

Article

Not peer-reviewed version

Enhancing Dermatological Diagnostics with EfficientNet: A Deep Learning Approach

[Ionela Manole](#)^{*}, [Alexandra-Irina Butacu](#), [Raluca Nicoleta Bejan](#), [George Sorin Tiplica](#)

Posted Date: 16 July 2024

doi: 10.20944/preprints202407.1322.v1

Keywords: artificial intelligence; benign lesions; classification; malignant lesions; neural networks; transfer learning



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Enhancing Dermatological Diagnostics with EfficientNet: A Deep Learning Approach

Ionela Manole ^{1,*}, Alexandra- Irina Butacu ¹, Raluca Nicoleta Bejan ² and George-Sorin Tiplica ¹

¹ 2nd Department of Dermatology, Colentina Clinical Hospital, Bucharest, Romania; e-mail ionela.manole@drd.umfcd.ro; adabutacu@gmail.com; george.tiplica@umfcd.ro

² AWS Certified Machine Learning Specialist, TensorFlow Certified Developer, Bucharest, Romania; e-mail raluca.ioan@gmail.com

* Correspondence: ionela.manole@drd.umfcd.ro; Tel.: 0040740014203

Abstract: Background: Despite recent advancements, medical technology has not yet reached its peak. Precision medicine is growing rapidly, thanks to machine learning breakthroughs powered by increased computational capabilities. This article explores a deep-learning application for computer-aided diagnosis in dermatology. **Methods:** Using a custom model based on EfficientNetB3 and deep learning, we propose an approach for skin lesion classification that offers superior results with smaller, cheaper, and faster inference times compared to other models. The skin images dataset used for this research includes 8,222 files selected from the authors' collection and the ISIC2019 archive, covering six dermatological conditions. **Results:** The model achieved 95.4% validation accuracy on four categories—melanoma, basal cell carcinoma, benign keratosis-like lesions, and melanocytic nevi—using an average of 1,600 images per category. Adding two categories with fewer images (about 700 each)—squamous cell carcinoma and actinic keratoses—reduced the validation accuracy to 88.8%. The model maintained accuracy on new clinical test images taken under the same conditions as the training dataset. **Conclusions:** The custom model demonstrated excellent performance on the diverse skin lesions dataset, with significant potential for further enhancements.

Keywords: artificial intelligence; benign lesions; classification; malignant lesions; neural networks; transfer learning

I. Introduction

As the most common cancer in the world, skin cancer has increased dramatically over the last few decades. Skin cancer is divided by dermatologists into two major types: non-melanoma skin cancers (NMSC), and melanoma, recognized for its aggressive behavior and risk to metastasize. While occurring less frequently than NMSC, melanoma is responsible for the majority of skin cancer mortality, pointing out the urgent need for accurate diagnosis [1].

The most common type of skin cancer is NMSC, which includes mainly basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). BCC rarely metastasizes but is associated with significant morbidity in those cases that are left untreated or treated with delay. SCC, although more aggressive than BCC in terms of metastasis, usually responds well to treatment if detected and treated on time. Increased incidence of NMSC has been associated with higher ambient UV, related to ozone depletion, and increased public awareness and screening activities [2].

Melanoma, although a relatively rare form of skin cancer, associates a highly metastatic behavior. Early detection is of extreme importance to support a high survival rate in effective melanoma management. The five-year survival rate for patients with early-stage melanoma is over 90%, but this figure decreases significantly with the higher stage at diagnosis [3].

Benign skin lesions, including actinic keratosis (AK), benign keratosis-like lesions, and melanocytic nevi, are widespread and typically non-malignant. Although with benign behavior, clinical differentiation from their malignant counterpart is sometimes challenging even for an

experienced dermatologist. When distinguishing malignant from benign disease, this diagnostic challenge may result in either unnecessary invasive procedures or delayed interventions.

The differential diagnosis of skin lesions is a complex task that reflects a cumulative knowledge of their morphology, distribution, and evolution over time. For malignant and benign skin lesions, dermoscopy, a technique that permits improved visualization, has provided a great deal of diagnostic accuracy. However, dermoscopic analysis can be subjective and experience-dependent, which can result in diagnostic variability and outcomes [4].

Early and accurate diagnosis of skin lesions is of utmost importance for treatment outcomes, especially when considering skin cancer. According to the World Health Organization, 1.5 million cases of skin cancers were reported in 2022, of which 330,000 were melanoma cases, leading to almost 60,000 deaths [5]. Therefore, the precise and timely diagnosis of skin lesions is essential, as it directly impacts patient management and prognosis. Although established diagnostic methods are effective, they can be subjective when using clinical evaluation as the main diagnostic tool. Such heterogeneity highlights the importance of moving toward diagnostic tools that are less subjective and more repeatable and scalable, which can also enhance early detection and improve clinical outcomes.

Today, the diagnostic landscape in dermatology is changing through the great impact of emerging technologies with depth in science, such as deep learning (DL) and artificial intelligence (AI). DL, a subset of machine learning (ML) has emerged as a powerful tool that is predicted to be a game changer in multiple fields of medicine, including dermatology. Computers and improved ML models can now solve hard, complicated diagnostic tasks with high accuracy. AI models, particularly when trained on large-scale databases, have a potential ability similar or superior to dermatologists in skin lesion classification. Moreover, these AI systems have a major impact in providing standardization in diagnostic interpretations and in the reduction of the inter-observer variability and frequency of diagnostic errors [6].

This consistent, repeatable interpretation may revolutionize skin lesion diagnostics as AI becomes integrated into clinical practice. Further, such integration allows for enhanced distribution of dermatologic expertise to underserved regions, democratizing access to expert-level diagnosis and care. To accommodate the vast diversity of skin types and conditions we can expect to handle in clinical practice, a durable deployment of AI tools in dermatology may require extensive validation and continuous training on multiple datasets [7].

Up to the present, in dermatology, DL models have outperformed standard diagnostic approaches in several studies. For instance, research performed by Esteva et al. (2017) has demonstrated that convolutional neural networks (CNNs) can classify skin cancer with a level of competence comparable to dermatologists, using vast datasets of dermoscopic images [8]. Similarly, Codella et al. (2017), reported that ensemble DL could remarkably improve melanoma detection in dermoscopy images [9]. In a more recent paper from Brinker et al. (2019), deep neural networks were found to outperform dermatologists after being trained on images of melanomas, indicating the potential of these models for clinical decision-making applications [10]. Moving forward, Liu et al. (2020) also reported an AI attempt to diagnose skin diseases with a differential diagnostic accuracy comparable to dermatologists [11]. These advances in AI are not only being transformed into plain research findings. Still, they are also being manifested into practical application results that could make use of saving lives through real clinical operations. These echo the belief within the medical field that AI, specifically DL algorithms can vastly improve both the accuracy and the consistency of diagnostic tests.

Nonetheless, as these advanced AI tools are being considered to be included in the clinic, significant challenges with deployment remain, driven by requirements in extensive training data and generalization across diverse patient populations. This includes also the validation process intended to demonstrate accuracy and reliability.

In this study, EfficientNetB3, an architecture that achieves state-of-the-art accuracy in image classification, was used. Developed by Tan and Le (2019), it is the ideal type of architecture for medical image analysis as it performs the best in benchmarks considering both size and computational efficiency [12]. The work presented in this article focuses on the usage of a custom

model based on EfficientNetB3 that uses transfer learning to analyze skin lesions on an extended dataset. This study aims to bridge the gap between the rapid advancements in technology and the clinical utility, in a scalable and efficient fashion to improve diagnostic accuracy and speed in dermatology by applying advanced AI methods. The model was trained and validated across six classes of skin diseases, encompassing benign and malignant conditions such as melanoma, basal cell carcinoma, squamous cell carcinoma, actinic keratoses, benign keratosis-like lesions, and melanocytic nevi.

By integrating the latest AI methods, progress can be achieved, such as improving diagnosis or reducing clinicians' workload, allowing more time for patient care and less for routine diagnosis. Combining AI strengths on the technical side with the nuanced understanding of skin pathology from expert dermatologists, we have created an interdependent mix that better the accuracy and applicability of the diagnostic process.

There are several novel contributions introduced to the dermatological research and clinical practice:

- Ranks among the top-performing models in the European region, indicating its potential to effectively address regional medical challenges.
- Achieves competitive results with our custom model based on EfficientNetB3, demonstrating efficient utilization of limited training data.
- Enhances practical feasibility and cost-effectiveness of deployment due to its modest computational requirements.
- It shows robust performance with fewer images compared to models that achieve similar or better results with larger datasets.

The following part of the study is organized as follows: Section II details the training dataset, data preprocessing steps, the model architecture and hyper-parameters, and techniques to combat overfitting, along with the training and validation processes. Section III presents the model's performance. Section IV interprets the results, highlighting their implications and limitations. Finally, the conclusion summarizes the findings and suggests directions for future research.

II. Methodology of Research

2.1. Description of the Training Dataset

The primary dataset used for training is a proprietary combination of files collected from the authors' collection of images and the International Skin Imaging Collaboration 2019 (ISIC2019) archive [13]. The dataset used in the study comprises a total of 8,222 images, covering six categories of dermatological lesions. These categories include three malignant skin conditions- melanoma, BCC, and SCC—and three non-malignant conditions—AK, benign keratosis-like lesions, and melanocytic nevi (Figure 1).

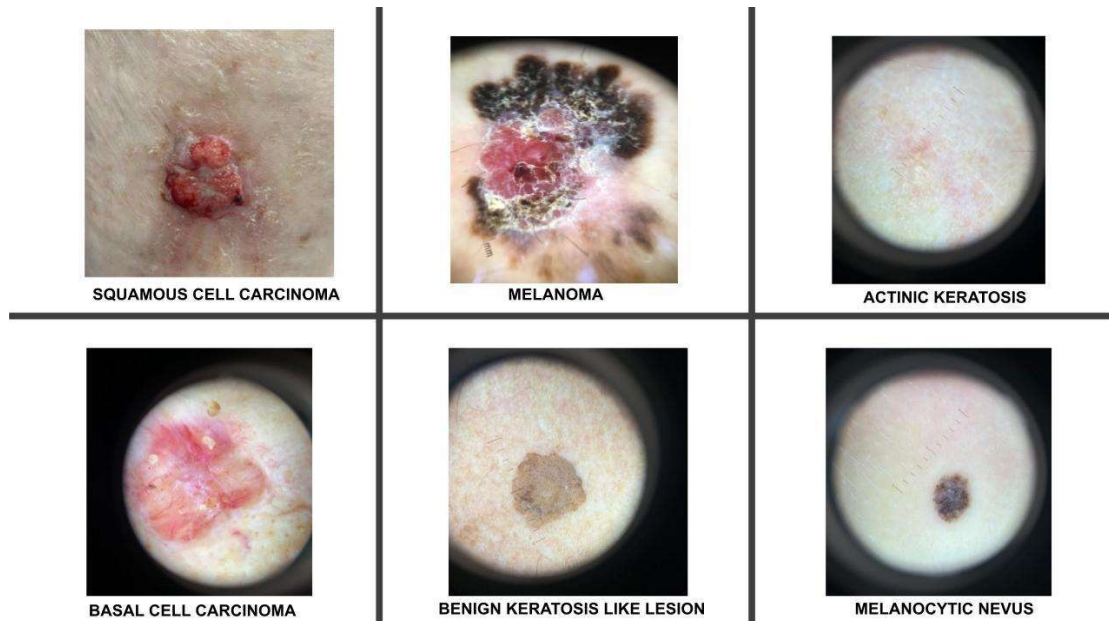


Figure 1. Examples of clinical and dermoscopic images used for training.

The data was divided into training, validation, and testing sets, with the training set consisting of 80% of the data (6,578 images) and the validation and test sets containing 20% (1,644 images). Table 1 includes the distribution of all images across the disease categories.

Table 1. Distribution of images across the skin conditions.

Classes	No of Images	No of augmented images	Total
Melanoma	1655	489	2144
BCC	1811	333	2144
Benign Keratosis-like lesions	1663	481	2144
Melanocytic Nevi	1686	458	2144
SCC	606	1538	2144
AK	801	1343	2144
Total	8222	4642	12864

Dermoscopic and close-up images were included in the dataset, ensuring that each image consistently belonged to one of the specified categories. The EfficientNetB3 model was initially trained on the ImageNet image database [14] using a resolution of 300x300 pixels, keeping the RGB color model. Consequently, all images in this project were resized to 300x300 pixels. This resizing process maintained the defining properties and shapes of the lesions, ensuring no distortion occurred. Additionally, this standardization reduced the dataset's size, enabling faster model training.

2.2. Data Pre-Processing

Before feeding images into the neural network model, the following preprocessing steps were applied to optimize model performance. This preparation ensured efficient and effective learning:

- Resizing: Images were resized to 300x300, ensuring uniformity in input sizes.
- Normalization: Pixel values were scaled to the [0, 1] interval to reduce data discrepancies and aid training convergence.

- Mean Subtraction and Standardization: Each pixel's value had the dataset's mean subtracted and was then divided by the standard deviation to further normalize the data, enhancing model convergence.
- Data Augmentation: This technique creates new images by modifying existing ones. We employ mirroring, translation, rotation, scaling, brightness adjustment, and noise addition to augment the existing pictures [15]. These augmented images are then added to the categories with less data, thereby balancing the training dataset.

2.3. The Usage and the Architecture of the Model

The EfficientNet neural network class of models, first proposed by Tan and Le [12], is a family of image classification models known for its feature extraction capabilities, that achieved very good accuracy while being 8.4x smaller and 6.1x faster on inference than the best existing ConvNet neural networks [16]. The decision to use a model based on EfficientNetB3 was driven by its effective balance between classification performance and resource efficiency. Compared to larger models like EfficientNet-B4, B5, and B6, EfficientNet-B3 offers competitive classification accuracy while requiring fewer parameters. This efficiency translates to reduced demands on memory and CPU usage, making it a practical choice for achieving high performance in classification tasks without excessive computational costs.

In this study, we developed a customized model based on EfficientNetB3, using transfer learning [17,18], and leveraged Imagenet to reduce computational cost and carbon footprint. We used TensorFlow [19] and Keras [20] to integrate EfficientNet with new dense layers, allowing the model to handle complex image classification tasks. The dense layers learned high-level representations and made final predictions based on features extracted by the convolutional layers, enhancing the model's ability to capture complex relationships and improve prediction accuracy. Fine-tuning our dataset further adapted the model to our project's needs. The architecture of the custom model includes the added layers described below and is represented in Figure 2:

- On top of the EfficientNetB3 model, adding a Batch Normalization [21] layer improved accuracy by enhancing convergence and helped reduce overfitting. Batch Normalization contributed to smoother training and improved generalization on unseen data by stabilizing and normalizing activations throughout the network.
- Two additional dense layers significantly enhanced classification performance by introducing non-linear features, extracting higher-level features, reducing parameter count and dimensionality of input images, and serving as a regularization technique.
- Finally, one dropout layer [22] randomly deactivated the neurons during training, which helped prevent overfitting by encouraging the model to generalize better. This technique improves the robustness and performance of the neural network on unseen data.

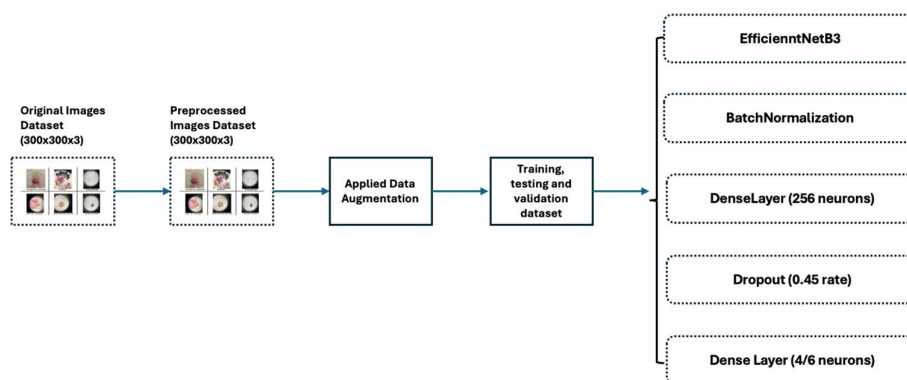


Figure 2. The architecture and the setup of the model.

2.4. Training and Validating the Model

During the study, we applied a sum of strategies and techniques to optimize our model's training process and performance. These included leveraging pre-trained weights from the 'Imagenet' dataset [14] to provide a strong starting point for learning and utilizing mixed precision policy [23] to enhance computational efficiency. We applied patience, stop patience, [24] and learning rate reduction [25] to fine-tune the training process, preventing overfitting and ensuring efficient learning rate adjustments. Transfer learning and unfreezing [26] allowed us to adapt the pre-trained model to our specific task, ensuring improved convergence and flexibility. Additionally, we saved the best weights [27] to guarantee optimal model performance for future use. Batch training [28] improved computational efficiency and promoted stable convergence through batch normalization.

2.4.1. Hyperparameters

The model's hyperparameters, listed in Table 2, fundamentally model the training process and overall performance, highlighting the importance of selecting the appropriate values.

Table 2. The hyperparameters of the model.

Hyperparameters	Values
Learning Rate	0.001
Batch size	32
Number of Epochs	19
Optimizer	Adamax
Dropout Rate	0.45
Activation Functions	Relu, Softmax
Regularization Parameters	Kernel Regularizer: L2 regularization with strength 0.016 Activity Regularizer: L1 regularization with strength 0.006 Bias Regularizer: L1 regularization with strength 0.006
Loss Function	Categorical Cross Entropy
Augmentation techniques	Rotate, Scale, Flip, Zoom

Training a custom model for skin lesion classification using a relatively small dataset poses challenges due to noisy gradients.

We chose the Adamax optimizer [29] for its robustness against gradient fluctuations using the infinity norm, leading to stable parameter updates. This helped to improve the performance of our research on skin lesion classification, where image variability affects gradient consistency. Adamax handles maximum gradients better than SGD or Adam, supporting smoother convergence and enhanced generalization.

A learning rate of 0.001 was initially chosen and adjusted during the training as a function of the validation accuracy and loss to facilitate model convergence. The optimal batch size was set at 32, influencing convergence speed and memory demands throughout the training process. The Categorical Cross-Entropy loss and Relu activation functions were employed for multi-class classification. The model demonstrated the best performance by adopting the combination of the fine-tuned hyperparameters listed above.

2.4.2.. Techniques Used to Combat Overfitting

Overfitting happens when a model is overtrained on its training data, leading it to perform poorly on new data. Essentially, the model strives to be as accurate as possible and it focuses too much on fine details and noise within its training dataset. These attributes are often not present in real-world data, so the model tends to not perform well. Overfitting can also occur when a model is too complex, relative to the amount of data. This can lead the model to hyper-focus on details present in the given data that may not be relevant to the general patterns the model must develop. Overfitting gives the illusion that a model is performing well, even though it has failed to make proper generalizations about the data provided [30].

To prevent overfitting, there were used several techniques, described below:

- *Dropout*: Dropout selectively deactivates neurons in neural network layers during training, simulating smaller networks within the model. This approach encourages the network to diversify its learning strategies, enhancing generalization and mitigating overfitting by preventing reliance on individual neurons. [31].
- *Batch Normalization*: Normalization adjusts data to a mean of zero and a standard deviation of one, aligning and scaling inputs. Batch Normalization speeds up training by preventing gradients from becoming too small, facilitating faster convergence with higher learning rates. It also acts as a regularizer, reducing overfitting and improving model generalization on new data. This stability reduces sensitivity to initial weight choices and simplifies experimenting with different architectures. [32].
- *Regularization*: We used the regularization techniques to reduce overfitting: L2 regularization with a strength of 0.016 for the kernel and L1 regularization at a strength of 0.006 for both activity and bias regularization. These methods were chosen to mitigate overfitting by penalizing large parameter values in the model, thereby promoting more straightforward and more generalized outcomes across varying datasets and scenarios.

III. Results

3.1. Training and Validation Accuracy and Loss

Monitoring training and validation accuracy and loss provides insights into how well a machine learning model generalizes. Lower training loss and higher accuracy indicate effective learning on seen data, while validation metrics assess performance on unseen data, ensuring the model's robustness and generalization. Figure 3 shows an accuracy of 95.4% when our custom model was trained on four classes, whereas Figure 4 shows an accuracy of 88.8% when our custom model was trained on six classes.

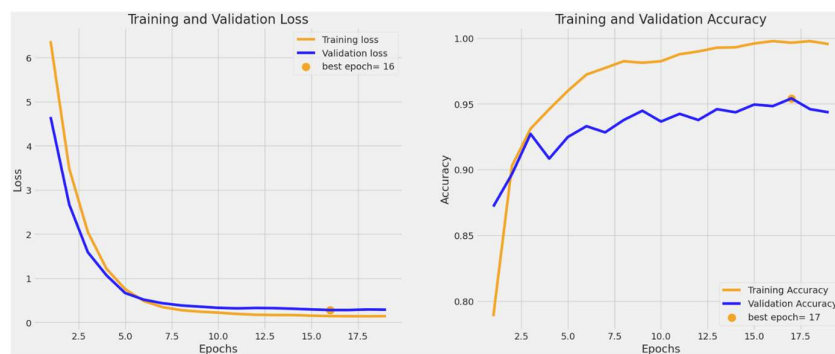


Figure 3. The validation accuracy and loss for the BCC, Benign Keratosis-like lesions, Melanocytic nevi and Melanoma classes

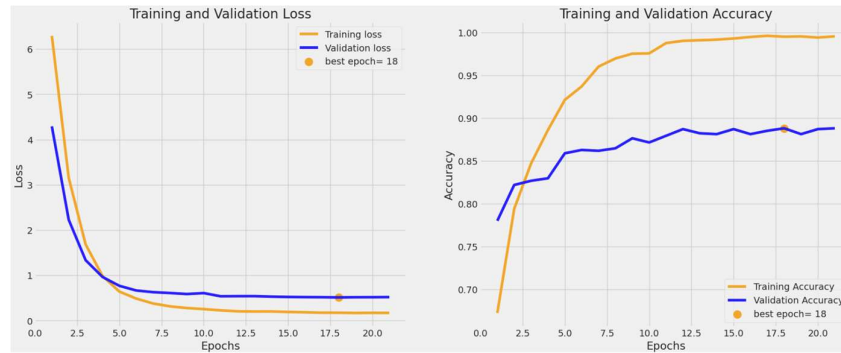


Figure 4. The validation accuracy and loss for BCC, Benign Keratosis-like lesions, Melanocytic nevi, Melanoma, SCC, and AK classes

3.2. Classification Performance

The test results show the performance metrics of the classification model, side by side, across the four different classes: BCC, benign keratosis-like lesions, melanocytic nevi, and melanoma (Table 3), and across the six different classes, the initial four, plus AK and SCC (Table 4). Performance metrics reported are precision (proportion of true positive predictions out of all positive predictions made by the model), recall (proportion of true positive predictions out of all actual positive cases in the data) and F1-score (harmonic mean of precision and recall, providing a single metric that balances both precision and recall). The model's overall accuracy, which measures the proportion of correctly classified instances out of the total instances and provides an overall assessment of the model's performance across all classes, is 95.4% (4 classes)/ 88.8% (6 classes).

Table 3. Testing results for four categories.

	Precision	Recall	F1-score	Support
Basal cell carcinoma	0.94	0.98	0.96	225
Benign keratosis-like lesions	0.94	0.89	0.91	208
Melanocytic nevi	0.95	0.97	0.96	210
Melanoma	1.00	0.99	1.00	207
Accuracy			0.96	850
Macro Avg	0.96	0.96	0.96	850
Weighted Avg	0.96	0.96	0.96	850

Table 4. Testing results for six categories.

	Precision	Recall	F1-score	Support
Actinic keratosis	0.74	0.77	0.75	100
Basal cell carcinoma	0.87	0.84	0.85	227
Benign keratosis-like lesions	0.85	0.85	0.85	208
Melanocytic nevi	0.94	0.97	0.96	210
Melanoma	1.00	1.00	1.00	207
Squamous cell carcinoma	0.69	0.54	0.61	76
Accuracy			0.89	1028

Macro Avg	0.85	0.84	0.85	1028
Weighted Avg	0.89	0.89	0.89	1028

3.3. Receiver Operating Characteristic (ROC) Curve

Figures 5 and 6 show the proposed model's ROC curves. The curve shows excellent performance, with AUC values ranging from 0.98 to 1.0 across four classes and from 0.93 to 1 across six classes. These results are relevant to the model's accuracy and promising clinical application potential.

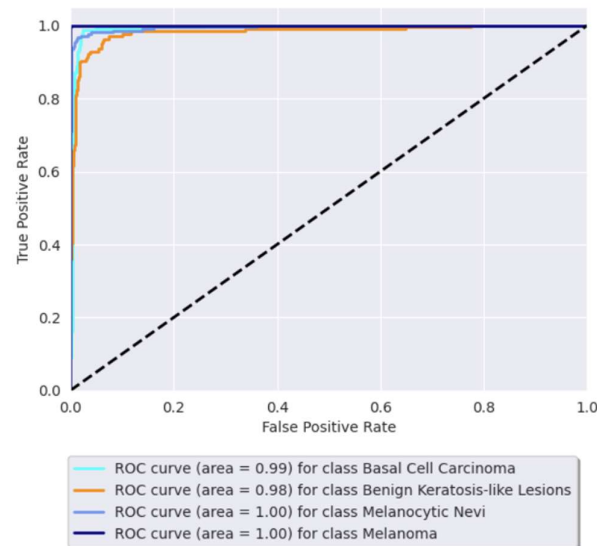


Figure 5. The ROC curve for the BCC, benign keratosis-like lesions, melanocytic nevi, and melanoma classes.

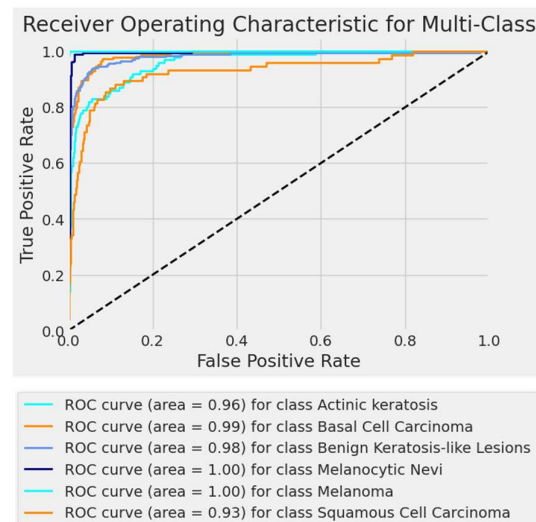


Figure 6. The ROC curve for BCC, benign keratosis-like lesions, melanocytic nevi, melanoma, SCC, and AK classes.

3.4. Confusion Matrix and Errors by Class

Figure 7 and Figure 8 below show the confusion matrix of the custom model for the test with four and, respectively, six classes. The confusion matrix visually represents the classification results of the test dataset per each class. The majority of the examples per class, represented on the diagonal

of the matrix, were accurately classified, i.e., for Melanoma, 205 images were correctly classified out of 207. The number of errors for each class is shown in Figures 9 and 10.

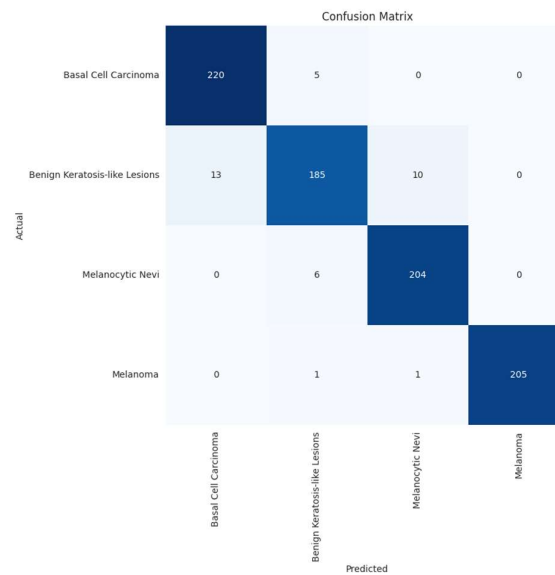


Figure 7. Confusion matrix for BCC, benign keratosis-like lesions, melanocytic nevi, melanoma classes.

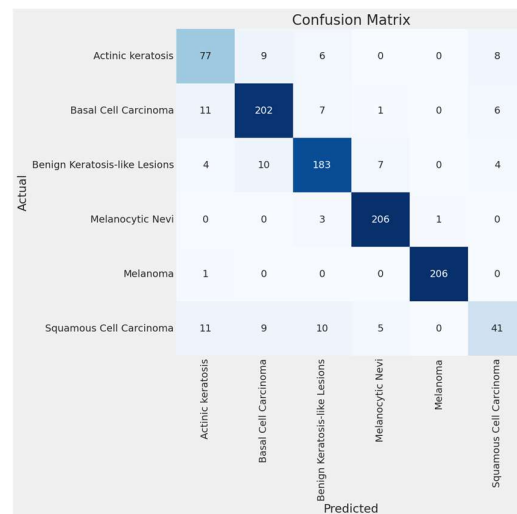


Figure 8. Confusion matrix for BCC, benign keratosis-like lesions, melanocytic nevi, melanoma, SCC, and AK classes.

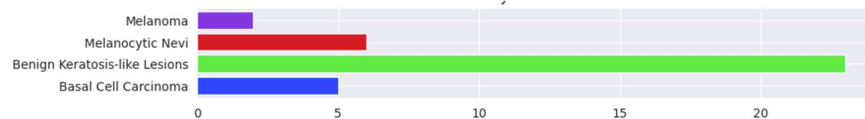


Figure 9. Errors per class for BCC, benign keratosis-like lesions, melanocytic nevi, and melanoma classes.

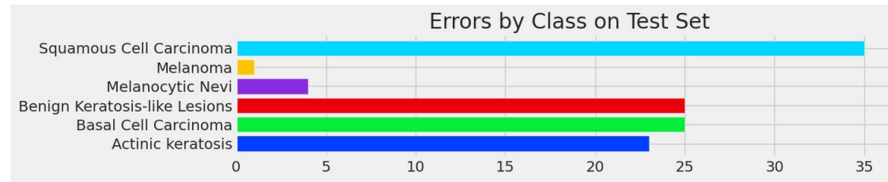


Figure 10. Errors per class for BCC, benign keratosis-like lesions, melanocytic nevi, melanoma, SCC, and AK classes.

IV. Discussion

AI integration in dermatology offers a substantial transformation of the medical field, and models such as EfficientNet certainly have the potential to revolutionize the diagnosis of skin lesions providing a better tool to assist with accuracy and efficiency.

The results of our study showed that the custom model based on EfficientNetB3 has an impressive ability to classify skin lesions with high accuracy. The model's performance was high in pathology categories well-represented by a large number of images (and therefore deep learnable data). For the categories having sufficient examples, an average accuracy of 95.4% was gained, which demonstrates a high recognition and classification capacity. Yet, the introduction of pathologies with a lower number of images led the model performance to drop to 88.8%. The reduced performance suggests the necessity for a balanced and representative dataset in each pathology category. These results stress the necessity to enhance data acquisition, allowing the model to be more generalizable and applied to a variety of clinical scenarios thereby providing reliable diagnostic outcomes.

To provide a comprehensive evaluation of our proposed model, its performance was compared with the results of several studies that utilize the EfficientNet architecture for skin lesion classification. Table 5 briefly summarizes comparative studies that used EfficientNet models, with a more detailed description of the studies' work presented below.

Table 5. Comparative studies using EfficientNet models.

Model	Year	Dataset	Model used	Scope	Accuracy
Karthik et al. [33]	2022	DermNet NZ, Derm7Pt, DermatoWeb, Fitzpatrick17k	EfficientNetV2 in conjunction with the Efficient Channel Attention block	Classification of 4 skin diseases: acne, AK, melanoma, and psoriasis.	84.7%
Ali et al. [34]	2022	HAM10000 dataset of dermatoscopic images	EfficientNet variants (results presented refer to EfficientNet B0)	Classification of 7 skin diseases	87.9%
Rafay et al. [35]	2023	Manually curated from Atlas Dermatology & ISIC Dataset	Fine-tuned EfficientNet-B2	Classification of 31 skin diseases	87.15%
Venugopal et al. [36]	2023	ISIC2019 dataset	EfficientNetV2-M	Binary classification:	95.49%

				malignant vs benign	
Venugopal et al. [36]	2023	ISIC2019 dataset	EfficientNet-B4	Binary classification: malignant vs benign	93.17%
Harahap et al. [37]	2024	ISIC2019 dataset	EfficientNet-B0 to EfficientNet-B7 (results reported to EfficientNet-B3)	Classification of 3 skin diseases: BCC, SCC, melanoma	77.6%
Harahap et al. [37]	2024	ISIC2019 dataset	EfficientNet-B0 to EfficientNet-B7 (results reported to EfficientNet-B4, the highest result obtained)	Classification of 3 skin diseases: BCC, SCC, melanoma	79.69%
Proposed model		ISIC2019 & personal images collection	EfficientNetB3	Classification of 4 skin diseases (benign& malign)	95.4%
Proposed model		ISIC2019 & personal images collection	EfficientNetB3	Classification of 6 skin diseases (benign& malign)	88.8%

Karthik et al. (2022) introduced in their study Eff2Net, a model designed to classify skin diseases with improved accuracy and reduced computational complexity. By integrating the Efficient Channel Attention (ECA) block into the EfficientNetV2 architecture, the authors replace the traditional Squeeze and Excitation (SE) block, thereby significantly reducing the number of trainable parameters. Eff2Net was trained on a diverse dataset comprising 4930 images, with data augmentation expanding this to 17,329 images across four skin disease categories: acne, AK, melanoma, and psoriasis. The model achieved a testing accuracy of 84.70%, outperforming other contemporary models like InceptionV3, ResNet-50, DenseNet-201, and EfficientNetV2 in overall accuracy with fewer parameters. Despite its strengths in reducing computational complexity and achieving high accuracy, Eff2Net has limitations, particularly in the accuracy of actinic keratosis [33].

Ali et al. (2022) explored the use of EfficientNet models (B0-B7) for classifying 7 classes of skin lesions using the HAM10000 dataset. Transfer learning from pre-trained ImageNet weights and fine-tuning on the HAM10000 dataset were applied to train the EfficientNet variants. Performance metrics like precision, recall, accuracy, F1 score, specificity, ROC AUC score, and confusion matrices were used to evaluate the models. The findings revealed that intermediate complexity models, such as EfficientNet B4 and B5, performed the best, with EfficientNet B4 achieving an F1 Score of 87% and a top-1 accuracy of 87.91% [34]. It is noteworthy that the accuracy of the EfficientNetB3 model in this study was reported to be 83.9% [34], which is lower than the accuracy achieved by our proposed model.

In their study, Rafay et al. (2023) aimed to perform the classification of a wide range of skin diseases (31 categories) using a novel dataset by blending two existing datasets, Atlas Dermatology and ISIC, resulting in 4910 images. The study utilized transfer learning with three types of convolutional neural networks: EfficientNet, ResNet, and VGG, and found that EfficientNet achieved the highest testing accuracy. The EfficientNet-B2 model was identified as the top performer, mainly due to its compound scaling and depth-wise separable convolutions, which enable efficient training with fewer parameters [35].

The study performed by Venugopal et al. (2023) focused on the binary classification of skin lesions (malignant vs. benign) using EfficientNet models (EfficientNetV2-M and EfficientNet-B4) and a database created by combining datasets from ISIC 2018, ISIC 2019, and ISIC 2020, totaling 58,032 images. The modified EfficientNetV2-M model achieved high performance, with an accuracy of 95.49% on the ISIC 2019 dataset, while the accuracy of the EfficientNet-B4 model was 93.17%. [36].

In their study, Harahap et al. (2024) investigate the use of EfficientNet models for classifying BCC, SCC, and melanoma using the ISIC 2019 dataset. The study implemented all eight EfficientNet variants (B0 to B7), with EfficientNet-B4 achieving the highest overall accuracy of 79.69%. The EfficientNet-B3 model achieved a validation accuracy of 74.87% and a testing accuracy of 77.60%, with a precision of 85.98%, recall of 73.44%, and F1-score of 79.21% [37]. Notably, these results are lower than the ones reported in our study, where we classified six diseases, including the three from the mentioned study, and achieved higher accuracy.

To summarize, these recent studies showcase models like EfficientNetB0, EfficientNetB2, EfficientNetV2-M, EfficientNet-B4, and EfficientNetB3. The datasets used are varied, including DermNet NZ, Derm7Pt, DermatoWeb, Fitzpatrick17k, HAM10000, ISIC2019, and proprietary collections, covering both public and private data sources. The scope of these studies is diverse, with some focusing on a broad range of skin diseases, such as 31 classes in EfficientSkinDis [35], while others concentrate on specific categories like 4 to 7 skin diseases, covering both benign and malignant lesions. Also, several studies tested the accuracy of EfficientNet models in comparison with other CNNs, which they surpassed in performance. The reported accuracies ranged from 84.7% to 95.49% (higher values only for binary classification), highlighting the variations in model performance depending on the dataset and classification task (binary or multi-class). Notably, the proposed model achieves 95.4% accuracy for classifying 4 skin diseases and 88.8% for 6 skin diseases, demonstrating competitive performance within this comparative framework.

Moving forward, the proposed model was also compared with other state-of-the-art classification models, all of them using images from ISIC or HAM10000 datasets (Table 6).

Table 6. Comparative studies using state-of-the-art CNN models.

Model	Year	Dataset	Model used	Scope	Accuracy
Bazgir et al. [38]	2024	Kaggle/ ISIC	Inception Network	Binary classification: malign vs benign	85.94%
Rahman et al. [39]	2024	Kaggle/ ISIC	NASNet	Binary classification: malign vs benign	86.73%
Anand et al. [40]	2022	Kaggle/ ISIC	Modified VGG16 architecture	Binary classification: malign vs benign	89.9%

Singh et al. [41]	2022	ISIC2018	Bayesian DenseNet-169	Classification of 7 skin diseases	73.65%
Ahmed et al. [42]	2024	ISIC2018	SCCNet derived from Xception architecture	Classification of 7 skin diseases	95.2%
Al-Rasheed et al. [43]	2022	HAM10000	Combination of VGG16, ResNet50, ResNet101	Classification of 7 skin diseases	93.5%
Naeem et al. [44]	2024	ISIC2019	SNC_Net	Classification of 8 skin diseases	97.81%
Naeem et al. [45]	2024	ISIC2019	DVFNet	Classification of 8 skin diseases	98.32%
Proposed model		ISIC2019	EfficientNetB3	Classification of 4 skin diseases	95.4%
Proposed model		ISIC2019	EfficientNetB3	Classification of 6 skin diseases	88.8%

Several published studies are focusing on binary classification (benign vs malignant) using the Kaggle/ISIC dataset [38–40]. More specifically, Bazgir and colleagues (2024) present an approach to classify skin cancer using an optimized InceptionNet architecture. The study focused on distinguishing between melanoma and non-melanoma skin lesions using a dataset of 2637 dermoscopic images, split into 1197 malignant and 1440 benign lesions. The InceptionNet model was evaluated using performance metrics, including precision, sensitivity, specificity, F1-score, and area under the ROC curve. The optimized InceptionNet achieved an accuracy score of 84.39% and 85.94% when using Adam and Nadam optimizers, respectively [38]. Using the same dataset, Rahman et al. (2024) present an approach to classify skin cancer using the NASNet architecture optimized for improved performance in detecting malignant versus benign lesions. The NASNet model's performance was evaluated using metrics like precision, sensitivity, specificity, F1-score, and area under the ROC curve. The optimized NASNet model achieved an accuracy of 86.73% [39]. In their study, Anand et al. (2022) focus on improving the VGG16 model using transfer learning for the classification of skin cancer into benign and malignant categories. The VGG16 model was enhanced by adding a flatten layer, two dense layers with the LeakyReLU activation function, and another dense layer with the sigmoid activation function. The improved model achieved an overall accuracy of 89.09% on a batch size of 128 using the Adam optimizer over ten epochs [40].

Singh et al. (2022) introduce a novel two-stage DL pipeline named SkiNet for the diagnosis of skin lesions. The framework integrates lesion segmentation followed by classification, incorporating uncertainty estimation and explainability to enhance model reliability and clinician trust. Using Bayesian MultiResUNet for segmentation and Bayesian DenseNet-169 for classification, the SkiNet pipeline achieves a diagnostic accuracy of 73.65%, surpassing the standalone DenseNet-169's accuracy of 70.01% [41]. Having the same image dataset and scope of classifying skin diseases into 7 categories, Ahmed et al. (2024) present a new deep learning model, SCCNet, based on the Xception architecture, with the inclusion of additional layers to enhance performance. These layers include convolutional layers for feature extraction, batch normalization layers for improved convergence, activation layers to introduce non-linearity, and dense layers for better classification performance. The model achieved an accuracy of 95.20%, with precision, recall, and F1-score values all above 95%, outperforming several state-of-the-art models such as ResNet50, InceptionV3, and Xception [42].

The article [43] by Al-Rasheed et al. presents a new approach to skin cancer classification using an ensemble of transfer learning models, specifically VGG16, ResNet50, and ResNet101. The study leverages Conditional Generative Adversarial Networks to augment the dataset, addressing class imbalance issues. The proposed models were trained on both balanced and unbalanced datasets, and their performance was evaluated using accuracy, precision, recall, and F1-score metrics. The ensemble approach achieved a superior accuracy of 93.5%, demonstrating a significant improvement over individual models, which had accuracies of around 92% [43].

Naeem et al. have published two studies using the ISIC 2019 dataset, focusing on the classification of 8 types of skin diseases [44,45]. In [44], DVFNNet achieved an impressive accuracy of 98.32%, outperforming several baseline CNN models like AlexNet, VGG-16, Inception-V3, and ResNet-50. In [45], the proposed SNC_Net model outperforms baseline models like EfficientNetB0, MobileNetV2, DenseNet-121, and ResNet-101, achieving an accuracy of 97.81%.

To sum up, the proposed model addresses a complex task of multi-class classification while still achieving an accuracy (95.4% for four classes and 88.8% for six classes) superior to the binary classification accuracies of the Inception Network, VGG16, and NASNet models. Also, the proposed EfficientNetB3 model outperforms Bayesian DenseNet-169, which achieved an accuracy of 73.65%. While SNC_Net and DVFNNet achieve higher accuracies (97.81% and 98.32%, respectively), it is essential to recognize that these models benefit from more specialized architectures and additional data preprocessing techniques. Overall, the proposed model using EfficientNetB3 demonstrates strong performance in multi-class classification tasks, particularly given its simpler architecture and lower computational requirements, with its competitive accuracy highlighting EfficientNetB3's capability to handle diverse and challenging dermatological datasets effectively.

Limitations of Current Research

Despite the presented advancements, our study also has several limitations that need to be recognized.

First, the dataset primarily contains images of individuals from selected demographics, skin diseases, or skin phototypes, which may not represent the population's global diversity. This aspect could limit the model's performance when tested on other skin types and conditions, consequently affecting its generalization.

Second, the total number of images is still relatively low, particularly for infrequent lesions such as certain types of melanoma. This might lead to overfitting, where the model overperforms on training data but underperforms if applied to new unseen data.

To mitigate these constraints, the research could evolve from the current work to widen the dataset in terms of size and diversity, with the inclusion of broader skin phototypes and less common skin conditions. It will allow the development of a model that is both accurate in common conditions and applicable with high fidelity in the early detection of less common (and more likely dangerous) lesions. For a broader range of data, more international dermatology centers are required to collaborate and compile a dataset that can lead to a model that is more representative of global diagnostic applications.

Furthermore, one of the biggest areas that needs to be addressed in dermatological AI research is the need for standardization when it comes to collecting images such as using high-resolution images as the benchmark. The efficacy of ML models, like EfficientNet, is heavily dependent on the quality of the input data. Higher-resolution images get the finer details of dermatological conditions which is important for accurate specification and diagnostics. Presently, the heterogeneity of imaging due to the differences in the devices used and the settings applied to collection centers remains one of the major obstacles. Creating a standard high-resolution image during the collection process would make the training data not only more uniform but also more descriptive. The standardization also would correct some of the dataset variability that was seen when models trained on mixed-quality image datasets do not perform as consistently or have reduced diagnostic performance in different clinical settings.

Future work should also focus on improving the interpretability properties of the model, offering more details related to AI's diagnostic rationale. This would be very helpful in educational settings and would increase the model's acceptability in clinical practice.

Also, incorporating multimodal data like the patient's history and demographics could improve the diagnostic accuracy of the model and make it more patient-specific. If it lays the groundwork for individual dermatological assessments, then this could be the first step in aligning more with the goals of precision medicine.

V. Conclusions

In our study, the implementation of a custom model based on EfficientNetB3 has demonstrated substantial potential for enhancing the diagnosis of skin lesions. Our model achieved a notably high accuracy rate (95.4%/ 88.8%), underscoring the critical role of a comprehensive and diverse dataset. Our findings revealed that the model's performance remains robust when there is an ample number of images for each pathology. However, there is a noticeable decline in accuracy when dealing with new or less common pathologies that have fewer representative images.

These results underscore the vital importance of ongoing efforts to expand and diversify dermatologic image datasets. Ensuring a broad, varied, and standardized dataset is essential to maintain the efficacy and applicability of AI diagnostic algorithms across a wide range of skin conditions. Continual dataset growth will support the model's ability to generalize effectively, thereby improving diagnostic precision and reliability across diverse clinical scenarios.

Author Contributions: Conceptualization- I.M., A.I.B, R.N.B, G.S.T.; Methodology- I.M., A.I.B, R.N.B, G.S.T.; Software- R.N.B.; Validation- I.M., A.I.B, R.N.B; Formal analysis- I.M., A.I.B, R.N.B; Writing—original draft preparation: I.M., A.I.B, R.N.B, G.S.T.; Writing—review and editing, I.M., A.I.B, R.N.B, G.S.T; Supervision- I.M., R.N.B., G.S.T.; Project administration- I.M., R.N.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. <https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/> (Last accessed 9June2024)
2. Reichrath J, Leiter U, Eigentler T, Garbe C. Epidemiology of skin cancer. Sunlight, vitamin D and skin cancer. 2014;120-140. 10.2174/1871520611307010083
3. <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2022.html> (Last accessed 9June2024)
4. Argenziano G, Soyer HP, Chimenti S, Talamini R, Corona R, Sera F, et al. Dermoscopy of pigmented skin lesions: results of a consensus meeting via the Internet. *J Am Acad Dermatol*. 2003;48(5):679-693. 10.1067/mjd.2003.281
5. <https://www.iarc.who.int/cancer-type/skin-cancer/> (Last accessed 9June2024)
6. Tschandl P, Codella N, Akay BN, Argenziano G, Braun RP, Cabo H, et al. Comparison of the accuracy of human readers versus machine-learning algorithms for pigmented skin lesion classification: an open, web-based, international, diagnostic study. *Lancet Oncol*. 2019;20(7):938-947. 10.1016/S1470-2045(19)30333-X
7. Marchetti MA, Liopyris K, Dusza SW, Codella NC, Gutman DA, Helba B, et al. Computer algorithms show potential for improving dermatologists' accuracy to diagnose cutaneous melanoma: Results of the International Skin Imaging Collaboration 2017. *J Am Acad Dermatol*. 2020;82(3):622-627. 10.1016/j.jaad.2019.07.016
8. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118. 10.1038/nature21056

9. Codella NC, Nguyen QB, Pankanti S, Gutman DA, Helba B, Halpern AC, et al. Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM J Res Dev.* 2017;61(4/5):5-1. 10.48550/arXiv.1610.04662
10. Brinker TJ, Hekler A, Enk AH, Klode J, Hauschild A, Berking C, et al. Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer.* 2019;119:11-17. 10.1016/j.ejca.2019.05.023
11. Liu Y, Jain A, Eng C, Way DH, Lee K, Bui P, et al. A deep learning system for differential diagnosis of skin diseases. *Nat Med.* 2020;26(6):900-908. 10.1038/s41591-020-0842-3
12. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In: *International conference on machine learning.* PMLR; 2019. p. 6105-6114. 10.48550/arXiv.1905.11946
13. The International Skin Imaging Collaboration: <https://gallery.isic-archive.com/>
14. ImageNet Website and Dataset - <https://www.image-net.org/>
15. Shorten C, Khoshgftaar TM. A survey on image data augmentation for deep learning. *J Big Data.* 2019;6(1):1-48. 10.1186/s40537-019-0197-0
16. Sharma N, Jain V, Mishra A. An analysis of convolutional neural networks for image classification. *Procedia Comput Sci.* 2018;132:377-384. 10.1016/j.procs.2018.05.198
17. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng.* 2009;22(10):1345-1359. 10.1109/TKDE.2009.191
18. Jain S, Singhanian U, Tripathy B, Nasr EA, Aboudaif MK, Kamrani AK. Deep learning-based transfer learning for classification of skin cancer. *Sensors.* 2021;21(23):8142. 10.3390/s21238142
19. An end-to-end platform for machine learning - www.tensorflow.org
20. Keras, a deep learning API written in Python - <https://keras.io/about/>
21. https://keras.io/api/layers/normalization_layers/batch_normalization/
22. https://www.tensorflow.org/api_docs/python/tf/keras/layers/Dropout
23. https://www.tensorflow.org/guide/mixed_precision
24. https://www.tensorflow.org/api_docs/python/tf/keras/callbacks/EarlyStopping
25. https://keras.io/api/callbacks/reduce_lr_on_plateau/
26. https://www.tensorflow.org/tutorials/images/transfer_learning
27. https://keras.io/api/callbacks/model_checkpoint/
28. <https://www.kaggle.com/code/residentmario/full-batch-mini-batch-and-online-learning>
29. Ruder S. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747.* 2016. 10.48550/arXiv.1609.04747
30. Goodfellow I, Bengio Y, Courville A. *Deep learning.* MIT Press; 2016.
31. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res.* 2014;15(1):1929-1958. 10.5555/2627435.2670313
32. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *International conference on machine learning.* PMLR; 2015. p. 448-456. 10.48550/arXiv.1502.03167
33. Karthik R, Vaichole TS, Kulkarni SK, Yadav O, Khan F. Eff2Net: An efficient channel attention-based convolutional neural network for skin disease classification. *Biomed Signal Process Control.* 2022;73:103406. 10.1016/j.bspc.2021.103406
34. Ali K, Shaikh ZA, Khan AA, Laghari AA. Multiclass skin cancer classification using EfficientNets—a first step towards preventing skin cancer. *Neurosci Inform.* 2022;2(4):100034. 10.1016/j.neuri.2021.100034
35. Rafay A, Hussain W. EfficientSkinDis: An EfficientNet-based classification model for a large manually curated dataset of 31 skin diseases. *Biomed Signal Process Control.* 2023;85:104869. 10.1016/j.bspc.2023.104869
36. Venugopal V, Raj NI, Nath MK, Stephen N. A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images. *Decis Anal J.* 2023;8:100278. 10.1016/j.dajour.2023.100278
37. Harahap M, Husein AM, Kwok SC, Wizley V, Leonardi J, Ong DK, et al. Skin cancer classification using EfficientNet architecture. *Bull Electr Eng Inform.* 2024;13(4):2716-2728. 10.11591/eei.v13i4.7159
38. Bazgir E, Haque E, Maniruzzaman M, Hoque R. Skin cancer classification using Inception Network. *World J Adv Res Rev.* 2024;21(2):839-849. 10.30574/wjarr.2024.21.2.0500
39. Rahman MA, Bazgir E, Hossain SS, Maniruzzaman M. Skin cancer classification using NASNet. *Int J Sci Res Arch.* 2024;11(1):775-785. 10.30574/ijrsra.2024.11.1.0106
40. Anand V, Gupta S, Altameem A, Nayak SR, Poonia RC, Saudagar AKJ. An enhanced transfer learning based classification for diagnosis of skin cancer. *Diagnostics.* 2022;12(7):1628. 10.3390/diagnostics12071628
41. Singh RK, Gorantla R, Allada SGR, Narra P. SkiNet: A deep learning framework for skin lesion diagnosis with uncertainty estimation and explainability. *PLoS One.* 2022;17(10). 10.1371/journal.pone.0276836
42. Ahmed T, Mou FS, Hossain A. SCCNet: An Improved Multi-Class Skin Cancer Classification Network using Deep Learning. In: *2024 3rd International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE);* 2024 Apr; IEEE. p. 1-5. 10.1109/ICAEEE62219.2024.10561672

43. Al-Rasheed A, Ksibi A, Ayadi M, Alzahrani AI, Zakariah M, Hakami NA. An ensemble of transfer learning models for the prediction of skin cancers with conditional generative adversarial networks. *Diagnostics*. 2022;12(12):3145. 10.3390/diagnostics12123145
44. Naeem A, Anees T, Khalil M, Zahra K, Naqvi RA, Lee SW. SNC_Net: Skin Cancer Detection by Integrating Handcrafted and Deep Learning-Based Features Using Dermoscopy Images. *Mathematics*. 2024;12(7):1030. 10.3390/math12071030
45. Naeem A, Anees T. DVFNet: A deep feature fusion-based model for the multiclassification of skin cancer utilizing dermoscopy images. *PLoS One*. 2024;19(3). 10.1371/journal.pone.0297667

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.