JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9781
Hal 973-984
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Implementation Of Attention Mechanism And Explainable Ai For Skin Lesion Classification Using CNN

Ilham Nur Fajri, Aditya Dwiputro Wicaksono, Lisda*

Fakultas Informatika, Program Studi Teknik Informatika, Universitas Telkom, Indonesia
Email: [1]ilhamnurfjri@student.telkomuniversity.ac.id, [2]adityaw@telkomuniversity.ac.id, [3*]lisdalis@telkomuniversity.ac.id ,
Email Penulis Korespondensi: adityaw@telkomuniversity.ac.id
Submitted **07-05-2026;** Accepted **09-06-2026;** Published **30-06-2026**

Abstract

Skin lesions are critical dermatological indicators that require early detection to prevent severe outcomes such as melanoma. Traditional Convolutional Neural Network (CNN) architectures employed for categorizing these lesions frequently encounter significant hurdles, notably disproportionate class distributions and a lack of transparency, functioning essentially as opaque "black boxes" during inferential processes. To mitigate these limitations, the current research deploys a ResNet-50 framework augmented by a Convolutional Block Attention Module (CBAM) to refine spatial and channel feature prioritization, alongside the integration of Gradient-Weighted Class Activation Mapping (Grad-CAM) to yield interpretable visualizations. The empirical analysis utilized the HAM10000 repository, incorporating a preprocessing pipeline that encompassed spatial resizing, pixel normalization, and data augmentation, subsequently trained via a bipartite transfer learning methodology. Quantitative metrics reveal that the CBAM-integrated architecture elevates the baseline global accuracy from 82.00% to 86.83%, while simultaneously augmenting the Macro F1-Score from 68.00% to 77.00%. Qualitative evaluation using Grad-CAM shows sharper and more localized heatmaps, indicating that the attention mechanism successfully guides the model to focus on clinically relevant lesion areas. These findings suggest that combining attention mechanisms with explainable AI not only enhances classification performance but also provides visual transparency, supporting clinical interpretation. This approach is expected to improve trust and reliability in automated skin lesion classification systems.

Keywords: Skin Lesion; CNN; Attention Mechanism; Explainable AI; Grad-CAM

1. INTRODUCTION

Skin lesions are significant dermatological conditions that can be indicative of both benign and malignant diseases. Early detection of malignant lesions, particularly melanoma, is critical due to its high metastatic potential and associated mortality rates [1], [2], [3]. Globally, skin diseases are among the top causes of non-fatal disability, affecting millions of people annually and imposing substantial clinical and economic burdens [1]. Conventional diagnosis relies heavily on the expertise of dermatologists, who evaluate lesion morphology, color, texture, and patient history [4]. However, limited availability of specialists, especially in underserved regions, has led to delays and inconsistencies in diagnosis [5], [6], [4].

Recent advances in artificial intelligence, particularly deep learning, have demonstrated potential in automating skin lesion classification. Convolutional Neural Networks (CNNs) have shown superior performance in recognizing complex visual patterns, including those present in dermatoscopic images, and can differentiate between subtle variations among lesion types [3], [7], [8], [4]. Despite these advancements, standard CNN architectures are prone to several challenges, including bias toward majority classes in imbalanced datasets and limited interpretability, often acting as "black boxes" in medical decision-making [7], [9], [10], [11]. This limitation reduces clinicians' trust in automated systems, hindering adoption in real-world clinical settings.

To resolve the inherent opacity of deep learning architectures, methodologies rooted in Explainable Artificial Intelligence (XAI), predominantly Gradient-Weighted Class Activation Mapping (Grad-CAM), have been integrated into analytical pipelines. Grad-CAM operationalizes this transparency by synthesizing topographical activation heatmaps that spatially delineate the topological regions exerting the highest statistical influence on the algorithmic prognostications, thereby furnishing medical practitioners with a decipherable heuristic of the computational reasoning. [12], [13], [14]. Through the visual rendering of the network's spatial attention, this technique facilitates the empirical corroboration that the predictive model is anchoring its classification on salient pathophysiological indicators, effectively circumventing spurious correlations with extraneous background noise, such as exogenous artifacts or adjacent non-lesional tissue.

Concurrently, the integration of attention paradigms has surfaced as a potent strategy for optimizing feature representation within Convolutional Neural Network (CNN) frameworks. Specifically, the Convolutional Block Attention Module (CBAM) functions as a computationally efficient, lightweight architecture that executes a dual-stage refinement process through sequential channel and spatial attention mechanisms. This hierarchical approach facilitates the accentuation of highly discriminative semantic features while systematically attenuating stochastic noise and non-informative spatial domains, thereby enhancing the model's representational fidelity. [15], [16], [17]. Integrating CBAM into CNN architectures has been shown to improve classification performance and feature localization in medical imaging tasks, including dermatology, radiology, and other biomedical applications [15], [16], [18], [19].

Previous research on skin lesion classification has demonstrated high accuracy using CNNs, transfer learning, and data augmentation techniques. For instance, Mawardi et al. [9] implemented a CNN-based mobile application achieving 97.38% accuracy, while Fathurrahman et al. [8] reported 87.14% accuracy using a conventional CNN. However, these studies either focused solely on accuracy without interpretability, or applied attention mechanisms on other medical

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9781
Hal 973-984
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

imaging modalities without integrating XAI [12], [15]. Therefore, there remains a research gap in combining attention-based CNN models with explainable AI for clinically interpretable skin lesion classification, particularly on benchmark datasets like HAM10000 [16].

The HAM10000 dataset, consisting of 10,015 dermatoscopic images across seven lesion categories, provides a robust benchmark for evaluating automated classification models [16]. Challenges inherent in this dataset include severe class imbalance and high intra-class variability, which complicate learning and generalization for conventional CNNs. Integrating CBAM into a ResNet-50 backbone allows the model to adaptively focus on informative features, while Grad-CAM provides visual explanations for model decisions, ensuring clinical relevance and interpretability.

This study aims to implement and evaluate a CNN model combining ResNet-50, CBAM, and Grad-CAM for multi-class skin lesion classification. The objectives are to: (1) enhance classification accuracy through attention-guided feature extraction, (2) provide interpretable visual explanations aligned with clinical evaluation, and (3) assess model performance quantitatively and qualitatively on the HAM10000 dataset. By addressing both performance and interpretability, this approach seeks to bridge the gap between automated diagnostic systems and clinical applicability in dermatology.

In summary, the integration of attention mechanisms and XAI in CNNs provides a dual benefit: improving predictive performance and offering transparent insights into model decision-making. This research contributes to the development of robust, interpretable, and clinically meaningful tools for automated skin lesion classification, addressing limitations of prior studies while leveraging state-of-the-art deep learning techniques [3]– [16].

2. RESEARCH METHODOLOGY

2.1 Research Stages

The overall research workflow for this study was designed to ensure a systematic and reproducible approach for multi-class skin lesion classification. The stages begin with literature review to understand the state-of-the-art in CNN-based skin lesion classification, attention mechanisms, and explainable AI methods . This is followed by data acquisition, preprocessing, dataset partitioning, model implementation and training, and finally evaluation and visualization. Each stage builds on the previous one to ensure consistency and reliability of the research outcomes. [16], [4], [14].

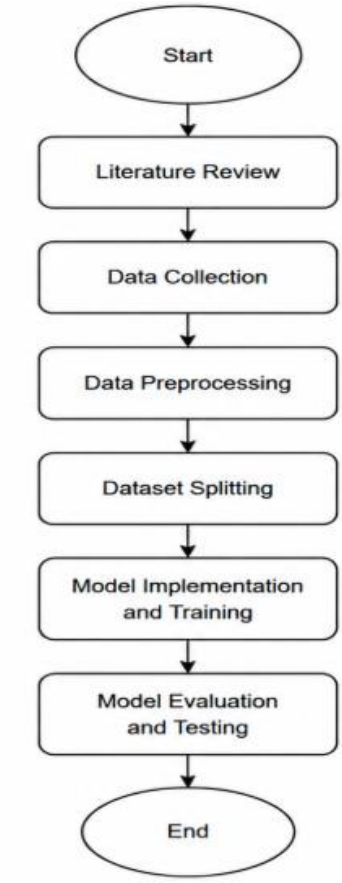

Figure 1. Research Stages

Figure 1 shows the overall research workflow, starting from literature review and data acquisition, continuing with preprocessing and dataset splitting, then moving to CNN model implementation and training, and concluding with

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9781
Hal 973-984
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

quantitative and qualitative evaluation. This flow ensures clarity and reproducibility, allowing readers to understand the stepwise approach applied in this study.

2.2 Data Preprocessing

Subsequent to the acquisition of the HAM10000 repository, a rigorous preprocessing pipeline was orchestrated to homogenize and optimize the dermoscopic imagery for subsequent neural network ingestion. This procedural stage commenced with the systematic extraction of metadata and raw image files, followed by a spatial re-dimensioning to a 224×224 pixel resolution to ensure architectural compatibility with the ResNet-50 backbone. To facilitate computational convergence, pixel intensities were subjected to min-max normalization, rescaling values to a unit interval [0, 1], while categorical target variables were transformed via one-hot encoding. Furthermore, to counteract the risk of algorithmic overfitting and bolster the model's inductive bias, a comprehensive data augmentation suite was deployed. This involved stochastic geometric and photometric transformationsnspecifically encompassing random rotations, multi-axial flipping, scaling, and luminance adjustments thereby augmenting the diversity of the training distribution [13], [16], [4].

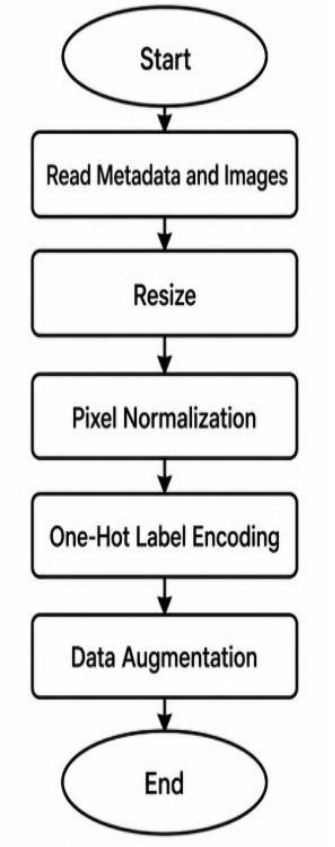

Figure 2. Preprocessing Pipeline

Figure 2 illustrates how each image from the dataset undergoes metadata reading, resizing, normalization, one-hot encoding, and augmentation. This ensures that the model receives consistent input data while enhancing its ability to generalize to unseen images. By explicitly detailing these steps, the preprocessing process becomes reproducible and transparent, forming a critical foundation for subsequent model training.
Following preprocessing, the dataset was split into training (70%), validation (15%), and testing (15%) subsets using stratified sampling. This preserved the class distribution across all subsets, ensuring fair evaluation and preventing data leakage [10].

2.3 Model Architecture and Implementation

.The computational framework was engineered utilizing the ResNet-50 architecture as the foundational backbone, augmented by the integration of the Convolutional Block Attention Module (CBAM) to refine feature extraction efficacy. By operationalizing a sequential dual-attention mechanism comprising both channel and spatial dimensions the CBAM facilitates the selective intensification of salient pathological topographies while concurrently attenuating non-informative background noise [14], [15], [20]. To optimize categorical stratification, a bespoke classification head was synthesized, incorporating a hybrid dual-pooling strategy (Global Average and Max Pooling) to preserve diverse feature representations. This was further fortified with batch normalization layers, ReLU activation functions, a dropout rate of 0.5, and L2 regularization, collectively engineered to stabilize the stochastic gradient descent process and mitigate the risk of model over-specialization.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9781
Hal 973-984
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

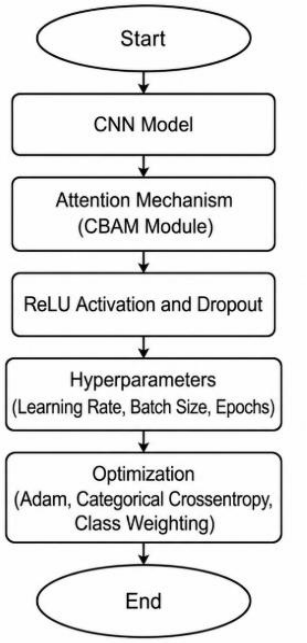

Figure 3. Model Training Flow

Figure 3 demonstrates the training flow of the model. Images pass through the ResNet-50 backbone, followed by the CBAM attention mechanism. Feature representations are processed through ReLU activation and dropout layers, followed by hyperparameter adjustment and optimization using Adam and categorical crossentropy loss. Class weighting is applied to address the class imbalance in HAM10000. This architecture allows the network to focus on clinically relevant lesion areas while maintaining robust learning across all classes.

Two CNN models were designed for comparison: the baseline ResNet-50 and ResNet-50 integrated with CBAM [8], [9], [20]. The CBAM module applies channel and spatial attention to emphasize relevant lesion features while suppressing irrelevant background information. A custom classification head includes dual-pooling, batch normalization, ReLU activation, dropout (0.5), and L2 regularization (1e-4) [15], [4].

To clarify the structural differences between these two architectures, Table 3 summarizes the comparison:

Table 1. Comparison of Baseline ResNet-50 and ResNet50 + CBAM Architecture

Component	Baseline ResNet-50	ResNet-50 + CBAM
Feature Extractor (Backbone)	ResNet-50 (ImageNet)	ResNet-50 (ImageNet)
Attention Mechanism	None	CBAM (Channel + Spatial)
Pooling Strategy	Global Average Pooling	Dual-Pooling (Avg + Max)
Activation Function	ReLU	ReLU
Normalization	BatchNorm	BatchNorm
Classification Layer	Dense 512 neurons	Dense 512 neurons
Regularization	Dropout 0.5	Dropout 0.5 + L2 1e-4
Output Layer	7 neurons (Softmax)	7 neurons (Softmax)
Optimizer	Adam	Adam
Loss Function	Categorical Crossentropy	Categorical Crossentropy
Class Weighting	Applied	Applied
Learning Rate Phase 1	1e-3	1e-3
Learning Rate Phase 2	5e-6	5e-6
Batch Size	32	32
Epoch Phase 1	20	20
Epoch Phase 2	50	50

Table 1 illustrates the design modifications in the CBAM-enhanced model. The attention mechanism, dual-pooling strategy, and additional regularization are introduced to guide the network toward clinically relevant lesion features, improve discrimination of minority classes, and prevent overfitting. Positioning this table in the methodology section clarifies the architectural framework prior to reporting training outcomes.

2.4 Model Evaluation and Deployment

To ensure a robust and holistic assessment of the model's classification efficacy, a bipartite evaluation strategy comprising quantitative and qualitative methodologies was operationalized [13], [16], [14]. The quantitative framework centered on the derivation of performance metrics from the confusion matrix, specifically leveraging True Positives (TP), True

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9781
Hal 973-984
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Negatives (TN), False Positives (FP), and False Negatives (FN). These fundamental constituents were utilized to compute Accuracy, Precision, Recall, and the F1-Score, providing a multi-dimensional perspective on the network's diagnostic reliability [13], [4]. The formal mathematical expressions for these metrics are articulated as follows: :

a. Accuracy - measures the proportion of correctly predicted samples over total samples:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

b. Precission - measures the proportion of correctly predicted positive samples over all predicted positive samples:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

c. Recall (Sensitivity) - measures the proportion of correctly predicted positive samples over all actual positive samples:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

d. F1-Score - harmonic mean of Precision and Recall, balancing both metrics:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

These metrics were computed for each lesion class to evaluate model performance on minority and majority classes, ensuring that the attention mechanism (CBAM) effectively mitigates class imbalance [10], [12], [25].

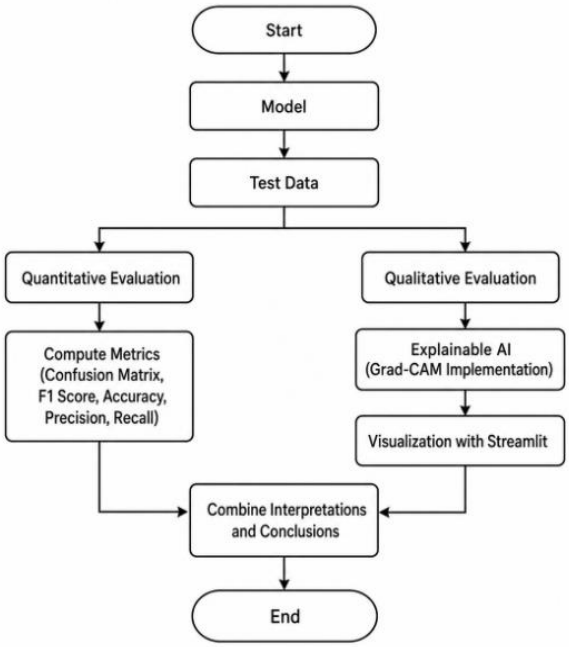

Figure 4. Evaluation Pipeline

Figure 4 illustrates the evaluation pipeline integrated with Streamlit for interactive visualization. Test images are processed through the trained model, quantitative metrics are computed, and Grad-CAM heatmaps highlight regions influencing model decisions. The Streamlit interface enables users to select test samples, view original images, predicted class outputs, and heatmaps in real time. This combination ensures transparency and interpretability while facilitating clinical validation [13], [14].

3. RESULT AND DISCUSSION

3.1 Dataset Preprocessing and Distribution

The empirical foundation of this research utilized the HAM10000 repository, a diverse collection encompassing 10,015 dermoscopic captures stratified across seven distinct pathological taxonomies: Melanoma, Melanocytic Nevi, Basal Cell Carcinoma, Actinic Keratoses, Benign Keratosis-like Lesions, Dermatofibroma, and Vascular Lesions. Exploratory data analysis identified a pronounced class asymmetry, wherein Melanocytic Nevi constituted the predominant statistical majority, while Dermatofibroma and Vascular Lesions were identified as the minority subsets. This distributional skew presents a substantial computational impediment for orthodox CNN architectures, as it frequently induces a predictive

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9781
Hal 973-984
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

bias toward majority classes, subsequently compromising the model's sensitivity and generalizability toward underrepresented categories.

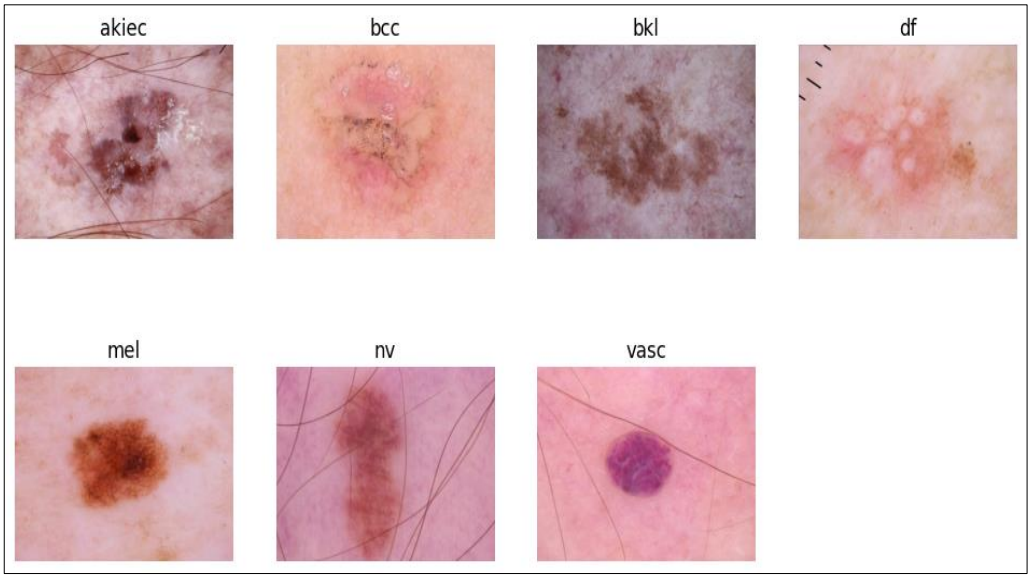

Figure 5. Representative Image From Each Lesion Class

Figure 5 shows representative images from each lesion class, highlighting the variations in color, shape, and texture inherent in the dataset. To ensure model robustness, preprocessing was applied to standardize the input. Images were resized to 224×224 pixels, normalized to the 0–1 range, and one-hot encoded. Data augmentation including random rotation, flipping, zooming, shifting, and brightness adjustment was applied to increase visual variability and reduce overfitting.

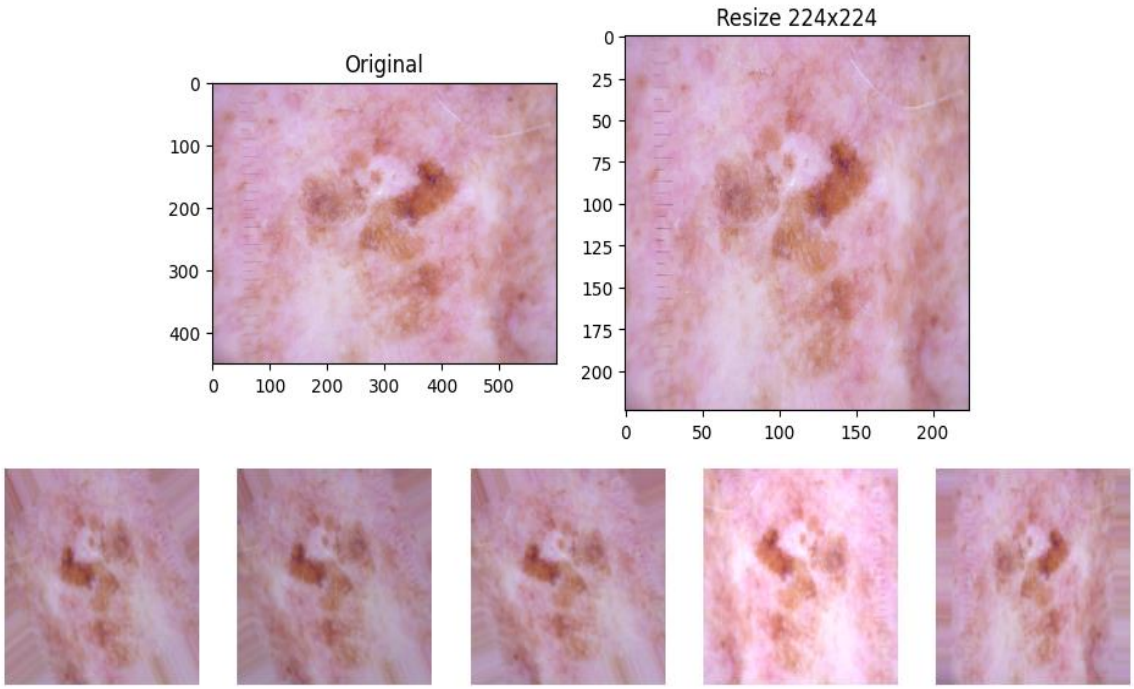

Figure 6. Preprocessing Aplied

Figure 6 illustrates the preprocessing pipeline applied to each image in the dataset. The dataset was then partitioned into training (70%), validation (15%), and testing (15%) subsets using stratified sampling to maintain class distributions across all subsets, preventing data leakage and ensuring reliable evaluation.

3.2 Model Training Result

The training process of both models was conducted in two phases: Phase 1 (Transfer Learning) with frozen early layers to retain general features, and Phase 2 (Fine-Tuning) with unfrozen deeper layers for adaptation to lesion-specific patterns. Training progress was monitored using accuracy and loss curves on both training and validation datasets.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9781
Hal 973-984
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

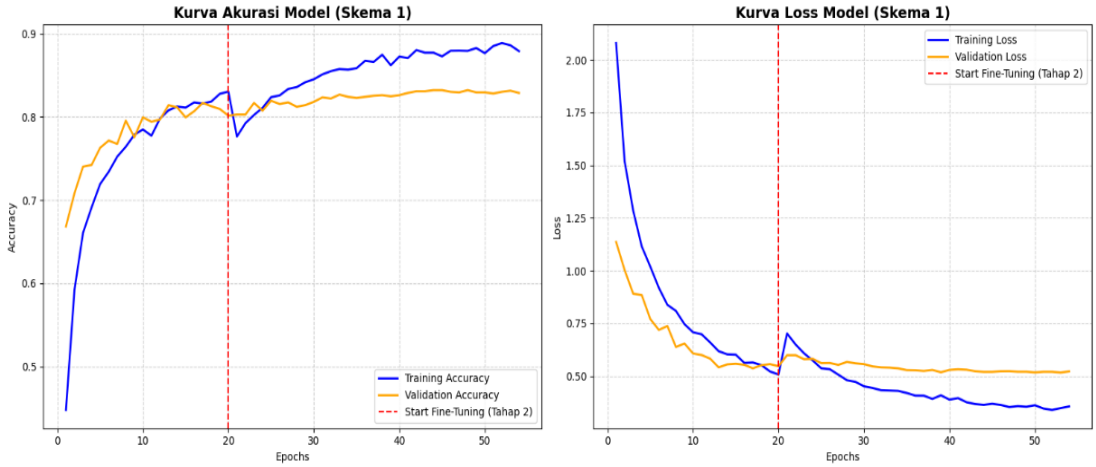

Figure 7. Training and Validation Accuracy and Loss – Baseline ResNet-50

During Phase 1, the training accuracy rapidly increased from approximately 45% to around 80%, while training and validation loss decreased sharply. This indicates that the frozen backbone effectively provides general visual features, enabling the classification head to adapt quickly to the HAM10000 dataset.

In Phase 2, slight fluctuations in validation accuracy were observed immediately after unfreezing deeper layers, reflecting fine-tuning adjustments. By the end of training, training accuracy reached approximately 88%, with validation accuracy stabilizing around 83%. The gap between training and validation metrics indicates minor overfitting, highlighting the need for attention mechanisms and enhanced feature selection in the next model.

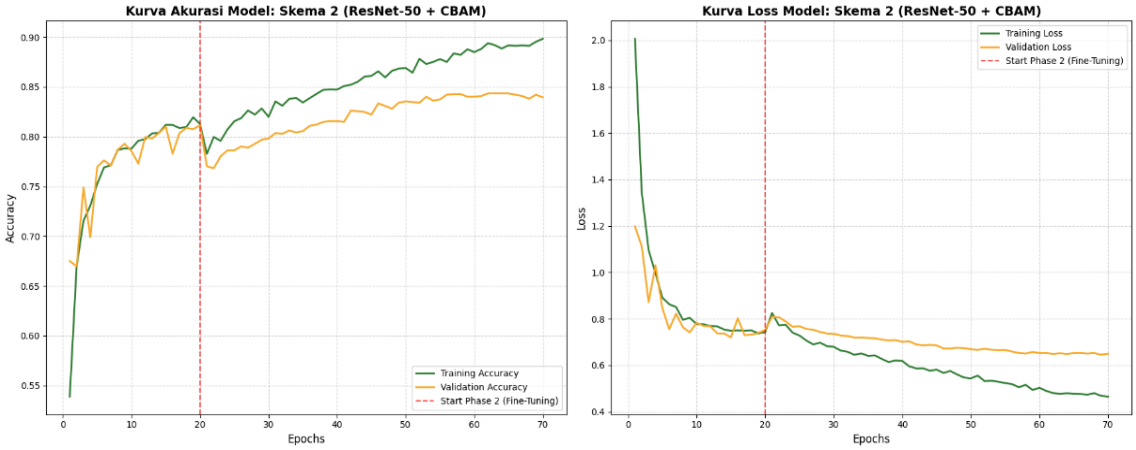

Figure 8. Training and Validation Accuracy and Loss – ResNet-50 + CBAM

The CBAM-enhanced model demonstrated a faster and more consistent convergence, with training accuracy approaching 90% and validation accuracy reaching 84–87%. The Dual-Pooling mechanism, attention layers, and regularization contribute to a more stable learning process, allowing the model to capture discriminative lesion features while reducing overfitting. Notably, the validation accuracy curve shows steady improvement during fine-tuning, suggesting enhanced generalization and better handling of minority classes.

3.3 Quantitative Evaluation

To ascertain the diagnostic efficacy of the proposed frameworks, a rigorous quantitative appraisal was performed utilizing the hold-out testing partition of the HAM10000 dataset. The evaluation encompassed a multifaceted analysis across the seven pathological classifications, leveraging Accuracy, Precision, Recall, and the F1-Score as primary performance indicators. To mitigate the statistical distortion caused by the inherent class asymmetry, these metrics were synthesized using both macro and weighted averaging techniques [16]. Furthermore, confusion matrices were meticulously constructed to provide a granular visualization of inter-class misclassification dynamics, thereby facilitating a deeper forensic identification of the specific lesion morphologies that presented the greatest computational challenge to the model's discriminative capabilities.

Table 2. Classification Report – Baseline ResNet-50

Class	Precision	Recall	F1-Score	Support
Actinic Keratoses	0.76	0.39	0.51	49
Basal Cell Carcinoma	0.66	0.58	0.62	77

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9781
Hal 973-984
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Benign Keratosis	0.58	0.73	0.65	165
Dermatofibroma	0.53	0.59	0.56	17
Melanocytic Nevi	0.93	0.91	0.92	1006
Melanoma	0.59	0.66	0.62	167
Vascular Lesions	0.90	0.82	0.86	22
Macro Avg	0.71	0.67	0.68	1503
Weighted Avg	0.83	0.82	0.82	1503

Analysis of the baseline ResNet-50 model indicates robust performance on the dominant class, Melanocytic Nevi, which achieved an F1-Score of 0.92. In contrast, minority classes and visually subtle lesions, such as Actinic Keratoses and Dermatofibroma, exhibited markedly lower recall values (0.39 and 0.59, respectively), reflecting misclassification tendencies attributable to class imbalance. The macro-average F1-Score of 0.68 underscores a pronounced bias toward majority classes, highlighting the necessity for mechanisms that prioritize feature-specific attention.

Figure 9. Confusion Matrix – Baseline ResNet-50

Inspection of the confusion matrix reveals frequent misclassification among minority and visually similar lesion classes. For instance, several Melanoma instances were erroneously predicted as Melanocytic Nevi or Benign Keratosis, demonstrating the baseline model's limitations in differentiating subtle pathological features in the absence of explicit attention guidance.

After evaluating the baseline ResNet-50 model, further analysis was conducted on the proposed ResNet-50 model integrated with the Convolutional Block Attention Module (CBAM). This evaluation aimed to determine whether the attention mechanism could improve the model's ability to recognize discriminative lesion features, particularly in minority and visually similar classes. The detailed classification performance of the ResNet-50 + CBAM model is presented in Table 3.

Table 3. Classification Report – ResNet-50 + CBAM

Class	Precision	Recall	F1-Score	Support
Actinic Keratoses	0.78	0.65	0.71	49
Basal Cell Carcinoma	0.79	0.71	0.75	77
Benign Keratosis	0.78	0.75	0.77	165
Dermatofibroma	0.83	0.59	0.69	17
Melanocytic Nevi	0.91	0.96	0.94	1006
Melanoma	0.73	0.60	0.66	167
Vascular Lesions	0.91	0.91	0.91	22
Macro Avg	0.82	0.74	0.77	1503
Weighted Avg	0.87	0.87	0.87	1503

As shown in Table 3, the ResNet-50 + CBAM model achieved improved classification performance across most lesion categories. The macro-average F1-score increased to 0.77, while the weighted-average F1-score reached 0.87, indicating that the proposed model provided more balanced performance across both majority and minority classes. Notable improvements were observed in Actinic Keratoses, Basal Cell Carcinoma, Benign Keratosis, and Vascular

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9781
Hal 973-984
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Lesions, suggesting that CBAM contributed to better feature refinement and class discrimination. Although Melanoma still showed classification challenges due to its visual similarity with Melanocytic Nevi, the overall results demonstrate that the integration of attention mechanisms strengthens the model's diagnostic robustness.

To further examine the prediction behavior of the ResNet-50 + CBAM model, a confusion matrix was generated to visualize the distribution of correct and incorrect classifications across the seven lesion classes. This analysis provides a more detailed understanding of the model's classification tendencies, including the extent to which misclassification occurs between visually similar lesion categories. The confusion matrix of the proposed model is shown in Figure 10.

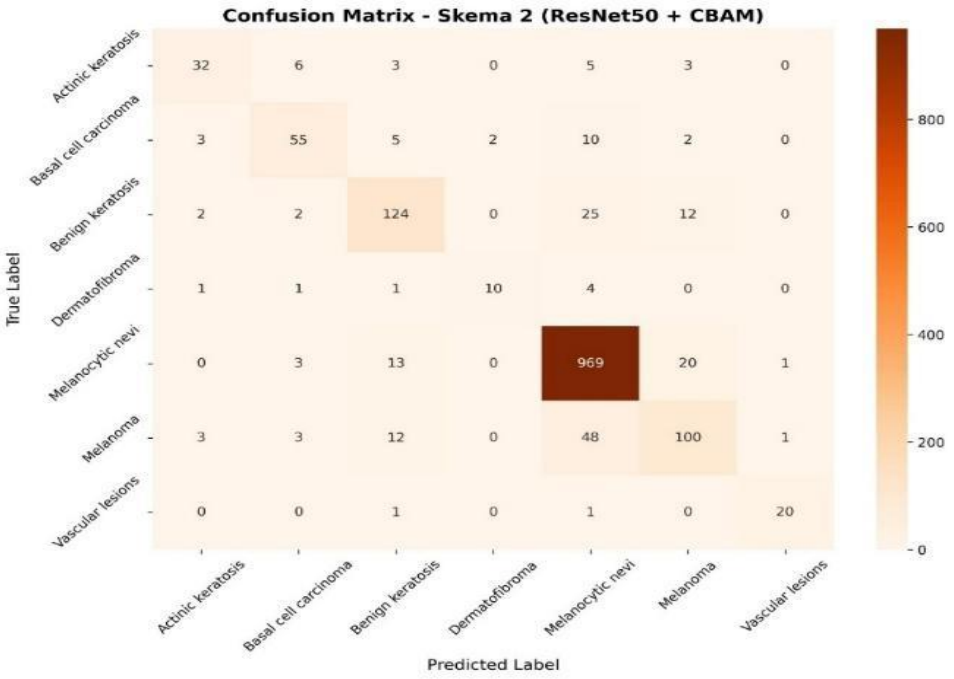

Figure 10. Confusion Matrix – ResNet-50 + CBAM

The CBAM-enhanced model demonstrates higher recall and F1-Score across minority classes, particularly Actinic Keratoses and Basal Cell Carcinoma. Macro F1-Score increased from 0.68 to 0.77, indicating that the attention mechanism mitigates class imbalance by guiding the model to focus on discriminative lesion features. Misclassifications of minority lesions decreased, while overall accuracy improved from 82% (baseline) to 86.83%. Some ambiguity remains for visually similar lesions, such as Melanoma vs Nevi, reflecting intrinsic challenges of dermatoscopic classification.

In conclusion, the quantitative evaluation substantiates that the integration of CBAM attention markedly enhances model sensitivity and robustness, particularly for underrepresented and visually complex lesion classes. Confusion matrices and classification reports collectively confirm that attention-guided feature selection improves multi-class classification performance while maintaining high predictive reliability for majority classes. These quantitative results provide a solid foundation for subsequent qualitative evaluation using Grad-CAM to further analyze model interpretability and spatial feature localization [15], [4], [19].

3.4 Qualitative Ecvaluation (Grad-CAM)

To augment the quantitative performance metrics, a qualitative forensic appraisal was executed through the implementation of Gradient-Weighted Class Activation Mapping (Grad-CAM). Grad-CAM facilitates interpretability by synthesizing topographic activation heatmaps that spatially isolate the specific input domains exerting the most profound influence on the model's categorical assertions. This visualization serves as a decipherable heuristic of the network's spatial attention, allowing for an empirical validation of the architectural reasoning. Such qualitative scrutiny was pivotal in substantiating the efficacy of the Convolutional Block Attention Module (CBAM) in steering the network's computational focus toward diagnostically salient lesion morphologies, effectively distinguishing them from non-informative background features [12], [13].

Figure 11 presents the Grad-CAM visualization for representative Melanoma samples, comparing the Baseline ResNet-50 and the ResNet-50 + CBAM model.

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 3, Juni 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v13i3.9781
Hal 973-984
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

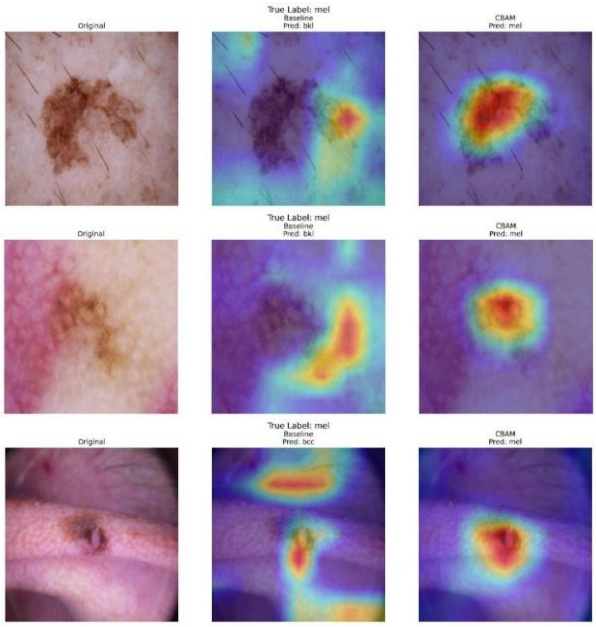

Figure 11. Grad-CAM Heatmap – Melanoma (Baseline vs CBAM)

In the baseline model, attention is often diffusely distributed across the image, with partial focus on surrounding healthy skin or imaging artifacts. In contrast, the CBAM-enhanced model exhibits a markedly concentrated activation over the lesion itself, highlighting salient pathological features such as irregular borders, heterogeneous pigmentation, and asymmetric morphology. This observation indicates that the CBAM module enhances spatial feature localization and reduces the network's misattention to irrelevant regions.

Similarly, Figure 12 presents Grad-CAM visualizations for Actinic Keratoses (AKIEC), which demonstrate analogous improvements in attention localization.

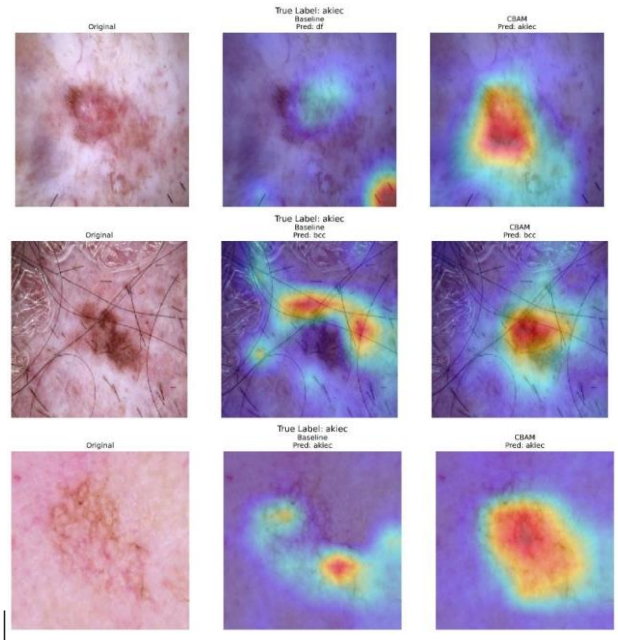

Figure 12. Grad-CAM Heatmap – Actinic Keratoses (Baseline vs CBAM)

In the baseline ResNet-50, activation is widely dispersed across both lesion and non-lesion areas. By contrast, the CBAM-enhanced model exhibits sharply focused attention centered on the lesion, enabling finer discrimination of subtle visual differences. This precise localization is particularly critical for early-stage, pre-cancerous lesion detection. The qualitative improvements observed in Grad-CAM heatmaps corroborate the quantitative performance gains reported in Section 3.3, indicating that attention-guided feature extraction substantially improves the model's capability to capture clinically relevant visual patterns.

The Grad-CAM analysis further confirms that the CBAM module effectively directs the network's computational focus toward clinically significant regions, enhancing interpretability of model predictions. For minority lesion classes, such as Actinic Keratoses, which are relatively small and visually subtle, the CBAM-enhanced model allocates attention

proportionally to feature relevance. Likewise, for visually similar categories, such as Melanoma versus Melanocytic Nevi, the integration of CBAM produces more distinct and precise heatmaps, thereby reducing misclassification risk.

Integration of Grad-CAM visualizations into the Streamlit interface enables real-time interactive exploration of model predictions. Clinicians and researchers can dynamically examine both the predicted class and the corresponding heatmap for each test image, providing immediate insight into whether the model attends to medically relevant features. This interactive evaluation mechanism is essential for establishing clinical trust in automated diagnostic systems, bridging the gap between high quantitative accuracy and practical usability in real-world healthcare applications [13].

In summary, the qualitative evaluation demonstrates that attention mechanisms, when combined with Grad-CAM, not only enhance the model's focus on lesion-specific regions but also provide interpretable visual explanations. These results complement the quantitative findings and substantiate that CBAM integration improves both predictive performance and transparency, particularly for minority or visually complex lesion classes, thereby supporting reliable clinical deployment of automated skin lesion classification systems.

4. CONCLUSION

This study successfully implemented a ResNet-50-based CNN model integrated with the Convolutional Block Attention Module (CBAM) and Grad-CAM for multi-class skin lesion classification using the HAM10000 dataset. The results demonstrate that the addition of CBAM improves the model's ability to extract more discriminative visual features by emphasizing clinically relevant lesion regions and reducing attention to non-informative background areas. Quantitatively, the proposed ResNet-50 + CBAM model achieved an accuracy of 86.83%, outperforming the baseline ResNet-50 model, which obtained 82.00% accuracy. The macro-average F1-score also increased from 68.00% to 77.00%, indicating better classification balance, particularly for minority and visually subtle lesion classes. Qualitative evaluation using Grad-CAM further confirms that the attention-enhanced model produces more focused and interpretable heatmaps, especially on lesion areas associated with irregular borders, pigmentation variation, and asymmetric patterns. These findings indicate that integrating attention mechanisms with explainable AI can improve both predictive performance and interpretability in automated dermatological image analysis. Nevertheless, several challenges remain, particularly in distinguishing visually similar lesion categories such as Melanoma and Melanocytic Nevi, as well as improving sensitivity for rare classes. Future studies may incorporate advanced imbalance-handling strategies, larger clinical datasets, and additional explainability methods to strengthen model robustness and support practical clinical validation.

REFERENCES

[1] World Health Organization, " Skin diseases as a global public health priority," 2025.
[2] World Health Organization, "Skin cancer, International Agency for Research on Cancer," 2022.
[3] American Cancer Society, "Skin cancer," 2022.
[4] I. Fathurrahman et al, "Pengembangan model CNN untuk klasifikasi penyakit kulit berbasis citra digital," vol. 8, no. 1, pp. 45–50, 2025.," Jurnal Infotek, vol. 8, no. 1, pp. 45-50, 2025.
[5] M. A. Richard et al, "Prevalence of most common skin diseases in Europe: A population-based study," Journal of the European Academy of Dermatology and Venereology, vol. 36, p. 1088–1096, 2022.
[6] K. P. Venkatesh et al, "Deep learning models across the range of skin disease," npj Digital Medicine, vol. 7, p. 32 , 2024.
[7] H. K. Jeong et al, "Deep learning in dermatology: A systematic review of current approaches, outcomes, and limitations," Journal of Investigative Dermatology Innovations, vol. 3, no. 1, pp. 100-150, 2023.
[8] E. S. Nugroho et al, "Boosting the performance of pretrained CNN architecture on dermoscopic pigmented skin lesion classification," Skin Research and Technology, vol. 29, no. 6, 2023.
[9] D. P. Mawardi et al, "Deteksi awal klasifikasi jenis penyakit kanker kulit dengan algoritma CNN berbasis mobile apps," Jurnal ATASI, vol. 6, no. 2, pp. 55-60, 2023.
[10] R. K. Duanti et al, "Klasifikasi kanker kulit jinak dan ganas menggunakan metode Xception berbasis Raspberry Pi," Jurnal-PTIIK Universitas Brawijaya, vol. 14, no. 2, pp. 55-63, 2025.
[11] G. M. A. Sihotang and J. Supardi, "Pengembangan model CNN ResNet-18 untuk klasifikasi kondisi gigi berbasis citra RGB sebagai solusi diagnostik digital," Jurnal Pendidikan dan Teknologi Indonesia, vol. 4, no. 2, pp. 747-758, 2024.
[12] M. Shafiq et al , "A novel skin lesion prediction and classification technique: ViT-GradCAM," Expert Systems, 2024.
[13] S. N. A. Kadir, E. S. Nugroho, and S. R. Dewi, "Fine-tuning of explainable CNNs for skin lesion classification based on dermatologists' feedback towards increasing trust," 2023.
[14] D. A. Agustina, "Klasifikasi citra jenis kulit wajah dengan algoritma CNN ResNet-50," Jurnal Ilmiah Sistem Informasi, vol. 8, no. 1, pp. 10-20, 2024.
[15] S. A. Hakim et al, "Klasifikasi Citra Generasi Artificial Intelligence menggunakan metode fine tuning pada residual network," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 11, no. 3, pp. 655-666, 2024.
[16] 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, 2018.
[17] Islam, W, Jones, M, Faiz, R dan Sadeghipour, N., Q, "Improving performance of breast lesion classification using a ResNet50 model optimized with a novel attention mechanism.," Tomography, vol. 8, no. 5, p. 2411–2425, 2022.
[18] Z. D. E. Putra and D. W. Utomo, "Penerapan deep learning dengan mekanisme attention untuk meningkatkan performa segmentasi liver dan tumor pada citra CT menggunakan ResUnet,"," Jurnal Nasional Teknologi dan Sistem Informasi, vol. 10, no. 3, pp. 231-239, 2024.

[19] Kılıç, Ş, "Deep feature engineering for accurate sperm morphology classification using CBAM-enhanced ResNet5," PLOS ONE, vol. 20, no. 9, 2025.
[20] V.-T. Nguyen, V.-T. Pham, and T.-T. Tran, "AC-MAMBASEG: An adaptive convolution and Mamba-based architecture for enhanced skin lesion segmentation," 2024.