PeleSENet: A Cutting-Edge Hybrid CNN for Enhanced Skin Cancer Detection with XAI and Adaptive Learning

Kian Lok Chin
Faculty Of Information Technology
Monash University Malaysia
Selangor, Malaysia
kchi0086@student.monash.edu

***Abstract*—Skin cancer is one of the most common and deadly forms of cancer worldwide. Early detection plays an important role in improving patient outcomes, but traditional diagnostic methods often require expert intervention, which can be time-consuming and subjective. In this paper, we present PeleSENet, a novel lightweight hybrid Convolutional Neural Network (CNN) model that combines the strengths of SeResNet and PeleeNet for skin cancer classification. The model provides real-time feedback to users with high-confidence results and incorporates Explainable AI (XAI) to ensure transparency in predictions. In cases of lower-confidence predictions, the image is sent for review by expert dermatologists to corporate ethical AI standards. Furthermore, our system implements adaptive retraining, in which the model is continuously improved based on new patient data, with automatic updates if performance improvements are observed. Experimental results demonstrate the effectiveness of PeleSENet in skin cancer detection, which offers a robust, scalable, and lightweight solution for clinical and patient-facing applications.**

***Index Terms*—Skin Cancer Detection, Convolutional Neural Networks (CNNs), SeResNet, PeleeNet, Explainable AI, Adaptive Learning, Lightweight CNN**

I. Introduction

Skin cancer is a global health concern, with increasing incidence rates and the need for accurate early detection. Traditional diagnostic methods, such as biopsies and expert visual inspection are often slow and require significant expertise [1]. Recent advancements in deep learning, particularly in the field of Convolutional Neural Networks (CNNs), have shown great promise for automating and enhancing the detection process [2]. However, existing models tend to be either computationally expensive or difficult to interpret. For instance, high parameters models like AlexNet [3], DenseNet [4], EfficientNet [5] are frequently used for skin cancer detection which may take up more storage and more time to process [6]. In this work, we propose PeleSENet, a novel lightweight hybrid CNN that combines the strengths of SEResNet [7] and PeleeNet [8] to create a robust framework for skin cancer detection. Our system includes Explainable AI (XAI) for model transparency and adaptive retraining to ensure continuous improvement over time. This paper presents a detailed description of the model architecture, the system's working principles, and its performance on standard skin cancer datasets.

II. Related Work

A. Explainable AI (XAI) in Medical Imaging

Explainable artificial intelligence (XAI) has gained significant sttention in medical imaging to enhance the transparency of deep learning models, particularly in high-stakes domains like cancer diagnosis. In skin cancer classification, where convolutional neural networks (CNNs) achieve dermatologist-level accuracy, the "black-box" nature of these models remains a critical barrier to clinical trust and adoption [9]. XAI methods like GRAD-CAM (Gradient-weighted Class Activation Mapping) have emerged as imperative tools to visualize regions of interest (ROIs) driving model predictions. For instance, GRAD-CAM [10] has been widely applied to dermoscopic and clinical skin lesion datasets (e.g., ISIC archive) [11], enabling clinicians to validate whether models focus on clinically relevant features such as pigment networks or irregular borders [9] [12]. This aligns with findings in broader medical imaging, where GRAD-CAM localizes critical regions in mammography for breast cancer detection [13] and highlights tumor boundaries in histopathology slides [14].

Beyond GRAD-CAM, LIME (Local Interpretable Model-agnostic Explanations) [15] has been employed to perturb input images and assess model sensitivity, albeit with mixed results. For example, studies on chest X-rays demonstrated LIME's utility in identifying spurious correlations that might compromise diagnostic reliability [13]. However, its computational inefficiency and fragmented visualizations limit its adoption in real-time clinical workflows compared to gradient-based methods like GRAD-CAM. Other techniques such as Layer-wise Relevance Propagation (LRP) [16] have also been explored, offering pixel-level relevance scores for dermatopathological images, however their complexity often requires additional validation by clinicians. Despite progress, systematic evaluations of XAI in clinical settings remain sparse. A systematic review of 37 studies on AI for skin cancer revealed that only three rigorously evaluated human-AI interaction, highlighting a gap in understanding how XAI

affects diagnostic confidence or accuracy [9]. Furthermore, while GRAD-CAM excels in localizing ROIs, it may overlook global context, as mentioned in histopathology analysis where multi-scale features are critical [12]. In a nutshell, XAI methods like GRAD-CAM are indispensable for bridging the trust gap in AI-driven dermatology. However, their clinical utility depends on rigorous validation and seamless integration into diagnostic workflows, making sure explanations align with medical expertise and reduce diagnostic uncertainty.

B. Adaptive Learning and Continuous Improvement in Medical AI

The application of adaptive learning in medical AI has gained significant attention across diverse clinical domains, offering insights relevant to skin cancer classification systems. Sparrow et al. [17] provide ethical frameworks for adaptive machine learning systems (MAMLS) in medicine, emphasizing their potential to evolve through continuous learning from clinical data streams. While their work primarily discusses regulatory implications rather than technical implementations, it establishes the foundational rationale for developing dynamic systems that improve through real-world deployment which is a principle central to our user-data-driven approach. In radiology and pathology applications, Singh et al. [18] demonstrate how continual learning enables models to sequentially assimilate new medical tasks while preserving prior knowledge. Their class-incremental framework achieved stable performance across multi-institutional datasets spanning X-rays, ultrasound, and histopathology images. This aligns with our objective to enable skin cancer classifiers to adapt to evolving user demographics and imaging device variations without catastrophic forgetting. The authors' rehearsal learning strategy using compressed feature buffers (REMIND) suggests methodologies that could be adapted for dermatoscopic data retention in resource-constrained environments. Yi et al. [19] further advance this through vision-language models (VLMs) employing continual learning for polyp detection and diabetic foot ulcer analysis. Their work demonstrates 15.7% performance improvement through rehearsal buffers when sequentially adapting to new medical domains, validated across six heterogeneous datasets. This technical approach informs our planned architecture where user-specific data will be selectively retained to personalize classification boundaries while maintaining baseline diagnostic accuracy. The authors' finding that rehearsal buffers restore 89% of lost generalization capability during sequential learning particularly supports our strategy of maintaining balanced skin lesion representations. Joshi [20] establishes theoretical parallels through adaptive educational systems, highlighting mechanisms for dynamic content personalization via learner modeling. Although focused on pedagogy, their feedback loop architecture where user interactions continuously refine predictive models directly mirrors our proposed system's operation. While none specifically address dermatology, these works collectively validate adaptive learning's potential for medical image analysis. Our innovation lies in applying these principles to skin cancer classification through CNN architectures optimized for user-specific data adaption, addressing unique challenges in dermatoscopic feature drift and personalized diagnostic thresholds.

C. Utilization of the ISIC Dataset in Skin Cancer Detection Models

Recent research explores various approaches to skin cancer detection using deep learning. For instance, Huang et al. [4] employed YOLOv5 with hyperspectral imaging, achieving 79.20% accuracy on a three-class skin cancer dataset. Teodoro et al. [21] combined a region of interest-based attention mechanism with a generative adversarial network, reporting 97.90% accuracy. Transfer learning strategies have also been explored, as demonstrated by Dandu et al. [22], who attained 90.96% accuracy using the ISIC 2018 dataset. Furthermore, Balaha and Hassan [23] integrated a sparrow search algorithm with deep transfer learning across multiple datasets, achieving high accuracies ranging from 85.87% to 98.83% depending on the dataset. Mukadam and Patil [24] used CNN combined with super resolution generative adversarial network for skin cancer detection and obtained an accuracy of 98.77%. Javid et al. [25] employed deep ensemble learning techniques, attaining 93.00% accuracy on the ISIC dataset for benign-malignant classification. Moreover, Kumar et al. [26] used deep neural networking framework to detect skin cancer and reported an accuracy of 93.26%. He et al. [27] employed CNN to predict skin cancer disease and reported an average accuracy of 76.80%. Apart from that, Aishwarya et al. [28] used the YOLO deep neural network and attained an average accuracy rate of 98.06%. Last but not least, In addition, Cassidy et al. [29] conducted a comprehensive investigation into various approaches employing CNN models. Their study meticulously established baseline performance metrics across multiple subsets of the ISIC datasets, offering critical insights and a robust framework that underpin the primary focus of this paper. These studies highlight the ongoing efforts to improve the accuracy and efficiency of automated skin cancer detection systems through diverse deep learning methodologies.

III. The Proposed CNN - PeleSENet

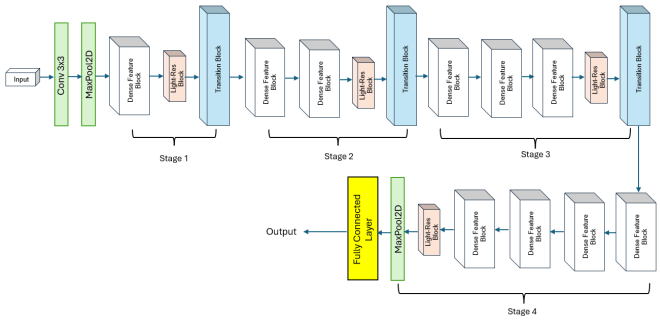

Fig. 1. PeleSENet Architecture

PeleSENet is a novel lightweight convolutional neural network designed specifically for skin cancer image detection.

With fewer than 3.5 million parameters, it is purposefully engineered for environments with limited computational resources, such as mobile devices or point-of-care systems. PeleSENet integrates the best aspects of PeleeNet [8] and SE-ResNet [7]: it leverages the dense connectivity and multi-branch design of PeleeNet to capture diverse feature representations while incorporating attention mechanisms inspired by SE-ResNet to recalibrate and emphasize salient features. The network's architecture is built on a stage-wise design that alternates between Dense Feature Blocks which they are responsible for aggregating local and global features while Transition Blocks help to reduce dimensionality and control model complexity. In addition, Light Residual Blocks equipped with a lightweight Convolutional Block Attention Module (Light CBAM) ensure that both channel-wise and spatial information are effectively refined [30]. This combination enables PeleSENet to extract subtle texture and colour variations critical for distinguishing between benign and malignant lesions. In summary, PeleSENet is a robust, efficient, and effective model for medical image classification that overcomes the limitations of its predecessors, ensuring enhanced detection accuracy while maintaining an extremely compact architecture.

A. Architecture

As shown in Figure 1, PeleSENet integrates three main components: **Dense Feature Blocks**, **Transition Blocks**, and a lightweight **LightCBAM** attention mechanism. Let

$$\mathbf{X} \in \mathbb{R}^{H\times W\times 3}$$

denote an input dermoscopy image.

The overall architecture is organized into several sequential stages that systematically extract, refine, and condense image features. The network begins with an initial convolutional stem that downsamples the input image. This module employs a 3×3 convolution followed by a max pooling layer, reducing the spatial dimensions and preparing the feature maps for subsequent processing. This operation is mathematically expressed as:

$$F_0 = \mathrm{MaxPool}_{2\times 2}\Big(\mathrm{ReLU}\Big(\mathrm{BN}\big(\mathrm{Conv}_{3\times 3}(\mathbf{X},\, \mathrm{stride}=2)\big)\Big)\Big) \in \mathbb{R}^{\frac{H}{4}\times\frac{W}{4}\times 3}. \tag{1}$$

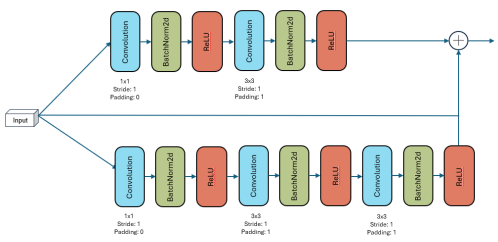

Fig. 2. Dense Feature Block

1) Dense Feature Block: Based on figure 2, each Dense Feature Block expands the feature set by aggregating outputs from parallel branches. Given an input feature map

$$x \in \mathbb{R}^{H\times W\times C},$$

the block employs two parallel paths:

- **Branch 1:** Applies a bottleneck transformation using a 1×1 convolution to reduce the channel dimension, followed by a 3×3 convolution.
- **Branch 2:** Uses a similar bottleneck approach but extends the path by incorporating an extra 3×3 convolution after the first 3×3 layer.

The outputs from these branches are concatenated with the original input along the channel dimension, enriching the feature representation by preserving the original features and adding complementary information. Mathematically, the operations are defined as follows:

Branch 1:

$$B_1(x) = \mathrm{conv}_{3\times 3}\Big(\mathrm{ReLU}\Big(\mathrm{conv}_{1\times 1}(x)\Big)\Big) \tag{2}$$

Branch 2:

$$B_2(x) = \mathrm{conv}_{3\times 3}\Big(\mathrm{ReLU}\Big(\mathrm{conv}_{3\times 3}\Big(\mathrm{ReLU}\big(\mathrm{conv}_{1\times 1}(x)\big)\Big)\Big)\Big) \tag{3}$$

Output:

$$y = \mathrm{Concat}\Big(x,\, B_1(x),\, B_2(x)\Big) \tag{4}$$

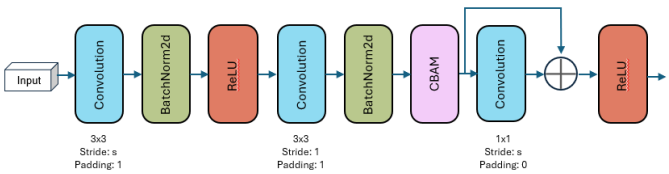

Fig. 3. Light Residual Block + CBAM

2) Light Residual Block with Light CBAM: Following the Dense Feature Block, a **Light Residual Block** refines the features. As shown in Figure 3, this block consists of two successive 3×3 convolution layers:

$$x' = \mathrm{conv}^{(2)}_{3\times 3}\Big(\mathrm{ReLU}\Big(\mathrm{conv}^{(1)}_{3\times 3}(x)\Big)\Big). \tag{5}$$

A residual connection adds either the original input or a transformed version (if channel dimensions differ) to the output:

$$y = \mathrm{ReLU}\Big(x' + \mathrm{Identity}(x)\Big). \tag{6}$$

Before the residual addition, the feature map is modulated by a lightweight attention mechanism, **Light CBAM**, which incorporates both channel and spatial attention which is shown in Figure 4.

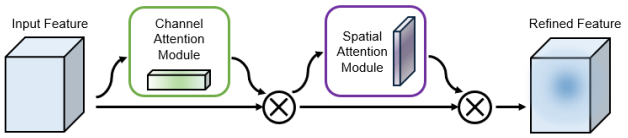

Fig. 4. Light CBAM

a) Channel Attention: The channel attention module computes attention weights $A_c(x)$ using both average pooling and max pooling. For an input x, this is defined as:

$$A_c(x) = \sigma\Big(\mathrm{MLP}\big(\mathrm{AvgPool}(x)\big) + \mathrm{MLP}\big(\mathrm{MaxPool}(x)\big)\Big), \tag{7}$$

where σ denotes the sigmoid activation function. The MLP comprises two 1×1 convolution layers with a ReLU activation between them. The feature map is then scaled by:

$$x_c = x \cdot A_c(x). \tag{8}$$

b) Spatial Attention: Spatial attention aggregates channel information using average and max operations:

$$A_s(x) = \sigma\Big(\mathrm{conv}_{3\times3}\Big(\big[\mathrm{Avg}(x);\, \mathrm{Max}(x)\big]\Big)\Big). \tag{9}$$

This spatial map highlights the important regions and is applied as:

$$x_s = x \cdot A_s(x). \tag{10}$$

The final output of Light CBAM is achieved by modulating the input feature map with both the channel and spatial attention maps.

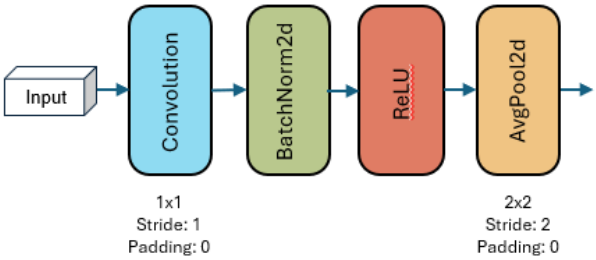

Fig. 5. Transition Block

3) Transition Block: Between stages, a Transition Block as shown in Figure 5 is utilized to compress the feature maps and reduce spatial resolution, thereby controlling model complexity. The Transition Block employs a 1×1 convolution to adjust the number of channels:

$$x_t = \mathrm{ReLU}\Big(\mathrm{BN}\big(\mathrm{Conv}_{1\times1}(x)\big)\Big).$$

This is followed by an average pooling layer that downsamples the spatial dimensions:

$$y_t = \mathrm{AvgPool}_{2\times2}(x_t).$$

This operation ensures that the network remains efficient and prevents overfitting by gradually reducing the feature map size as the network deepens.

a) Stage-Wise Organization.: The network is constructed by stacking several such stages. In each stage, the number of Dense Feature Blocks increases, promoting deeper feature fusion. After the final dense stage, a global adaptive average pooling layer aggregates the spatial dimensions to produce a compact feature vector:

$$x_{\text{final}} = \mathrm{AdaptiveAvgPool}(x).$$

This vector is then passed through a dropout layer (to mitigate overfitting) and finally through a fully connected layer for classification:

$$\hat{y} = \mathrm{FC}(x_{\text{final}}).$$

During training, standard weight initialization techniques such as Kaiming normalization are applied to ensure stable convergence.

b) Parameter Efficiency.: A main advantage aspect of PeleSENet is its parameter efficiency. By carefully designing each block, using bottleneck layers in the Dense Feature Block, lightweight convolutional operations in the Light Residual Block, and efficient attention mechanisms, PeleSENet maintains a parameter count of less than 3.5 million. This is achieved without compromising the network's ability to learn complex representations, which is particularly important for the fine-grained distinctions required in skin cancer detection.

c) Summary of Data Flow.: To summarize, the data flow in PeleSENet can be expressed as follows:

$$\begin{aligned}
x_0 &\rightarrow \text{Stem Block} \rightarrow x_1, \\
x_1 &\rightarrow \text{n * Dense Block} \rightarrow \text{Light Res Block (CBAM)} \\
&\rightarrow \text{Transition Block} \\
&\rightarrow \text{(Repeated across stages)} \rightarrow x_n, \\
x_n &\rightarrow \mathrm{AdaptiveAvgPool} \rightarrow x_{\text{pool}}, \\
x_{\text{pool}} &\rightarrow \text{Dropout} \rightarrow \mathrm{FC} \rightarrow \hat{y}.
\end{aligned}$$

The architecture progresses through four hierarchical stages with channel dimensions [32, 64, 128, 256], each stage containing increasing Dense Feature Blocks (1→2→3→4). Final classification uses adaptive average pooling followed by dropout (p=0.2) and linear projection. This hierarchical and modular design not only supports the robust learning of discriminative features but also ensures that the model remains computationally tractable and deployable in real-world clinical settings.

B. Motivation

The primary motivation behind PeleSENet is to address and overcome the inherent limitations found in its progenitor models which are PeleeNet and SE-ResNet respectively, especially when applied to the domain of skin cancer image classification. PeleeNet, while remarkably efficient in terms of computational cost and parameter count, tends to fall short in

its ability to extract sufficiently diverse features from complex medical images. Its densely connected architecture, although beneficial for feature propagation, but it may inadvertently lead to redundant computations and may not capture subtle differences in skin lesion textures and colours that are crucial for accurate diagnosis.

On the other hand, SE-ResNet introduces a sophisticated channel attention mechanism through squeeze-and-excitation blocks, significantly enhancing the network's capacity to focus on relevant features. However, this comes at the cost of increased complexity and a higher parameter count, which can be prohibitive in resource-constrained settings such as mobile applications or real-time clinical environments. Moreover, SE-ResNet's design may not be optimally tuned to handle the fine-grained and often noisy patterns present in skin cancer images.

PeleSENet is conceived to merge the strengths of both architectures while mitigating their weaknesses. By integrating the dense feature propagation of PeleeNet with an efficient attention mechanism inspired by SE-ResNet, PeleSENet is capable of capturing a richer set of discriminative features without incurring a heavy computational burden. The introduction of Light CBAM ensures that both channel-wise and spatial attention are dynamically adjusted, enabling the network to highlight the most critical regions in an image such as irregular borders, varied textures, and subtle colour differences that often characterize malignant lesions.

On the other hand, the modular design of PeleSENet, which strategically interleaves Dense Feature Blocks with Transition Blocks and Light Residual Blocks, promotes efficient gradient flow and prevents overfitting. This is particularly important for skin cancer detection, where datasets can be limited in size and imbalanced. The resulting architecture is not only robust and accurate but also lightweight enough to be deployed in practical medical applications. Thus, the design decisions underlying PeleSENet directly address the shortcomings of earlier models and provide an optimised solution for the challenges inherent in skin cancer image analysis.

C. Strengths

PeleSENet offers several key advantages over its predecessor architectures. First, its parameter-efficient design—with fewer than 3.5 million parameters ensures rapid inference and low computational overhead, making it ideal for deployment on mobile devices and in clinical settings. The dense connectivity pattern inherited from PeleeNet allows for effective feature reuse and diverse representation learning, while the integration of a lightweight attention mechanism enhances the network's ability to focus on crucial diagnostic features.

Furthermore, the network's modular design, which alternates between Dense Feature Blocks, Light Residual Blocks, and Transition Blocks, facilitates both deep feature extraction and efficient downsampling. This not only improves gradient propagation and stabilizes training but also ensures that subtle variations in skin lesions are accurately captured and processed. Overall, PeleSENet strikes an optimal balance between efficiency and accuracy, outperforming both PeleeNet and SE-ResNet in the context of skin cancer image classification.

D. Novelties and Contributions

In this work, we introduce a completely new lightweight CNN architecture, PeleSENet which is specifically optimized for medical image analysis in the field of skin cancer detection. Our contributions are threefold. First and foremost, we propose a novel integration of Dense Feature Blocks and Light Residual Blocks with an efficient attention mechanism, combining the dense connectivity of PeleeNet with the recalibration capabilities of SE-ResNet. Second, our design achieves high classification performance with an extremely compact model size, enabling real-time inference on resource-limited devices. Third, we demonstrate that our designed architecture not only improves accuracy in detecting subtle pathological patterns in skin lesions but also offers significant computational benefits over existing models. This innovative approach opens new avenues for deploying deep learning in clinical diagnostics and other medical imaging applications.

IV. Experiment

In this experiment, a Tesla GPU P100 was utilized to train the model, due to its high computational power for faster processing of deep learning tasks. The operating system used during the experiment is Windows 11, and the programming language employed for model development is Python, with the aid of PyTorch framework. PyTorch was selected for its flexibility, efficiency, and strong community support in implementing machine learning models. This setup ensures smooth training and evaluation, especially when working with large image datasets like those in this project.

A. Dataset

The dataset used in this experiment is a subset of the ISIC skin cancer dataset [11], a well-known dataset in medical imaging for classifying skin cancer. The dataset consists of two main subfolders: one for training and another for testing. It contains images from two distinct classes: benign and malignant.

Benign: These are non-cancerous growths that are generally harmless and can be treated with minimal intervention. Malignant: These are cancerous growths that can spread to other parts of the body and require more aggressive treatment, often including surgery, chemotherapy, or radiation [31]. Before encoding, the unique labels for the dataset were ['benign', 'malignant']. After encoding, the distribution of labels was as follows:

Class	Train Size	Test Size
Benign	1440	360
Malignant	1197	300
Total	2637	660

TABLE I
DATASET DISTRIBUTION

Based on table IV-A, the dataset is relatively small, with a total of 2637 images in the training set and 660 images in the

testing set. Given the relatively limited number of images, data augmentation techniques were applied to artificially increase the size of the dataset and introduce diversity into the training process. Although the dataset exhibits a slight imbalance, we have implemented an effective approach to address this issue, which will be demonstrated in the Training Process.

B. Pre-processing

The pre-processing stage is crucial for preparing the dataset and making sure that the model can generalize well during training. The steps include creating a train-validation split, applying data augmentation, and converting images into a suitable format for the neural network.

The training set was split into a new train set and a validation set. Specifically, 15% of the training data was reserved for validation, while the remaining 85% was used for training. The splitting was stratified based on the label distribution to ensure that both classes (benign and malignant) were represented proportionally in both the training and validation sets.

Training Set: *2241 images*
Validation Set: *396 images*

To address the issue of a small dataset, data augmentation was applied. Augmentation techniques artificially increase the dataset size and introduce variations in the images, allowing the model to generalize better [32]. For better comparison and visualisation the original image is shown as fig 6. The augmentations performed include:

1) Random Rotation: Rotating the image by 90 degrees to introduce rotational invariance as shown in Figure 7.
2) Random Horizontal Flip: Flipping the image horizontally to mimic potential real-world variations as shown in Figure 8.
3) Random Vertical Flip: Flipping the image vertically, providing additional variability in orientation as shown in Figure 9.
4) Color Jitter: Modifying the brightness, contrast, saturation, and hue of the image, simulating lighting variations in real-world scenarios as shown in Figure 10.
5) Random Affine Transform: Applying transformations such as rotation, scaling, translation, and shearing to mimic different perspectives and distortions as shown in Figure 11.
6) Random Perspective: Adding perspective distortion to simulate changes in the viewpoint of the image as shown in Figure 12.

By applying these augmentations, the dataset was increased from 2241 images to 11205 images. The main reason for augmenting the dataset is that smaller datasets can lead to poor generalization in models, especially when training deep neural networks like CNNs. Augmentation provides more training examples, thereby reducing overfitting and enabling the model to learn more robust features. This is particularly important when using a lightweight CNN, which may struggle to learn complex patterns from a limited dataset.

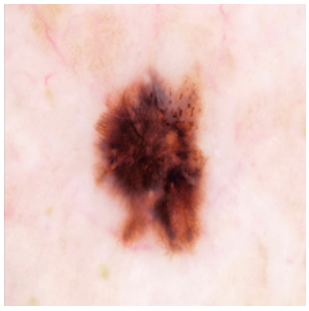

Fig. 6. Non-augmented image

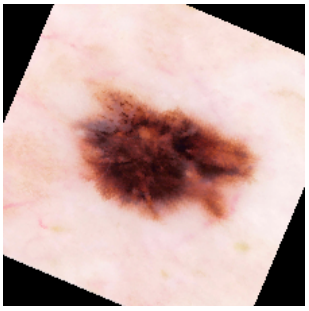

Fig. 7. Rotated 90 degrees

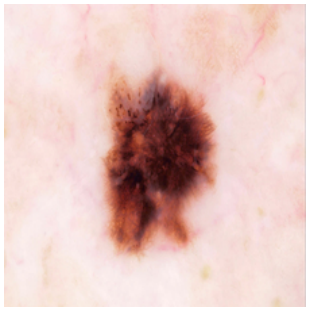

Fig. 8. Flipped horizontally

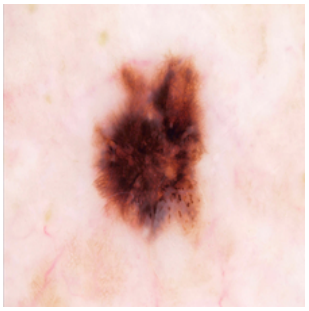

Fig. 9. Flipped vertically

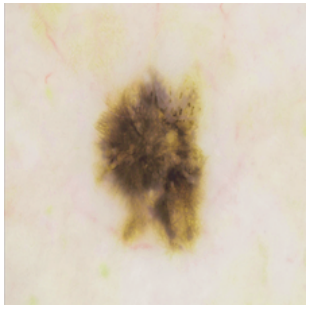

Fig. 10. Color jitter (brightness, contrast, saturation, hue)

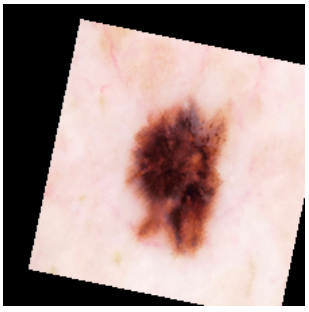

Fig. 11. Affine transformation (scaling, translation, shearing)

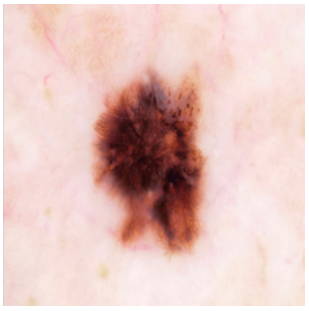

Fig. 12. Random perspective distortion

C. Training Process

In this experiment, the training process was carried out with detailed steps to ensure effective learning. Below are the key parts of the training code:

- Device Selection: The model was transferred to the GPU if available. This ensures that training can take place much faster compared to using a CPU, particularly when dealing with large datasets and deep networks.
- Class Weights: Since the dataset has an imbalanced class distribution, class weights were calculated. The weight for each class is inversely proportional to the class frequency, helping to handle class imbalance during training. The criterion used for loss calculation was CrossEntropyLoss with these class weights.
- Training Loop: The model was trained for 35 epochs, where the model's performance was evaluated on the validation set after each epoch. The best model weights were saved, and the final model was loaded for evaluation on the test set.
- Model Evaluation: After training, the model was evaluated on the test set by calculating the accuracy. The accuracy metric was computed by comparing the model's predicted labels to the ground truth labels.

This training pipeline ensures that the model is trained efficiently and can generalize well to unseen data.

D. Ablation

In our ablation study, we systematically modified key components of PeleSENet to quantify their individual contributions to the network's performance, as shown in the Figure 9. First, we removed the entire CBAM module which fuses both channel and spatial attention mechanisms, resulting in an accuracy of 86.21%. This drop highlights that the attention module is critical for enhancing feature representation. Furthermore, we simplified the Dense Feature Block by using only a single branch instead of the dual-branch design. The accuracy fell further to 82.73%, indicating that the dual-branch structure plays an important role in capturing richer, multi-scale features through complementary processing paths.

We also examined the impact of the spatial attention component by retaining only the ChannelAttention module. With this modification, the model achieved an accuracy of 86.67%, suggesting that while channel attention contributes significantly to performance, the additional spatial attention further refines the feature maps by focusing on informative regions. Finally, the full PeleSENet architecture, which integrates both the two-branch Dense Feature Block and the complete CBAM (combining channel and spatial attention), reached the highest accuracy of 89.70%. These results collectively demonstrate that each component, especially the combined attentions, plays a vital role in boosting the network's discriminative power.

For context, our design is inspired by earlier works such as PeleeNet and SEResNet. PeleeNet, with its efficient network architecture, contributed valuable insights into simplifying deep models while maintaining high accuracy, leading to an accuracy of 87.58%. SEResNet, on the other hand, introduced the powerful residual connections that significantly enhanced the feature flow, which is reflected in its accuracy of 86.36%. By building upon these foundational models, PeleSENet incorporates novel enhancements like the attention mechanism and dual-branch design, offering improved performance and robustness.

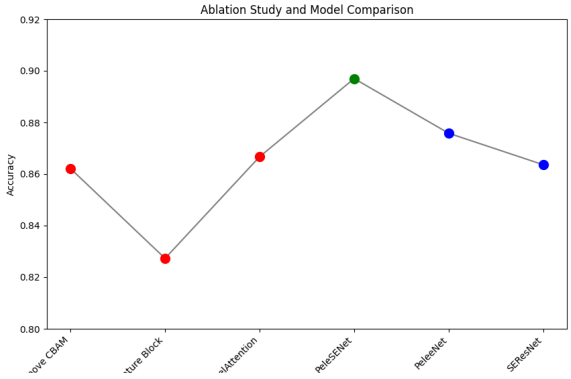

Fig. 13. Ablation Study

E. Results

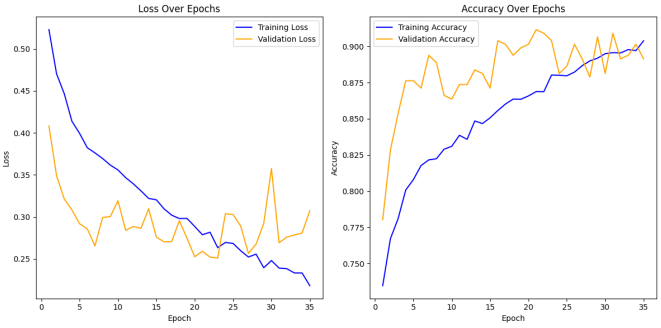

Fig. 14. Accuracy/Loss over Epochs

Based on the training curve as shown in Figure 14, during the 35-epoch training process, the model demonstrated progressive improvement across multiple performance metrics. Initially, the training loss was relatively high at 0.5228, accompanied by a training accuracy of 73.47%. Over the epochs, the loss steadily decreased to 0.2181, while the accuracy increased to 90.39%. The validation metrics reflected similar trends, with the loss stabilizing near 0.3074 and the accuracy reaching around 91.16%, indicating that the model generalized well to unseen data.

The confusion matrix as shown in fig 15 offers additional insight into the model's performance. For the positive class, the precision is calculated as 257/(257+25), which is approximately 91.1%. This means that when the model predicts a positive class, it is correct 91.1% of the time. Sensitivity, or recall, is determined as 257/(257+43) and comes out to roughly 85.7%, indicating the model's ability to identify 85.7% of

actual positive cases. The balance between high precision and robust sensitivity illustrates that the model effectively minimizes false positives while still capturing a significant proportion of true positives. Overall, these results confirm the model's strong performance and its capacity for reliable binary classification.

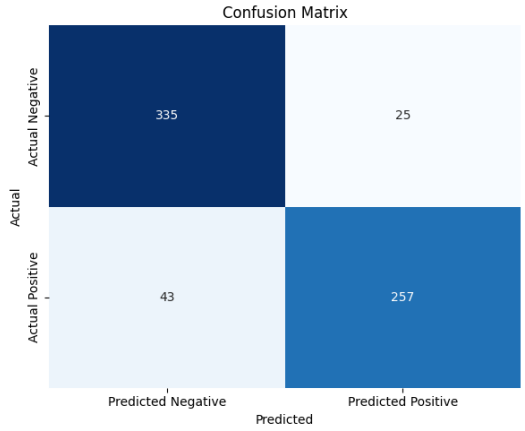

Fig. 15. Confusion Matrix

F. Comparison

a) Non-pretrained models: In the comparison table II, we observe a wide range of models with varying numbers of parameters and accuracy levels. While traditional models such as AlexNet (58M parameters) and ResNet (23.5M parameters) show high accuracy, they come with significant computational and storage costs. This makes them less feasible for deployment in resource-constrained environments, such as rural areas where computational and storage resources are limited. In contrast, our proposed model achieves a commendable accuracy of 0.8970 with only 3.45M parameters.

By focusing on creating a lightweight CNN architecture, our method strikes a balance between performance and efficiency. This is particularly important for real-world deployment in rural areas where devices may have limited processing power and storage. Compared to heavier models such as ResNet or EfficientNet-B0, our method not only reduces resource consumption but also achieves higher accuracy, showing that it is possible to optimize CNNs for lower computational cost without sacrificing performance. The ability to maintain accuracy with fewer parameters ensures that our model is suitable for edge devices and low-power systems, making it an ideal solution for applications where both efficiency and accuracy are crucial.

b) Comparison of our proposed method with other state-of-the-art methods using similar dataset: The comparative analysis presented in the table underscores the challenge of balancing performance and computational efficiency in CNN architectures. Many state-of-the-art methods, such as those based on VGG16 (Angelina and Ulfitria, Anand et al.) [77] [78] or ensemble approaches (Maurya et al.) [79], achieve high accuracy scores; however, they do so at the cost of significantly increased parameter counts—often exceeding 130 million. Even sophisticated frameworks like the CNN-based two-phase evolutionary method and binary residual feature fusion exhibit parameter counts that are not ideal for resource-limited environments.

In contrast, our proposed CNN model attains an accuracy of 89.70% with only 3,454,906 parameters. This lightweight design is particularly advantageous for scenarios where computational power and storage are constrained, such as rural

TABLE II
COMPARISON OF OUR PROPOSED METHOD WITH OTHER NON-PRETRAINED CNN MODELS USING THE SAME DATASET.

Model	Number of Params	Accuracy
SqueezeNet [33]	736450	0.8424
ShuffleNet [34]	956090	0.8424
MnasNet [35]	3104874	0.8121
MobileNetV2 [36]	2227520	0.8121
CondenseNet [37]	3743010	0.7788
AIRnet [38]	25380962	0.8348
AlexNet [3]	58289538	0.8409
BagNet [39]	13643842	0.8409
CBAMResNet [40]	11267418	0.8500
ChannelNet [41]	2852162	0.8242
DarkNet [42]	19819426	0.8303
DIA-ResNet [43]	5785378	0.8515
EfficientNetB0 [44]	4010110	0.8530
ESPNetV2 [45]	647122	0.8545
FBNet [46]	3591170	0.8242
FishNet [47]	15574018	0.8561
HRNet [48]	19254102	0.8470
IBN(b)-ResNet [49]	23513666	0.8227
IBN-DenseNet [49]	6955906	0.8485
IBN-ResNet [49]	23512130	0.8364
AirNeXt [50]	25559394	0.8530
BN-Inception [51]	10272290	0.8485
DIA-PreResNet [43]	5783714	0.8606
DiracNetV2 [52]	10999810	0.8333
DLA-46c [53]	1044914	0.8348
DPN68 [54]	11780268	0.8212
DRN [55]	30722770	0.8879
IBN-ResNeXt101 [49]	42132802	0.8318
IGCV3 [56]	707090	0.8182
i-RevNet [57]	122053502	0.8182
iSQRTCOVResNet [58]	43090754	0.8500
MENet [59]	**222382**	0.8136
MixNet [60]	3480456	0.8167
MobileNetV3 [61]	881162	0.8000
MSDNet [62]	17338224	0.8500
NASNet [63]	4235092	0.8470
Oct-ResNet [64]	4911042	0.8561
PreResNet [65]	4905154	0.8515
PyramidNet [66]	40761464	0.8576
ResAttNet [67]	29765826	0.8576
ResNet-50 [68]	23512130	0.8303
ResNeX-14 [69]	5082434	0.8470
SparseNet-121 [70]	3089762	0.8333
SENet [7]	29321266	0.8682
ShaResNet-18 [71]	15328322	0.8530
SqueezeNext [72]	595314	0.7924
BN-VGG-11 [73]	128777282	0.8621
WRN-50 [74]	66804226	0.8273
XDenseNet [75]	6954114	0.8394
ZFNet [76]	58268802	0.8333
Our method	3454906	**0.8970**

deployments. By reducing the model's complexity without sacrificing performance, our approach demonstrates a viable pathway for implementing efficient deep learning solutions in environments with limited resources, ensuring both scalability and accessibility.

TABLE III
COMPARISON OF OUR PROPOSED METHOD WITH OTHER STATE-OF-THE-ART METHODS USING THE SAME DATASET.

Study	Model	Accuracy (%)	Number of Params
Angelina and Ulfitria [77]	VGG16	83.75	134,268,738
Maurya et al. [79]	CNN ensemble models	87.88	212,715,792
Ghosh et al. [80]	CNN-based two-phase evolutionary framework	83.40	39,646,615
Nivyashree and Pramila [81]	CNN	80.80	59,932,738
Shekar and Hailu [82]	DenseNet-169, local binary pattern	**89.70**	22,730,124
Hussein et al. [83]	ResNet-18	89.39	11,177,538
Salian and Sawarka [84]	EfficientNetB3	87.12	10,699,306
Abdar et al. [85]	Binary residual feature fusion	89.24	232,271,466
Anand et al. [78]	VGG16	89.09	139,160,954
Gupta et al. [86]	InceptionV3	83.20	26,210,882
Ours	CNN	**89.70**	**3,454,906**

V. THE PROPOSED FRAMEWORK

A. Architecture

The proposed framework is a hybrid system that combines a lightweight convolutional neural network which is PeleSENet with advanced modules for explainability and adaptive learning, exclusively designed for skin cancer detection. The system not only classifies skin lesions as benign or malignant but also incorporates expert feedback to continuously improve its performance.

a) Image Acquisition and Preprocessing: The process begins by capturing skin lesion images, which are then preprocessed through resizing, normalization, and data augmentation. These steps ensure that the input images are optimized for feature extraction and robust against variations.

b) Feature Extraction and Initial Classification: Once pre-processed, the image is given to PeleSENet as input. Within PeleSENet, multiple convolutional layers, enhanced with CBAM attention blocks gives an edge in extracting hierarchical features and recalibrate channel-wise responses to highlight the most informative aspects of the lesion. The model computes the classification probabilities, and the highest probability is taken as the confidence level.

c) Decision-Making Based on Confidence Threshold:

- High Confidence ($(\geq T)$): If the maximum probability exceeds the preset confidence threshold in this case where it will be set at 98%, the system immediately classifies the image:
  - For malignant predictions, a Grad-CAM heatmap is generated to visually indicate the regions influencing the decision. The malignant result, along with its heatmap, is displayed and stored as a high-confidence case. This is shown in Figure 16.
  - For benign predictions, a similar process is followed: a Grad-CAM heatmap is produced, the benign classification is displayed, and the result is stored as high-confidence benign. This is shown in Figure 17.
- Low Confidence ($< T$): If the confidence level is below the threshold, the image is flagged for further review:
  - The system sends the image and a preliminary result to a dermatologist dashboard.
  - It then waits for expert feedback, denoted as F, which is used to update the final classification.
  - If the model's initial prediction aligns with the dermatologist's feedback, the final decision is displayed along with a Grad-CAM heatmap. Otherwise, the result is presented without the heatmap.

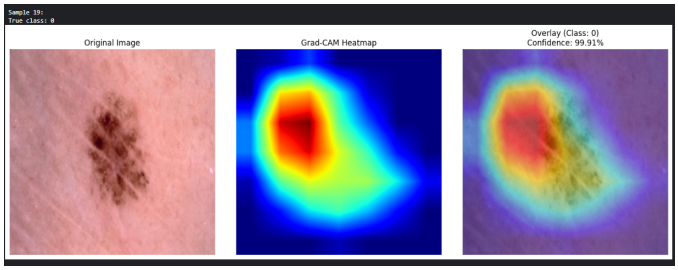

Fig. 16. Malignant Prediction with Confidence Level and Explainability

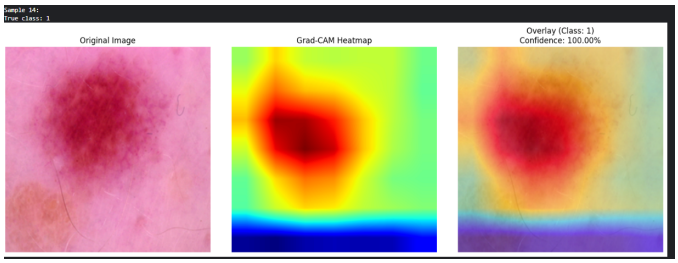

Fig. 17. Benign Prediction with Confidence Level and Explainability

d) Logging and Adaptive Learning: Every image, along with its final classification and associated confidence score, is logged into a database. When new feedback is received, it is appended to a data buffer. At specified retraining intervals (R), the system combines the original training dataset (D) with the accumulated feedback (buffer B) to form a new training set (T_{new}). A new model, M_{new}, is retrained on this updated dataset and evaluated on D:

- If M_{new}'s accuracy meets or exceeds that of the current model, it replaces the old model and the update is logged.
- Otherwise, the existing model is retained, the new feedback is discarded, and a rollback is logged.

e) Continuous Monitoring and Integration: The framework continually monitors system performance and adjusts its confidence threshold based on ongoing feedback. It also integrates seamlessly with external systems, such as hospital databases and mobile health applications, ensuring scalability and robust performance in clinical settings. This adaptive algorithm, outlined in Algorithm 1: Adaptive Skin Lesion Classification with Expert Feedback, underscores the system's commitment to precision and continuous improvement, making it both innovative and practical for real-world dermatological applications.

Algorithm 1 Adaptive Skin Lesion Classification with Expert Feedback

BEGIN
INPUT: Image I, model M, confidence threshold T (e.g., 98%), data buffer B, retraining interval R, original training dataset D.
PREPROCESS I (resize, normalize, augment) and **EXTRACT** features using PeleSENet(I).
COMPUTE classification probabilities $P = M(I)$, set $C = \max(P)$.
if $C \geq T$ **then**
    if predicted class is *malignant* **then**
        $H \leftarrow$ Grad-CAM(I)
        DISPLAY malignant classification **with** heatmap H
        STORE as high-confidence malignant
    else
        $H \leftarrow$ Grad-CAM(I)
        DISPLAY benign classification **with** heatmap H
        STORE as high-confidence benign
    end if
else
    FLAG I for dermatologist review
    SEND I + preliminary result to dashboard
    WAIT for dermatologist feedback F
    UPDATE final classification based on F
    if model prediction matches F **then**
        $H \leftarrow$ Grad-CAM(I)
        DISPLAY final classification **with** H
    else
        DISPLAY final classification **without** Grad-CAM
    end if
end if
LOG I, final result, and C in database
if new feedback F received **then**
    APPEND (I, F) to buffer B
end if
if mod(current time, R) $= 0$ **then**
    PREPARE $T_{\text{new}} = D \cup B$
    RETRAIN M_{new} on T_{new}
    EVALUATE M_{new} on D
    if accuracy(M_{new}) $\geq$ accuracy(M) **then**
        UPDATE $M \leftarrow M_{\text{new}}$ & **LOG** update
    else
        RETAIN M; **DISCARD** B; **LOG** rollback
    end if
    CLEAR buffer B
end if
MONITOR system performance
if dermatologist review completed **then**
    UPDATE database with final decision
    ADJUST T based on feedback
end if
SAVE logs & checkpoints
END

B. Motivation

The primary motivation behind this work is to create a portable, affordable, and highly effective skin cancer detection system that can be deployed in resource-limited settings, particularly in rural areas where access to specialized dermatological services is limited. By developing a lightweight CNN like PeleSENet with less than 3.5 million parameters, we aim to significantly reduce computational and storage requirements, thereby enabling rapid and accurate diagnoses even on mobile or edge devices.

This research is driven by the urgent need for early detection of skin cancer, which is critical for improving patient outcomes and reducing healthcare costs. Traditional diagnostic methods often require expensive equipment and specialist intervention, limiting their availability in underserved regions. Our approach leverages state-of-the-art deep learning techniques combined with eXplainable AI (XAI) methodologies, such as Grad-CAM, to provide clear visual explanations of model decisions. This transparency not only escalates user confidence but also ensures that dermatologists can verify and trust the system's predictions, thus bridging the gap between automated detection and human expertise.

Furthermore, the incorporation of an adaptive learning mechanism represents a forward-thinking strategy. By periodically retraining the model with new user inputs, the system continuously improves and adapts to changing data distributions. This ensures that the model remains robust over time and can effectively handle novel cases, reducing false negatives and false positives. The framework also includes safeguards that allow for the rollback of model updates if performance deteriorates, thereby maintaining reliability and consistency in diagnostic results.

Ethical considerations play a significant role in our motivation. In an era where AI transparency is important, our system is designed to operate with full accountability. The combination of explainable decision-making and human oversight through dermatologist reviews prevents the dissemination of false information, ensuring that patient care remains both safe and effective. Ultimately, our goal is to democratize access to high-quality skin cancer detection, reduce diagnostic costs, and create an environment of continuous learning and improvement in medical AI.

C. Strengths

The proposed framework exhibits numerous strengths that position it as a state-of-the-art solution for skin cancer detection. One of its primary advantages is its lightweight design; with fewer than 3.5 million parameters, PeleSENet is highly efficient and can be deployed on low-resource devices, making it ideal for rural and remote healthcare settings. This computational efficiency not only reduces the cost of hardware but also ensures rapid processing times, which is critical for timely diagnosis.

Another significant strength lies in the integration of eXplainable AI through Grad-CAM. This feature provides visual heatmaps that highlight regions of the image most influential in

the decision-making process. Such transparency is invaluable for both end-users and medical professionals, as it enhances trust in the system and allows dermatologists to verify automated diagnoses, reducing the likelihood of misdiagnosis.

The adaptive learning component further enhances the framework's robustness by enabling continuous improvement. The system periodically retrains the model with new user data, ensuring that it adapts to emerging patterns and maintains high accuracy over time. Importantly, the framework incorporates a safeguard mechanism that reverts to the previous model if the new one performs worse on a standardized testing dataset, thus ensuring reliability and stability in clinical applications.

Additionally, the modular and scalable design of the framework facilitates easy integration with existing healthcare infrastructures. Its flexibility allows for future upgrades, such as the incorporation of additional diagnostic tools or integration with telemedicine platforms. By combining affordability, efficiency, transparency, and continuous learning, this framework not only improves diagnostic accuracy but also democratizes access to advanced medical technology, ultimately reducing healthcare disparities and costs.

Finally, ethical considerations are deeply embedded in the system's design. The framework ensures that decisions are made transparently and with human oversight, aligning with best practices in ethical AI. This balanced approach of automated efficiency and expert verification makes the proposed framework a reliable, trustworthy, and cutting-edge solution for enhanced skin cancer detection.

VI. Conclusion

In conclusion, our work presents PeleSENet, a novel lightweight convolutional neural network designed for skin cancer detection, which successfully balances computational efficiency and diagnostic accuracy. By fusing the dense connectivity of PeleeNet with the channel recalibration capabilities of SE-ResNet through a dual-branch design and a streamlined attention mechanism, PeleSENet achieves robust feature extraction with fewer than 3.5 million parameters. The integration of Grad-CAM provides visual insights, enhancing explainability and fostering trust between AI systems and clinical practitioners. Moreover, the incorporation of adaptive learning strategies allows the model to continuously improve by assimilating new data, thereby addressing challenges such as feature drift and dataset imbalances. Our experimental results on the ISIC dataset, supported by rigorous ablation studies, demonstrate that the proposed architecture not only outperforms several existing models but also remains ideally suited for deployment in resource-constrained settings. This work paves the way for accessible, efficient, and transparent AI-driven diagnostic tools in skin cancer screening. Future research will explore extending these methodologies to other medical imaging domains and further refining adaptive mechanisms to enhance long-term model performance and reliability.

References

[1] Hugh M Gloster Jr and David G Brodland. The epidemiology of skin cancer. *Dermatologic Surgery*, 22(3):217–226, 1996.

[2] Aarushi Shah, Manan Shah, Aum Pandya, Rajat Sushra, Ratnam Sushra, Manya Mehta, Keyur Patel, and Kaushal Patel. A comprehensive study on skin cancer detection using artificial neural network (ann) and convolutional neural network (cnn). *Clinical eHealth*, 6:76–84, 2023.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[4] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. *arXiv [cs.CV]*, 2016.

[5] Mingxing Tan and Quoc V Le. EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv [cs.LG]*, 2019.

[6] Maryam Naqvi, Syed Qasim Gilani, Tehreem Syed, Oge Marques, and Hee-Cheol Kim. Skin cancer detection using deep learning—a review. *Diagnostics*, 13(11), 2023.

[7] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks, 2019.

[8] Robert J. Wang, Xiang Li, and Charles X. Ling. Pelee: A real-time object detection system on mobile devices. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[9] Katja Hauser, Alexander Kurz, Sarah Haggenmüller, Roman C. Maron, Christof von Kalle, Jochen S. Utikal, Friedegund Meier, Sarah Hobelsberger, Frank F. Gellrich, Mildred Sergon, Axel Hauschild, Lars E. French, Lucie Heinzerling, Justin G. Schlager, Kamran Ghoreschi, Max Schlaak, Franz J. Hilke, Gabriela Poch, Heinz Kutzner, Carola Berking, Markus V. Heppt, Michael Erdmann, Sebastian Haferkamp, Dirk Schadendorf, Wiebke Sondermann, Matthias Goebeler, Bastian Schilling, Jakob N. Kather, Stefan Fröhling, Daniel B. Lipka, Achim Hekler, Eva Krieghoff-Henning, and Titus J. Brinker. Explainable artificial intelligence in skin cancer recognition: A systematic review. *European Journal of Cancer*, 167:54–69, 2022.

[10] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.*, 128(2):336–359, February 2020.

[11] ISIC Challenge — challenge.isic-archive.com. https://challenge.isic-archive.com/data/. [Accessed 13-02-2025].

[12] Bas HM Van der Velden, Hugo J Kuijf, Kenneth GA Gilhuijs, and Max A Viergever. Explainable artificial intelligence (xai) in deep learning-based medical image analysis. *Medical Image Analysis*, 79:102470, 2022.

[13] Deepshikha Bhati, Fnu Neha, and Md Amiruzzaman. A survey on explainable artificial intelligence (xai) techniques for visualizing deep learning models in medical imaging. *Journal of Imaging*, 10(10):239, 2024.

[14] Katarzyna Borys, Yasmin Alyssa Schmitt, Meike Nauta, Christin Seifert, Nicole Krämer, Christoph M Friedrich, and Felix Nensa. Explainable ai in medical imaging: An overview for clinical practitioners–beyond saliency-based xai approaches. *European journal of radiology*, 162:110786, 2023.

[15] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should I trust you?": Explaining the predictions of any classifier. 2016.

[16] Alexander Binder, Grégoire Montavon, Sebastian Bach, Klaus-Robert Müller, and Wojciech Samek. Layer-wise relevance propagation for neural networks with local renormalization layers. 2016.

[17] Justin Oakley Robert Sparrow, Joshua Hatherley and Chris Bain. Should the use of adaptive machine learning systems in medicine be classified as research? *The American Journal of Bioethics*, 24(10):58–69, 2024. PMID: 38662360.

[18] Amritpal Singh, Mustafa Burak Gurbuz, Shiva Souhith Gantha, and Prahlad Jasti. Class-incremental continual learning for general purpose healthcare models, 2023.

[19] Huahui Yi, Ziyuan Qin, Qicheng Lao, Wei Xu, Zekun Jiang, Dequan Wang, Shaoting Zhang, and Kang Li. Towards general purpose medical ai: Continual learning medical foundation model, 2023.

[20] Meet Joshi. Adaptive learning through artificial intelligence. *International Journal on Integrated Education*, 7:41–43, 02 2024.

[21] Arthur A M Teodoro, Douglas H Silva, Renata L Rosa, Muhammad Saadi, Lunchakorn Wuttisittikulkij, Rao Asad Mumtaz, and Demóstenes Z Rodríguez. A skin cancer classification approach using

GAN and RoI-based attention mechanism. *J. Signal Process. Syst.*, 95(2-3):211–224, March 2023.

[22] Ravi Dandu, M Vinayaka Murthy, and YB Ravi Kumar. Transfer learning for segmentation with hybrid classification to detect melanoma skin cancer. *Heliyon*, 9(4), 2023.

[23] Hossam Magdy Balaha and Asmaa El-Sayed Hassan. Skin cancer diagnosis based on deep transfer learning and sparrow search algorithm. *Neural Comput. Appl.*, 35(1):815–853, January 2023.

[24] Sufiyan Bashir Mukadam and Hemprasad Yashwant Patil. Skin cancer classification framework using enhanced super resolution generative adversarial network and custom convolutional neural network. *Applied Sciences*, 13(2), 2023.

[25] Muhammad Hasnain Javid, Waqas Jadoon, Haris Ali, and Muhammad Danish Ali. Design and analysis of an improved deep ensemble learning model for melanoma skin cancer classification. In *2023 4th International Conference on Advancements in Computational Sciences (ICACS)*, pages 1–6, 2023.

[26] Amit Kumar, TY Satheesha, Bizotto Beatriz Lucia Salvador, Sathiyanarayanan Mithileysh, and Syed Thouheed Ahmed. Augmented intelligence enabled deep neural networking (audnn) framework for skin cancer classification and prediction using multi-dimensional datasets on industrial iot standards. *Microprocessors and Microsystems*, 97:104755, 2023.

[27] Xiaoyu He, Yong Wang, Shuang Zhao, and Xiang Chen. Co-attention fusion network for multimodal skin cancer diagnosis. *Pattern Recognition*, 133:108990, 2023.

[28] N Aishwarya, K Manoj Prabhakaran, Frezewd Tsegaye Debebe, M Sai Sree Akshitha Reddy, and Posina Pranavee. Skin cancer diagnosis with yolo deep neural network. *Procedia Computer Science*, 220:651–658, 2023. The 14th International Conference on Ambient Systems, Networks and Technologies Networks (ANT) and The 6th International Conference on Emerging Data and Industry 4.0 (EDI40).

[29] Bill Cassidy, Connah Kendrick, Andrzej Brodzicki, Joanna Jaworek-Korjakowska, and Moi Hoon Yap. Analysis of the isic image datasets: Usage, benchmarks and recommendations. *Medical Image Analysis*, 75:102305, 2022.

[30] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module, 2018.

[31] Ph.D. Jill Seladi-Schulman. Benign Skin Tumors: What They Mean — healthline.com. https://www.healthline.com/health/cancer/benign-skin-cancer. [Accessed 24-02-2025].

[32] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning, 2017.

[33] Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and ¡0.5mb model size, 2016.

[34] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices, 2017.

[35] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, Mark Sandler, Andrew Howard, and Quoc V. Le. Mnasnet: Platform-aware neural architecture search for mobile, 2019.

[36] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks, 2019.

[37] Gao Huang, Shichen Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Condensenet: An efficient densenet using learned group convolutions, 2018.

[38] Evelyn Chee and Zhenzhou Wu. Airnet: Self-supervised affine registration for 3d medical images using neural networks, 2018.

[39] Wieland Brendel and Matthias Bethge. Approximating cnns with bag-of-local-features models works surprisingly well on imagenet, 2019.

[40] Yana Luo and Zhongsheng Wang. An improved resnet algorithm based on cbam. In *2021 International Conference on Computer Network, Electronic and Automation (ICCNEA)*, pages 121–125, 2021.

[41] Hongyang Gao, Zhengyang Wang, and Shuiwang Ji. Channelnets: Compact and efficient convolutional neural networks via channel-wise convolutions, 2018.

[42] Joseph Redmon and Ali Farhadi. Yolo9000: Better, faster, stronger, 2016.

[43] Zhongzhan Huang, Senwei Liang, Mingfu Liang, and Haizhao Yang. Dianet: Dense-and-implicit attention network, 2019.

[44] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking model scaling for convolutional neural networks, 2020.

[45] Sachin Mehta, Mohammad Rastegari, Anat Caspi, Linda Shapiro, and Hannaneh Hajishirzi. Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation, 2018.

[46] Bichen Wu, Xiaoliang Dai, Peizhao Zhang, Yanghan Wang, Fei Sun, Yiming Wu, Yuandong Tian, Peter Vajda, Yangqing Jia, and Kurt Keutzer. Fbnet: Hardware-aware efficient convnet design via differentiable neural architecture search, 2018.

[47] Shuyang Sun, Jiangmiao Pang, Jianping Shi, Shuai Yi, and Wanli Ouyang. Fishnet: A versatile backbone for image, region, and pixel level prediction, 2019.

[48] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao. Deep high-resolution representation learning for visual recognition, 2020.

[49] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net, 2020.

[50] Lu Yang, Qing Song, Yingqi Wu, and Mengjie Hu. Attention inspiring receptive-fields network for learning invariant representations. *IEEE Transactions on Neural Networks and Learning Systems*, 30(6):1744–1755, 2019.

[51] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015.

[52] Sergey Zagoruyko and Nikos Komodakis. Diracnets: Training very deep neural networks without skip-connections, 2018.

[53] Fisher Yu, Dequan Wang, Evan Shelhamer, and Trevor Darrell. Deep layer aggregation, 2019.

[54] Yunpeng Chen, Jianan Li, Huaxin Xiao, Xiaojie Jin, Shuicheng Yan, and Jiashi Feng. Dual path networks, 2017.

[55] Fisher Yu, Vladlen Koltun, and Thomas Funkhouser. Dilated residual networks, 2017.

[56] Ke Sun, Mingjie Li, Dong Liu, and Jingdong Wang. Igcv3: Interleaved low-rank group convolutions for efficient deep neural networks, 2018.

[57] Jörn-Henrik Jacobsen, Arnold Smeulders, and Edouard Oyallon. i-revnet: Deep invertible networks, 2018.

[58] Peihua Li, Jiangtao Xie, Qilong Wang, and Zilin Gao. Towards faster training of global covariance pooling networks by iterative matrix square root normalization, 2018.

[59] Zheng Qin, Zhaoning Zhang, Shiqing Zhang, Hao Yu, and Yuxing Peng. Merging and evolution: Improving convolutional neural networks for mobile applications, 2018.

[60] Mingxing Tan and Quoc V. Le. Mixconv: Mixed depthwise convolutional kernels, 2019.

[61] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, Quoc V. Le, and Hartwig Adam. Searching for mobilenetv3, 2019.

[62] Gao Huang, Danlu Chen, Tianhong Li, Felix Wu, Laurens van der Maaten, and Kilian Q. Weinberger. Multi-scale dense networks for resource efficient image classification, 2018.

[63] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. Learning transferable architectures for scalable image recognition, 2018.

[64] Yunpeng Chen, Haoqi Fan, Bing Xu, Zhicheng Yan, Yannis Kalantidis, Marcus Rohrbach, Shuicheng Yan, and Jiashi Feng. Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution, 2019.

[65] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks, 2016.

[66] Dongyoon Han, Jiwhan Kim, and Junmo Kim. Deep pyramidal residual networks, 2017.

[67] Fei Wang, Mengqing Jiang, Chen Qian, Shuo Yang, Cheng Li, Honggang Zhang, Xiaogang Wang, and Xiaoou Tang. Residual attention network for image classification, 2017.

[68] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[69] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks, 2017.

[70] Ligeng Zhu, Ruizhi Deng, Michael Maire, Zhiwei Deng, Greg Mori, and Ping Tan. Sparsely aggregated convolutional networks, 2019.

[71] Alexandre Boulch. Sharesnet: reducing residual network parameter number by sharing weights, 2017.

[72] Amir Gholami, Kiseok Kwon, Bichen Wu, Zizheng Tai, Xiangyu Yue, Peter Jin, Sicheng Zhao, and Kurt Keutzer. Squeezenext: Hardware-aware neural network design, 2018.

[73] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.
[74] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks, 2017.
[75] Ameya Prabhu, Girish Varma, and Anoop Namboodiri. Deep expander networks: Efficient deep networks from graph theory, 2018.
[76] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks, 2013.
[77] C Angelina and RU Ulfitria. Classification of skin cancer using resnet and vgg deep learning network. In *Proceedings of the 11th International Applied Business and Engineering Conference, ABEC*, 2023.
[78] Vatsala Anand, Sheifali Gupta, Ayman Altameem, Soumya Ranjan Nayak, Ramesh Chandra Poonia, and Abdul Khader Jilani Saudagar. An enhanced transfer learning based classification for diagnosis of skin cancer. *Diagnostics*, 12(7), 2022.
[79] Ritesh Maurya, Anant Krisn Bais, T Gopalakrishnan, Malay Kishore Dutta, Nageshwar Nath Pandey, and Srinivasa Murthy YV. Skin lesion classification using deep feature fusion and selection using xgboost classifier. In *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pages 1–5. IEEE, 2024.
[80] Arjun Ghosh, Nanda Dulal Jana, Swagatam Das, and Rammohan Mallipeddi. Two-phase evolutionary convolutional neural network architecture search for medical image classification. *IEEE Access*, 11:115280–115305, 2023.
[81] T. Nivyashree and P. V. Pramila. Detection of malignant and benign skin lesions using the influence of activation function and accuracy analysis in densely connected convolutional network compared over convolutional neural network. In *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)*, pages 1–6, 2023.
[82] B.H. Shekar and Habtu Hailu. An efficient stacked ensemble model for the detection of covid-19 and skin cancer using fused feature of transfer learning and handcrafted methods. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11(3):878–894, 2023.
[83] Hadeer Hussein, Ahmed Magdy, Rehab Abdel-Kader, and Khaled Ali. Binary classification of skin cancer using pretrained deep neural networks. *Suez Canal Engineering, Energy and Environmental Science*, 1(2):10–14, 2023.
[84] Saumya R Salian and Sudhir D Sawarkar. Melanoma skin lesion classification using improved efficientnetb3. *Jordanian Journal of Computers and Information Technology*, 8(1), 2022.
[85] Moloud Abdar, Mohammad Amin Fahami, Satarupa Chakrabarti, Abbas Khosravi, Paweł Pławiak, U. Rajendra Acharya, Ryszard Tadeusiewicz, and Saeid Nahavandi. Barf: A new direct and cross-based binary residual feature fusion with uncertainty-aware module for medical image classification. *Information Sciences*, 577:353–378, 2021.
[86] Siddharth Gupta, Avnish Panwar, and Kishor Mishra. Skin disease classification using dermoscopy images through deep feature learning models and machine learning classifiers. In *IEEE EUROCON 2021 - 19th International Conference on Smart Technologies*, pages 170–174, 2021.