RESEARCH ARTICLE

Skin Cancer Detection Using Data Fusion and Deep Transfer Learning Techniques

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

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

Abstract

Skin cancer remains a major healthcare challenge. Early detection is important because it increases the chances of successful treatment. Deep learning has become an effective approach for automated skin lesion classification. However, many existing studies rely on a single dataset, which limits the diversity of training images and may reduce model generalization. This study proposes a transfer learning framework for binary skin cancer classification using dataset-level fusion. Two public datasets, PAD-UFES-20 and ISIC 2016, are combined to create a fused dataset that contains both clinical smartphone images and dermoscopic images. The images are resized to 224×224 pixels, converted to RGB format, and divided into training, validation, and testing sets using stratified splitting. Data augmentation and class weighting are applied to improve model learning. Two pretrained models, ResNet50 and DenseNet121, are fine-tuned using a two-phase training strategy. Model performance is evaluated using accuracy, sensitivity, specificity, positive and negative predictive values, AUC-ROC, and confusion matrices. Experimental results show that ResNet50 achieved an accuracy of 82.19%, while DenseNet121 achieved the best accuracy of 82.98%. The findings suggest that dataset-level fusion improves training data diversity and provides reliable performance for binary skin cancer classification. The proposed framework is organized into a Training Layer and a Testing Layer, which separate model development from inference. This design provides a flexible framework for future medical image classification applications.

Keywords: Skin cancer, binary classification, dataset fusion, transfer learning, ResNet50, DenseNet121, PAD-UFES-20, ISIC 2016.

1 Introduction

Skin cancer is one of the major health concerns worldwide. It affects millions of people every year and remains a major public health concern. According to the GLOBOCAN report, the global burden of cancer continues to increase, including various types of skin cancer [1]. Skin cancer develops when abnormal skin cells grow without control. The main cause is long-term exposure to ultraviolet (UV) radiation from sunlight. The disease mainly includes basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma [2]. Melanoma is less common than the other types, but it is the most aggressive form

Submitted: 11 June 2026
Revised: 21 July 2026
Accepted: 04 August 2026
Published: 09 October 2026

Vol. 2, **No.** 4, 2026.
10.62762/JIAP.2026.243543

***Corresponding author:**
✉Shabib Aftab
shabib.aftab@gmail.com

Citation
Mubarik, T., & Aftab, S. (2026). Skin Cancer Detection Using Data Fusion and Deep Transfer Learning Techniques. *ICCK Journal of Image Analysis and Processing*, 2(4), 218–230.

© 2026 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

because it can spread to other organs if it is not detected early [3, 4]. Early diagnosis plays an important role in reducing the death rate and improving patient survival. Therefore, developing accurate and reliable computer-aided systems for early skin cancer detection has become an important research area.

Skin cancer is usually diagnosed through clinical examination, dermoscopy, and biopsy. Dermoscopy helps dermatologists examine skin lesions in more detail and improves diagnostic accuracy [5, 6]. However, these diagnostic approaches rely on structured clinical criteria, such as the seven-point checklist [7] and the ABCD rule of dermoscopy [8], whose correct application still requires considerable specialist expertise. Similar visual features between benign and malignant lesions can also make diagnosis difficult. In some cases, small changes in lesion color, size, or texture are hard to identify with the human eye. Biopsy provides reliable results, but it is an invasive procedure and requires additional time and cost [9]. These challenges may delay diagnosis and treatment. Therefore, there is a growing need for automated systems that can support dermatologists and improve the accuracy of skin cancer detection.

Recent advances in artificial intelligence (AI) have changed the field of medical image analysis [2]. AI-based systems can automatically learn important image features and support disease diagnosis with high accuracy. Among different AI techniques, deep learning has shown promising results for image classification [10, 11]. Convolutional Neural Networks (CNNs) can automatically extract useful features from medical images. This reduces the need for manual feature engineering. Transfer learning has further improved the performance of deep learning models [12]. It uses knowledge learned from large image datasets and applies it to medical imaging tasks. This approach reduces training time and performs well even when medical datasets are limited. Well-known CNN architectures, such as VGG [13] and EfficientNet [14], have been widely used as backbone models for transfer learning in various medical image classification tasks, including skin cancer diagnosis. As a result, transfer learning has become a popular choice for skin cancer classification and other healthcare applications.

The performance of a deep learning model depends on the quality and diversity of the training data. Public skin lesion datasets have played an important role in developing computer-aided diagnosis systems. Among these datasets, ISIC 2016 is one of the most widely used benchmark datasets for skin lesion classification [15]. It contains 900 dermoscopic images and has been used to evaluate many deep learning models. Another publicly available dataset, PAD-UFES-20, includes 2,298 smartphone-acquired clinical images together with patient clinical information [16]. Other public datasets, such as HAM10000 [17], have also been widely used for benchmarking skin lesion classification models. Collectively, such datasets have contributed to skin cancer research. However, each dataset has different image characteristics, acquisition conditions, and data distributions. These differences may limit the generalization ability of models trained on a single dataset. Therefore, combining information from multiple datasets has become an important direction for developing more reliable skin cancer classification systems.

Recent studies reported that combining different data sources improves feature learning and increases classification accuracy. Similar approaches have also been explored for skin lesion analysis by combining clinical information with skin images [18–20]. However, most existing studies focus on image–metadata fusion instead of combining different public skin lesion datasets. Therefore, the potential of dataset-level fusion for skin cancer classification is still not fully explored. This limitation provides the main motivation for the proposed study.

Although many studies have reported promising results for skin cancer classification, several challenges still remain. Most existing methods are developed using a single public dataset, which limits the diversity of training samples. Some studies combine clinical metadata with skin images, but very few investigate dataset-level fusion by integrating different public skin lesion datasets. This limitation may affect the robustness and generalization of deep learning models when they are tested on unseen data. To address this issue, this study proposes a dataset fusion framework by combining the PAD-UFES-20 and ISIC 2016 datasets. The fused dataset is used to train two transfer learning models, ResNet50 and DenseNet121, for binary skin cancer classification. Standard image preprocessing, data augmentation, and stratified data splitting are applied to improve model learning. The effectiveness of the proposed framework is evaluated through extensive experiments. The proposed framework provides a simple and effective approach for developing reliable computer-aided skin cancer

classification systems. The proposed framework follows a two-layer architecture, consisting of a Training Layer and a Testing Layer. The Training Layer is responsible for model development, while the Testing Layer performs skin lesion classification using the trained model. This organization provides a clear workflow for binary skin cancer classification and supports future applications in medical image analysis.

The main contributions of this study are as follows:

- This study proposes a dataset-level fusion approach by combining the PAD-UFES-20 and ISIC 2016 datasets for binary skin cancer classification. The fused dataset increases the diversity of the training data.
- Two transfer learning models, ResNet50 and DenseNet121, are trained using a two-phase fine-tuning strategy.
- The proposed framework is evaluated using accuracy, sensitivity, specificity, positive and negative predictive values, AUC-ROC, and confusion matrices. The experimental results show that DenseNet121 achieved the highest accuracy of 82.98%.

2 Related Work

Skin cancer classification has become one of the most active research areas in medical image analysis. In the past, researchers manually extracted image features and used traditional machine learning methods. This required extensive manual effort and did not give very good results. The introduction of deep learning changed this process. Convolutional Neural Networks (CNNs) automatically learn important image features and have achieved better results in skin lesion classification. Later, transfer learning became a popular approach because it reduces training time and improves performance, especially when medical datasets are small.

Esteva et al. [10] trained a deep CNN on a large set of skin lesion images and reached accuracy close to expert dermatologists. Haenssle et al. [11] compared a CNN with dermatologists for melanoma diagnosis and found the model performed just as well, sometimes better. Brinker et al. [21] tested a CNN against a large group of dermatologists and got reliable results. Kassem et al. [12] used transfer learning on the ISIC dataset and improved accuracy while training faster. More recently, de Lima and Krohling [22] explored the use of transformer-based models alongside CNNs for skin lesion diagnosis. Their research demonstrated that transformers can achieve strong performance even when trained on small datasets. These studies demonstrate that deep learning models perform effectively in skin cancer diagnosis. However, each study was trained and tested on only a single dataset. Consequently, this approach limits the ability of models to generalize to unseen data.

The PAD-UFES-20 dataset is one of the most widely used public datasets for clinical skin lesion analysis. It contains 2,298 smartphone-acquired skin lesion images collected under real clinical conditions [16]. The dataset also includes patient information such as age, gender, lesion location, and medical history. These characteristics make the dataset useful for developing practical skin cancer classification systems. For this reason, many researchers have used PAD-UFES-20 to evaluate deep learning and transfer learning models. Representative studies that fuse PAD-UFES-20 images with patient metadata [18–20] are discussed together with other fusion approaches in the data-fusion subsection below.

The ISIC 2016 dataset [15] is one of the earliest benchmark datasets developed for skin lesion classification. It contains high-quality dermoscopic images collected from different clinical centers. Because of its standardized images and publicly available annotations, the dataset has been widely used to evaluate deep learning models. Gutman et al. [15] introduced the ISIC 2016 challenge to give researchers a shared benchmark, which helped many teams compare their methods fairly. Many models have since been tested on ISIC 2016 using CNN and transfer learning techniques [10–12]; however, each model was trained on this dataset alone, without combining it with other sources. These studies show that the ISIC 2016 dataset has played an important role in the development of skin cancer classification models. Nevertheless, because most methods are trained and evaluated using only one dataset, the diversity of training data remains limited, which may reduce the generalization ability of deep learning models.

Data fusion has become an important research direction in skin cancer classification. It combines information from different sources to improve feature learning and classification performance. In skin cancer research, fusion can be performed using clinical images, dermoscopic images, patient metadata,

or multiple public datasets. By learning from complementary information, deep learning models can better capture lesion characteristics and improve diagnostic performance. As a result, many recent studies have explored different fusion strategies for automated skin cancer diagnosis.

Most skin cancer fusion studies combine lesion images with clinical metadata. Pacheco and Krohling [18] used an attention mechanism to merge images and metadata and obtained better accuracy than image-only models, and in later work [20] they confirmed that adding clinical data improves accuracy. Yin et al. [19] built a DenseNet model that fused metadata with images for tumor classification. Tang et al. [23] went further with a joint-individual fusion structure that learned image and metadata features together, which improved the results on public datasets. Ge et al. [24] combined dermoscopy and clinical images using deep saliency features. Kawahara et al. [25] and Yap et al. [26] merged clinical images with dermoscopic images into a single model. Their research showed that using more than one image type helps improve diagnosis. Nunnari et al. [27] studied how fusing raw pixels with patient metadata affects CNN-based classification and confirmed that even simple fusion strategies can improve results. Chen et al. [28] proposed MDFNet, which fused skin images with clinical data and reported strong accuracy gains. Ou et al. [29] built a similar fusion model using real smartphone-collected images and metadata, designed for low-resource clinical settings.

Recent studies have explored transformer-based fusion methods instead of simple concatenation. Cai et al. [30] built a transformer with separate encoders for images and metadata, joined by a decoder. Tang et al. [31] proposed FusionM4Net, a multi-stage fusion network for multi-label skin lesion classification. Pedro and Oliveira [32] studied how attention and self-attention mechanisms affect skin lesion classification and showed that attention helps the model focus on the most useful image regions. Zhang et al. [33] proposed TFormer, which fuses image and metadata features across the whole network instead of at just one point. Xu et al. [34] designed RemixFormer to combine imaging and non-imaging patient data for tumor diagnosis. Omeroglu et al. [35] designed a soft attention-based multimodal framework that improved multi-label classification by weighting image and metadata features differently. Vachmanus et al. [36] introduced DeepMetaForge, a vision-transformer model with a dedicated fusion module. Cheslerean-Boghiu et al. [37] added an interpretability layer so that clinicians can understand the model's decisions rather than relying on the output alone.

Several further studies have advanced dataset-level and multimodal fusion specifically for skin cancer classification. Georgiadis et al. [38] investigated the effect of combining multiple public skin cancer image datasets into hyperdatasets and showed that dataset-level fusion consistently improves classification accuracy for CNN and transfer learning models. Remya et al. [39] proposed a transfer learning framework that fused dermoscopic images with patient metadata using a Vision Transformer and channel attention mechanism, achieving 99% accuracy on skin lesion classification. Adebiyi et al. [40] conducted a systematic review of transformer-based models for skin lesion classification, showing how fusion techniques have evolved over time. More recently, Das et al. [41] compared several fusion designs for skin lesion detection and found cross-attention worked best. In contrast, Marchetti et al. [42] conducted an in-depth comparison of deep learning algorithms and dermatologists using the ISIC 2016 dermoscopic image dataset, further establishing it as a standard benchmark for skin lesion classification. Still, almost all of these studies fuse images with metadata.

Overall, the reviewed studies show that data fusion improves skin cancer classification by combining images, metadata, or multiple imaging modalities. However, most existing methods focus on image–metadata fusion rather than dataset-level fusion. Based on the reviewed literature, very limited attention has been given to combining two public skin lesion datasets into a unified training set. In particular, the fusion of PAD-UFES-20 and ISIC 2016 remains largely unexplored. This limitation provides the main motivation for the proposed framework.

3 Materials and Methods

3.1 Datasets

This study uses two publicly available skin lesion datasets: PAD-UFES-20 [16] and ISIC 2016 [15]. These datasets were selected because they contain different types of skin lesion images, which allows the model to learn from more diverse visual features. PAD-UFES-20 consists of 2,298 clinical images captured using smartphone cameras under real clinical conditions. The dataset also provides patient metadata and

diagnostic labels for each image. The diagnostic labels include six classes. Melanoma (MEL), basal cell carcinoma (BCC), and squamous cell carcinoma (SCC) are considered malignant and assigned a label of 1. Actinic keratosis (ACK), nevus (NEV), and seborrheic keratosis (SEK) are considered benign and assigned a label of 0.

ISIC 2016 is a public dataset that includes 900 high-quality dermoscopic images collected from different clinics. Every image has a ground-truth label identifying it as either benign (0) or malignant (1). Unlike PAD-UFES-20, which uses clinical smartphone photos, ISIC 2016 uses images from specialized medical devices. By combining these two datasets, we create a more diverse collection for training. This mixture adds important variations in lighting, image quality, and the overall appearance of the skin lesions.

3.2 Image Preprocessing

Images from both datasets are read and resized to 224×224 pixels. They are also converted from BGR to RGB format, as most models require RGB input. To preserve original details, we do not apply any Gaussian filtering or smoothing at this stage. Before training, all images go through several preprocessing steps to ensure consistency. Any images that fail to load or lack a matching label in the metadata are excluded. Finally, fixed random seeds are set for Python, NumPy, and TensorFlow to ensure reproducibility. Figure 1 shows sample images from the PAD-UFES-20 and ISIC 2016 datasets before and after preprocessing. The preprocessing step includes image resizing to 224×224 pixels and RGB color conversion.

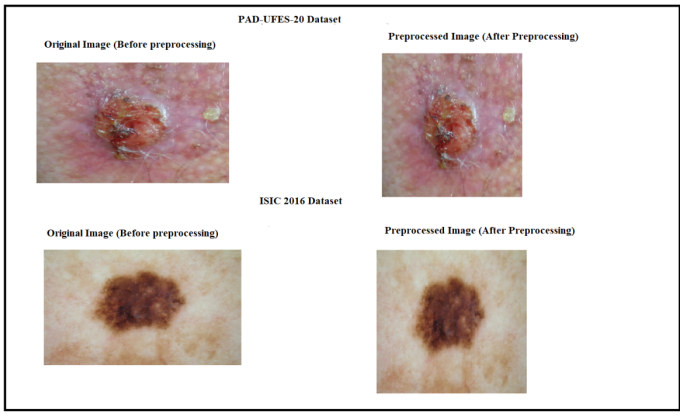


Figure 1. Sample images from the PAD-UFES-20 and ISIC 2016 datasets before and after image preprocessing.

3.3 Dataset Fusion and Splitting

After preprocessing, the images and labels from PAD-UFES-20 and ISIC 2016 are combined into a single dataset. Dataset-level fusion increases the total number of training samples. This approach also introduces greater diversity in image appearance, acquisition devices, and lighting conditions, which helps the model learn more robust and generalized features. The combined dataset is then shuffled to remove any ordering bias. It is split into training and test sets, with 70% used for training (2,238 images) and 30% for testing (960 images). The training set is further split into 80% training (1,790 images) and 20% validation (448 images). We use stratified splitting at both stages to ensure a consistent ratio of benign to malignant images across all subsets. All images are converted to floating-point format before training.

3.4 Data Augmentation and Class Weighting

Since medical image datasets are often limited in size and imbalanced between classes, data augmentation is applied to the training set only. Each training image is transformed using random rotations (up to $30°$) and width/height shifts (up to 20%). We also apply shear transformations, zooming (up to 20%), and horizontal and vertical flips. Two separate augmented copies of the training set are generated using different random seeds. These two augmented sets are then combined with the original training images to create a training set three times larger than the original. As a result, the training set grows from 1,790 to 5,370 images. This process helps the model see more variation during training, which reduces the risk of overfitting.

To address the remaining class imbalance, we implement a class weighting strategy. The class weights are defined as $\{0 : 0.826,\ 1 : 1.268\}$, where the benign class receives a weight of 0.826 and the malignant class receives a weight of 1.268. These weights are applied during training to ensure that the minority class has a fair impact on the learning process. Finally, the augmented dataset is shuffled to ensure random ordering during training.

3.5 Proposed Framework

The framework architecture consists of two layers: a Training Layer, which handles data ingestion, preprocessing, and model development; and a Testing Layer, which performs inference and delivers the final classification decision. The two layers communicate through a single, well-defined interface: the trained model weights. These weights act as the output artifact of the Training Layer and serve as the essential input for the Testing Layer. Figure 2 illustrates the complete architecture of the proposed framework.

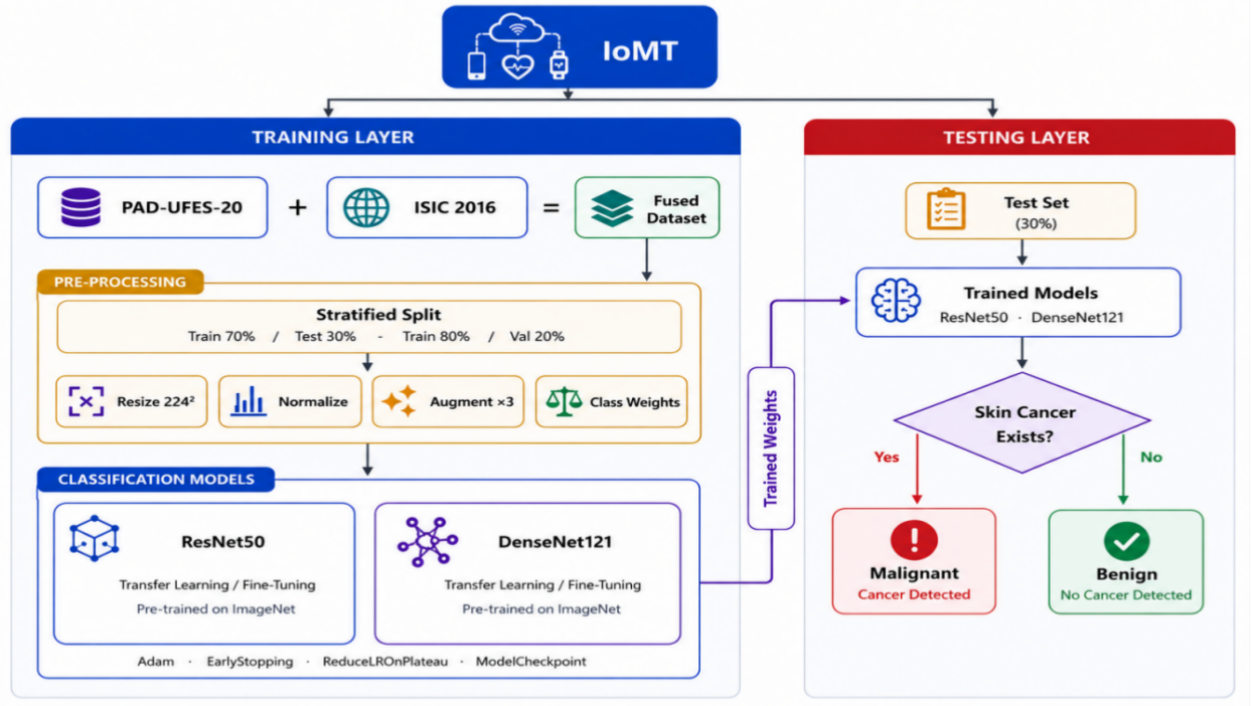


Figure 2. Proposed framework for binary skin cancer classification.

As shown in Figure 2, each layer performs a distinct set of operations. In the Training Layer, we first combine the PAD-UFES-20 and ISIC 2016 datasets into one. We then preprocess the data by resizing images to 224×224 pixels, normalizing values, augmenting the training set, and calculating class weights. We train two models, ResNet50 and DenseNet121, using transfer learning with ImageNet weights. During training, we use the Adam optimizer along with callbacks such as ModelCheckpoint, ReduceLROnPlateau, and EarlyStopping to ensure efficient and stable learning.

In the Testing Layer, the trained models are evaluated on the 30% test set to classify lesions as either malignant or benign. The trained model uses the learned weights to classify each skin lesion as benign or malignant. This pipeline is designed for an Internet of Medical Things (IoMT) setting. In this environment, data collected from connected devices can be processed and analyzed automatically.

3.6 Model Architecture and Fine-Tuning

Two pretrained convolutional neural network architectures are used for binary skin cancer classification: ResNet50 [43] and DenseNet121 [44]. Both models are initialized with weights pretrained on ImageNet and are fine-tuned using a two-phase training strategy to improve learning stability [45]. In the first phase, the pretrained backbone remains frozen and only the newly added classification layers are trained. In the second phase, selected layers of the pretrained network are unfrozen and fine-tuned using a smaller learning rate, allowing the model to better adapt to the skin lesion dataset while preserving the useful features learned from ImageNet.

For the ResNet50 model, we implement a two-phase training process. In Phase 1, the ResNet50 base remains frozen while we train a new classification head. This head includes a global average pooling layer, a dropout layer (rate 0.3), and a dense layer with 512 units and ReLU activation. It finishes with another dropout layer (rate 0.3) and a final dense layer with sigmoid activation for binary output. We run this phase for 15 epochs using the Adam optimizer with a learning rate of 0.001. In Phase 2, we unfreeze the last 50 layers of the ResNet50 base for fine-tuning. However, the batch normalization layers remain frozen to preserve stable statistics. We then recompile the model with a lower learning rate of 0.00005 and train it for up to 40 additional epochs.

For the DenseNet121 model, we follow the same two-phase training approach. In Phase 1, the base remains frozen while we train a new classification head. This head consists of a global average pooling layer, a dropout layer (rate 0.5), and a dense layer with 256 units (using ReLU activation and L2 regularization). It also includes batch normalization, a second dropout layer (rate 0.4), and a final dense layer with sigmoid activation. We train this configuration for 20 epochs

using a learning rate of 0.001. In Phase 2, we unfreeze the last 60 layers of the DenseNet121 base for fine-tuning, while keeping the batch normalization layers frozen. The model is then trained for up to 40 additional epochs with a reduced learning rate of 0.00005.

For both models, images are preprocessed using the model-specific preprocessing functions provided by Keras, matching the normalization used during the original ImageNet training.

3.7 Training Configuration

Both models are trained with a batch size of 32. Three callbacks are used to improve training stability. First, a model checkpoint saves the best-performing weights based on validation accuracy. Second, ReduceLROnPlateau, a learning rate reducer, automatically lowers the learning rate when validation loss stops improving. Finally, EarlyStopping halts training if the validation loss does not improve for 12 consecutive epochs. This callback also automatically restores the best weights found during the training process. Class weights are applied throughout training to handle class imbalance.

4 Results and Discussion

There are a total of 2,298 images in the PAD-UFES-20 dataset, whereas there are 900 images in the ISIC 2016 dataset. The fusion of these two datasets results in a total of 3,198 images, with 1,936 benign lesions and 1,262 malignant lesions. The dataset is split into a 70% training set and a 30% testing set. From the training set, 20% was taken out for the validation set. This makes a total of 1,790 images in the training set, 448 images in the validation set, and 960 images in the testing set. Before training the model, the training set was augmented, increasing the total to 5,370 images. Both transfer learning models, ResNet50 and DenseNet121, were trained on the training set to predict the presence or absence of skin cancer. To evaluate the performance, the following measures were used:

$$\text{Misclassification Rate} = \frac{\text{OP}_1/\text{AR}_0 + \text{OP}_0/\text{AR}_1}{\text{AR}_0 + \text{AR}_1} \quad (1)$$

$$\text{Accuracy} = \frac{\text{OP}_0/\text{AR}_0 + \text{OP}_1/\text{AR}_1}{\text{AR}_0 + \text{AR}_1} \quad (2)$$

$$\text{Positive Prediction Value} = \frac{\text{OP}_1/\text{AR}_1}{\text{OP}_1/\text{AR}_1 + \text{OP}_1/\text{AR}_0} \quad (3)$$

$$\text{Negative Prediction Value} = \frac{\text{OP}_0/\text{AR}_0}{\text{OP}_0/\text{AR}_0 + \text{OP}_0/\text{AR}_1} \quad (4)$$

$$\text{Specificity} = \frac{\text{OP}_0/\text{AR}_0}{\text{OP}_0/\text{AR}_0 + \text{OP}_1/\text{AR}_0} \quad (5)$$

$$\text{Sensitivity} = \frac{\text{OP}_1/\text{AR}_1}{\text{OP}_1/\text{AR}_1 + \text{OP}_0/\text{AR}_1} \quad (6)$$

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

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

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

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

In these equations, AR_0 represents the actual result of the benign class, while AR_1 represents the actual result of the malignant class. Similarly, OP_0 represents the output predicted as benign, and OP_1 represents the output predicted as malignant. The combined term OP_j/AR_i does not represent a mathematical division. Instead, it denotes the corresponding cell of the confusion matrix; it represents the number of samples whose actual class is i and whose predicted class is j. For example, OP_0/AR_0 represents the number of true negatives, while OP_1/AR_1 represents the number of true positives. Similarly, OP_1/AR_0 represents the number of false positives, and OP_0/AR_1 represents the number of false negatives.

Table 1. Training confusion matrix for ResNet50.

$N = 5370$	OP_0 **(Neg.-0)**	OP_1 **(Pos.-1)**
Actual AR_0 (Benign-0)	3250	2
Actual AR_1 (Malignant-1)	0	2118

With a total of 5,370 training images, ResNet50 correctly predicted 3,250 images as benign and 2,118 as malignant, at the epoch where the best validation weights were saved. Table 1 shows the training confusion matrix. ResNet50 achieved a training accuracy of 99.96%.

Table 2. Training confusion matrix for DenseNet121.

$N = 5370$	OP_0 **(Neg.-0)**	OP_1 **(Pos.-1)**
Actual AR_0 (Benign-0)	3075	177
Actual AR_1 (Malignant-1)	43	2075

Similarly, DenseNet121 was trained using the same 5,370 images. At its best-weight epoch, it correctly predicted 3,075 images as benign and 2,075 as malignant. Table 2 shows these results. DenseNet121 achieved a training accuracy of 95.90%.

Table 3. Testing confusion matrix for ResNet50.

$N = 960$	OP_0 **(Neg.-0)**	OP_1 **(Pos.-1)**
Actual AR_0 (Benign-0)	505	76
Actual AR_1 (Malignant-1)	95	284

After training, both models were evaluated on the independent test set of 960 images. ResNet50 correctly predicted 505 benign and 284 malignant images. Table 3 presents the testing confusion matrix. ResNet50 achieved a testing accuracy of 82.19%, with an AUC-ROC of 0.8800.

Table 4. Testing confusion matrix for DenseNet121.

$N = 960$	OP_0 **(Neg.-0)**	OP_1 **(Pos.-1)**
Actual AR_0 (Benign-0)	501	80
Actual AR_1 (Malignant-1)	83	296

When DenseNet121 was evaluated on the same test set, it correctly predicted 501 benign and 296 malignant images. Table 4 presents the corresponding testing confusion matrix. DenseNet121 achieved a testing accuracy of 82.98%.

The complete set of performance metrics for both models, in both the training and testing phases, is summarized in Table 5.

DenseNet121 performed slightly better than ResNet50 on the test set. Its testing accuracy was 82.98%, compared to 82.19% for ResNet50. The larger difference, however, is in the gap between training and testing accuracy. ResNet50 reached a training accuracy of 99.96%, which dropped to 82.19% on the test set—a gap of about 17.8 percentage points. This shows strong overfitting: the model learned the training data too well but struggled with new images. DenseNet121 reached a training accuracy of 95.90% and a test accuracy of 82.98%. The gap here is smaller, around 12.9 percentage points, which means DenseNet121 generalizes better to new, unseen images.

The AUC-ROC value for DenseNet121 was 0.8880, while ResNet50 achieved 0.8800. These values indicate that both models can effectively distinguish between benign and malignant skin lesions. The slightly higher AUC-ROC of DenseNet121 is consistent with its better classification performance on the test set. The ROC curves for ResNet50 and DenseNet121 are shown in Figures 3 and 4, respectively.

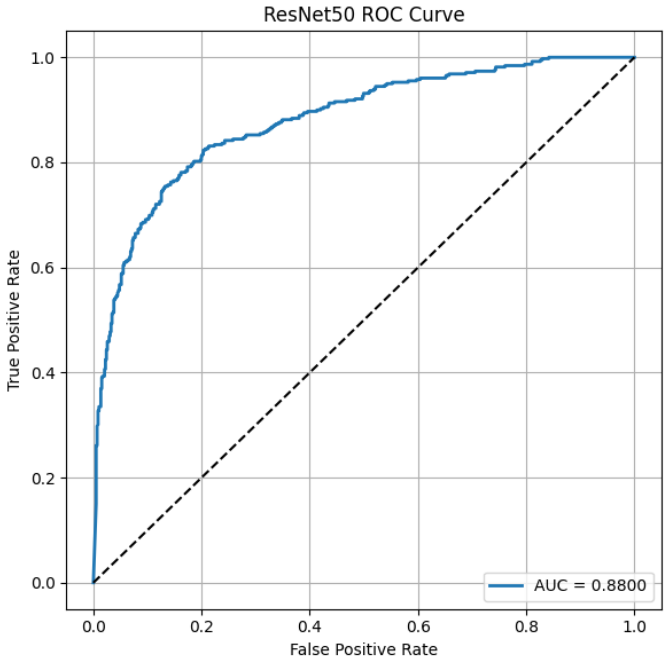


Figure 3. ROC curve for the ResNet50 model.

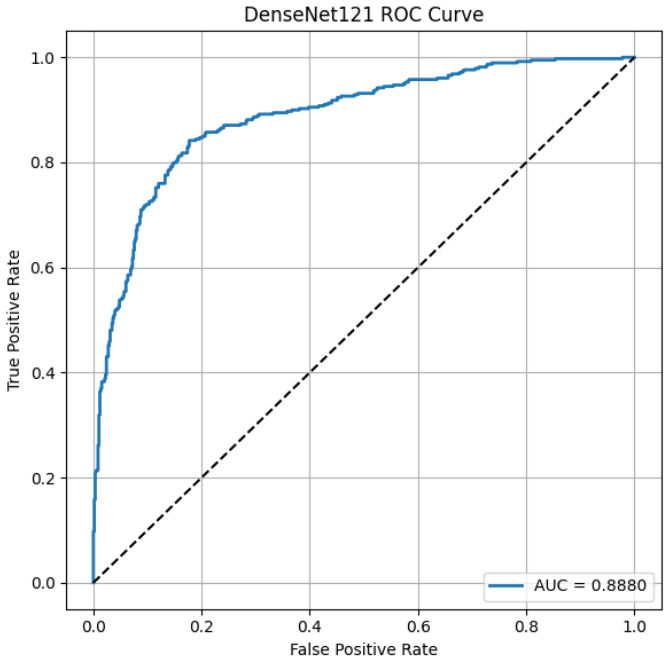


Figure 4. ROC curve for the DenseNet121 model.

This difference comes from how each network is built. DenseNet121 connects every layer to every other

Table 5. Performance metrics comparison of ResNet50 and DenseNet121.

Model	Phase	Acc.	Miscl.	Spec.	Sens.	PPV	NPV	FPR	FNR	LR+	LR−
ResNet50	Training	0.9996	0.0004	0.9994	1.0000	0.9991	1.0000	0.0006	0.0000	1626.00	0.0000
	Testing	0.8219	0.1781	0.8692	0.7493	0.7889	0.8417	0.1308	0.2507	5.7285	0.2884
DenseNet121	Training	0.9590	0.0410	0.9456	0.9797	0.9214	0.9862	0.0544	0.0203	17.9999	0.0215
	Testing	0.8298	0.1702	0.8623	0.7810	0.7872	0.8579	0.1377	0.2190	5.6720	0.2540

Table 6. Performance comparison of the proposed framework with previous studies.

Reference	Methodology	Accuracy
[10]	Deep CNN (Inception-v3), 3-class‡	72.1%
[19]	MD-Net (DenseNet-169 + metadata fusion), 6-class‡	81.4%
[22]	PiT Transformer + MetaBlock, 6-class‡	80.0%
[46]	DiffMIC (diffusion-based, 6-class)†	64.5%
[47]	Multi-layer CNN (PAD-UFES-20), 6-class‡	70.0%
[48]	Siamese neural network (clinical and dermoscopic images)	74.3%
[49]	EfficientNetB3 + XGBoost (image-only)	76.0%
[50]	Multimodal deep learning (PAD-UFES-20)	78.2%
Proposed	**ResNet50 (fused dataset)**	**82.19%**
Proposed	**DenseNet121 (fused dataset)**	**82.98%**

†This method addresses a 6-class skin and oral cancer classification task and is not directly comparable to the binary classification results in this table; it is included for contextual reference only.
‡This method addresses a multi-class classification task and is not directly comparable to the binary classification results in this table; it is included for contextual reference only.

layer, which helps the model reuse features and avoid overfitting. ResNet50 also uses skip connections, but each layer learns more on its own. When only a few layers are fine-tuned, this can make the model memorize the training data. Since the fused dataset used here is still moderate in size, DenseNet121's feature-reuse design helps it avoid overfitting more effectively.

The training and validation accuracy and loss curves of both models are shown in Figures 5 and 6. Both models converged successfully during training. DenseNet121 showed more stable validation performance than ResNet50, which indicates better generalization. These observations agree with the smaller gap between its training and testing accuracy.

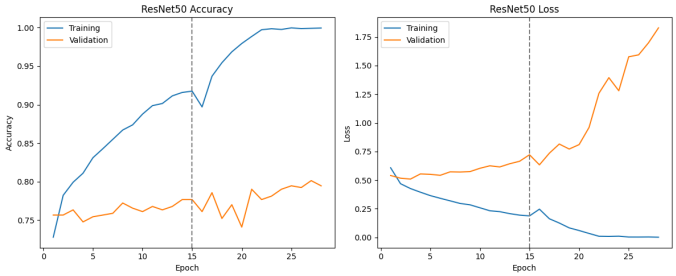


Figure 5. Training and validation accuracy and loss curves for ResNet50.

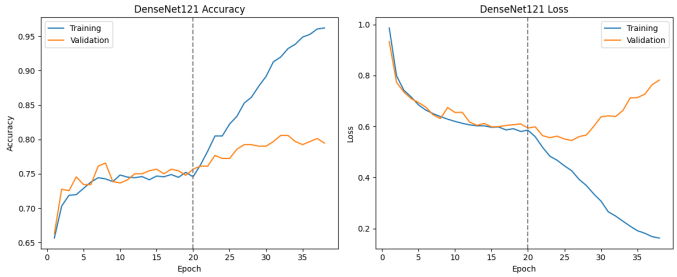


Figure 6. Training and validation accuracy and loss curves for DenseNet121.

These results also matter in real clinical use. DenseNet121 had a sensitivity of 78.10%, which means it correctly identified most malignant cases. A missed malignant case, called a false negative, is the most dangerous error because it delays treatment for the patient. DenseNet121 had a false negative rate of 21.90%, whereas ResNet50 had a higher rate of 25.07%. This means DenseNet121 missed fewer cancer cases, making it more suitable for clinical use. In comparison, a false positive is generally less critical, as it usually results in additional clinical examinations rather than delayed treatment. Both models had similar false positive rates: 13.08% for ResNet50 and 13.77% for DenseNet121.

Table 6 compares the proposed framework against

several previously published skin cancer classification methods. The proposed ResNet50 and DenseNet121 models achieved 82.19% and 82.98% accuracy, respectively, outperforming the selected comparison methods. These results demonstrate the effectiveness of the proposed dataset-level fusion framework. The fused dataset increases training data diversity and improves the generalization of the transfer learning models.

5 Conclusion

This study proposed a dataset-level fusion framework for binary skin cancer classification using the PAD-UFES-20 and ISIC 2016 datasets. The two datasets were combined to increase the diversity of the training data by including both clinical smartphone images and dermoscopic images. The fused dataset was preprocessed using image resizing, stratified data splitting, data augmentation, and class weighting before training the models. Two transfer learning models, ResNet50 and DenseNet121, were fine-tuned using a two-phase training strategy. Experimental results showed that both models achieved good classification performance on the fused dataset. DenseNet121 achieved the highest accuracy of 82.98%, while ResNet50 achieved 82.19%. These results demonstrate that dataset-level fusion improves training data diversity and helps the models learn more robust and generalized features. The proposed framework also provides a simple and effective approach for automated skin cancer classification and has the potential for deployment in Internet of Medical Things (IoMT) environments. This study has some limitations: the proposed framework is limited to binary skin cancer classification and was evaluated only on publicly available datasets. Further validation on additional datasets is required to better assess its generalization capability. Future work will investigate the integration of additional public skin lesion datasets, ensemble learning techniques, attention-based fusion methods, and transformer architectures to further improve classification performance. The framework will also be extended to support multi-class skin lesion classification and IoMT-based clinical decision support systems for real-time skin cancer screening.

Data Availability Statement

The datasets used in this study are publicly available. PAD-UFES-20 is available at https://data.mendeley.com/datasets/zr7vgbcyr2/1 and ISIC 2016 is available at https://challenge.isic-archive.com/data/. Further information is available from the corresponding author upon reasonable request.

Funding

This work was supported without any funding.

Conflicts of Interest

The 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

This study did not involve any new experiments on humans or animals. It used only publicly available, de-identified skin lesion datasets (PAD-UFES-20 and ISIC 2016), whose original data collection and patient consent were approved by the ethics committees of the institutions that created them. Therefore, no additional ethical approval was required for the present study.

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] 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]

[3] Fried, L., Tan, A., Bajaj, S., Liebman, T. N., Polsky, D., & Stein, J. A. (2020). Technological advances for the detection of melanoma: Advances in diagnostic techniques. *Journal of the American Academy of Dermatology, 83*(4), 983-992. [CrossRef]

[4] Janda, M., Cust, A. E., Neale, R. E., Aitken, J. F., Baade, P. D., Green, A. C., ... & Whiteman, D. C. (2020). Early detection of melanoma: a consensus report from the Australian skin and skin cancer research centre melanoma screening Summit. *Australian and New Zealand journal of public health, 44*(2), 111-115. [CrossRef]

[5] Davis, S., Piggott, C., Lyon, C., & DeSanto, K. (2020). Effectiveness of dermoscopy in skin cancer diagnosis. *Canadian Family Physician, 66*(10), 739-740. [CrossRef]

[6] 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]

[7] 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]

[8] 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]

[9] Jung, J. M., Cho, J. Y., Lee, W. J., Chang, S. E., Lee, M. W., & Won, C. H. (2021). Emerging minimally invasive technologies for the detection of skin cancer. *Journal of personalized medicine, 11*(10), 951. [CrossRef]

[10] 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]

[11] 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]

[12] 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]

[13] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. [CrossRef]

[14] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning* (pp. 6105-6114). PmLR. https://proceedings.mlr.press/v97/tan19a.html?ref=ji

[15] Gutman, D., Codella, N. C., Celebi, E., Helba, B., Marchetti, M., Mishra, N., & Halpern, A. (2016). Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC). *arXiv preprint arXiv:1605.01397*. [CrossRef]

[16] 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]

[17] 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]

[18] 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]

[19] 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]

[20] 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]

[21] 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]

[22] de Lima, L. M., & Krohling, R. A. (2022, November). Exploring advances in transformers and CNN for skin lesion diagnosis on small datasets. In *Brazilian Conference on Intelligent Systems* (pp. 282-296). Cham: Springer International Publishing. [CrossRef]

[23] Tang, P., Yan, X., Nan, Y., Hu, X., Menze, B. H., Krammer, S., & Lasser, T. (2024). Joint-individual fusion structure with fusion attention module for multi-modal skin cancer classification. *Pattern Recognition, 154*, 110604. [CrossRef]

[24] Ge, Z., Demyanov, S., Chakravorty, R., Bowling, A., & Garnavi, R. (2017, September). Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images. In *International conference on medical image computing and computer-assisted intervention* (pp. 250-258). Cham: Springer International Publishing. [CrossRef]

[25] Kawahara, J., Daneshvar, S., Argenziano, G., & Hamarneh, G. (2018). Seven-point checklist and skin lesion classification using multitask multimodal neural nets. *IEEE journal of biomedical and health informatics, 23*(2), 538-546. [CrossRef]

[26] Yap, J., Yolland, W., & Tschandl, P. (2018). Multimodal skin lesion classification using deep learning. *Experimental dermatology, 27*(11), 1261-1267. [CrossRef]

[27] Nunnari, F., Bhuvaneshwara, C., Ezema, A. O., & Sonntag, D. (2020, August). A study on the fusion of pixels and patient metadata in CNN-based

classification of skin lesion images. In *International cross-domain conference for machine learning and knowledge extraction* (pp. 191-208). Cham: Springer International Publishing. [CrossRef]

[28] Chen, Q., Li, M., Chen, C., Zhou, P., Lv, X., & Chen, C. (2023). MDFNet: application of multimodal fusion method based on skin image and clinical data to skin cancer classification. *Journal of Cancer Research and Clinical Oncology, 149*(7), 3287-3299. [CrossRef]

[29] Ou, C., Zhou, S., Yang, R., Jiang, W., He, H., Gan, W., ... & Li, J. (2022). A deep learning based multimodal fusion model for skin lesion diagnosis using smartphone collected clinical images and metadata. *Frontiers in Surgery, 9*, 1029991. [CrossRef]

[30] Cai, G., Zhu, Y., Wu, Y., Jiang, X., Ye, J., & Yang, D. (2023). A multimodal transformer to fuse images and metadata for skin disease classification. *The Visual Computer, 39*(7), 2781-2793. [CrossRef]

[31] Tang, P., Yan, X., Nan, Y., Xiang, S., Krammer, S., & Lasser, T. (2022). FusionM4Net: A multi-stage multi-modal learning algorithm for multi-label skin lesion classification. *Medical Image Analysis, 76*, 102307. [CrossRef]

[32] Pedro, R., & Oliveira, A. L. (2022, July). Assessing the impact of attention and self-attention mechanisms on the classification of skin lesions. In *2022 international joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE. [CrossRef]

[33] Zhang, Y., Xie, F., & Chen, J. (2023). TFormer: A throughout fusion transformer for multi-modal skin lesion diagnosis. *Computers in biology and medicine, 157*, 106712. [CrossRef]

[34] Xu, J., Gao, Y., Liu, W., Huang, K., Zhao, S., Lu, L., ... & Chen, X. (2022, September). RemixFormer: a transformer model for precision skin tumor differential diagnosis via multi-modal imaging and non-imaging data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 624-633). Cham: Springer Nature Switzerland. [CrossRef]

[35] Omeroglu, A. N., Mohammed, H. M., Oral, E. A., & Aydin, S. (2023). A novel soft attention-based multi-modal deep learning framework for multi-label skin lesion classification. *Engineering Applications of Artificial Intelligence, 120*, 105897. [CrossRef]

[36] Vachmanus, S., Noraset, T., Piyanonpong, W., Rattananukrom, T., & Tuarob, S. (2023). DeepMetaForge: A deep vision-transformer metadata-fusion network for automatic skin lesion classification. *IEEE Access, 11*, 145467-145484. [CrossRef]

[37] Cheslerean-Boghiu, T., Fleischmann, M. E., Willem, T., & Lasser, T. (2023). Transformer-based interpretable multi-modal data fusion for skin lesion classification. *arXiv preprint arXiv:2304.14505*. [CrossRef]

[38] Georgiadis, P., Gkouvrikos, E. V., Vrochidou, E., Kalampokas, T., & Papakostas, G. A. (2025). Building better deep learning models through dataset fusion: A case study in skin cancer classification with hyperdatasets. *Diagnostics, 15*(3), 352. [CrossRef]

[39] Remya, S., Anjali, T., & Sugumaran, V. (2024). A novel transfer learning framework for multimodal skin lesion analysis. *IEEE Access, 12*, 50738-50754. [CrossRef]

[40] Adebiyi, A., Abdalnabi, N., Simoes, E. J., Becevic, M., Hoffman Smith, E., & Rao, P. (2024). Transformers in skin lesion classification and diagnosis: A systematic review. *MedRxiv*, 2024-09. [CrossRef]

[41] Das, A., Agarwal, V., & Shetty, N. P. (2025). Comparative analysis of multimodal architectures for effective skin lesion detection using clinical and image data. *Frontiers in Artificial Intelligence, 8*, 1608837. [CrossRef]

[42] Marchetti, M. A., Codella, N. C., Dusza, S. W., Gutman, D. A., Helba, B., Kalloo, A., ... & International Skin Imaging Collaboration. (2018). Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. *Journal of the American Academy of Dermatology, 78*(2), 270-277. [CrossRef]

[43] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778). [CrossRef]

[44] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017, July). Densely connected convolutional networks. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 2261-2269). IEEE. [CrossRef]

[45] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning?. *IEEE transactions on medical imaging, 35*(5), 1299-1312. [CrossRef]

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

[47] 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]

[48] Battle, M. L., Atapour-Abarghouei, A., & McGough, A. S. (2022, December). Siamese neural networks for skin cancer classification and new class detection using clinical and dermoscopic image datasets. In *2022 IEEE International Conference on Big Data (Big Data)* (pp. 4346-4355). IEEE. [CrossRef]

[49] 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]

[50] Saxena, S., Kaushik, K., & Chaudhary, A. (2025). AI-Driven Early Detection of Skin Cancer using PAD-UFES-20: TRL 1 & 2 Achievements in Multimodal Deep Learning. [CrossRef]



Tuba Mubarik is currently pursuing her M.S. degree in Computer Science at 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 at the 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 an academician and 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. He has authored more than 70 research publications in international journals and conferences. He is a Senior Member of IEEE. (Email: shabib.aftab@gmail.com)