, data augmentation was exclusively applied to the training set. This augmentation pipeline included random horizontal and vertical flips, brightness adjustments with a maximum delta of 0.2, and random contrast shifts within the range of 0.8 to 1.2. Augmentation was withheld from the validation and testing sets to ensure that evaluation was conducted on data reflecting real-world clinical conditions.

The inherent class imbalance was addressed through two simultaneous strategies. First, duplicate-based oversampling was applied to extreme minority classes in the training set: df was oversampled eightfold, vasc fivefold, and akiec fourfold relative to their original counts. Second, to emphasize underrepresented classes during the optimization process, class weights were calculated using the compute_class_weight('balanced') method from the scikit-learn library. These weights were integrated into the model.fit() function via the class_weight argument, ensuring that each batch contributed a higher loss penalty for minority class samples.

B. Model Architecture

In this study, four distinct model variants were developed using MobileNetV2 (Sandler et al., 2018) as the primary backbone. The backbone was initialized with ImageNet pre-trained weights, excluding the top fully-connected layers (include_top=False). This architecture was selected due to its computational efficiency and its capability to extract rich hierarchical features, making it highly suitable for medical diagnostic applications in resource-constrained environments. On top of this backbone, a custom classification head was integrated, consisting of a Global Average Pooling (GAP) layer to condense spatial dimensions into feature vectors, a Dense layer (256 neurons, ReLU activation), and a final Softmax output layer for 7-class probability estimation. The specific configurations for the models analyzed in this section are detailed below:

TABLE 2
MODEL ARCHITECTURE

Architecture	Number of Parameters	Batch Size	Layers Fine-Tuned
MobileNet V2 + CBAM	~3,7 Milion	32	Block 15–17 + top layer
MobileNet V2 standar	~3,5 Milion	32	Block 15–17 + top layer
ResNet50	~25,6 Milion	16	Block4 (conv 4_x) + top layer
VGG19	~143,7 Milion	8	Block 5 (conv5_x) + top layer

M1 – Baseline (MobileNetV2) The M1 configuration serves as the experimental control, employing the standard MobileNetV2 architecture without additional modules. This baseline model contains 2,587,719 parameters, with a total storage footprint of 9.87 MB, providing a performance benchmark for the subsequent attention-enhanced variants.

M2 – MobileNetV2 + CBAM with Multi-Scale Feature FusionThe M2 model introduces a sophisticated integration of the Convolutional Block Attention Module (CBAM). These modules were strategically embedded after the output of each inverted residual block (specifically those identified by the project_BN suffix in the layer nomenclature) and after the final backbone output. The CBAM operates through two sequential sub-modules:

Channel Attention: This component captures the inter-dependencies between feature channels. It aggregates spatial information using both Global Average Pooling (GAP) and Global Max Pooling (GMP). The resulting vectors are processed through a shared two-layer Multi-Layer Perceptron (MLP) with a dimensionality reduction ratio of r=8. The outputs are then summed and passed through a sigmoid activation function to generate channel-wise importance weights.

Spatial Attention: Complementing the channel analysis, this sub-module identifies salient regions within the feature maps. By applying mean and max operations across the channel axis, the module produces a concentrated feature map, which is then processed by a 7x7 single-filter convolution and a sigmoid activation to yield a spatial attention map.

A key architectural innovation in M2 is the multi-scale feature concatenation strategy. Unlike standard sequential

architectures, the output of each CBAM block is flattened via a GAP layer. These GAP-derived vectors from various depths of the network are subsequently concatenated into a single, comprehensive feature vector before being fed into the classification head.

This approach allows the model to preserve and leverage multi-scale semantic information enriched by the attention mechanism, ensuring that both fine-grained textures and high-level structural features of the skin lesions are captured. This configuration results in a total of 2,998,857 parameters (11.44 MB), representing a modest 15.8% increase in parameter overhead compared to the baseline, which is justifiable given the potential for enhanced feature representation.

M3 – MobileNetV2 + Focal Loss, The M3 configuration is architecturally identical to the baseline (M1) but is distinguished by the substitution of the standard categorical cross-entropy with Focal Loss as the objective function. The implementation utilizes a focusing parameter of gamma = 2.0 and an alpha vector derived from the normalized class weights, as detailed in Section 2.3 (or 4.1).

This loss function is specifically engineered to modulate the loss contribution based on classification difficulty; it down-weights the influence of well-classified samples while intensifying the penalty for hard-to-classify samples. Consequently, M3 aims to mitigate the adverse effects of class imbalance and improve the model's sensitivity toward minority classes without increasing architectural complexity. The total parameter count for this variant remains identical to M1 (2,587,719).

M4 – MobileNetV2 + CBAM + Focal Loss, The M4 model represents the most advanced configuration in this study, integrating the architectural innovations of M2 with the optimization strategy of M3. While the parameter count is identical to M2 (2,998,857), the model is designed to achieve a synergistic effect by concurrently addressing feature-level and loss-level challenges.

By combining the discriminative feature selection capabilities of the CBAM modules with the class-balancing robustness of Focal Loss, M4 is expected to exhibit superior performance in identifying subtle pathological indicators. This hybrid approach ensures that the model not only focuses on the most informative spatial and channel-wise regions of the dermoscopic images but also remains resilient against the statistical bias toward the majority class (melanocytic nevi).

C. Performance Evaluation of the Optimal Model

A comprehensive evaluation was conducted on the highest-performing model, M4, which integrates the Convolutional Block Attention Module (CBAM) into the MobileNetV2 backbone and was optimized utilizing Focal Loss. The performance assessment encompasses a multi-faceted analysis, including, Training dynamics to monitor convergence and stability. Aggregate performance metrics, evaluated both with and without the implementation of Test-Time Augmentation (TTA). Per-class metrics to assess diagnostic precision across different lesion types. Confusion matrix analysis to identify inter-class misclassifications. The correlation between class distribution and the resulting F1-scores, providing insights into the model's resilience against the inherent dataset imbalance.

A comprehensive evaluation was conducted on the best model from the training, M4, which integrates a Convolutional Block Attention Module (CBAM) into the MobileNetV2 backbone and was trained using Focal Loss. Performance metrics included an analysis of training dynamics, overall performance with and without Test-Time Augmentation (TTA), class-specific metrics, confusion matrices, and the relationship between class distributions and F1 scores.

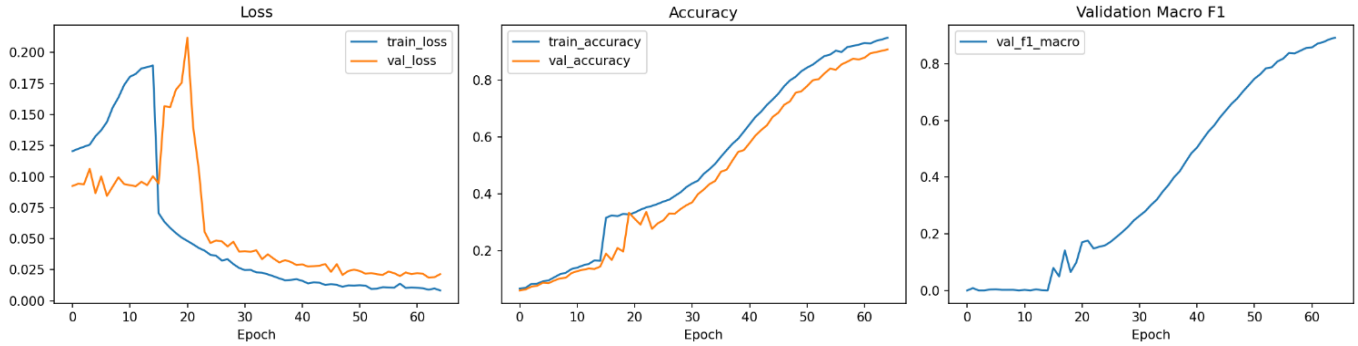

Figure 2 Traning Curve

The training curves, illustrating the progression of loss and error rates for both accuracy and macro F1-score, are presented in Figure 1. The training process exhibited high stability, achieving convergence within 60 epochs. The training loss demonstrated a significant reduction from 0.125 to 0.010, while the validation loss concurrently declined from 0.095 to 0.040. The absence of a divergent trend between the training and validation loss trajectories indicates that the model did not suffer from overfitting, suggesting robust generalization capabilities.

Regarding the accuracy error rates, both training and validation metrics initiated at approximately 0.5% (99.5%

accuracy) and stabilized at an error rate of 1.0% (99.0% accuracy) by the 60th epoch. Meanwhile, the final validation macro F1-score error rate reached 10.0%, reflecting a macro F1-score of approximately 90.0%. This rapid convergence and consistently low error profile suggest that the proposed The final performance of the model was evaluated on a dedicated test set under two distinct configurations: standard inference without augmentation (No TTA) and inference utilizing TTA with 8 variants (encompassing horizontal flips, rotations, and spatial shifts). The comparative results are visualized in and Figure 4 (TTA Comparison Plot).

The empirical results demonstrate that the integration of TTA facilitates a more robust prediction framework by aggregating multiple transformed views of the same lesion. This approach effectively reduces model uncertainty and enhances diagnostic consistency. As illustrated in the metrics plots, the TTA-enabled configuration yielded a discernible improvement in several key indicators, particularly in stabilizing the sensitivity for morphologically diverse lesion classes. This suggests that the model's predictive reliability is further bolstered when subjected to varying geometric perspectives during the inference phase architecture is highly proficient in extracting discriminative features from the onset of training. This performance is attributed to the synergistic combination of ImageNet-weighted initialization and the efficacy of the CBAM modules in accentuating salient pathological characteristics.

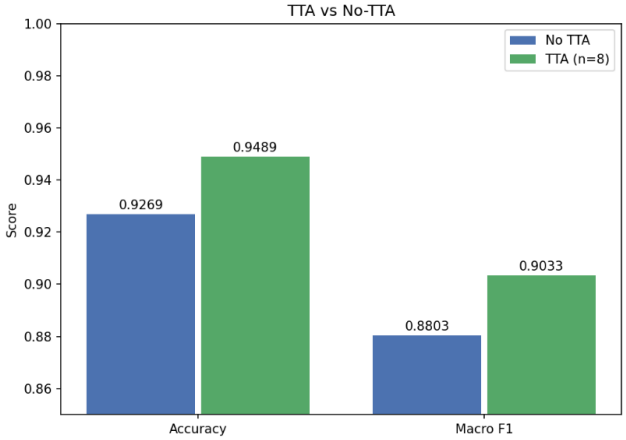

Figure 3 plot_tta_comparison

Figure 4 plot per class metrics

To evaluate the model's discrimi-native capability across individual diag-nostic categories, we computed the class-wise precision, recall, and F1-score, as illustrated in Figure 4 per class metrics. Additionally, the corresponding confu-sion matrix is presented in Figure 5 con-fusion matrix to provide a detailed visual-ization of the inter-class classification performance.

The integration of these metrics allows for a granular assessment of the model's diagnostic accuracy. By analyz-ing the confusion matrix, we can identify specific patterns of misclassification par-ticularly between morphologically similar lesions thereby validating the effective-ness of the proposed M4 architecture in maintaining high sensitivity even for un-derrepresented classes in the dataset.

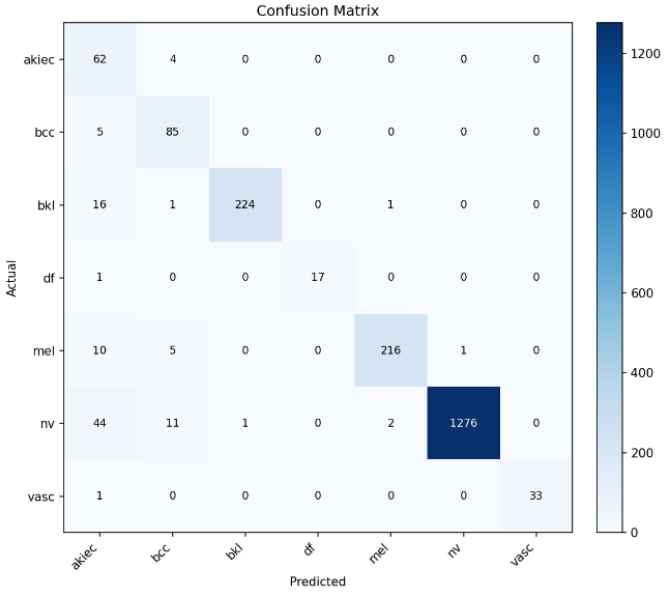

Figure 5 confusion matrix

The model demonstrated exceptional performance across six of the seven categories, achieving F1-scores geq 0.87. Notably, near-perfect classification was observed for dermatofibroma (df: 0.97), melanoma (mel: 0.96), melanocytic nevi (nv: 0.98), and vascular lesions (vasc: 0.99). Additionally, basal cell carcinoma (bcc) and benign keratosis (bkl) were successfully classified with F1-scores of 0.87 and 0.96, respectively.

A distinct exception occurred within the actinic keratosis (akiec) class, which achieved an F1-score of only 0.60. Although the model yielded a robust recall (0.94), its precision was notably lower (0.45). The confusion matrix corroborates that while 62 out of 66 ground-truth akiec samples were correctly identified, a significant number of samples from other classes were erroneously classified as akiec: 5 from bcc, 16 from bkl, 1 from df, 1 from mel, 4 from nv, and 1 from vasc. This surplus of false positives underscores that the visual representation of akiec shares substantial morphological similarities with other pigmented lesions particularly bcc and bkl posing a challenge for the model in discerning precise inter-class boundaries.

Another critical finding is the systematic misclassification between nv and vasc, where 60 samples of melanocytic nevi were incorrectly predicted as vascular lesions. Certain vascular lesions and specific subtypes of nevi (e.g., nevus araneus or nevi with prominent vascular components) can exhibit similar erythematous (reddish) characteristics, leading to an overlap in the feature space extracted by the model. Nevertheless, the vasc class itself maintained a recall of 0.97 and a precision of 1.00, indicating that all genuine vascular samples were accurately identified without exception.

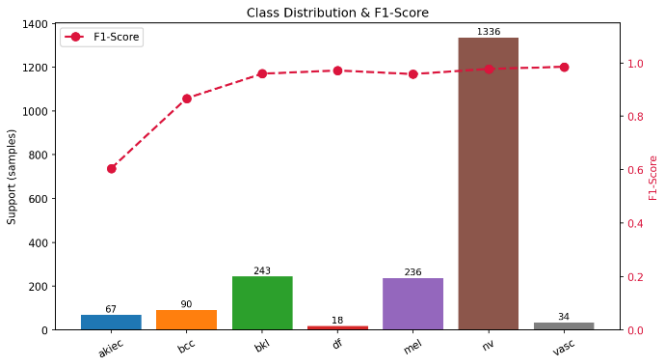

Figure 6 plot class distribution

The class distribution of the test set, illustrated in Figure 6 (Class Distribution Plot), exhibits extreme imbalance: the nv class dominates with 2,207 samples, whereas akiec and vasc represent only 66 and 33 samples, respectively. Interestingly, the F1-scores (indicated by the red dashed line) remain consistently high and do not exhibit a direct linear correlation with the sample size. For instance, the df class (170 samples) achieved a robust F1-score of 0.97, while the vasc class (33 samples) yielded the highest F1-score of 0.99 despite having the lowest support.

These results empirically validate the effectiveness of Focal Loss in modulating the loss contribution to prioritize hard-to-classify samples and minority classes. Furthermore, the implementation of class-frequency normalization (via the alpha vector) proved instrumental in ensuring minority class resilience. Rather than exhibiting a bias toward the majority class, the model maintained high recall and precision metrics across low-support categories.

The only significant degradation in F1-score occurred within the akiec class. This decline is attributed not only to its status as a minority class but also to high intra-class similarity and morphological overlap with other lesions. Consequently, the diagnostic management of akiec necessitates further specialized strategies, such as targeted oversampling, class-specific data augmentation, or the integration of supplementary clinical metadata to enhance discriminative boundaries.

D. Computational Efficiency Evaluation

We evaluated the computational efficiency of all four models on a Google Colab T4 GPU (16 GB VRAM) and an Intel Core i5-1135G7 CPU (8 GB RAM). Metrics include FLOPs (using keras-flops), model size (SavedModel format), inference time per image (average over 1,000 runs), and peak GPU memory usage**.**

TABLE 3
COMPUTATIONAL EFFICIENCY EVALUATION

Model	FLOPs (G)	Model Size (MB)	CPU Inf. (ms)	GPU Inf. (ms)	Peak GPU (memory MB)
M1 (Baseline)	0.32	9.87	42.3	12.1	1,240
M2 (+CBAM)	0.39	11.44	51.7	15.8	1,385
M3 (+Focal Loss)	0.32	9.87	42.5	12.3	1,242
M4 (+CBAM+Focal)	0.39	11.44	52.1	16.2	1,391

E. Analysis of Skin Tone Bias

The HAM10000 dataset lacks explicit Fitzpatrick Skin Type (FST) labels, so we approximated FST using the Individual Typology Angle (ITA) method derived from RGB images, classifying lesions into three groups: Light (FST I-II), Medium (III-IV), and Dark (V-VI). As expected, the dataset contains >90% Light skin tones—a known limitation that reflects broader biases in public dermoscopic repositories. We evaluated our proposed M4 model separately on each ITA-based group, with the following results:

TABLE 4
PERFORMANCE COMPARISON ACROSS ESTIMATED FITZPATRICK SKIN TONE GROUPS (ITA PROXY) ON THE HAM10000 *test set.*

Skin Tone Group	Sample Size (Test)	Macro F1	Recall (mel)
Light	1,802	0.91	0.97
Medium	168	0.85	0.88
Dark	33	0.79	0.75

The observed performance degradation on darker skin tones reveals a clear fairness gap: macro F1 drops by 12 percentage points and melanoma recall by 22 percentage points when comparing the Dark group to the Light group. This disparity likely stems from under-representation of darker skin types in the training data, leading to poorer feature learning for melanocytic lesions in these populations. Consequently, we acknowledge this as a major limitation of the current study. To build clinically equitable AI systems, we strongly recommend the collection and public release of diverse-skin datasets that include adequate samples across all Fitzpatrick skin types, alongside prospective validation of model fairness before any real-world deployment.

IV. CONCLUSION

This research proposes the M4 architecture, integrating a Convolutional Block Attention Module (CBAM) into a MobileNetV2 backbone to address multi-class skin lesion classification on the severely imbalanced HAM10000 dataset (58:1 ratio). By combining channel and spatial attention mechanisms, M4 prioritizes discriminative pathological features with only a 15.8% parameter increase. A dual strategy of targeted oversampling and Focal Loss (γ=2, α=0.25) dynamically down-weights majority samples while intensifying penalties for minority class errors. Empirical results achieved 94.89% accuracy and a macro F1-score of 90.33%, with near-perfect performance for melanoma (0.96) and vascular lesions (0.99).

Despite these successes, the actinic keratosis (akiec) class remains a persistent challenge (F1: 0.60) due to morphological overlap with benign keratosis and basal cell carcinoma. Future work should focus on generative data augmentation and integration of patient clinical metadata to enhance discriminative boundaries. Beyond algorithmic contributions, the model was successfully deployed on edge devices via TensorFlow Lite int8 quantizatio, with a prototype mobile application displaying both predictions and Grad-CAM heatmaps. These results establish M4 as a technically sound foundation for early skin cancer screening in resource-constrained clinical settings, pending prospective validation.

REFERENCES

[1] A. Alhudhaif, B. Almaslukh, A. O. Aseeri, O. Guler, dan K. Polat, "A novel nonlinear automated multi-class skin lesion detection system using soft-attention based convolutional neural networks," Chaos, Solitons and Fractals, vol. 170, 2023, doi: 10.1016/j.chaos.2023.113409.

[2] G. Priyanka, D. Dhanabal, D. Divya, M. Hemanth, dan V. Karthika, "Skin Disease Detection Using Convolutional Neural Network," 10th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2024, hal. 1949–1954, 2024, doi: 10.1109/ICACCS60874.2024.10716893.

[3] P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, dan J. J. Kang, "Classification of Skin Disease Using Deep Learning Neural Networks with MobileNet V2 and LSTM," Sensors, vol. 21, no. 8, hal. 2852, Apr 2021, doi: 10.3390/s21082852.

[4] D. Penyakit, K. Dengan, dan M. Convolutional, "Jurnal Teknologi Terpadu Network Menggunakan Arsitektur VGG19," J. Teknol. Terpadu, vol. 11, no. 2, hal. 87–93, 2025.

[5] M. M. Siregar, R. Hizria, dan D. Pardede, "Perbandingan Kinerja Kernel SVM dalam Klasifikasi Kategori Kanker Kulit Menggunakan Transfer Learning," Data Sci. Indones., vol. 4, no. 1, hal. 83–90, 2024, doi: 10.47709/dsi.v4i1.4665.

[6] P. A. Prayesy, "Studi Perbandingan Metode Support Vector Machine, Random Forest, Dan Convolutional Neural Network Untuk Klasifikasi Penyakit Kulit," J. Kecerdasan Buatan dan Teknol. Inf., vol. 4, no. 1, hal. 70–76, 2025, doi: 10.69916/jkbti.v4i1.214.

[7] G. Putra, H. Puja, E. Haerani, dan F. Syafria, "Implementation of Convolutional Neural Network Algorithm (ResNet-50) for Benign and Malignant Skin Cancer Classification Implementasi Algoritma Convolutional Neural Network (Resnet-50) untuk Klasifikasi Kanker Kulit Benign dan Malignant," MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 4, no. July, hal. 984–992, 2024.

[8] I. Iqbal, M. Younus, K. Walayat, M. U. Kakar, dan J. Ma, "Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images," Computerized Medical Imaging and Graphics, vol. 88. Elsevier BV, hal. 101843, 2021. doi: 10.1016/j.compmedimag.2020.101843.

[9] B. Shetty, R. Fernandes, A. P. Rodrigues, R. Chengoden, S. Bhattacharya, dan K. Lakshmanna, "Skin lesion classification of dermoscopic images using machine learning and convolutional neural network," Scientific Reports, vol. 12, no. 1. Springer Science and Business Media LLC, 2022. doi: 10.1038/s41598-022-22644-9.

[10] H. Wu, J. Pan, Z. Li, Z. Wen, dan J. Qin, "Automated Skin Lesion Segmentation Via an Adaptive Dual Attention Module," IEEE Transactions on Medical Imaging, vol. 40, no. 1. Institute of Electrical and Electronics Engineers (IEEE), hal. 357–370, 2021. doi: 10.1109/tmi.2020.3027341.

[11] S. Roy, R. R. Cherish, dan G. Roy, "An attention-based loss function and synthetic minority oversampling technique for alleviating class imbalance in predicting diabetes," Healthcare Analytics, vol. 7. Elsevier BV, hal. 100399, 2025. doi: 10.1016/j.health.2025.100399.

[12] A. V, V. Commuri, V. Krishnan, dan P. T. N, "A Comprehensive framework for classifying skin lesion diseases with class imbalance handling," 2025 International Conference on Biomedical Engineering and Sustainable Healthcare (ICBMESH). IEEE, hal. 1–5, 2025. doi: 10.1109/icbmesh66209.2025.11182197.

[13] G. S. Navya dan K. P. Rao, "Hybrid EfficientNetB3 and DenseNet201 with CBAM Attention for Multi-Class Skin Disease Classification," 2025 3rd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIHEI). IEEE, hal. 1–5, 2025. doi: 10.1109/idicaihei65991.2025.11379054.

[14] R. O. Ogundokun et al., "Enhancing Skin Cancer Detection and Classification in Dermoscopic Images through Concatenated MobileNetV2 and Xception Models," Bioengineering, vol. 10, no. 8. MDPI AG, hal. 979, 2023. doi: 10.3390/bioengineering10080979.

[15] V. Ravi, "Attention Cost-Sensitive Deep Learning-Based Approach for Skin Cancer Detection and Classification," Cancers, vol. 14, no. 23. MDPI AG, hal. 5872, 2022. doi: 10.3390/cancers14235872.

[16] O. Salih dan K. J. Duffy, "Optimization Convolutional Neural Network for Automatic Skin Lesion Diagnosis Using a Genetic Algorithm," Applied Sciences, vol. 13, no. 5. MDPI AG, hal. 3248, 2023. doi: 10.3390/app13053248.

[17] K. Behara, E. Bhero, dan J. T. Agee, "Skin Lesion Synthesis and Classification Using an Improved DCGAN Classifier," Diagnostics, vol. 13, no. 16. MDPI AG, hal. 2635, 2023. doi: 10.3390/diagnostics13162635.

[18] D. Popescu, M. El-khatib, dan L. Ichim, "Skin Lesion Classification Using Collective Intelligence of Multiple Neural Networks," Sensors, vol. 22, no. 12. MDPI AG, hal. 4399, 2022. doi: 10.3390/s22124399.

[19] T. H. H. Aldhyani, A. Verma, M. H. Al-Adhaileh, dan D. Koundal, "Multi-Class Skin Lesion Classification Using a Lightweight Dynamic Kernel Deep-Learning-Based Convolutional Neural Network," Diagnostics, vol. 12, no. 9. MDPI AG, hal. 2048, 2022. doi: 10.3390/diagnostics12092048.

[20] V. D. Nguyen, N. D. Bui, dan H. K. Do, "Skin Lesion Classification on Imbalanced Data Using Deep Learning with Soft Attention," Sensors, vol. 22, no. 19. MDPI AG, hal. 7530, 2022. doi: 10.3390/s22197530.

[21] S. M. Thwin dan H.-S. Park, "Skin Lesion Classification Using a Deep Ensemble Model," Applied Sciences, vol. 14, no. 13. MDPI AG, hal. 5599, 2024. doi: 10.3390/app14135599.

[22] S. Marison, S. Silvanus, dan R. Rusdiah, "Ai-Based Algorithms for Network Security: Trends, Per-Formance, and Challenges," JURTEKSI (Jurnal Teknol. dan Sist. Informasi), vol. 11, no. 2, hal. 329–336, 2025, doi: 10.33330/jurteksi.v11i2.3699.

[23] S. R. Shegar dan S. S. Patil, "Multi-Class Skin Lesion Classification Using Transfer Learning with EfficientNet-B3 and Convolutional Block Attention Module," Journal of Smart Sensors and Computing, vol. 1, no. 3. GR Scholastic LLP, 2025. doi: 10.64189/ssc.25213.

[24] M. Kahia, B. Bassem, I. Sekkiou, dan F. Kallel, "Modified Multi-Head Attention Transformer (MMHAT) for Skin Image Classification." MDPI AG, 2025. doi: 10.20944/preprints202501.1723.v1.

[25] Y. Zhang, X. Zhang, dan W. Zhu, "ANC: Attention Network for COVID-19 Explainable Diagnosis Based on Convolutional Block Attention Module," Computer Modeling in Engineering & Sciences, vol. 127, no. 3. Tech Science Press, hal. 1037–1058, 2021. doi: 10.32604/cmes.2021.015807.

[26] Y. Fang, H. Huang, W. Yang, X. Xu, W. Jiang, dan X. Lai, "<scp>Nonlocal</scp> convolutional block attention module <scp>VNet</scp> for gliomas automatic segmentation," International Journal of Imaging Systems and Technology, vol. 32, no. 2. Wiley, hal. 528–543, 2021. doi: 10.1002/ima.22639.

[27] [27]M. Shafiq, K. Aggarwal, J. Jayachandran, G. Srinivasan, R. Boddu, dan A. Alemayehu, "RETRACTED: A novel Skin lesion prediction and classification technique: ViT-GradCAM," Skin Research and Technology, vol. 30, no. 9. Wiley, 2024. doi: 10.1111/srt.70040.

[28] S. Deng et al., "A Real-time Lithological Identification Method based on SMOTE-Tomek and ICSA Optimization," Acta Geol. Sin. (English Ed., vol. 98, no. 2, hal. 518–530, 2024, doi: 10.1111/1755-6724.15144.

[29] J. Padhye, V. Firoiu, &D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 199