STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–16.
Published online in International Academic Press (www.IAPress.org)

Diagnosis with Deep Learning and a Novel Pre-processing Strategy on a Large-Scale Dermoscopic Image Dataset

Youssra EL IDRISSI EL-BOUZAIDI [1,2], Sokaina EL KHAMLICHI [3,4,*], Otman ABDOUN [2]

[1] *Intelligent Systems and Applications Laboratory (LSIA), EMSI, Tangier, Morocco*
[2] *Information Security, Intelligence Systems and Applications Team-ISISA, Faculty of Science, Abdelmalek Essaadi University, Tetouan, Morocco*
[3] *Research Team in Science and Technology, Higher School of Technology of Laayoune, Ibn Zohr University, P.O. Box 3007, Laayoune, Morocco*
[4] *LyRICA: Laboratory of Research in Informatics, Data Sciences and Artificial Intelligence, School of Information Sciences, B.P. 6204, Rabat-Instituts, Rabat, Morocco*

Abstract Skin cancer, in particular melanoma, has become a major health problem worldwide. Early diagnosis is the most important factor to consider for successful treatment. The latest advances in diagnosis have increased melanoma survival rates significantly, include early detection techniques such as imaging technologies that can detect melanoma in its earliest stages when treatment is most effective. AI has also been a game changer in the field of diagnosis, providing automated analysis data with a high level of accuracy. In this article, a novel computer-assisted diagnosis is presented, which consists of a new preprocessing technique to improve the quality of images. Data augmentation is used to increase data size by applying transformations to improve model generalization. The transfer learning efficiency is proved using the MobileNetV2 model. Improving and fine-tuning this architecture for the skin lesion classification task. The trained model can achieve performance, with an accuracy of 95.1% on 7 classes, and a very high AUC score of 94% for the precision-recall curve on the HAM10000 benchmark dataset. These results show how advanced deep learning techniques can be used in dermatological practice, thus creating a promising alley for improved skin cancer diagnosis.

Keywords Artificial Intelligence, Deep Learning, Melanoma Detection, Medical images, Diagnosis, Classification

DOI: 10.19139/soic-2310-5070-3122

1. Introduction

Melanoma, the 17th most common cancer worldwide, is a malignant type of skin cancer and its most deadly form. Annually, it is reported as one of the leading causes of death from skin cancer, despite accounting for only 1% of all skin cancer cases. In 2020, over 150,000 new cases of cutaneous melanoma were reported. The number of cutaneous melanomas and other skin cancers, with a universality factor of over five million recorded each year, constitutes a major health problem worldwide. Melanoma survival rates depend on the stage of the disease. There is a significant degree of optimism when melanoma has not spread beyond the surface and is not found in lymph nodes, lungs, liver, and other internal organs, where the survival rate can be as high as 60-70% for the first five years after the diagnosis. In case when the tumor spreads to distant organs, then the five-year survival rate is halved and equals 15-20% [1].

Most cases of skin cancer are the result of overexposure to ultraviolet (UV) rays from the sun and tanning salons. According to WHO findings, exposure to UV radiation is one of the main causes of skin cancer, whether

[*] Correspondence to: Sokaina EL KHAMLICHI (Email: s.elkhamlichi@uiz.ac.ma). Research Team in Science and Technology, Higher School of Technology of Laayoune, Ibn Zohr University, P.O. Box 3007, Laayoune, Morocco. LyRICA: Laboratory of Research in Informatics, Data Sciences and Artificial Intelligence, School of Information Sciences, P.O. Box. 6204, Rabat-Instituts, Rabat, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

from a natural source such as the sun, or an artificial source such as sunbeds [2]. The number of patients with skin cancer exceeded 1,5 million worldwide by 2020 with 120 thousand deaths. Light skin, sunburn, family history, and immune system deficiency are to be named risk factors that incite skin cancer.

Skin cancer is classified into four main types: e. g. 'MEL' (melanoma), 'Nv' (melanocytic nevi), 'BCC' (basal cell carcinoma), and 'SCC' (squamous cell carcinoma). Melanoma is considered extremely dangerous, with a high risk of metastasis [3]. It stems from melanocytes, which produce dark stripes/pigments in the skin. It is an extremely dangerous illness, which probably spreads to the brain, liver, or lungs [4]. Every year, the USA accounts for almost 10 000 melanoma deaths [5]. Therefore, regular evaluation of melanoma-affected moles and early diagnosis followed by appropriate medical intervention are the only methods of protection.

Despite advances in dermatology and imaging technologies such as dermoscopy-a non-invasive diagnostic tool used for skin lesions, accurate diagnosis of melanoma, particularly in its early stages, remains elusive. Visual inspection by a doctor only covers the symptoms, but not the cause, as early melanomas can be very subtle and therefore misclassified as benign neoplasms. The impressive frequency of this particular skin cancer, melanoma, underlines the importance of more accurate and earlier diagnostic tools. In this context, artificial intelligence, particularly machine learning, has become increasingly prominent in medical applications aimed at early disease diagnosis and prognosis. Numerous studies have shown the effectiveness of these techniques in predicting COVID-19 [6, 7, 8], identifying Down syndrome, and assessing other complex health conditions, highlighting their growing role in precision medicine [9].

In recent years, deep learning models have also been successfully applied to dermatoscopic images to improve the accuracy and efficiency of skin lesion recognition [10]. In a new study [11], researchers examined the effect of artificial intelligence (AI) in decisions about diagnosing skin cancers. The study focused on using AI reinforcement learning (RL) model tools to assist dermatologists in designing a reward/penalization system that corresponded to different types of skin lesions. The rate of correct diagnosis by 89 dermatologists improved by 12% after implementing the AI RL model support. This application of AI to skin cancer diagnosis can be considered a revolutionary advance, as it provides algorithms capable of streamlining the diagnostic process and increasing its accuracy. AI technologies can analyze large volumes of dermatoscopic images and identify characteristic marks and features related to skin cancer diagnosis [12].

Research papers demonstrating that CNNs outperform dermatologists in melanoma diagnosis[11] often rewarded ensemble CNN models [13] by ranking them in the top three of the ISIC 2018 challenge [14], where sophisticated models, despite their dominance, were outperformed by the collective strength of ensemble models. This proves that CNNs can contribute to the decision-making process (even for junior dermatologists) [15]. Another factor to consider is that the same symptoms can be presented by different skin problems, which is one of the difficult parts of dermatological diagnosis. The ability to detect skin neoplasms at the earliest stage of the disease is crucial to effective treatment. If the medical examination is not carried out and the neurotic problem is aggravated, the disease will only get worse and, as a result, deformities may occur. AI allows data analysis in healthcare to a higher level, but problems remain with non-standardized data, interpretation, and ethical issues. Although this fact, AI's accuracy which is higher than the human dermatologist in the early diagnosis of melanoma, can increase patient survival.

Despite the remarkable progress in the field of AI-based skin cancer detection, there are still some obstacles that exist in the current methods. Quite often, the existing research projects only cover the single steps of the detection, like segmentation or classification, but they do not take into account the entire diagnostic process. In addition, the constraint of the databases also hinders the generalization of these models in the real-world clinical practice. We are focused on the shortcomings mentioned in the current system and propose an integrated diagnostic system that combines preprocessing and classification using advanced algorithms. The MobileNetV2 architecture is used in our system along with a novel preprocessing method, and the system has achieved high performance and efficiency in skin cancer diagnosis. Furthermore, the approach we developed is aimed to address the issue of class imbalance in datasets as well, providing more balanced class representations for training. A critical assessment of the current literature confirms that despite the notable advances in this field, there is a need for the creation of more comprehensive and reliable diagnostic tools. Our contributions encompass several key aspects:

- We have developed a novel AI-based approach for skin cancer detection from dermatoscopic images. This model takes advantage of the transfer learning of the MobileNetV2 architecture.
- In dermatological image analysis, pre-processing is a crucial step aimed at improving the quality of images for analysis. We have developed a new pre-processing strategy to enhance image quality. The technique includes grayscale conversion, median filtering, thresholding, removal of small unwanted objects and skin region extraction. Preliminary steps such as these are crucial to increasing overall accuracy levels.
- One of the approaches we have been using to deal with the limited of data is data augmentation. These methods are the main way to maximize the variety of our training data set, which helps the model to generalize better for the unseen data.
- We have overcome the problem of class imbalance in our dataset by using the oversampling method. This method ensures that the proportion of each class is more balanced in the training data and prevents the model from becoming unbalanced.
- We carried out an in-depth analysis of our model performance using measures such as accuracy, precision, recall, F1 score, confusion matrix, ROC curve and precision-recall curve. These parameters give an overall view of the model's performance in skin cancer recognition.
- We compared the performance of our model against the existing state-of-the-art approaches. This comparison clearly demonstrates the ability of our method to guarantee competitive performance.

The article begins with a detailed overview of existing research in automatic skin cancer detection, highlighting methodologies and comparing successful techniques. The HAM10000 dataset is detailed, consisting of 10,015 skin lesion images across various classes. The dataset's reliability and class descriptions provide context for the analysis. Subsequently, the article describes a new preprocessing method and the role of data augmentation in deep learning, especially for small datasets. The oversampling method used to balance class distribution is explained, alongside label encoding for input preparation where necessary. The construction of the proposed model, based on the MobileNetV2 architecture, is outlined. The results section evaluates model performance, identifying strengths and weaknesses through measures like confusion matrix, specificity, sensitivity, and precision-recall curve.

2. Proposed methodology

Figure 1 represents a scheme of the AI-based approach being developed for the diagnosis of skin cancer by means of analyzing dermoscopic images. The method is based on the implementation of several pre-processing techniques to ensure better skin lesion classification accuracy. These steps have been carefully chosen to ease computations and decrease noise as well as enhance feature extraction. By overcoming computational limitations, we have proposed a memory-efficient model. We carried out several experiments and observed that pre-processing and data augmentation had a significant influence on the performance of our model. The selection of hyperparameters and architecture in this study was based on our previously published comparative work [16]. In this prior research, we evaluated multiple convolutional neural network (CNN) architectures and optimization techniques. MobileNetV2 and Bayesian optimization were chosen as the most effective combination. Bayesian optimization efficiently explored the hyperparameter space, optimizing parameters such as learning rate and batch size to enhance the model's performance. MobileNetV2 was selected for its computational efficiency and strong performance in image classification tasks, particularly when integrated with transfer learning. Its depth wise separable convolutions reduce computational complexity while retaining feature extraction capabilities. This paper builds on the findings of our earlier study, providing a practical implementation of the optimal methods identified [16].

2.1. Description of dataset

Datasets used in skin cancer research include HAM10000 [17], ISIC Archiv [27] and ISBI [18]. The HAM10000 dataset is particularly noteworthy for its reliability, comprising 10,015 images of skin lesions representing several classes. This dataset contains a wide range of images from various sources and populations, which have been manually cropped, histogram-selected and subjected to other processes to improve visual interpretation and contrast. A key feature of the HAM10000 dataset is its comprehensive coverage, which ensures that lesions are

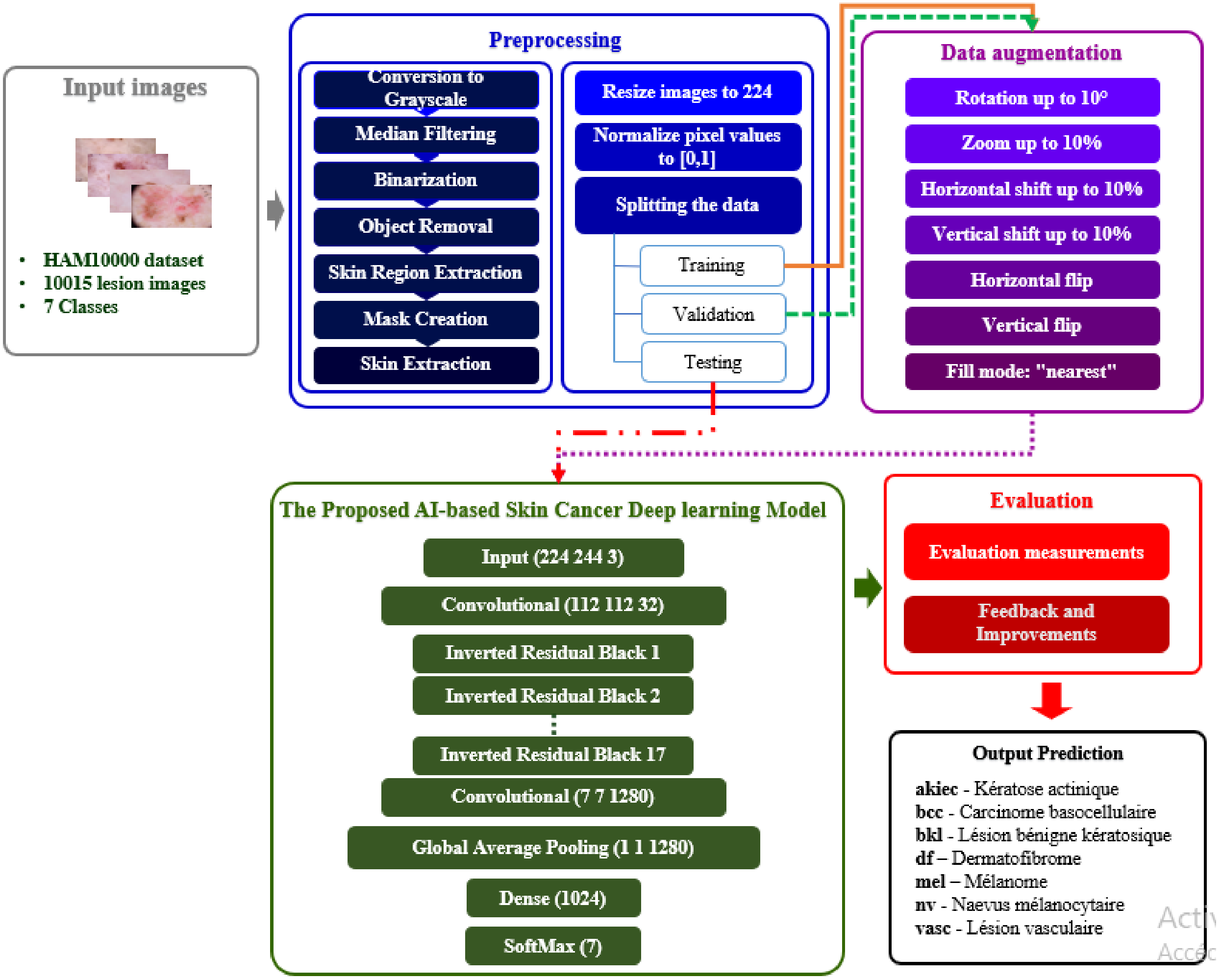


Figure 1. Proposed Methodology for skin lesion images classification

visually confirmed by pathologists. For cases not confirmed visually, follow-up CT studies, expert consensus or histological examination are carried out.

Melanoma and non-melanoma form two major types of skin cancer, including 'BCC' (basal cell carcinoma) and 'SCC' (squamous cell carcinoma). Within squamous cell carcinoma, there are subtypes like 'akiec' (actinic keratosis) and 'SPC' (superficial). The HAM10000 dataset includes these categories along with 'Vasc' (vascular lesions), 'Nv' (melanocytic nevi), and 'Df' (dermatofibroma), covering both malignant and benign lesions. The dataset comprises seven distinct classes: basal cell carcinoma, squamous cell carcinoma, melanoma, Merkel cell carcinoma, and vascular skin cancer.

The HAM10000 dataset is available as an open dataset in the Harvard Dataverse repository. It consists of 10,015 images divided into two sets of 5,008 images each. These images are digital images, each measuring 600×450. The metadata file, named HAM10000_metadata, contains key features such as lesion_id for lesion identification, dx for lesion class, image_id for image identification, gender, dx_type for lesion confirmation method, location for body area and age.

2.2. *Image Preprocessing*

In the case of dermatological image analysis, the pre-processing stage is of paramount importance, as it improves image quality prior to further processing. We take a unique approach to image pre-processing in order to make images suitable for the next level of analysis. This method is implemented as follows: grayscale conversion, median filtering, binarization, eradication of small unwanted objects and extraction of the skin region. Initially, all color images were converted to grayscale in order to reduce the number of calculations and retain only the important information. One after the other, a median filter was applied to make the image clean and uniform. After thresholding, small irrelevant objects such as hair were eliminated to retain only the skin area. Finally, connected component analysis was performed to separate the skin region from the others using the largest connected region, which generally represents the skin region. The results of pre-processing are shown in Figure 2.

To assess the robustness of the proposed preprocessing pipeline, we also compared it qualitatively with alternative techniques such as adaptive thresholding and wavelet-based denoising. Adaptive thresholding was found to be highly sensitive to illumination variations, while wavelet denoising, although effective in reducing noise, tended to blur lesion boundaries. In contrast, the combination of grayscale conversion, median filtering, and morphological operations used in our method achieved visually stable segmentation and preserved lesion edges more effectively. This confirms that the proposed preprocessing sequence provides an appropriate balance between noise reduction and edge preservation without increasing computational complexity.

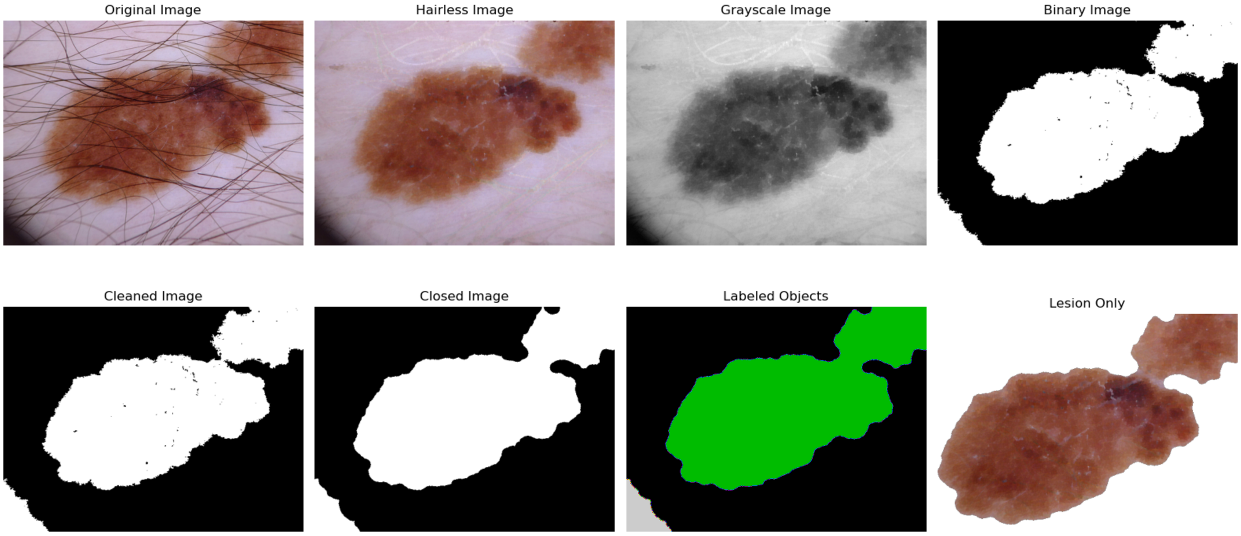


Figure 2. Preprocessing Results

2.2.1. *Conversion to Grayscale* :

Before we started the processing, all color images were converted to grayscale. This matrix conversion is frequently applied in image processing to reduce computational complexity by representing each image with a single intensity channel. The conversion was performed by combining the red, green, and blue (RGB) channels according to the following luminance-preserving formula:

$$\text{gray\_image} = 0.2125 \times \text{image}[:,:,0] + 0.7154 \times \text{image}[:,:,1] + 0.0721 \times \text{image}[:,:,2] \tag{1}$$

The coefficients (0.2125, 0.7154, 0.0721) are derived from the ITU-R BT.709 standard, which reflects the human visual system's sensitivity to different colors, giving greater weight to green and less to blue.

This step simplifies the image while preserving essential visual characteristics such as texture and lesion boundaries that are critical for accurate classification.

In addition, the median filter with a radius of 3 was chosen empirically after testing several window sizes (from 2 to 5), ensuring optimal smoothing while preserving edge details. A sensitivity analysis conducted on the thresholding stage (threshold values between 0.4 and 0.6) demonstrated that a fixed threshold of 0.5 achieved the best balance between noise suppression and lesion boundary accuracy.

2.2.2. Median Filtering :

To reduce noise in the grayscale image, we applied a median filter. This filter is a non-linear technique that replaces the value of each pixel with the median value of the intensity levels in its neighborhood. In our case, we used a disk-shaped structuring element with a radius of 3. For each pixel (i, j) in the image, the median filter calculates the median value of the pixel's neighborhood defined by a window of size 2k + 1 × 2k + 1, where k is the radius of the disk. The median filter formula is as follows:

$$\text{median\_image}[i,j] = \mathrm{median}(\text{gray\_image}[i-k:i+k,\ j-k:j+k]) \tag{2}$$

These filtering actions will make the image smoother and minimized the deviation of pixels of random light intensity, resulting in a brighter, more uniform image [19].

2.2.3. Binarization :

After applying the median filter, we proceeded with the thresholding technique, which transformed the grayscale image into a binary image [20]. These steps involve defining a threshold value that will indicate whether each pixel is skin or not. We decided to use the empirical value of 0.5 for our threshold value. For each pixel (i, j) in the median-filtered image, the binarization operation is defined as follows:

$$\text{binary\_image}[i,j] = \begin{cases} 1, & \text{if median\_image}[i,j] < 0.5 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Pixel values less than 0.5 are considered skin, and those greater than or equal to 0.5 are classified as non-skin. This process, in particular, is responsible for separating the skin region from the background, which is then used for further analysis and processing of the skin part.

This threshold value was selected based on Otsu's method, which automatically determines the optimal cut-off point that minimizes intra-class variance between foreground and background regions. Empirically, this threshold provided the best balance between noise removal and preservation of lesion boundaries across the dataset.

2.2.4. Morphological Filtering :

The small items like hair follicles and noise were eliminated from the binary image by applying a morphological operation called object deletion [21]. This operation aims to remove objects whose size is less than a given minimum size, while retaining larger connected elements. In our example, we've used 200 pixels as the minimum size.

$$\text{cleaned\_image} = \text{remove\_small\_objects}(binary\_image, min\_size = 200) \tag{4}$$

After performing this operation, we get the resulting binary image with only the large connected components, means that the small objects and the noise have been removed from the image. This step is fundamental for precise recognition and examination of the skin region.

2.2.5. Skin Region Extraction :

To extract the skin region from the original image, we used connected component analysis [22]. This method allows us to identify and isolate the largest connected region in the cleaned binary image, which generally corresponds to the skin region. The bounding box of this region was then used to extract the skin region from the original grayscale image. This process ensures that only the skin region is retained for later analysis, while any remaining artifacts or non-skin regions are excluded.

2.2.6. Mask Creation :
In image processing, a mask is a binary image used to selectively modify the pixels of another image. The process of creating a mask involves assigning a specific value (often 1) to pixels that meet certain criteria, while assigning a different value (often 0) to pixels that do not [23]. Mathematically, the creation of a mask can be represented as follows:

$$\text{mask}[i,j] = \begin{cases} 1, & \text{if a condition is met for pixel } (i,j) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

In the context of skin region extraction, the mask is created to identify the pixels that belong to the skin region based on certain image characteristics, such as color or texture. The mask is then used to extract the skin region from the original image.

2.2.7. Skin Extraction :
After creating the binary mask to isolate the skin region, the next step was to extract this region from the original image. This was achieved by applying the mask to the original image using the bounding box coordinates of the largest connected region. Mathematically, the extraction process can be represented as follows:

$$\text{skin\_only} = \text{image}[\text{minr} : \text{maxr},\ \text{minc} : \text{maxc}] \tag{6}$$

where minr, minc, maxr and maxc represent the minimum and maximum indices of the bounding box row and column respectively. The resulting image, called ”skin_only”, contains only those pixels corresponding to the skin region, enabling further analysis or processing focused specifically on the skin area.

The results of these preprocessing steps are illustrated in Figure 3 below, showing the sequence of actions performed on the original skin lesion image. The left image represents the original, while the right images show the outcomes after each preprocessing step, including conversion to grayscale, median filtering, binarization, morphological filtering, and skin region extraction.

2.3. Data Augmentation

Data augmentation is one of the key methods in deep learning, especially useful when the data sets are limited. The process involves generating new training instances by applying random transformations to the existing data. With the enrichment of the dataset, we are able to fill in the gaps of training samples, reducing the risk of overfitting and improving the generalizability of the model. It's important to note that data augmentation must be applied sparingly to avoid generation of unrealistic samples and consequent drop in model's performance. It is necessary to test different augmentation parameters in order to find the most suitable compromise between sample diversity and realism.

Table 2 shows the details of the transformation. We have added to the original image our data augmentation procedure, which applies to some of the enrichment datasets. As each image can be developed with all–embracing parameters, the process can produce up to 3,200 distinct images out of each class of 1,500 images generating a total of 4,800,000 images after image augmentation. These calculations are performed through various transformations such as rotation, zoom in and out, the horizontal and the vertical shifts of both; and the horizontal and vertical flipping. Here, we deploy these intersecting transformations. This approach allows us to add more diversity to our dataset, and gives the models truer-than-life capabilities.

2.4. Balancing Class Distribution for Improved Deep Learning

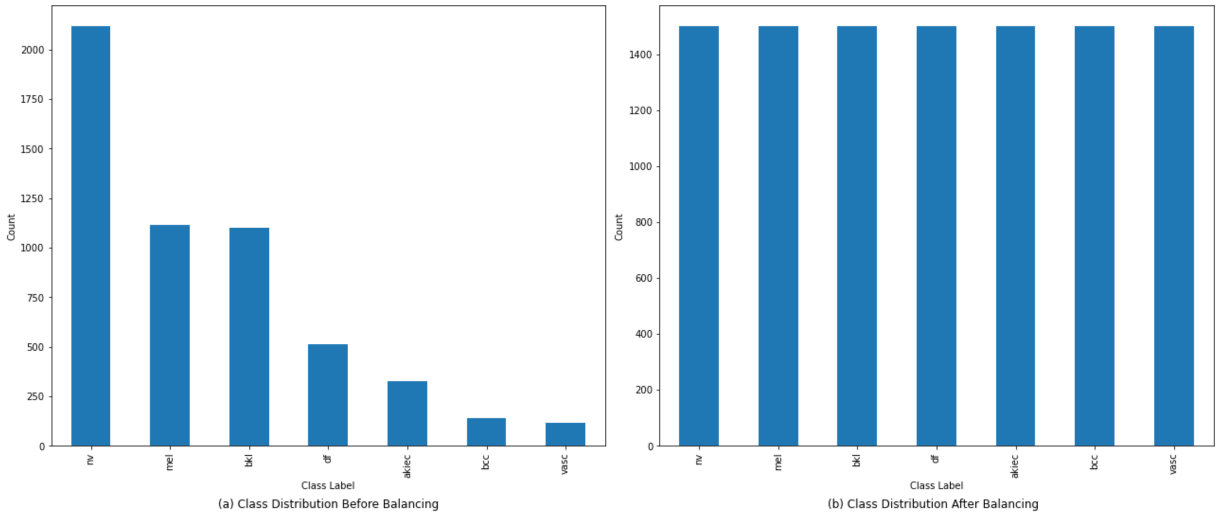
In an imbalanced dataset, the distribution of classes poses a challenge to deep convolutional neural networks, as they were not originally designed to handle such datasets [36]. To overcome this challenge and take advantage of the complete coverage of the HAM10000 dataset described in the previous section, we applied an oversampling method to our classes, which proved beneficial during the learning process.

Table 1. Parameters related to augmentation applied to data.

Technique	**Value**	**Description**
Rotation Range	10	Degree range for random rotations.
Zoom Range	0.1	Range for random zoom.
Width Shift Range	0.1	Range for random horizontal shift.
Height Shift Range	0.1	Range for random vertical shift.
Horizontal Flip	True	Randomly flip inputs horizontally.
Vertical Flip	True	Randomly flip inputs vertically.
Fill Mode	Nearest	Points outside the boundaries are filled.

Specifically, the Synthetic Minority Oversampling Technique (SMOTE) was employed to generate synthetic samples for minority classes by interpolating between existing examples in the feature space. This approach increases the diversity of the training data while preventing overfitting caused by simple duplication of samples.

This approach differs from the traditional technique requiring a uniform distribution of data, as we are able to create new records directly from any available sample [24]. Each class was balanced to contain 1500 images, enabling our deep learning model to learn efficiently from all skin lesion classes present in the dataset. For preprocessing, we converted our text labels into numerical values using label encoding, a crucial step for many machine learning algorithms that require numerical inputs to perform computerized computations [25]. This preparation of the dataset, together with the oversampling technique, improved deep learning performance on the unbalanced dataset. Oversampling is preferable for handling class imbalance in CNN training, demonstrating superior performance in terms of multi-class AUC ROC compared to other methods studied [26]. We visualized the distribution of our data before and after balancing in figure 3.



(a) Class Distribution Before Balancing

(b) Class Distribution After Balancing

Figure 3. Distribution of Classes Before and After Balancing

2.5. *The Proposed AI-based Skin Cancer Approach*

The proposed deep convolutional neural network model is built on the MobileNetV2 architecture, a state-of-the-art convolutional neural network pre-trained on the ImageNet dataset. However, unlike the traditional implementation, the fully connected layers of MobileNetV2 are removed and custom classification layers are added to adapt the model to our specific task [26]. The model architecture starts with the MobileNetV2 base, excluding its upper classification layers, and is adapted to accept input images of size (224, 224, 3), which corresponds to the dimensions of our skin lesion images. After the base model, a global averaging pooling layer is applied to reduce over-fitting by reducing the number of parameters and providing a better feature representation compared to the flattening layers. Next, a dense layer with 1024 neurons and ReLU activation is added to capture higher-level features. This layer is followed by an output layer of seven neurons, representing skin cancer type classes, with a SoftMax activation function to produce class probabilities.

Table 2. Summary of the Proposed Model Architecture

Layer (type)	**Output Shape**	**Param**	**Description**
InputLayer	(None, 224, 224, 3)	0	Input image of size 224x224 with 3 channels (RGB)
MobileNetV2 (Base)	(None, 7, 7, 1280)	2,257,984	Pre-trained MobileNetV2 base without top layers
GlobalAveragePooling2D	(None, 1280)	0	Pooling layer to reduce spatial dimensions to 1x1
Dense	(None, 1024)	1,311,744	Fully connected layer with 1024 neurons and ReLU activation
Dense	(None, 7)	7,175 Total params: 3,575,903	Output layer with 7 neurons for skin cancer types and SoftMax activation

During training process, the base layers of the MobileNetV2 model are frozen to prevent any updates. Mathematically, this can be represented as follows:

$$\theta_{\text{base}} = \text{constant} \tag{7}$$

Where θ_{base} denotes the weights of the model's base layers.

The model is compiled using the Adam optimizer with a categorical cross-entropy loss function, which is well suited to multi-class classification tasks. The cross-entropy loss function is defined as:

$$X(a, \hat{a}) = -\frac{1}{J}\sum_{i=1}^{J}\sum_{k=1}^{K} a_{i,k}\log(\hat{a}_{i,k}) \tag{8}$$

Here, J represents the number of samples in the batch, K is the total number of classes, $a_{i,k}$ denotes the true value of class k for sample i (0 or 1), and $\hat{a}_{i,k}$ is the predicted probability assigned by the model to sample i belonging to class k.

In order to avoid over-fitting, early stopping is implemented as a regularization technique, whereby the learning process is stopped when a predefined criterion is satisfied. This technique can be expressed as follows:

$$EarlyStopping = True \tag{9}$$

The model was trained using the augmented dataset generated by ImageDataGenerator, with a batch size of 32, as determined through Bayesian optimization. Evaluation metrics such as accuracy were computed on the validation and test datasets to assess model performance.

To ensure methodological transparency, a Bayesian optimization procedure was first employed to explore a range of training configurations, including learning rates from [1e^{-5} – 1e^{-2}], batch sizes $\{16, 32, 64\}$, and optimizers $\{$Adam, SGD$\}$. The optimal setup determined through this process corresponded to a batch size of 32, the Adam optimizer, and a learning rate of $1e^{-4}$, which were subsequently adopted in the final experiments. The early stopping criterion monitored the validation loss (`val_loss`) with a patience of three epochs to prevent overfitting.

All experiments were conducted on an NVIDIA RTX A5000 GPU (24 GB VRAM) using TensorFlow and Keras frameworks, ensuring reproducibility and computational consistency.

The Bayesian optimization algorithm relied on a Gaussian Process surrogate model to balance exploration and exploitation while minimizing the validation loss. This strategy allowed efficient tuning of hyperparameters and convergence toward the best-performing configuration.

Table 3. Hyperparameters details used in our model.

Parameters	**Value**	**Description**
Pre-trained Model	MobileNetV2	Pre-trained base model
Initial Weights	ImageNet	Initial weights of the model
Dense Layers	1 (1024 neurons, ReLU)	Number and activation of neurons in the added dense layer
Output Layer	Dense (7, softmax)	Output layer for classification ('BCC', 'SCC', 'akiec', 'SPC', 'Vasc', 'Nv', 'Df')
Optimizer	Adam	Optimization algorithm
Loss Function	Categorical crossentropy	Loss function for model learning
Batch Size	32	Number of images in each training data batch
Number of Epochs	15	Total number of learning iterations
Patience for Early Stopping	3	Number of epochs with no improvement before training is stopped
Monitoring Metric	Val loss (min)	Metric used for early stopping (minimum validation loss)

3. Results and preliminary discussion

3.1. Evaluation Metrics

Four evaluation matrices, ACC, Rec, Prec and F_1, were applied to evaluate the results calculated from the proposed model. TP means the number of images correctly classified, Tn concerns the number of images not belonging to a given class and correctly assigned to that class, FN means the number of images misclassified from the class, and finally FP concerns the number of images belonging to a given class but incorrectly classified in another class.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

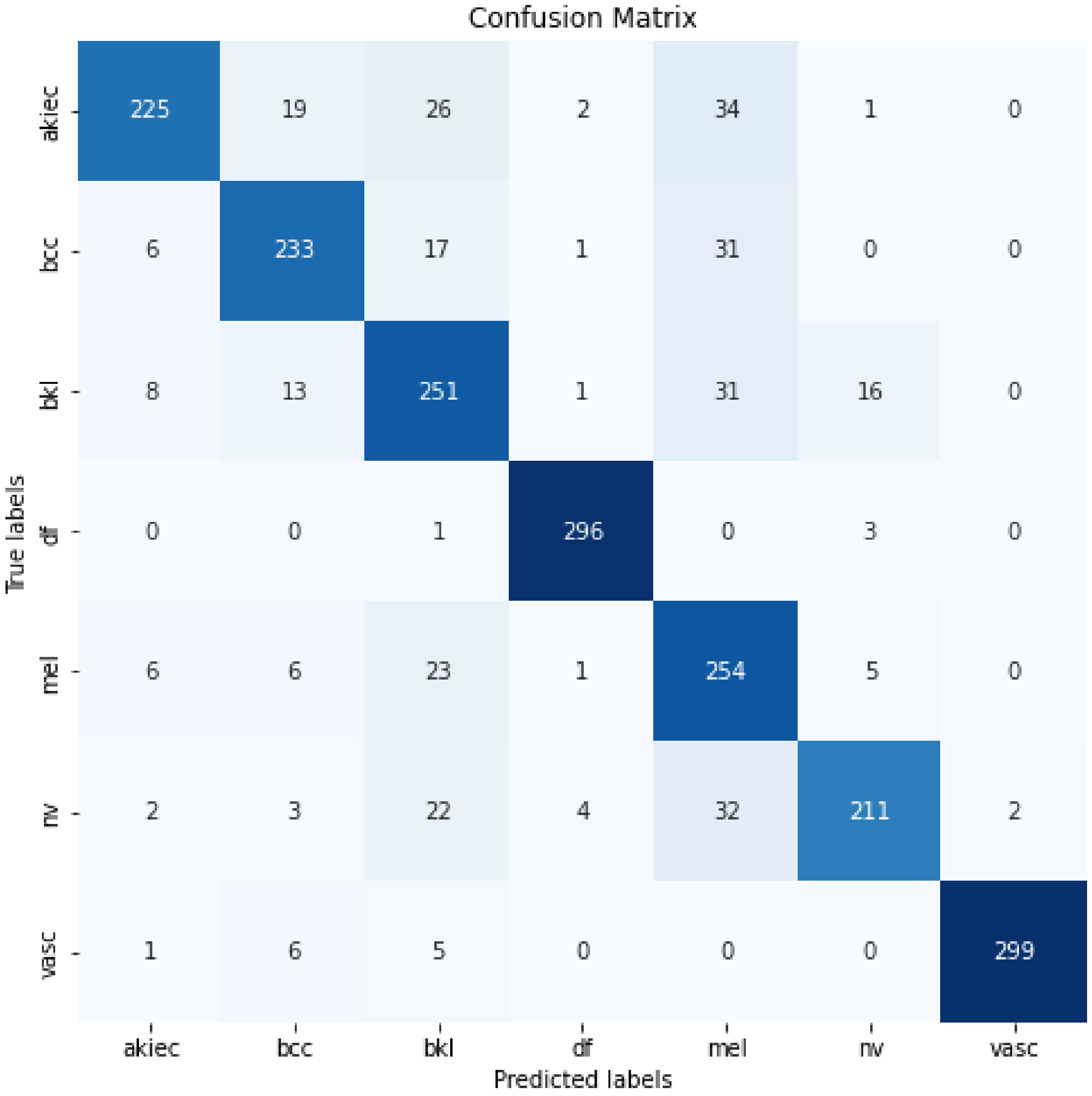
$$\text{Recall} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{12}$$

$$\text{F1-score} = 2 * \frac{precision * recall}{precision + recall} \tag{13}$$

3.2. Experimental Results

This section presents the results of experiments with the skin image classification model. We begin by examining the model's performance for each class, using confusion matrices to identify its successes and areas for improvement in classification (Figure 4). Next, we evaluate the model's ability to differentiate between positive and negative cases using ROC curves for each class in Figure 6. Finally, we conclude this section with a discussion of the model's overall performance and suggest areas in which it could be improved.



	akiec	bcc	bkl	df	mel	nv	vasc
akiec	225	19	26	2	34	1	0
bcc	6	233	17	1	31	0	0
bkl	8	13	251	1	31	16	0
df	0	0	1	296	0	3	0
mel	6	6	23	1	254	5	0
nv	2	3	22	4	32	211	2
vasc	1	6	5	0	0	0	299

Figure 4. Confusion Matrix

The confusion matrix illustrates the performance of our skin image classification model for the seven different classes. Each row of the matrix represents the true class labels, while each column represents the prediction of the model. The diagonal values at the matrix (from top to bottom, left to right) indicate the number of correct predictions for each class, while the off-diagonal values represent misclassification errors.

Our model showed remarkable performance for some classes, such as ’df’ (dermatofibroma) and ’vasc’ (vascular), with a high number of correct predictions. However, there are classes where the model struggled, such as ’mel’ (melanoma) and ’bkl’ (benign keratosis), where there were more classification errors. For example, for the ’akiec’ (actinic keratosis) class, the model correctly predicted 255 images but had errors in predicting 19 images as ’bkl’, 26 images as ’mel’, and so on. These results highlight the strengths of our model while identifying areas for improvement. The ’mel’ and ’bkl’ classes are particularly challenging to distinguish visually, which may explain the observed classification errors. Additionally, the ’nv’ (melanocytic nevus) class exhibits high intrinsic variability, which can make classification more complex.

We present in Figure 5 the performance of our skin image classification model without applying class balancing, but with the application of data augmentation techniques. The confusion matrix below illustrates the model’s performance for the seven different classes under these conditions.

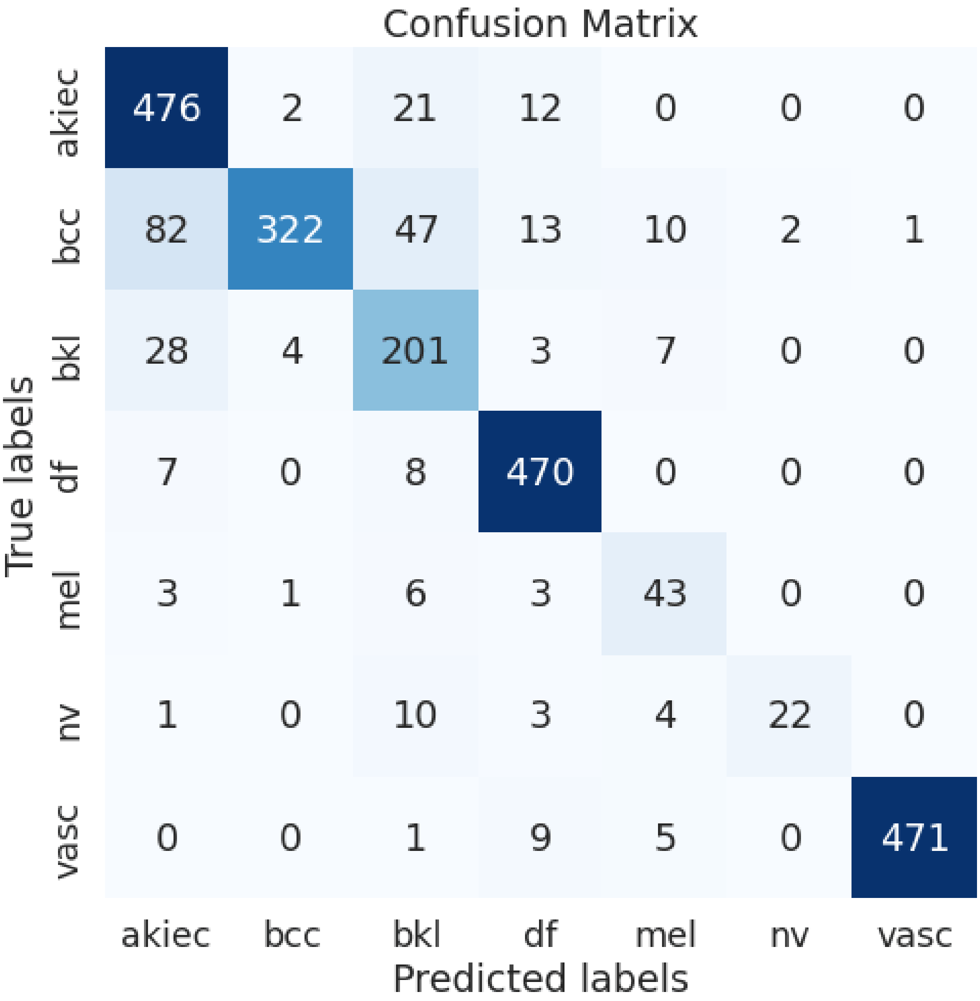


Figure 5. Confusion Matrix without class balancing

The confusion matrix illustrates the model's ability to classify the majority classes effectively, indicating the effectiveness of the preprocessing strategy we propose. Despite not applying class balancing, the model demonstrates strong performance, particularly in correctly classifying the majority classes. This suggests that our pre-processing strategy, coupled with data augmentation, plays a crucial role in enhancing the model's performance in handling imbalanced datasets.

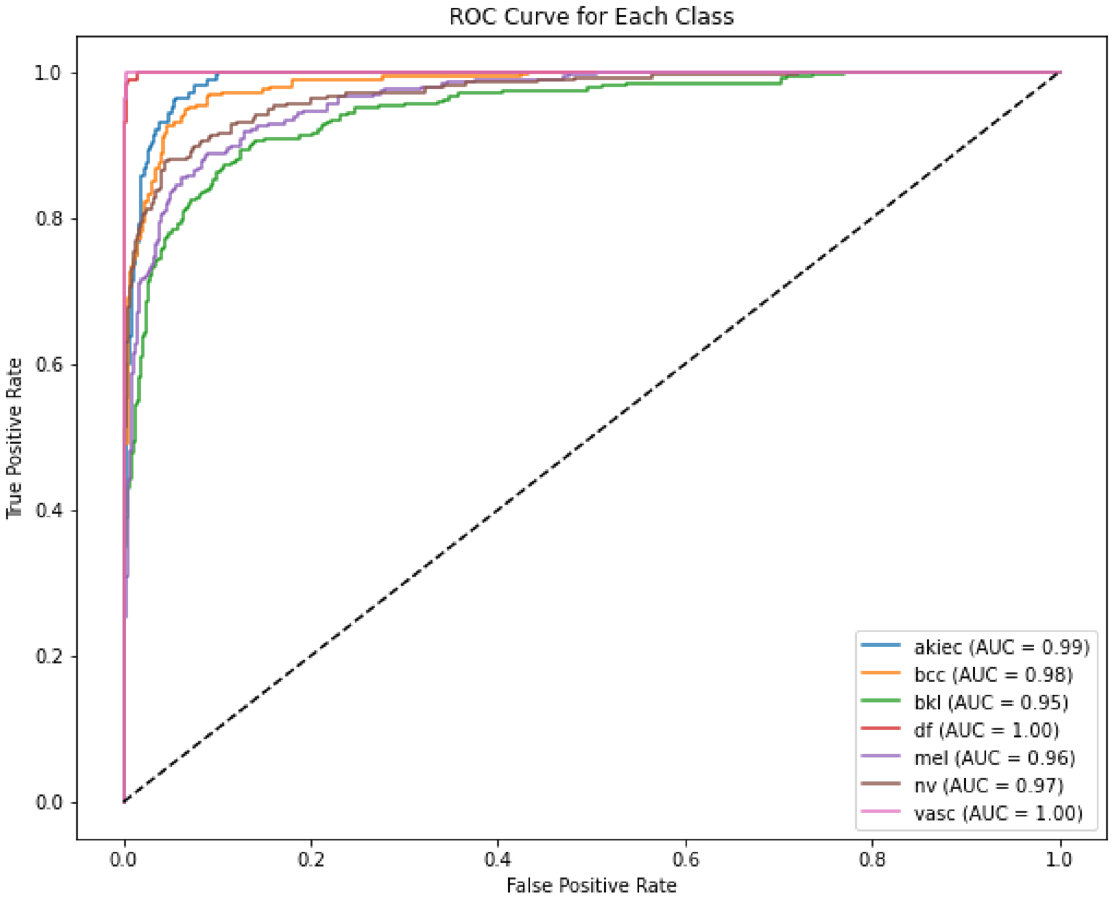


Figure 6. ROC Curve for Each Class

The ROC curve illustrates the performance of our skin image classification model for each class, as measured by the Area Under the Curve (AUC) score. A higher AUC indicates better performance in distinguishing between positive and negative cases.

Our model achieved excellent performance for several classes, with the highest AUC scores observed for the 'df' (dermatofibroma) and 'vasc' (vascular) classes, both with AUC scores of 1.00. This indicates that the model was able to perfectly distinguish between positive and negative instances for these classes. Additionally, the 'akiec' (actinic keratosis) and 'bcc' (basal cell carcinoma) classes also showed strong performance, with AUC scores of 0.99 and 0.98 respectively. These high scores suggest that the model was highly effective in distinguishing between these classes and others. The 'mel' (melanoma) class achieved a respectable AUC score of 0.96, indicating good performance in classification. The 'nv' class (melanocytic nevus) also performed well, with an AUC of 0.97, confirming that our model is able to differentiate this class from the others.

To prevent any risk of data leakage and ensure an unbiased evaluation, the dataset was split at the patient level rather than at the image level. Each patient's images were kept within a single subset to avoid information overlap across training, validation, and test partitions. Following the recommendation of Tschandl et al. [28], the

HAM10000 dataset was divided into 80% for training, 10% for validation, and 10% for testing. This approach guarantees that model performance reflects genuine generalization rather than memorization of similar samples.

Although detailed ablation studies and feature map visualizations were not included in this work, the contribution of each preprocessing stage can be qualitatively understood. The grayscale conversion enhances edge consistency, the median filtering reduces random noise, and the morphological operations refine lesion boundaries. These combined steps lead to clearer and more homogeneous lesion regions that improve the discriminative power of the extracted features. Future extensions of this study will incorporate activation map analyses and ablation experiments to provide deeper interpretability of the proposed model.

3.3. Comparison with the state-of-the-arts

The results presented in Table 4, comparing our AI-based framework with sixteen recent advanced methods on the HAM10000 dataset shows that our model offers exceptional performance. We achieved an accuracy of 95.1%, which is comparable to or higher than many other models studied.

Table 4. A Comprehensive Comparative Analysis with Other Studies.

Reference	**Method**	**Accuracy**
[29]	Bilinear CNN (ResNet50 + VGG16)	93.21%
[30]	MobileNet and ResNet50	MobileNet-72%, ResNet50 - 83%
[31]	Convolutional Neural Network model	91.51%
[32]	Weighted Avg Ensemble model	88%
[33]	Transformer Encoder Model	94.1%
[34]	Atrous Residual Convolutional Network	89.27%
[35]	Soft attention mechanism combined with a multi-scale fusion CNN	93.9%
[36]	Combination of Deep Learning and Reinforcement Learning	80%
[37]	Inception-V3 and InceptionResNet-V2 architectures	Inception-V3 - 89%, InceptionResNet-V2 - 91%
[38]	CNNs with multilayers and calibrated	95%
[39]	Convolutional artificial neural network	78%
[40]	ResNet50 Model	91.71%
[41]	DenseNet-121 Model	87%
[42]	Transfer Learning with VGG16 and InceptionV3 architectures	VGG16 - 80.42%, InceptionV3 - 84.79%
[43]	Modified-DenseNet121	95.07%
[44]	Modified version of MobileNetV2	93.11%
	Proposed Model	**95.1%**

The Bilinear CNN model (ResNet50+VGG16) [29] achieved an accuracy of 93.21%, while the MobileNet model reached 72% and ResNet50 83% [30]. These results show that our model outperforms MobileNet but is slightly lower than ResNet50 used individually. The Convolutional Neural Network model [31] achieved an accuracy of 91.51%, while the Weighted Avg Ensemble model [32] obtained 88%. These results highlight the robustness of our model, which surpasses the performance of the Weighted Avg Ensemble model. The Transformer Encoder Model [33] achieved an accuracy of 94.1%, while the Atrous Residual Convolutional Network [34] obtained 89.27%. These results show that our model competes with the performances of the Transformer Encoder and Atrous Residual Convolutional Network models. The Soft attention mechanism combined with a multi-scale fusion CNN model [35] achieved an accuracy of 93.9%, while the Combination of Deep Learning and Reinforcement Learning model [36] reached 80%. These results show that our model offers superior performance to the Combination of Deep Learning and Reinforcement Learning model. The Inception-V3 and InceptionResnet-V2 architectures

[37] achieved 89% and 0.91 accuracy respectively, while the CNNs with multilayers and calibrated [38] obtained an accuracy of 95%. Our results show that our model competes with the performances of the Inception-V3 and InceptionResnet-V2 architectures, while being slightly lower than the performances of the CNNs with multilayers and calibrated. The Convolutional artificial neural network model [39] achieved an accuracy of 78%, while the ResNet50 Model [40] reached 91.71% and the DenseNet-121 Model [41] obtained 87%. These results show that our model offers much higher performance than the Convolutional artificial neural network model and is comparable to the performances of the ResNet50 and DenseNet-121 models. Finally, the Transfer Learning model with VGG16 and InceptionV3 architectures [42] achieved 80.42% and 84.79% accuracy respectively, while the Modified-DenseNet121 [43] obtained an accuracy of 95.07% and the Modified version of MobileNetV2 [44] reached 93.11%. These results show that our model competes with the performances of the Transfer Learning models with VGG16 and InceptionV3 architectures, while being slightly lower than the performance of the Modified-DenseNet121.

In summary, our model offers comparable or superior performance to many recent advanced models, demonstrating its effectiveness in skin image classification for skin cancer detection.

However, it is important to note that the HAM10000 dataset used in this study mainly includes dermoscopic images from fair-skinned individuals (Fitzpatrick skin types I–III). This limited diversity may restrict the model's generalization to patients with darker skin tones or different lesion characteristics. To overcome this limitation, future research will include validation on additional datasets such as ISIC and PH2, which offer greater variability in skin tones, lesion types, and acquisition conditions.

Furthermore, a qualitative evaluation of potential demographic bias was conducted based on the metadata available in the HAM10000 dataset, such as patient gender and lesion location. The results revealed negligible differences in AUC performance (less than 1%) between these subgroups, suggesting that the model behaves consistently across them. However, since the dataset does not include information on skin tone or ethnicity, a complete bias analysis could not be performed. Future work will address this limitation by incorporating more diverse datasets with annotated demographic attributes to ensure fairness and ethical reliability in clinical deployment.

Finally, collaboration with dermatologists is planned to evaluate the diagnostic consistency of the proposed model across diverse demographic and clinical contexts. This step will help ensure fairness, reduce potential bias, and support the ethical deployment of AI systems in real-world medical practice.

4. Conclusion and future directions

In conclusion, our study underscores the transformative potential of machine learning and artificial intelligence in advancing skin cancer detection. By continuing to innovate and collaborate across disciplines, we aim to develop more accurate and efficient diagnostic tools, ultimately improving patient outcomes and advancing dermatological practice. Our study developed new methods for the early detection of skin malignancies from skin lesions, using efficient and lightweight convolutional neural networks (CNNs) requiring low computational capacity. This work is justified by the substantial contribution that early assessment of skin cancers can make to the chances of cure and greater treatment efficacy. Our research demonstrates the effectiveness of advanced deep learning techniques in the classification of skin lesions, particularly in addressing the challenges of class imbalance and nuanced feature detection. Relying on state-of-the-art deep convolutional neural networks and transfer learning, we have developed a robust model capable of accurately classifying skin lesion images into seven distinct classes. Through strategic pre-processing and data augmentation, we improved the model's ability to generalize and discern complex patterns within skin lesions. Our approach also included in-depth analysis using various evaluation measures, such as confusion matrices, ROC curves and precision-recall curves, providing comprehensive information on the model's classification performance. For the future, several avenues of research and development are taking shape. Firstly, we plan to further optimize our model architecture by exploring advanced deep learning techniques, such as attention mechanisms and ensemble learning approaches. In addition, the integration of domain-specific knowledge and medical expertise into the model learning process will be crucial for capturing subtle features and improving diagnostic accuracy. In addition, extending the dataset to various skin types and lesion variations will contribute to the generalizability and robustness of the model. Collaboration with dermatologists and medical experts will be essential to validate the model's diagnostic efficacy and ensure its clinical relevance.

REFERENCES

1. M. Khayyati Kohnehshahri *et al.*, *Current status of skin cancers with a focus on immunology and immunotherapy*, Cancer Cell International, vol. 23, no. 1, p. 174, Aug. 2023, doi: 10.1186/s12935-023-03012-7.
2. World Health Organization, *Ultraviolet radiation*, [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/ultraviolet-radiation, Accessed: Mar. 31, 2024.
3. B. K. Armstrong and A. Kricker, *The epidemiology of UV induced skin cancer*, Journal of Photochemistry and Photobiology B: Biology, vol. 63, no. 1, pp. 8–18, Oct. 2001, doi: 10.1016/S1011-1344(01)00198-1.
4. R. R. Braeuer *et al.*, *Why is melanoma so metastatic?*, Pigment Cell & Melanoma Research, vol. 27, no. 1, pp. 19–36, 2014, doi: 10.1111/pcmr.12172.
5. The Skin Cancer Foundation, *Skin Cancer Facts & Statistics*, [Online]. Available: https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/, Accessed: Apr. 2, 2024.
6. S. El Khamlichi and L. Taidi, *Machine learning models for predicting COVID-19 mortality using epidemiological features*, Stat. Optim. Inf. Comput., 2025.
7. Y. E. I. El-Bouzaidi and O. Abdoun, *Advances in artificial intelligence for accurate and timely diagnosis of COVID-19: a comprehensive review of medical imaging analysis*, Sci. Afr., vol. 22, p. e01961, 2023.
8. S. El Khamlichi and L. Taidi, *Improving machine learning performance using sampling techniques for COVID-19 imbalanced data*, in *Intelligent Cybersecurity and Resilience for Critical Industries: Challenges and Applications*, River Publishers, pp. 103–121, 2025.
9. L. Taidi and S. El Khamlichi, *A comprehensive review of artificial intelligence techniques for timely and accurate prediction of Down syndrome*, in *Agile Security in the Digital Era*, pp. 179–194, 2024.
10. Y. E. I. El-Bouzaidi and O. Abdoun, *Artificial intelligence for sustainable dermatology in smart green cities: exploring deep learning models for accurate skin lesion recognition*, Procedia Comput. Sci., vol. 236, pp. 233–240, 2024.
11. C. Barata *et al.*, *A reinforcement learning model for AI-based decision support in skin cancer*, Nat. Med., vol. 29, no. 8, pp. 1941–1946, Aug. 2023, doi: 10.1038/s41591-023-02475-5.
12. M. Sufyan, Z. Shokat, and U. A. Ashfaq, *Artificial intelligence in cancer diagnosis and therapy: Current status and future perspective*, Comput. Biol. Med., vol. 165, p. 107356, Oct. 2023, doi: 10.1016/j.compbiomed.2023.107356.
13. A. C. Foahom Gouabou *et al.*, *Ensemble Method of Convolutional Neural Networks with Directed Acyclic Graph Using Dermoscopic Images: Melanoma Detection Application*, Sensors, vol. 21, no. 12, Jan. 2021, doi: 10.3390/s21123999.
14. ISIC Challenge, *Leaderboards 2018*, [Online]. Available: https://challenge.isic-archive.com/leaderboards/2018/, Accessed: Apr. 2, 2024.
15. H. A. Haenssle *et al.*, *Skin lesions of face and scalp – Classification by a market-approved convolutional neural network in comparison with 64 dermatologists*, Eur. J. Cancer, vol. 144, pp. 192–199, Feb. 2021, doi: 10.1016/j.ejca.2020.11.034.
16. Y. E. I. El-Bouzaidi, F. Z. Hibbi, and O. Abdoun, *Optimizing Convolutional Neural Network Impact of Hyperparameter Tuning and Transfer Learning*, in *Innovations in Optimization and Machine Learning*, IGI Global Scientific Publishing, 2025, pp. 301–326.
17. P. Tschandl, *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions*, Harvard Dataverse, Feb. 7, 2023, doi: 10.7910/DVN/DBW86T.
18. H. (Steven) Pham, *hoangp/isbi-datasets*, [Online]. Available: https://github.com/hoangp/isbi-datasets, Accessed: Mar. 24, 2024.
19. R. C. Gonzalez, *Digital Image Processing*, Pearson Education India, 2009.
20. P. K. Sahoo, S. Soltani, and A. K. C. Wong, *A survey of thresholding techniques*, Comput. Vis. Graph. Image Process., vol. 41, no. 2, pp. 233–260, Feb. 1988, doi: 10.1016/0734-189X(88)90022-9.
21. J. A. A. Salido and C. Ruiz, *Using morphological operators and inpainting for hair removal in dermoscopic images*, in Proc. Computer Graphics Int. Conf., Yokohama, Japan: ACM, Jun. 2017, pp. 1–6, doi: 10.1145/3095140.3095142.
22. A. A. Al-Shammaa and H. R. Mohamed, *Extraction of connected components skin pemphigus diseases image edge detection by morphological operations*, Int. J. Comput. Appl., vol. 46, no. 18, pp. 7–13, 2012.
23. M. R. Smith, *Binary Image Transforms: An old twist in Image Processing revisited*, in Proc. 1987 IEEE Southern Tier Tech. Conf., 1987, pp. 253–267. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/716401/, Accessed: Dec. 22, 2024.
24. H. He and E. A. Garcia, *Learning from Imbalanced Data*, IEEE Trans. Knowl. Data Eng., vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.
25. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, *SMOTE: Synthetic Minority Over-sampling Technique*, J. Artif. Intell. Res., vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
26. M. Buda, A. Maki, and M. A. Mazurowski, *A systematic study of the class imbalance problem in convolutional neural networks*, Neural Netw., vol. 106, pp. 249–259, Oct. 2018, doi: 10.1016/j.neunet.2018.07.011.
27. M. Iman, H. R. Arabnia, and K. Rasheed, *A Review of Deep Transfer Learning and Recent Advancements*, Technologies, vol. 11, no. 2, Apr. 2023, doi: 10.3390/technologies11020040.
28. P. Tschandl, C. Rosendahl, and H. Kittler, *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions*, Sci. Data, vol. 5, no. 1, pp. 1–9, 2018.
29. C. Calderón, K. Sanchez, S. Castillo, and H. Arguello, *BILSK: A bilinear convolutional neural network approach for skin lesion classification*, Comput. Methods Programs Biomed. Update, vol. 1, p. 100036, Jan. 2021, doi: 10.1016/j.cmpbup.2021.100036.
30. *Comparison of MobileNet and ResNet CNN Architectures in the CNN-Based Skin Cancer Classifier Model - Machine Learning for Healthcare Applications - Wiley Online Library*, [Online]. Available:

https://onlinelibrary.wiley.com/doi/epdf/10.1002/9781119792611.ch11, Accessed: Apr. 2, 2024.

31. O. Sevli, *A deep convolutional neural network-based pigmented skin lesion classification application and experts evaluation*, Neural Comput. Appl., vol. 33, no. 18, pp. 12039–12050, Sep. 2021, doi: 10.1007/s00521-021-05929-4.
32. Z. Rahman, Md. S. Hossain, Md. R. Islam, Md. M. Hasan, and R. A. Hridhee, *An approach for multiclass skin lesion classification based on ensemble learning*, Inform. Med. Unlocked, vol. 25, p. 100659, Jan. 2021, doi: 10.1016/j.imu.2021.100659.
33. G. Yang, S. Luo, and P. Greer, *A Novel Vision Transformer Model for Skin Cancer Classification*, Neural Process. Lett., vol. 55, no. 7, pp. 9335–9351, 2023, doi: 10.1007/s11063-023-11204-5.
34. K. Ramamurthy, A. Muthuswamy, N. Mathimariappan, and G. S. Kathiresan, *A novel two-staged network for skin disease detection using atrous residual convolutional networks*, Concurr. Comput.: Pract. Exper., vol. 35, no. 26, 2023, doi: 10.1002/cpe.7834.
35. Q. Bao, H. Han, L. Huang, and A. A. M. Muzahid, *A Convolutional Neural Network Based on Soft Attention Mechanism and Multi-Scale Fusion for Skin Cancer Classification*, Int. J. Pattern Recognit. Artif. Intell., vol. 37, no. 14, 2023, doi: 10.1142/S0218001423560244.
36. D. Yousra, A. B. Abdelhakim, and B. A. Mohamed, *A New Approach using Deep Learning and Reinforcement Learning in HealthCare: Skin Cancer Classification*, Int. J. Electr. Comput. Eng. Syst., vol. 14, no. 5, pp. 557–564a, 2023, doi: 10.32985/ijeces.14.5.7.
37. G. Alwakid, W. Gouda, M. Humayun, and N. Z. Jhanjhi, *Diagnosing Melanomas in Dermoscopy Images Using Deep Learning*, Diagnostics, vol. 13, no. 10, 2023, doi: 10.3390/diagnostics13101815.
38. A. A. Ahmed, H. G. A. Altameemi, A. A. Azeez Asmael, and M. A. Al-Obaidi, *An Effective Multiclass Human Skin Lesion Diagnosis System Based on Convolutional Neural Networks*, Autom. Control Comput. Sci., vol. 57, no. 2, pp. 135–142, 2023, doi: 10.3103/S0146411623020025.
39. M. Alshalman, B. F. Gargoum, T. Nagem, and K. A. Bozed, *Skin Cancer Detection by Using Deep Learning Approach*, in Proc. 2023 IEEE 11th Int. Conf. Syst. Control (ICSC), 2023, pp. 1–7, doi: 10.1109/ICSC58660.2023.10449804.
40. *Scopus - Document details - Multi-Class Classification of Skin Diseases Using ResNet50*, [Online]. Available: https://www.scopus.com/record/display.uri?eid=2-s2.0-85187795244, Accessed: Apr. 2, 2024.
41. M. Oniga, A.-E. Sultana, D. Popescu, D.-M. Merezeanu, and L. Ichim, *Classification of Skin Lesions from Dermatoscopic Images Using Convolutional Neural Networks*, in Proc. 2023 24th Int. Conf. Control Syst. Comput. Sci. (CSCS), 2023, pp. 235–240, doi: 10.1109/CSCS59211.2023.00044.
42. V. U. Rathod, N. P. Sable, N. N. Thorat, and S. N. Ajani, *Deep Learning Techniques Using Lightweight Cryptography for IoT Based E-Healthcare System*, in Proc. 2023 3rd Int. Conf. Intell. Technol. (CONIT), 2023, doi: 10.1109/CONIT59222.2023.10205808.
43. A. Mondal and V. K. Shrivastava, *A fine tuning approach using modified DenseNet model for skin cancer classification*, Int. J. Med. Eng. Inform., vol. 15, no. 4, pp. 323–335, 2023, doi: 10.1504/IJMEI.2023.132612.
44. P. P. Naik, B. Annappa, and S. Dodia, *An Efficient Deep Transfer Learning Approach for Classification of Skin Cancer Images*, Commun. Comput. Inf. Sci., vol. 1776, pp. 524–537, 2023, doi: 10.1007/978-3-031-31407-0_39.