RESEARCH ARTICLE

A Deep Learning Based Framework for Skin Cancer Detection Using a Layered Software Architecture with Transfer Learning

Tuba Mubarik[1,*] **and Shabib Aftab**[1]

[1] Department of Computer Science, Virtual University of Pakistan, Lahore 54000, Pakistan

Abstract

Skin cancer has become a serious public health issue worldwide. The number of cases is rising due to higher UV exposure and changing lifestyle patterns. Early detection is the best way to save lives. However, many areas still lack specialized doctors and equipment needed for a quick diagnosis. In this paper, we develop and test an automated skin cancer detection method based on standard smartphone photography. It provides a practical solution for resource-limited settings by removing the need for expensive clinical dermoscopy equipment, making diagnosis accessible in areas with limited infrastructure. The system was trained and tested on PAD-UFES-20, which contains 2,298 clinical smartphone images of skin lesions. The images were grouped into two categories: malignant and benign. Malignant cases involved Basal Cell Carcinoma, Squamous Cell Carcinoma, and Melanoma. Benign lesions consisted of Actinic Keratosis, Nevus, and Seborrheic Keratosis. Transfer learning was used to fine-tune two models: EfficientNetB3 and VGG16. Both were pre-trained on ImageNet. EfficientNetB3 achieved a test accuracy of 82.61%. It outperformed VGG16, which achieved an accuracy of 79.78%. The system follows a modular pipeline covering data ingestion, preprocessing, model training, and inference. Unlike many existing approaches that require high-end hardware, our approach provides high accuracy using standard technology. This design allows easy adaptation in clinical settings with limited infrastructure. The proposed four-layer modular pipeline embodies software engineering principles of separation of concerns and component independence, enabling maintainable and scalable deployment in resource-constrained environments.

Keywords: skin cancer, binary classification, transfer learning, EfficientNetB3, VGG16, PAD-UFES-20, deep learning, medical image analysis, modular pipeline architecture, layered system design.

1 Introduction

Skin cancer is one of the most diagnosed cancers worldwide. It poses a significant public health challenge due to its increasing incidence. The

Submitted: 22 June 2026
Revised: 29 July 2026
Accepted: 04 August 2026
Published: 19 August 2026

Vol. 2, **No.** 3, 2026.
10.62762/JSE.2026.541211

***Corresponding author:**
✉ Tuba Mubarik
mubariktuba.1205@gmail.com

Citation
Mubarik, T., & Aftab, S. (2026). A Deep Learning Based Framework for Skin Cancer Detection Using a Layered Software Architecture with Transfer Learning. *ICCK Journal of Software Engineering*, 2(3), 185–196.

disease develops when abnormal skin cells grow uncontrollably. It is often caused by prolonged exposure to ultraviolet (UV) radiation. Early detection is essential because timely treatment can significantly improve patient outcomes. According to global cancer statistics, cancer incidence continues to rise worldwide, emphasizing the need for effective screening and diagnostic systems [1]. Skin cancer manifests in multiple types, such as Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC), Melanoma (MEL), Actinic Keratosis (ACK), Nevus (NEV), and Seborrheic Keratosis (SEK), as described in standard clinical dermatology references [2]. When detected early, most of these conditions are treatable. However, melanoma is particularly aggressive. It is difficult to manage once it progresses beyond the primary lesion stage.

Dermatologists usually look at skin problems directly to diagnose them. They often use a dermatoscope to obtain a magnified view of skin lesions. A dermatoscope is a handheld imaging device that magnifies the skin surface and reveals structures not visible to the naked eye [3]. They often follow standardized visual assessment criteria for early melanoma detection. The ABCDE rule (Asymmetry, Border irregularity, Color variation, Diameter, and Evolution) is among the most widely applied [4, 5]. Additional structured approaches include the seven-point checklist [6] and the seven-feature dermoscopic algorithm [7], all of which guide dermatologists in identifying suspicious skin lesions. The classification follows the standard clinical categories defined in the World Health Organization (WHO) Classification of Skin Tumours, where lesions are broadly divided into benign and malignant categories for initial screening purposes. These traditional methods depend heavily on personal experience. Consequently, diagnostic accuracy can vary from person to person. There are not enough skin specialists in many rural or low-income areas. The state of non-invasive imaging for cutaneous melanoma remains limited outside specialized centers [8], and existing computer-aided diagnostic systems have highlighted the persistent gap between available technology and clinical accessibility [9]. As a result, patients frequently wait until the disease is quite advanced before seeking assistance.

The PAD-UFES-20 dataset is a publicly available skin lesion dataset developed by the Federal University of Espírito Santo, Brazil. It contains clinical skin lesion images captured using smartphone cameras. Unlike dermoscopic datasets collected under controlled clinical conditions, PAD-UFES-20 better reflects real-world scenarios [10]. These images can be captured by patients or healthcare workers without specialized equipment. The widespread availability of smartphone cameras and recent advances in deep learning have enabled the development of accessible diagnostic tools [11]. These tools are practical and suitable for real-world deployment. These tools do not require expensive imaging equipment. They also do not require immediate access to experts. Convolutional neural networks (CNN) have shown promising results in dermatology. They are frequently coupled with transfer learning techniques. They have demonstrated excellent classification performance [12–15]. Transfer learning involves a pre-trained model structure. It uses an extremely large generic dataset for initialization. The model is then fine-tuned on task-specific data. This approach is highly appropriate for medical imaging applications. It works well because annotated data is often limited [16, 17]. The effectiveness of transfer learning has also been demonstrated across a wide range of skin lesion analysis studies. For example, lightweight federated transfer learning models have been proposed for skin cancer detection while minimizing data and computational requirements [18]. Attention-based methods combined with multi-dataset fusion have further improved skin disease classification [19]. Deep learning techniques have also been applied to the automated diagnosis of skin lesions captured in smartphone images [20], and hybrid CNN-transformer architectures with feature fusion have achieved strong performance in skin cancer classification [21]. These studies highlight the adaptability of deep learning models across different skin imaging conditions and support their application in automated skin cancer diagnosis.

High-resolution dermoscopic images are used in most common benchmark datasets. These images are captured strictly in clinical settings. Prominent examples include the ISIC [22, 23] and HAM10000 [24] datasets. However, these datasets do not reflect real-world smartphone imaging conditions used in clinical practice. This includes images taken by patients or community health workers. The PAD-UFES-20 dataset addresses this specific limitation [10]. It consists of 2,298 clinical smartphone images. The Dermatological and Surgical Assistance Program at the Federal University of Espírito Santo, Brazil collected this dataset. Each image includes

valuable clinical metadata about the patient. However, this study uses only the images, not the metadata.

Despite the success of deep learning approaches, most existing studies rely on dermoscopic images, multimodal inputs, or patient metadata. However, limited research has focused on image-only binary classification using the PAD-UFES-20 dataset. This gap motivates the development of a lightweight and deployable diagnostic framework.

In this paper, we outline a binary classification framework. This framework is applied directly to the PAD-UFES-20 dataset. It is implemented using two fine-tuned architectures. These pre-trained models are EfficientNetB3 and VGG16. A benign/malignant binary problem is created from six-class data. Binary classification was chosen for two reasons. First, clinical decision-making often requires only distinguishing between benign and malignant lesions. Second, binary models are simpler and easier to use when data is limited. A class-weighted training method is deployed for this. We also use a triple-fold data augmentation technique. One key contribution is demonstrating effective binary classification using only smartphone images. Pacheco and Krohling [16] used images with patient metadata and got an accuracy of 0.764. Khan et al. [25] used EfficientNetB3 with metadata and got an accuracy of 0.78. This study uses only clinical images without patient metadata, while achieving a higher accuracy of 82.61%. These results demonstrate that competitive performance can be achieved without relying on patient metadata. This is useful because metadata is often not available in low resource areas. The trained EfficientNetB3 model delivers clinically relevant results. The framework follows a modular layered architecture that improves maintainability, scalability, and deployment flexibility. This makes the system easier to maintain and update for deployment in resource-limited clinical environments.

This paper has four main objectives. First, to test if transfer learning works well on smartphone images without metadata. Second, to check if grouping six lesion types into two classes gives good results. Third, to compare EfficientNetB3 and VGG16 under the same conditions. Fourth, to demonstrate that a modular pipeline design can support deployment in low-resource clinical settings.

2 Related Work

Research on automated skin cancer diagnosis has grown significantly. Over the past decade, deep learning has driven this growth. The growing interest in automated melanoma recognition through neural networks has been well documented in recent literature [26]. A landmark study by Esteva et al. [13] demonstrated a powerful deep CNN. It classified keratinocyte carcinomas and seborrheic keratoses highly accurately. The model performed at a level comparable to 21 board-certified dermatologists. This work used the ISIC and Dermofit datasets. It established the potential of AI tools for clinical dermatology.

Haenssle et al. [14] extended this line of work. They compared a Google Inception V4 model against 58 dermatologists. The study used 100 patient cases involving dermoscopic images. Their results showed that deep learning models can compete with clinical specialists. This was further supported by Brinker et al. [15]. Their CNN was trained on dermoscopic images. It outperformed 145 dermatologists in classifying clinical melanoma cases. This finding was further reinforced by Pham et al. [27], whose optimized deep CNN architecture with a custom loss function outperformed every participating dermatologist in dermoscopic melanoma diagnosis.

Pacheco and Krohling introduced a diagnostic tool [16] that uses smartphone images and patient metadata to classify six types of skin cancer, while simultaneously making the PAD-UFES-20 dataset publicly available [10]. Their combined image-metadata model achieved an accuracy of 0.764. Most benchmark datasets use dermoscopic images taken in controlled environments. In contrast, PAD-UFES-20 uses standard smartphone cameras. This makes it closer to real-world, low-resource conditions.

Several studies achieved high performance on public datasets. Kadampur and Al Riyaee [17] implemented a cloud-based deep learning architecture on the HAM10000 dataset. Jinnai et al. [28] developed a skin cancer classification system for pigmented lesions using deep learning on dermoscopic images, further demonstrating the effectiveness of CNN-based approaches across different imaging conditions. Wei et al. [29] employed pre-trained MobileNet and DenseNet models on the ISIC 2016 dataset to extract robust features. Similarly, Kassem et al. [30] utilized a modified GoogLeNet architecture with transfer learning to classify eight distinct skin cancer categories from the ISIC 2019 dataset. Anjum et al. [31] further combined deep semantic segmentation

with multi-class lesion classification using CNNs, achieving robust performance on dermoscopic image datasets. Goceri [32] demonstrated that CNN-based architectures can classify dermatological conditions from clinical images with high accuracy, further supporting their suitability for automated skin lesion analysis tasks.

Gessert et al. [33] built an ensemble of EfficientNet models [34] using multiple input resolutions combined with patient metadata. Similarly, Karki et al. [35] utilized a multi-resolution EfficientNet ensemble to improve melanoma detection. Khan et al. [25] combined EfficientNetB3 with a gradient boosting classifier [36] and integrated skin lesion images with clinical metadata, achieving 0.78 accuracy on the PAD-UFES-20 dataset. In contrast, the proposed study achieves higher accuracy using EfficientNetB3 alone with image data only, without relying on metadata or auxiliary classifiers.

Several advanced neural network approaches have been explored for automated diagnosis. Yin et al. [37] investigated clinical metadata for skin tumor classification. They combined this metadata with deep CNNs in their study. Another research group developed an attention-based multimodal framework [38]. This framework improves classification performance and enhances model interpretability. Azeem et al. [39] contributed SkinLesNet to this field. SkinLesNet is a multi-layer deep CNN designed for robust melanoma detection. Finally, Uliana and Krohling [40] applied generative diffusion models to skin cancer classification using the PAD-UFES-20 dataset, providing a novel generative baseline for comparison on the same benchmark used in this study.

Many researchers focus specifically on automated skin analysis. Oztel et al. [41] explored skin disease classification using smartphones. This method targets mobile health monitoring through standard device cameras. Jain et al. [42] presented a transfer learning approach for skin cancer classification. They performed a comprehensive comparison of six pre-trained networks. This comparison helped find the best approach for lesion classification. Their study also applied data replication to manage dataset imbalances. These systems highlight the value of computer-aided tools for early skin cancer screening.

Despite recent advancements, class imbalance remains a significant challenge in real-world smartphone datasets. In the PAD-UFES-20 dataset, Melanoma contains only 52 images, while Basal Cell Carcinoma (BCC) contains 845 images. This extreme class imbalance can cause standard deep learning models to favor majority classes. To address this issue, various techniques like weighted loss functions, data augmentation, and specialized sampling are required.

To overcome these data constraints, this study proposes a streamlined, end-to-end deep learning framework. Unlike existing multimodal systems, our approach relies strictly on image data. This eliminates the need for clinical metadata collection in low-resource environments. By integrating targeted data augmentation with class-weighted optimization, our model successfully handles minority classes. The proposed image-only architecture achieved 82.61% accuracy on the PAD-UFES-20 dataset, suitable for mobile-based skin cancer screening.

3 Material and Methods

This section describes the dataset, preprocessing pipeline, binary labeling scheme, data augmentation strategy, model architectures, and training configuration. The overall framework is divided into two stages: a training stage and a testing stage.

3.1 Dataset Description

The PAD-UFES-20 dataset was built under the Dermatological and Surgical Assistance Program of Federal University of Espírito Santo, Brazil [10]. It consists of 1,373 patient records and a total of 2,298 clinical skin lesion images captured with standard smartphone cameras. The six categories in which these diagnoses are classified are Actinic Keratosis (ACK), Basal Cell Carcinoma (BCC), Melanoma (MEL), Nevus (NEV), Squamous Cell Carcinoma (SCC), and Seborrheic Keratosis (SEK). Besides images, each patient record contains 26 clinical variables (although this paper only uses images from the records). There is a considerable class imbalance. BCC is the class with more images (845), while ACK is the second class (730). On the opposite end, the class MEL represents only 52 images. NEV, SCC, and SEK have 244, 192, and 235 images respectively. This imbalance is addressed explicitly during training. PAD-UFES-20 is a publicly available dataset. It was collected under ethical approval by the Federal University of Espirito Santo, Brazil. Patient consent was obtained by the original dataset authors. This study only used the publicly released images. No new patient data was collected.

3.2 Binary Label Assignment

The six diagnostic categories were mapped into two groups. The malignant group (label 1) includes BCC, SCC, and MEL, which are malignant conditions requiring immediate treatment. The benign group (label 0) includes ACK, NEV, and SEK, which are non-cancerous and generally do not require urgent intervention. This binary mapping reflects standard clinical practice for assessing lesion severity.

3.3 Preprocessing

All images were resized to 224×224×3 pixels to meet the input requirements of EfficientNetB3 and VGG16. Pixel values were normalized by a factor of 1/255, mapping the range to [0,1]. The dataset was split into 80% training and a 20% testing set. Of the 80% training data, 20% was further allocated as the validation set, resulting in an 80:20 training-validation split. This resulted in 1,470 training images, 368 validation images, and 460 testing images. This split ratio is commonly used in deep learning studies. It gives enough data for training while keeping an unbiased test set. The split was stratified by class to ensure each set had a similar proportion of benign and malignant images, which was important due to the dataset's class imbalance. K-fold cross-validation was not used in this study because the dataset was small and training time was a concern. Additionally, data augmentation tripled the training set size, which reduced the need for cross-validation. A fixed stratified split was therefore considered sufficient for reliable evaluation.

3.4 Class Weight and Data Augmentation

The original six diagnostic categories exhibited substantial class imbalance; for example, Melanoma comprised only 52 images whereas Basal Cell Carcinoma contained 845. To address this, class weights were computed using the scikit-learn library and incorporated into the binary cross-entropy loss function during training. Because the six categories were consolidated into two binary groups, the resulting class-level imbalance was considerably reduced, yielding weights close to unity: the benign class received a weight of 0.95 and the malignant class a weight of 1.06. To further mitigate the impact of within-group imbalance at the original six-class level, data augmentation was applied exclusively to the training set, ensuring that minority classes such as Melanoma contributed more effectively to model learning. Augmentation was implemented using the Keras `ImageDataGenerator` with the following transformations:

- Rotation: up to 30°
- Horizontal and vertical flipping
- Width and height shift: $\pm$10%
- Shear: $\pm$10%
- Zoom: $\pm$10%

These transformations were carefully selected to preserve the diagnostic characteristics of the lesion images while sufficiently increasing the diversity of the training samples.

3.5 System Architecture

This framework is structured as a four-layer software pipeline.

1. Data Ingestion Layer: It loads clinical image records.
2. Preprocessing Layer: It is responsible for resizing, normalization, and augmentation.
3. Model Layer: This layer is responsible for fine-tuning the transfer learning models.
4. Inference Layer: The layer that produces the final binary classification output.

Each layer functions independently because of this modular design. Individual components can be changed or replaced without affecting the system as a whole. During development, the preprocessing layer was tested separately before connecting to the model layer. The inference layer was also replaced once without disrupting other components. The layered architecture does not directly impact classification accuracy; instead, it improves maintainability, modularity, and ease of future development. The classification performance is primarily determined by the EfficientNetB3 and VGG16 models and their training configurations. This was shown when the inference layer was changed without touching the other layers. EfficientNetB3 was selected for this system. It offers high computational efficiency. It is ideal for resource-constrained deployment environments.

The Data Ingestion Layer takes raw smartphone images as input. It loads and organizes image files with their binary labels. The Preprocessing Layer receives these raw images. It resizes, normalizes, and augments them. The output is a clean, ready-to-train image tensor. The Model Layer takes this tensor as input. It passes it through EfficientNetB3 or VGG16. The output

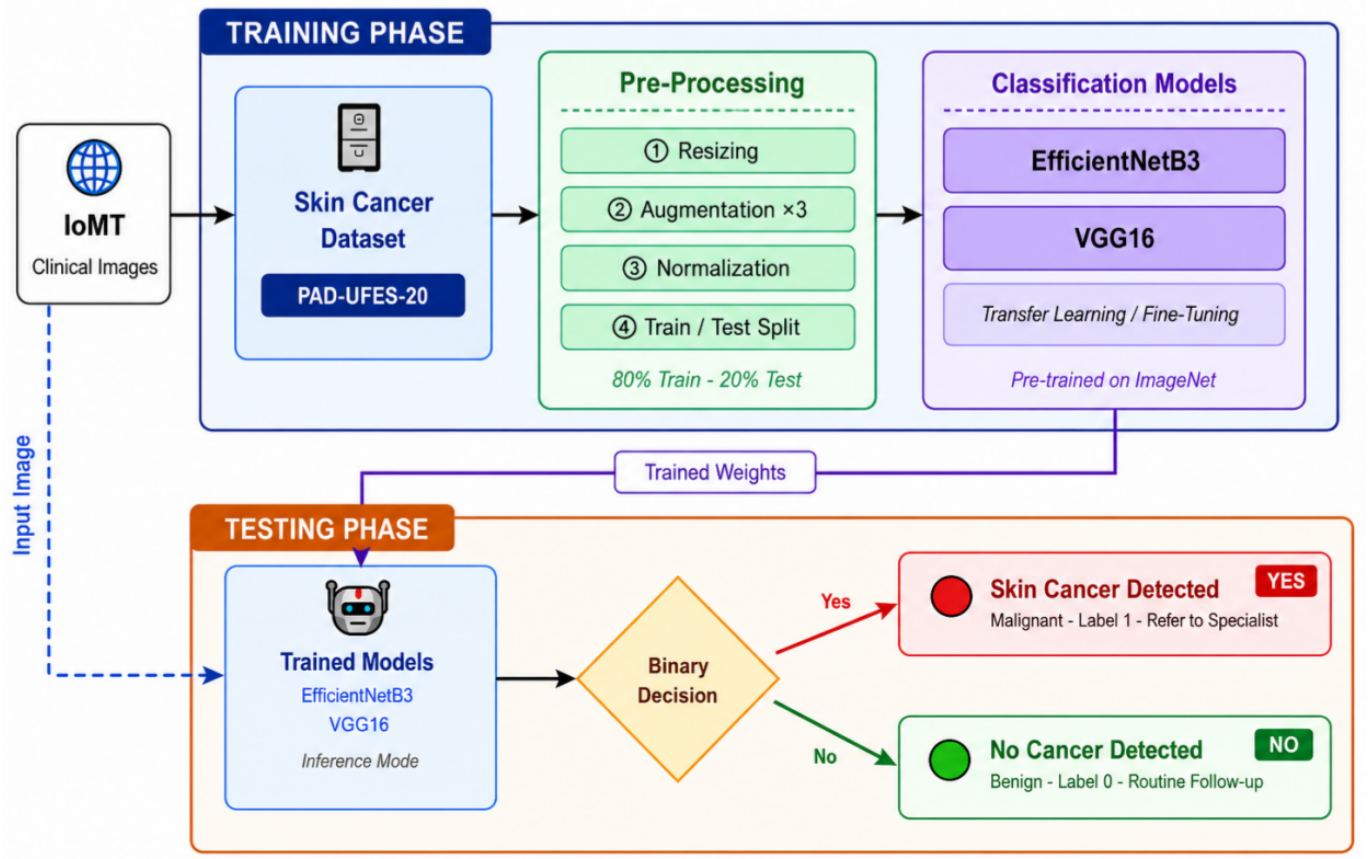

Figure 1. Proposed binary skin cancer detection framework using EfficientNetB3 and VGG16 trained and evaluated on the PAD-UFES-20 dataset.

is a probability score between 0 and 1. The Inference Layer receives this score. It applies a 0.5 threshold to produce the final binary label. A score above 0.5 means malignant. Below 0.5 means benign. Figure 1 illustrates the overall architecture of the proposed framework, showing the data flow from input images to the final binary classification output.

3.6 Model Architecture

Two fine-tuned pre-trained models were used: EfficientNetB3 and VGG16. These two models were selected for specific reasons. EfficientNetB3 is known for high accuracy with fewer parameters. It is suitable for small medical datasets. VGG16 is a well-established baseline model. It is widely used in skin lesion classification studies. Using both allows a direct comparison between a lightweight modern model and a classic deep architecture. Both models were initialized using ImageNet pre-trained weights. The EfficientNet family has demonstrated strong transfer learning performance, partly attributed to advances in self-training strategies such as the Noisy Student method [44], which significantly enhanced ImageNet feature representations. We fine-tuned EfficientNetB3 and VGG16 for binary classification. EfficientNetB3 scales network depth, width, and resolution uniformly [34]. Only its first 100 layers were kept trainable. In contrast, VGG16 consists of 13 convolutional layers in 5 blocks [45]. All VGG16 layers were fine-tuned to adapt the model to the target domain. Both models used the same custom classification head. This head includes Flatten and BatchNormalization layers. Next were Dense layers with 128 and 64 units using ReLU. Dropout rates of 0.3 and 0.2 were applied between them. Finally, a single Dense unit with Sigmoid activation performed binary classification.

3.7 Training Configuration

Both models were trained using the Adam optimizer. The initial learning rate was set to $5x10^{-5}$. We used the binary cross-entropy loss function. Training was limited to a maximum of 50 epochs, with early stopping controlling the actual training length. A fixed batch size of 32 was used, as it is standard for medical image classification and balances memory usage with training stability. Three callbacks were run simultaneously during training. This helped

improve generalization and avoid overfitting. First, the ModelCheckpoint callback was used. It saved only the best weights based on validation accuracy. Secondly, an EarlyStopping callback was implemented. It monitored validation loss with a patience of 15 epochs. This ended training if no improvement was observed. Finally, a ReduceLROnPlateau callback was added. It monitored validation loss during training. It reduced the learning rate by a factor of 0.5. This triggered when validation loss was stationary for 4 epochs. The minimum learning rate limit was $1x10^{-6}$. A custom weighted F1-score metric was used. It was tracked carefully during model evaluation. Standard validation metrics were also monitored. This specific tracking was necessary due to unbalanced classes.

4 Result and Discussions

The PAD-UFES-20 skin lesion dataset served as the basis for the training and testing of both VGG16 and EfficientNetB3. The independent test set (460 images) and the augmented training set (4,410 images) were used to evaluate performance. The following performance measures were used to evaluate the proposed models. The performance metrics were computed using the standard confusion matrix notation, where TP (True Positives), TN (True Negatives), FP (False Positives), and FN (False Negatives) denote the counts of correctly and incorrectly classified samples for the positive and negative classes, respectively.

$$\text{Misclassification Rate} = \frac{FP + FN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$\text{Positive Predictive Value} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Negative Predictive Value} = \frac{TN}{TN + FN} \quad (4)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{False Positive Ratio} = 1 - \text{Specificity} \quad (7)$$

$$\text{False Negative Ratio} = 1 - \text{Sensitivity} \quad (8)$$

$$\text{Likelihood Ratio Positive} = \frac{\text{Sensitivity}}{1 - \text{Specificity}} \quad (9)$$

$$\text{Likelihood Ratio Negative} = \frac{1 - \text{Sensitivity}}{\text{Specificity}} \quad (10)$$

With a total of 4410 images, EfficientNetB3 correctly predicted 2317 images as negative and 2089 images as positive during training. The comparison between predicted and actual results is shown in Table 1. EfficientNetB3 achieved a training accuracy of 99.91%.

Table 1. Confusion matrix of training using EfficientNetB3.

Samples N=4410		Predicted	
		Negative-0	Positive-1
INPUT	Actual Negative (Benign-0)	TN = 2317	FP = 2
	Actual Positive (Malignant-1)	FN = 2	TP = 2089

Similarly, VGG16 model was also trained using these 4410 images. It predicted 2314 images as negative and 2075 images as positive during the training procedure. Table 2 explains these results in detail. VGG16 achieved a training accuracy of 99.52%.

Table 2. Confusion matrix of training using VGG16.

Samples N=4410		Predicted	
		Negative-0	Positive-1
INPUT	Actual Negative (Benign-0)	TN = 2314	FP = 5
	Actual Positive (Malignant-1)	FN = 16	TP = 2075

After the models were trained, they were used to evaluate the test set. The independent test set consisted of 460 images. During testing, EfficientNetB3 correctly predicted 197 negative and 183 images as positive. Table 3 provides a comparison of actual and predicted results of the model. EfficientNetB3 achieved a test accuracy of 82.61%.

When the VGG16 model was used to evaluate the test set, it correctly predicted 204 images as negative and 163 as positive. Table 4 reflects the details of the results. VGG16 achieved a testing accuracy of 79.78%.

Table 3. Confusion matrix of testing using EfficientNetB3.

Samples N=460		Predicted	
		Negative-0	Positive-1
INPUT	Actual Negative (Benign-0)	TN = 197	FP = 45
	Actual Positive (Malignant-1)	FN = 35	TP = 183

Table 4. Confusion matrix of testing using VGG16.

Samples N=460		Predicted	
		Negative-0	Positive-1
INPUT	Actual Negative (Benign-0)	TN = 204	FP = 38
	Actual Positive (Malignant-1)	FN = 55	TP = 163

Table 5. Performance metrics of transfer learning models.

Metrics	EfficientNetB3		VGG16	
	Training	Testing	Training	Testing
Accuracy	0.9991	0.8261	0.9952	0.7978
Misclassification Rate	0.0009	0.1739	0.0048	0.2022
Specificity	0.9991	0.8140	0.9978	0.8430
Sensitivity	0.9990	0.8394	0.9923	0.7477
Positive Predictive Value	0.9990	0.8026	0.9976	0.8109
Negative Predictive Value	0.9991	0.8491	0.9931	0.7876
False Positive Ratio	0.0009	0.1860	0.0022	0.1570
False Negative Ratio	0.0010	0.1606	0.0077	0.2523
Likelihood Ratio Positive	1158.3910	4.5129	460.2511	4.7625
Likelihood Ratio Negative	0.0010	0.1973	0.0077	0.2993

Detailed results of performance metrics of both models are given in Table 5. The results indicate that the EfficientNetB3 model showed better performance than VGG16 in predicting and classifying skin cancer. The training accuracy of EfficientNetB3 was 99.91%. However, the test accuracy was 82.61%. This gap of approximately 17.3 percentage points indicates substantial overfitting, which is a known challenge when fine-tuning large pretrained models on small medical datasets such as PAD-UFES-20. VGG16 also showed a similar pattern. Its training accuracy was 99.52% but test accuracy dropped to 79.78%, representing a gap of approximately 19.7 percentage points, which is even more pronounced than that observed for EfficientNetB3. This is common when the dataset is small. The PAD-UFES-20 dataset has only 2,298 images. Data augmentation was used to reduce this effect. Early stopping and dropout layers were also applied. These steps helped but could not fully close the gap. Other methods could help close this gap further. Mixup can blend two images together during training. Cutout can hide random parts of an image during training. Stronger weight decay can also be used. These methods were not used in this study but can be tried in future work. It should be noted that all results reported in this study are based on a single experimental run with a fixed random seed. Future work should report results averaged over multiple runs with confidence intervals to better assess model stability. The AUC-ROC value for EfficientNetB3 was 0.90. The AUC-ROC value for VGG16 was 0.86. Both models can tell benign and malignant cases apart well. Figure 2 shows the ROC curve for EfficientNetB3. Figure 3 shows the ROC curve for VGG16.

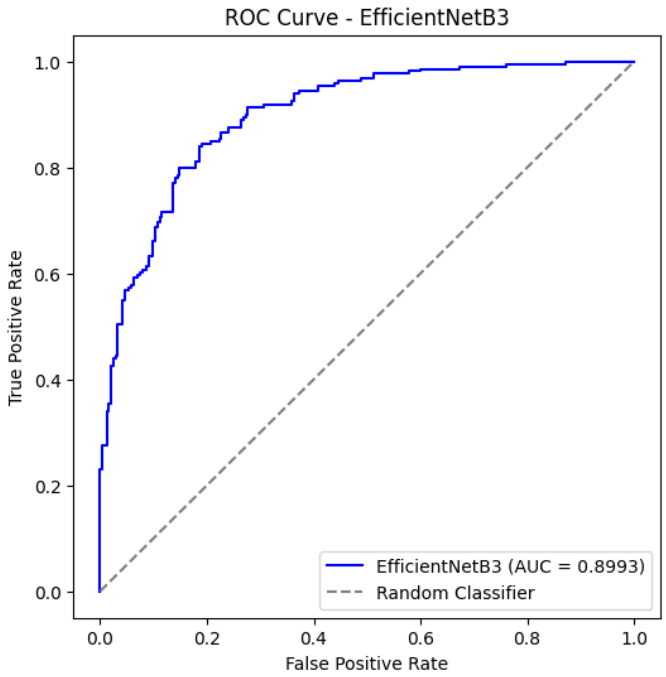

Figure 2. ROC curve graph for EfficientNetB3 Model.

EfficientNetB3 outperformed VGG16 on this dataset for two primary reasons. First, EfficientNetB3's compound scaling method balances depth, width, and resolution, making it more parameter-efficient and better suited for small medical datasets. Second, we froze the first 100 layers of EfficientNetB3 to preserve useful ImageNet features and fine-tuned only the later layers. In contrast, all VGG16 layers were trainable, which increased the risk of overfitting on the limited training data. These differences explain the 2.83% accuracy gap between the two models.

These results have practical meaning in clinical settings. A sensitivity of 83.94% means the model correctly identified most malignant cases. This is useful for early screening. The model can be used by community health workers. They can take a smartphone photo of a skin lesion. The model gives an instant result. If the result is malignant, the patient is referred to a specialist. This reduces delay in diagnosis. It is especially helpful

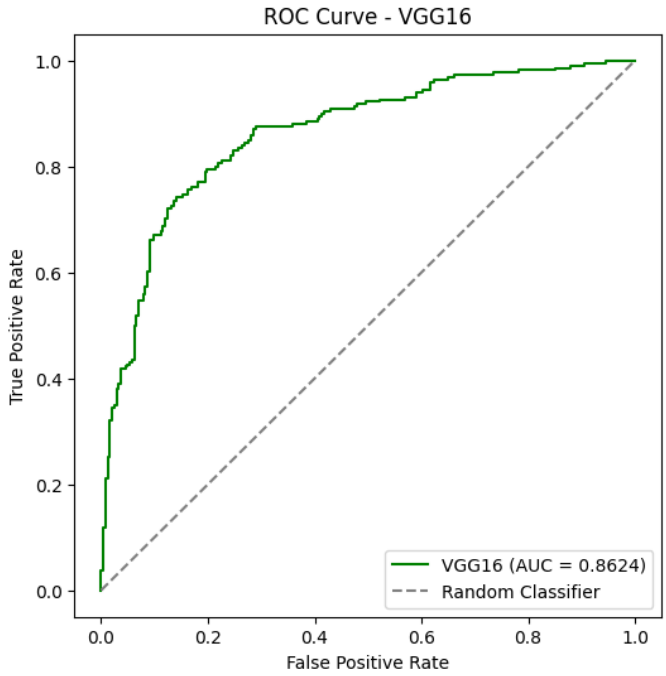

Figure 3. ROC curve graph for VGG16 Model.

in rural areas where dermatologists are not available. On a test machine equipped with an NVIDIA GTX 1080 Ti GPU, the EfficientNetB3 model (236.93 MB) required 5.23 ms per image, while VGG16 (206.14 MB) required 4.25 ms per image. Both models are sufficiently lightweight for deployment on standard clinical workstations. Both models are fast enough to run on a smartphone or a small clinic computer.

A false negative means a malignant lesion is missed. This is more dangerous. The patient does not get treatment on time. Cancer can progress to a later stage. A false positive means a benign lesion is wrongly flagged as malignant. This causes unnecessary hospital visits and patient anxiety. However, false positives are less harmful than false negatives in cancer screening. EfficientNetB3 had a false negative ratio of 16.06%. VGG16 had a higher false negative ratio of 25.23%. This suggests that EfficientNetB3 has greater potential for clinical decision support due to its lower false-negative rate. Missing fewer malignant cases is the priority in cancer detection. But both types of errors have clinical cost.

Table 6 compares the proposed model with other published methods on the same dataset. The proposed EfficientNetB3 model reached 82.61% accuracy, which is higher than all other methods listed. Pacheco and Krohling combined images with patient metadata and reached 76.4%. Yin et al. [37] also used metadata along with images and reached 81.4%. Both of these

Table 6. Comparative performance with benchmark methodologies on PAD-UFES-20 dataset.

Reference	Methodologies	Accuracy
[43]	MobileNet + MetaBlock Attention Mechanism	74.2%
[16]	Combined Clinical Fusion Model (Images + Fused Patient Metadata)	76.4%
[40]	DiffMIC (Diffusion-Based Baseline Architecture)	64.5%
[39]	SkinLesNet (ResNet-50 Pipeline Variant)	76.4%
[37]	MD-Net (DenseNet-169 + Clinical Metadata Fusion)	81.4%
[25]	EfficientNetB3 + Gradient Boosting (Images + Clinical Metadata)	78.0%
[41]	MobileNetV2 (Smartphone Optimized Network)	79.2%
[42]	Standard Transfer Learning Ensemble (Xception/ResNet)	79.5%
Proposed	EfficientNetB3 (Proposed Optimized Network)	**82.61%**

methods needed extra patient information, which is not always available. The proposed model reached a higher accuracy using only images. This shows that a well-tuned image-only model can match or beat models that use extra metadata. Oztel et al. [41] used MobileNetV2 for smartphone images and reached 79.2%. Jain et al. [42] used a transfer learning ensemble and reached 79.5%. The proposed model performed better than both. This may be because EfficientNetB3 balances depth, width, and image resolution together, which helps it learn better from a small dataset like PAD-UFES-20.

This suggests that with appropriate architectural choices and training strategies (including targeted data augmentation and class-weighted loss), an image-only model can achieve performance comparable to or better than methods that rely on additional patient metadata. The advantage of our approach is its applicability in resource-limited settings where clinical metadata is often unavailable.

5 Conclusion

This paper presents a binary classification framework for skin cancer detection. The framework uses smartphone images from the PAD-UFES-20 dataset. Two transfer learning models were used: EfficientNetB3 and VGG16. EfficientNetB3 achieved a test accuracy of 82.61%. It outperformed VGG16, which scored 79.78%. The AUC score of 0.90 also confirms that EfficientNetB3 gives a strong and reliable separation between benign and malignant cases. This study makes three primary contributions.

First, strong results were achieved using only images, without any metadata. Second, a four-layer modular software pipeline was designed following software engineering principles of modularity and separation of concerns, enabling independent testing, replacement, and scaling of each component without affecting the system as a whole. Third, EfficientNetB3 outperformed VGG16 on this dataset. Despite these encouraging results, several limitations should be acknowledged. The dataset contains only 2,298 images. This is a small number for deep learning. Melanoma had only 52 images, causing class imbalance. This can affect how well the model learns rare cases, even after using class weights and augmentation. The model was tested on one dataset only. This means its performance on images from other sources or cameras has not yet been evaluated. No patient metadata was used, which could have helped improve accuracy. Future work can further improve this framework in several ways. The model can be tested on additional datasets, such as ISIC. This will help evaluate its performance on different skin lesion images. Patient metadata, such as age, may also be incorporated to improve classification accuracy. Larger and more balanced datasets could further reduce the impact of class imbalance. Model compression techniques can be applied to support deployment on mobile devices. Explainable AI techniques, such as Grad-CAM, can also be incorporated. These techniques can highlight the image regions that influenced the model's prediction. This can help doctors better understand and trust the system. Future work could further assess the stability of the results by running experiments with multiple random seeds and reporting confidence intervals. The framework should also be evaluated in real-world clinical environments with dermatologists. In addition, a simple mobile application can be developed for community health workers. It can assist them in identifying suspicious skin lesions. This would be especially useful in areas where dermatology specialists are not readily available.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

Shabib Aftab served as an Editor-in-Chief of the *ICCK Journal of Software Engineering* at the time of manuscript submission. To ensure the integrity of the peer-review process, Shabib Aftab was not involved in the editorial handling, peer review, or decision-making process for this manuscript, which was handled independently by another editor. The remaining authors declare no conflicts of interest.

AI Use Statement

The authors declare that no generative AI was used in the preparation of this manuscript.

Ethical Approval and Consent to Participate

Ethical approval and informed patient consent were obtained by the original collectors of the PAD-UFES-20 dataset (Federal University of Espírito Santo, Brazil). The present study utilized only the publicly released, de-identified image subset and did not involve any new recruitment of human participants or collection of personal data; therefore, separate ethical approval was not required under standard secondary data use policies.

References

[1] Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians, 71*(3), 209-249. [CrossRef]

[2] Wolff, K., Johnson, R. A., Saavedra, A. P., & Roh, E. K. (2017). Eczema/Dermatitis. *Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology*, 25-49. Available at: https://doctorlib.org/medical/fitzpatrick-atlas-dermatology/5.html

[3] Sinz, C., Tschandl, P., Rosendahl, C., Akay, B. N., Argenziano, G., Blum, A., ... & Kittler, H. (2017). Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin. *Journal of the American Academy of Dermatology, 77*(6), 1100-1109. [CrossRef]

[4] Nachbar, F., Stolz, W., Merkle, T., Cognetta, A. B., Vogt, T., Landthaler, M., ... & Plewig, G. (1994). The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions. *Journal of the American Academy of Dermatology, 30*(4), 551-559. [CrossRef]

[5] Robinson, J. K., & Turrisi, R. (2006). Skills training to learn discrimination of ABCDE criteria by those at risk of developing melanoma. *Archives of dermatology, 142*(4), 447-452. [CrossRef]

[6] Healsmith, M. F., Bourke, J. F., Osborne, J. E., & Graham-Brown, R. A. C. (1994). An evaluation of the revised seven-point checklist for the early diagnosis

of cutaneous malignant melanoma. *British Journal of Dermatology, 130*(1), 48-50. [CrossRef]

[7] Dal Pozzo, V., Benelli, C., & Roscetti, E. (1999). The seven features for melanoma: a new dermoscopic algorithm for the diagnosis of malignant melanma. *European Journal of Dermatology, 9*(4), 303-8. https://www.jle.com/en/revues/ejd/e-docs/the_seven_features_for_melanoma_a_new_dermoscopic_algorithm_for_the_diagnosis_of_malignant_melanma_100088/article.phtml?cle_doc=000186F8

[8] Smith, L., & MacNeil, S. (2011). State of the art in non-invasive imaging of cutaneous melanoma. *Skin Research and Technology, 17*(3), 257-269. [CrossRef]

[9] Masood, A., & Al-Jumaily, A. A. (2013). Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms. *International journal of biomedical imaging, 2013*(1), 323268. [CrossRef]

[10] Pacheco, A. G., Lima, G. R., Salomao, A. S., Krohling, B., Biral, I. P., De Angelo, G. G., ... & De Barros, L. F. (2020). PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones. *Data in brief, 32*, 106221. [CrossRef]

[11] Udrea, A., Mitra, G. D., Costea, D., Noels, E. C., Wakkee, M., Siegel, D. M., ... & Nijsten, T. E. C. (2020). Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. *Journal of the European Academy of Dermatology and Venereology, 34*(3), 648-655. [CrossRef]

[12] Goyal, M., Knackstedt, T., Yan, S., & Hassanpour, S. (2020). Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Computers in biology and medicine, 127*, 104065. [CrossRef]

[13] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115-118. [CrossRef]

[14] Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., ... & Zalaudek, I. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of oncology, 29*(8), 1836-1842. [CrossRef]

[15] Brinker, T. J., Hekler, A., Enk, A. H., Klode, J., Hauschild, A., Berking, C., ... & Schrüfer, P. (2019). A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task. *European journal of cancer, 111*, 148-154. [CrossRef]

[16] Pacheco, A. G., & Krohling, R. A. (2020). The impact of patient clinical information on automated skin cancer detection. *Computers in biology and medicine, 116*, 103545. [CrossRef]

[17] Kadampur, M. A., & Al Riyaee, S. (2020). Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. *Informatics in Medicine Unlocked, 18*, 100282. [CrossRef]

[18] Khullar, V., Kaur, P., Gargrish, S., Mishra, A. M., Singh, P., Diwakar, M., ... & Gupta, I. (2025). Minimal sourced and lightweight federated transfer learning models for skin cancer detection. *Scientific reports, 15*(1), 2605. [CrossRef]

[19] Sasithradevi, A., Kanimozhi, S., Sasidhar, P., Pulipati, P. K., Sruthi, E., & Prakash, P. (2025). EffiCAT: A synergistic approach to skin disease classification through multi-dataset fusion and attention mechanisms. *Biomedical Signal Processing and Control, 100*, 107141. [CrossRef]

[20] Sakl, M., Essid, C., Salah, B. B., & Sakli, H. (2023, June). Dl methods for skin lesions automated diagnosis in smartphone images. In *2023 International Wireless Communications and Mobile Computing (IWCMC)* (pp. 1142-1147). IEEE. [CrossRef]

[21] Agarwal, S., & Mahto, A. K. (2025). Skin cancer classification: Hybrid CNN-transformer models with KAN-Based fusion. *arXiv preprint arXiv:2508.12484*. [CrossRef]

[22] Codella, N. C., Gutman, D., Celebi, M. E., Helba, B., Marchetti, M. A., Dusza, S. W., ... & Halpern, A. (2018, April). Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic). In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)* (pp. 168-172). IEEE. [CrossRef]

[23] ISIC Archive. (2019). ISIC 2019: Skin lesion analysis towards melanoma detection. Retrieved from https://challenge2019.isic-archive.com/

[24] Tschandl, P., Rosendahl, C., & Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data, 5*(1), 180161. [CrossRef]

[25] Khan, I. U., Aslam, N., Anwar, T., Aljameel, S. S., Ullah, M., Khan, R., ... & Akhtar, N. (2021). Remote diagnosis and triaging model for skin cancer using EfficientNet and extreme gradient boosting. *Complexity, 2021*(1), 5591614. [CrossRef]

[26] Efimenko, M., Ignatev, A., & Koshechkin, K. (2020). Review of medical image recognition technologies to detect melanomas using neural networks. *BMC bioinformatics, 21*(Suppl 11), 270. [CrossRef]

[27] Pham, T. C., Luong, C. M., Hoang, V. D., & Doucet, A. (2021). AI outperformed every dermatologist in dermoscopic melanoma diagnosis, using an optimized deep-CNN architecture with custom mini-batch logic and loss function. *Scientific Reports, 11*(1), 17485. [CrossRef]

[28] Jinnai, S., Yamazaki, N., Hirano, Y., Sugawara, Y., Ohe,

Y., & Hamamoto, R. (2020). The development of a skin cancer classification system for pigmented skin lesions using deep learning. *Biomolecules, 10*(8), 1123. [CrossRef]

[29] Wei, L., Ding, K., & Hu, H. (2020). Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network. *IEEE Access, 8*, 99633-99647. [CrossRef]

[30] Kassem, M. A., Hosny, K. M., & Fouad, M. M. (2020). Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning. *IEEE Access, 8*, 114822-114832. [CrossRef]

[31] Anjum, M. A., Amin, J., Sharif, M., Khan, H. U., Malik, M. S. A., & Kadry, S. (2020). Deep semantic segmentation and multi-class skin lesion classification based on convolutional neural network. *IEEE Access, 8*, 129668-129678. [CrossRef]

[32] Goceri, E. (2021). Deep learning based classification of facial dermatological disorders. *Computers in Biology and Medicine, 128*, 104118. [CrossRef]

[33] Gessert, N., Nielsen, M., Shaikh, M., Werner, R., & Schlaefer, A. (2020). Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. *MethodsX, 7*, 100864. [CrossRef]

[34] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PmLR.

[35] Karki, S., Kulkarni, P., & Stranieri, A. (2021, February). Melanoma classification using EfficientNets and Ensemble of models with different input resolution. In *Proceedings of the 2021 Australasian Computer Science Week Multiconference* (pp. 1-5). [CrossRef]

[36] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). [CrossRef]

[37] Yin, W., Huang, J., Chen, J., & Ji, Y. (2022). A study on skin tumor classification based on dense convolutional networks with fused metadata. *Frontiers in Oncology, 12*, 989894. [CrossRef]

[38] O'Reilly, C. (2024). Enhancing dermatological skin lesion classification with multi-modal attention-based models and explainability [Doctoral dissertation, Technological University Dublin]. Arrow@TU Dublin. [CrossRef]

[39] Azeem, M., Kiani, K., Mansouri, T., & Topping, N. (2023). SkinLesNet: classification of skin lesions and detection of melanoma cancer using a novel multi-layer deep convolutional neural network. *Cancers, 16*(1), 108. [CrossRef]

[40] Uliana, J. J., & Krohling, R. A. (2025). Diffusion models applied to skin and oral cancer classification. *arXiv preprint arXiv:2504.00026*. [CrossRef]

[41] Oztel, I., Yolcu Oztel, G., & Sahin, V. H. (2023). Deep learning-based skin diseases classification using smartphones. *Advanced Intelligent Systems, 5*(12), 2300211. [CrossRef]

[42] Jain, S., Singhania, U., Tripathy, B., Nasr, E. A., Aboudaif, M. K., & Kamrani, A. K. (2021). Deep learning-based transfer learning for classification of skin cancer. *Sensors, 21*(23), 8142. [CrossRef]

[43] Pacheco, A. G., & Krohling, R. A. (2021). An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification. *IEEE journal of biomedical and health informatics, 25*(9), 3554-3563. [CrossRef]

[44] Xie, Q., Luong, M. T., Hovy, E., & Le, Q. V. (2020, June). Self-training with noisy student improves imagenet classification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 10684-10695). IEEE. [CrossRef]

[45] Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In *Proceedings of the International Conference on Learning Representations (ICLR)*. https://ora.ox.ac.uk/objects/uuid:60713f18-a6d1-4d97-8f45-b60ad8aebbce

Tuba Mubarik is currently pursuing her M.S. degree in Computer Science from the Virtual University of Pakistan, Lahore. Her research interests include Machine Learning, Deep Learning, and Medical Image Analysis. (Email: mubariktuba.1205@gmail.com)

Shabib Aftab completed his Ph.D. in Computer Science from National College of Business Administration and Economics, Lahore. He previously received the M.S. degree in Computer Science from Comsats University, Islamabad, the M.Sc. degree in Information Technology from the Punjab University College of Information Technology, Lahore, and the Bachelor's degree from Government College University, Lahore. He is a dedicated academician and accomplished researcher with over 16 years of academic and research experience. Currently, he is serving as an Assistant Professor in the Department of Computer Science at the Virtual University of Pakistan. His research interests include Software Process Improvement, Data Mining, Predictive Analytics, Applied Machine Learning, and Applied Data Science. Dr. Aftab has authored more than 70 research publications in prestigious international journals and conferences. His work focuses on developing innovative, data-driven solutions to real-world challenges across domains such as healthcare, finance, and software engineering. He is also a Senior Member of IEEE, reflecting his ongoing commitment to advancing research and contributing to the global scientific community. (Email: shabib.aftab@gmail.com)