FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 9 No. 2, February, 2025, pp 158 - 169
DOI: https://doi.org/10.33003/fjs-2025-0902-3138

DERMATOLOGICAL DIAGNOSIS OF SKIN LESIONS USING ATTENTION MECHANISMS

***Balogun Ghaniyyat Bolanle, Makassa Mariam Masire, Woru Mariam Mero and Sadiku Peter Ogirima**

Department of Computer Science, University of Ilorin, Nigeria

*Corresponding authors' email: balogun.gb@unilorin.edu.ng

ABSTRACT

Every year, both in developed and developing nations, the prevalence of skin cancer rises. Due to a lack of resources and medical knowledge, diagnosing skin lesions is more difficult in third-world nations. Unusual growths or alterations in the skin are known as skin lesions, and they can result from a number of causes, including cancer, inflammatory diseases, infections, traumas, and heredity. Malignant (cancerous) and benign (non-cancerous) skin lesions are both possible. This disease is contracted when the pigments that produce skin color become cancerous. Dermatologists find it difficult to diagnose skin cancer since the pigments of various skin conditions might look identical. This led to the goal of this work, which is to use attention mechanisms to design a system for dermatological diagnoses of skin lesions. Modern network architectures ResNet and EfficientNet, enhanced with specially designed patch-based attention heads, are the approach used to accomplish this. The HAM10000 dataset, a thorough compilation of dermatoscopic pictures of typical pigmented skin lesions, was used in the investigation. In order to improve the model's capacity to recognize minute yet crucial variations among lesion types, attention heads were created to highlight and identify important characteristics within patches of the dermatoscopic pictures. According to the experimental results, the model that accurately classifies images into different lesion types had the lowest accuracy of 72% on a dataset of over 10,000 image instances, while the model that determines whether a lesion is cancerous or non-cancerous had the highest accuracy of 98%, demonstrating its robustness and reliability for the binary classification task. At 0.54, DF had the lowest F-1 score, whereas at 1.0, Normal had the highest precision. As a result, it is advised that skin lesion detection and categorization be modeled using attention mechanisms.

Keywords: Attention Mechanisms, EfficientNet, Modern network

INTRODUCTION

The skin is the main barrier that protects the body's critical organs. It serves as a barrier to keep various threats away from our interior organs. However, skin damage can result from diseases brought on by viruses, fungi, or even dust. A small skin lesion has the potential to develop into a major health issue. A person's skin health is influenced by a variety of things during their productive life, including exposure to sunlight, smoking, drinking, sports, infections, and the workplace. In addition to impairing the integrity of skin function, these variables also harm the skin, negatively impact human health, and in extreme situations, endanger human life. (Montagna, 2012).

Effective treatment and a lower chance of serious repercussions depend on the early and precise diagnosis of skin lesions, particularly malignant ones (Wolff et al., 2017). Numerous industries have been touched by the development of machine learning techniques. Some of the effects that machine learning approaches have demonstrated in the health sector in recent years include the detection and classification of brain tumours, diabetes detection and prediction, and breast cancer detection and classification (Debelee et al., 2020).

This research endeavour aims to investigate these dermatological diagnostic techniques, with a specific emphasis on the utilization of attention mechanisms within artificial intelligence models to pinpoint significant regions in skin lesion imagery, thereby improving the accuracy of dermatological diagnostics, which is essential for the effective management and treatment of skin diseases. (Vaswani et al., 2017). The application of attention mechanisms to emphasize key regions in skin lesions offers several benefits. By directing attention to the most relevant portions of the lesion, these models can significantly enhance the accuracy of detection and classification. This targeted strategy is especially advantageous for the identification of early-stage malignant lesions, which may be easily missed by conventional diagnostic approaches.

A major advancement in medical diagnostics has been made with the integration of deep learning and attention mechanisms in dermatology, specifically in the identification and analysis of skin lesions. According to systematic reviews by Jeong et al. (2022), deep learning, a subset of artificial intelligence (AI), has been thoroughly investigated for dermatological applications. These works demonstrate different deep learning techniques and their increasing use in dermatology. AdaViT, an adaptive computation framework for vision transformers based on self-attention mechanisms, was presented by Meng et al. in 2021. The function of spatial and channel-wise attention processes in image emotion identification is highlighted by Li et al. (2021), who also show how adaptable these mechanisms are across a range of image analysis domains. The use of attention mechanisms in particular domains, such as surgical action recognition and fine-grained picture recognition, is further demonstrated by Nwoye et al. (2021).

The current research struggles with low multi-class accuracy (72%) and inconsistent performance across lesion types. My work will enhance feature discrimination and classification consistency using advanced attention mechanisms.

An attention-driven lightweight model for detecting pigmented skin lesions was introduced by Hu and Yang in 2023. This model maintains great performance while lowering computing complexity by utilizing ghosted features and the DFC attention mechanism. The precise identification of lesion characteristics is the goal of the SANet, a superpixel attention network put forth by He, Lei, and Wang (2019). This task is still difficult because of uneven class distribution and a lack of data.

An attention-enabled ensemble-based deep learning method for multiclass categorization of skin lesion images, DermAI

1.0, was presented by Sanga et al. (2023). This method greatly improves diagnostic accuracy by fusing attention processes with previously established transfer learning models. A new multi-modal contextual fusion model called MSHA was created by Eze et al. (2023) to detect Varicella Zoster Virus skin lesions early. The model outperforms state-of-the-art models in accuracy and loss by combining attention-based contextual information, mixed-scale dense convolution, self-attention processes, and hierarchical feature fusion.

A dermatological classification technique using a combination of SVMs and Back Propagation Neural Networks (BPNNs) was proposed in a paper by Mohammed et al. (2020). They started their methodology by using regularized Random Forests to alter the elements, and then they applied image enhancement algorithms. With an astounding classification accuracy of 99.7% and sensitivity of 99.4%, the research's results were outstanding. It is important to remember, though, that the dataset they looked at only included 400 photos, highlighting the necessity of additional validation using a larger dataset to firmly prove the validity of this approach. In order to simultaneously diagnose and localize the skin lesions in dermoscopy pictures, Yang et al. (2019) used a multi-task CNN with CAMs. The region of interest in the image, which typically corresponds to the entire lesion, is highlighted by the obtained CAMs. Zhang et al. (2018) enhanced feature mappings between successive layers by adding attention modules to residual networks. These authors also employ CAMs to demonstrate how the suggested network ignores the majority of the surrounding skin in favour of concentrating on the skin lesion. Even though CAMs can provide some interpretability, they often only show the full lesion and cannot identify smaller areas of interest that might be connected to various dermoscopic features.

Grignaffini et al. (2022) undertook a systematic review of existing literature concerning the application of machine learning methodologies for the detection and classification of skin cancer across a variety of datasets, including MedNode, ISIC2017, HAM10000, ISIC2016, PH2, DermIS, DermQuest, ISIC archive, IDS, ISIC 2019, ISIC2020, ISIC2018, a 7-point checklist, and DermNZ. By analyzing 68 research articles, the authors delivered a comprehensive overview of the machine learning techniques employed in skin cancer categorization, along with their respective performance metrics. The structure of the review is notably organized, with distinct sections dedicated to methods, findings, and the introduction. The reliability of the review is enhanced by the authors' detailed explanation of the inclusion and exclusion criteria applied in selecting the studies. Additionally, a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram was utilized to illustrate the study selection process. The Results Section offers an in-depth analysis of the various machine learning techniques applied to skin cancer categorization, accompanied by performance metrics.

In a comprehensive analysis, Hauser et al. (2022) explored the application of explainable artificial intelligence (XAI) in the detection of skin cancer. XAI encompasses artificial intelligence frameworks that articulate their decision-making processes, thereby assisting medical professionals in comprehending and interpreting the predictions generated by these models. In their research, Jeong et al. (2023) sought to investigate the prevailing methodologies, results, and limitations associated with the application of deep learning in the field of dermatology. Their analysis encompassed studies published from 2015 to 2021, leading to the identification of 65 relevant papers that satisfied their inclusion criteria. The authors elaborated on the diverse deep learning methodologies employed, the dermatological conditions examined, and the performance metrics utilized to evaluate the models. This comprehensive examination of the implementation of deep learning techniques in dermatology represents a notable advancement in the discipline. The review article also provided references to additional datasets that could be beneficial for researchers. However, it is possible that significant studies were missed due to the focus being limited to publications from 2015 to 2021. While the authors recognized the constraints of deep learning in dermatology, offering more targeted suggestions for future research to mitigate these limitations would have been advantageous.

An examination of various methods for categorizing skin diseases through machine learning techniques was conducted by Mohammed and Al-Tuwaijari (2021). The study explored several methodologies, including support vector machines, decision trees, random forests, artificial neural networks, and deep learning. Additionally, it addressed the techniques and performance metrics utilized within these systems. The paper provided a comprehensive explanation of the machine learning algorithms employed for skin disease classification. However, it fell short of delivering an exhaustive analysis of the methodologies reviewed. It is possible that the survey did not encompass all existing classification methods for skin diseases. Moreover, a more in-depth analysis and critique of the methodologies discussed would have enhanced the overall contribution of the paper.

Balaji et al. (2020) introduced a technique for the detection and segmentation of skin diseases utilizing the dynamic graph cut algorithm, coupled with classification via a Naive Bayes classifier. Initially, the authors employed the dynamic graph cut algorithm to segment the skin lesion from its background, followed by the classification of the lesion into various categories based on texture and colour features using the Naive Bayes classifier.

In certain remote regions, the limited availability of dermatologists necessitates that non-specialists undertake the diagnosis and treatment of dermatological conditions, despite their insufficient training and expertise in this specialized area. Although dermatology textbooks are accessible as reference sources, these practitioners encounter considerable obstacles in achieving precise diagnoses. The uneven distribution of dermatologists and healthcare resources exacerbates the challenges associated with accurate diagnosis in these locales (Hoang et al., 2022). Additionally, cultural stigmas related to skin cancer and lesions can lead to social ostracism and discrimination, further complicating the situation. The repercussions of skin cancer and lesions extend beyond the individual, impacting families, communities, and the broader economy (Journal of Clinical and Aesthetic Dermatology, 2018).

The existing literature highlights notable progress and potential uses of deep learning and attention mechanisms within dermatology. The incorporation of these technologies into dermatological diagnostics, although still in its nascent stages, demonstrates encouraging outcomes and underscores the necessity for continued research and innovation. These developments signify a vital advancement toward more precise, efficient, and interpretable diagnostic methodologies in dermatology. Consequently, this research endeavors to explore a novel approach to dermatological diagnostics through the implementation of an attention mechanism system.

MATERIALS AND METHODS

Modern network designs like ResNet and EfficientNet, reinforced with specially designed patch-based attention heads, was used in this study on "Dermatological Diagnosis of Skin Lesions via Attention Mechanisms." The HAM10000 dataset, a comprehensive collection of dermatoscopic pictures of common pigmented skin lesions, was used in this investigation. The suggested solution ~~will~~ focus on the parts of the skin where the current technologies are less accurate, enabling more efficient detection in those areas. Through the application of attention techniques, the suggested system will let the detection model to concentrate on important areas of the image, resulting in more precise legions detection.

In order to manage big datasets and process images in real-time, the suggested model was made to be efficient and easily incorporated into current clinical procedures. In order to create a user-friendly skin lesion identification tool, the suggested system was designed to accommodate the preferences of both patients and professionals.

This system uses attention to focus on relevant regions of the image, it extracts features using attention mechanisms while focusing on both classification and diagnosis. The system is structured into several interconnected components designed to handle various aspects of the segmentation process from data handling to model deployment. To facilitate seamless access to high-quality imaging data, the system will utilize the Kaggle API for downloading the HAM10000 datasets directly into the Google Colab environment. This integration ensures that the data used for training and validation is up-to-date and readily accessible without manual downloading and uploading, enhancing workflow efficiency. Given the large size of the imaging data, efficient data handling strategies was implemented. The system will manage data storage transiently in the Colab environment, utilizing cloud storage solutions if necessary for scalability and data persistence. Data was processed in batches to optimize memory usage and computational resource allocation.

Description of the dataset

The dataset used in this project comprises various images of skin lesions, which include both benign and malignant types. The primary sources of the dataset are:

HAM10000: A dataset consisting of 10,015 dermatoscopic images of pigmented lesions.

ISIC 2019: An expanded dataset including multiple types of skin lesions with annotations.

Additional Datasets: Normal skin images sourced from multiple additional datasets to provide a comprehensive training set. The lesion types present in the dataset include: Actinic Keratoses and Intraepithelial Carcinoma (AKIEC) (total number), Basal Cell Carcinoma (BCC) (total number), Benign Keratosis-like Lesions (BKL), Dermatofibroma (DF), Melanoma (MEL), Melanocytic Nevi (NV), Vascular Lesions (VASC) and Normal (non-lesion) skin images.Each image is labeled with its corresponding lesion type, and the dataset also includes annotations for each image.

Data Processing and Organization

The data processing phase involved a series of steps to prepare the dataset for training and validation. This section discusses each step and the outcomes achieved. The research used standard image preprocessing techniques on the HAM10000 dataset, including resizing images to a uniform input size suitable for ResNet and EfficientNet architectures. Data augmentation methods, such as rotation, flipping, scaling, and color normalization, were applied to increase data diversity and reduce overfitting. Additionally, patch-based preprocessing was implemented to divide dermatoscopic images into smaller regions, enabling the attention heads to focus on localized features. This patch-based approach aimed to enhance the model's ability to capture subtle variations among different lesion types.

Load and Move Dataset Files

The initial step was to define the source and target directories for the dataset files. The source directory contained all the images, while the target directory was designated to organize these images based on their classes. This step involved moving all images from the source directory to a specified target directory for better organization.By systematically organizing the images, each file was correctly placed in its respective class directory, facilitating the subsequent steps of data processing and model training.

Load Dataset and Create Class Labels

Important details like picture filenames and the class labels that go with them were included in the dataset metadata, which was loaded from a CSV file. In order to train the classification model, class labels for the photos were made using this information. Ensuring that every image was appropriately identified and classed required the creation of precise class labels based on the metadata. This labeling procedure was essential to the training process's efficacy.

Create Directories for Each Class and Move Images

To properly organize the dataset, directories were made for each type of lesion in addition to "Normal." Images were separated and categorized into their appropriate classifications thanks to this framework. Following the creation of class labels in the preceding phase, images were then relocated to the appropriate folders. This stage made sure that every class had its own directory and that the dataset was properly organized. The training process was eventually more effective as a result of the organization's facilitation of data processing and management. Building a strong classification model requires efficient data administration and processing, which was established by arranging the dataset into clearly defined directories. Figure 1 show Creating Directories for Each Class

Figure 1: Creating Directories for Each Class

Restricting and Splitting Data

To ensure a balanced training process, the dataset was restricted to a maximum number of images per class, with a portion of these images reserved for validation. The restricted dataset resulted in the following counts for the training and validation sets:

Training Set Class Counts:

- NV: 270 files
- MEL: 270 files
- BKL: 270 files
- DF: 104 files
- AKIEC: 270 files
- BCC: 270 files
- VASC: 128 files
- Normal: 270 files

Validation Set Class Counts:

- NV: 30 files
- MEL: 30 files
- BKL: 30 files
- DF: 11 files
- AKIEC: 30 files
- BCC: 30 files
- VASC: 14 files
- Normal: 30 files

The data split ensured that there was a balanced representation of each class in both the training and validation sets, which is crucial for training a model that generalizes well.

Calculating Class Weights

Class weights were calculated to address the imbalance in the dataset. These weights were used during model training to ensure that the model did not become biased towards the more frequent classes. The calculated class weights were as follows:

Class weights:

NV: 0.8574074074074074
MEL: 0.8574074074074074
BKL: 0.8574074074074074
DF: 2.2259615384615383
AKIEC: 0.8574074074074074
BCC: 0.8574074074074074
VASC: 0.8574074074074074
Normal: 1.80859375

These weights helped to balance the influence of each class during the training process, improving the model's ability to accurately classify less frequent classes. Figure 2: Calculating the class weight of the dataset

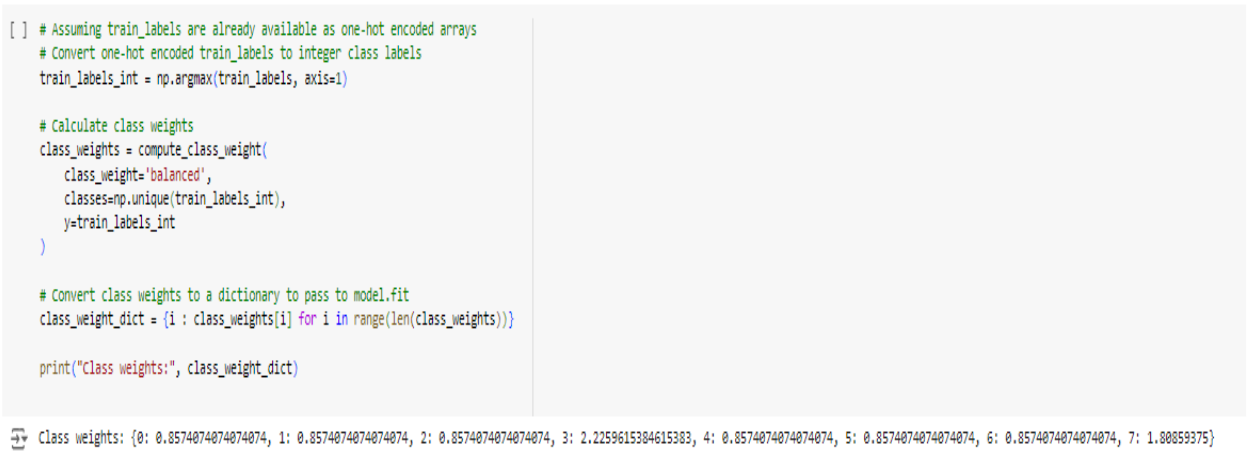

Figure 2: Calculating the class weight of the dataset

Model Architecture

The model architecture used for both skin lesion classification and cancer diagnosis was based on the EfficientNetB1 model, enhanced with an attention mechanism and regularized dense layers. The architecture is summarized as follows:

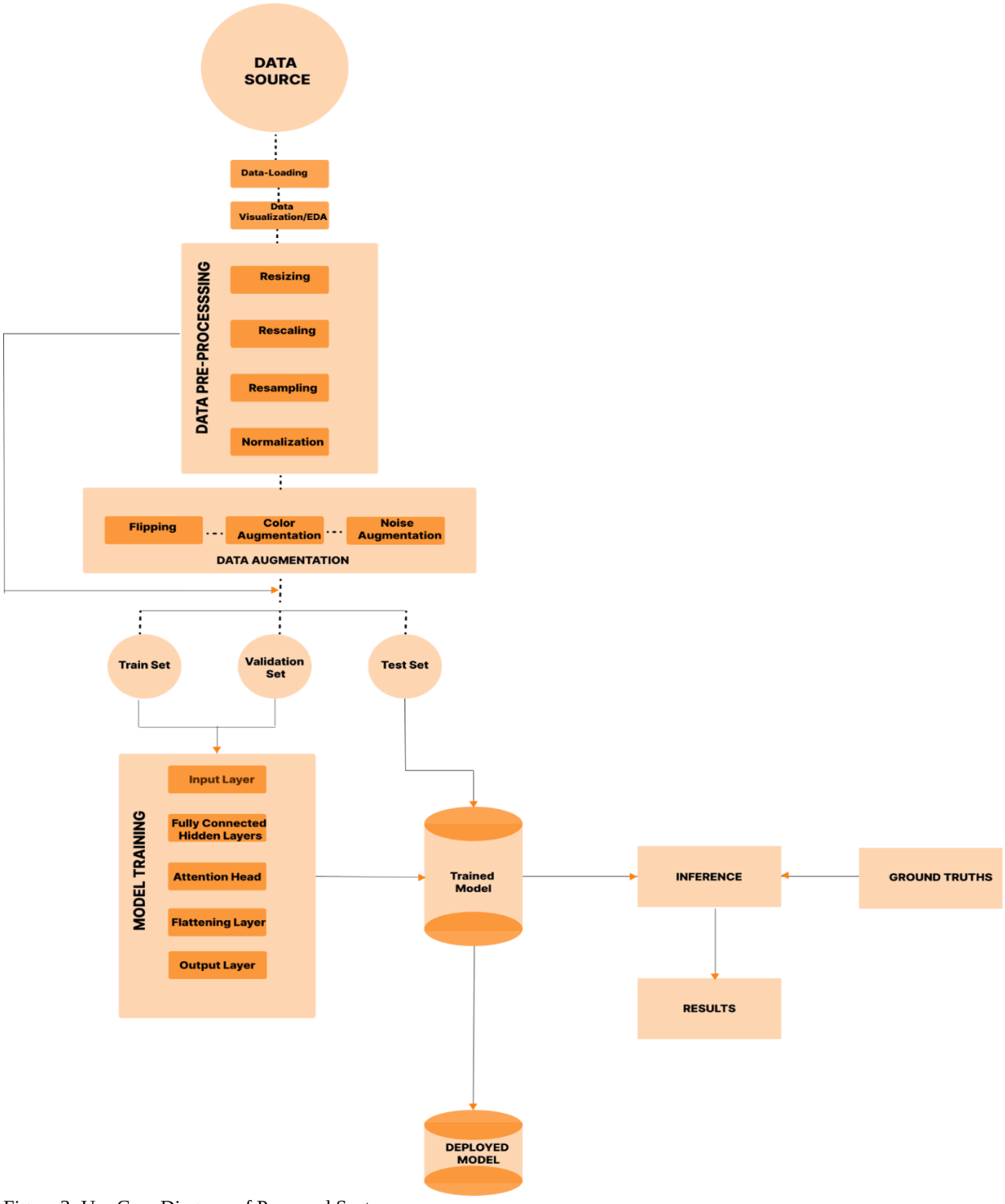

Figure 3: Use Case Diagram of Proposed System

1. Base Model: EfficientNetB1: A pre-trained EfficientNetB1 model was used as the base for feature extraction. This model is known for its efficiency and performance on various image classification tasks. Figure 3: Use Case Diagram of Proposed System

2. Attention Mechanism: Attention Layer: A single-channel attention map was applied using a Conv2D layer with a sigmoid activation function. This attention mechanism helps the model focus on the most relevant parts of the image.

3. Global Average Pooling: GlobalAveragePooling2D: This layer reduced the spatial dimensions of the feature maps to a single vector, summarizing the information from the entire image.

4. Regularized Dense Layers:

Dense Layers: Two dense layers with L2 regularization were added, each followed by Batch_Normalization and Dropout layers to prevent overfitting.

Output Layer: A final dense layer with a softmax activation function was used to produce the class probabilities.

The detailed architecture is illustrated in the table below, describing each layer and its corresponding hyperparameters.

Table 1: Architecture of Each Layer and its Parameters

	Layer Type	Layer Details
	Input	Shape: (256, 256, 3)
	EfficientNetB1 Base Model	Pre-trained on ImageNet, include_top=False
	Attention Mechanism	Conv2D(1, (1, 1), activation='sigmoid', padding='same')
	Multiplication	Multiply() with attention map
	Global Average Pooling	GlobalAveragePooling2D()
7.	Dense Layer 1	Dense(256, activation='relu', kernel_regularizer=l2(0.01))
8.	Batch Normalization 1	BatchNormalization()
9.	Dropout 1	Dropout(0.5)
10.	Dense Layer 2	Dense(128, activation='relu', kernel_regularizer=l2(0.01))
11.	Batch Normalization 2	BatchNormalization()
12.	Dropout 2	Dropout(0.5)
13.	Output Layer (Multi-class)	Dense(num_classes, activation='softmax', kernel_regularizer=l2(0.01))
14.	Output Layer (Binary)	Dense(2, activation='softmax', kernel_regularizer=l2(0.01))

Model development

The model's operation can be mathematically represented through a series of transformations applied to the input data. Let \(x \) represent the input image tensor. Figure 4 show Mathematical Representation of the model

The operations are as follows:

EfficientNetB1 Feature Extraction:

$$FEffNet(x) = EfficientNetB1(x)$$

Attention Mechanism:

$$A(x) = \sigma\left(Conv2D1 \times 1\left(FEffNet(x)\right)\right)$$

$$FAtt(x) = FEffNet(x) \odot A(x)$$

Global Average Pooling:

$$G(x) = GlobalAveragePooling2D\left(FAtt(x)\right)$$

Dense Layers with Regularization:

$$D1(x) = Dropout\left(BatchNorm\left(Dense256\left(ReLU\left(G(x)\right)\right)\right)\right)$$

$$D2(x) = Dropout\left(BatchNorm\left(Dense128\left(ReLU\left(D1(x)\right)\right)\right)\right)$$

Output Layer:

$$ymulti = Softmax\left(Densenum_{classes(D2(x))}\right)$$

$$ybinary = Softmax\left(Dense2\left(D2(x)\right)\right)$$

Where:

σ *denotes the sigmoid activation function..*

$\odot$ *represents the element − wise multiplication.*

$Conv2D1 \times 1$ *is a convolution operation with a* $1 \times 11 \times 1$ *filter.*

$Densek$ *represents a dense layer with kk units.*

$ReLU$ *is the rectified linear activation function.*

$BatchNorm$ *is the batch normalization layer.*

$Dropout(p)$*is the dropout operation with a dropout rate of* pp.

$Softmax$ *is the softmax activation function.*

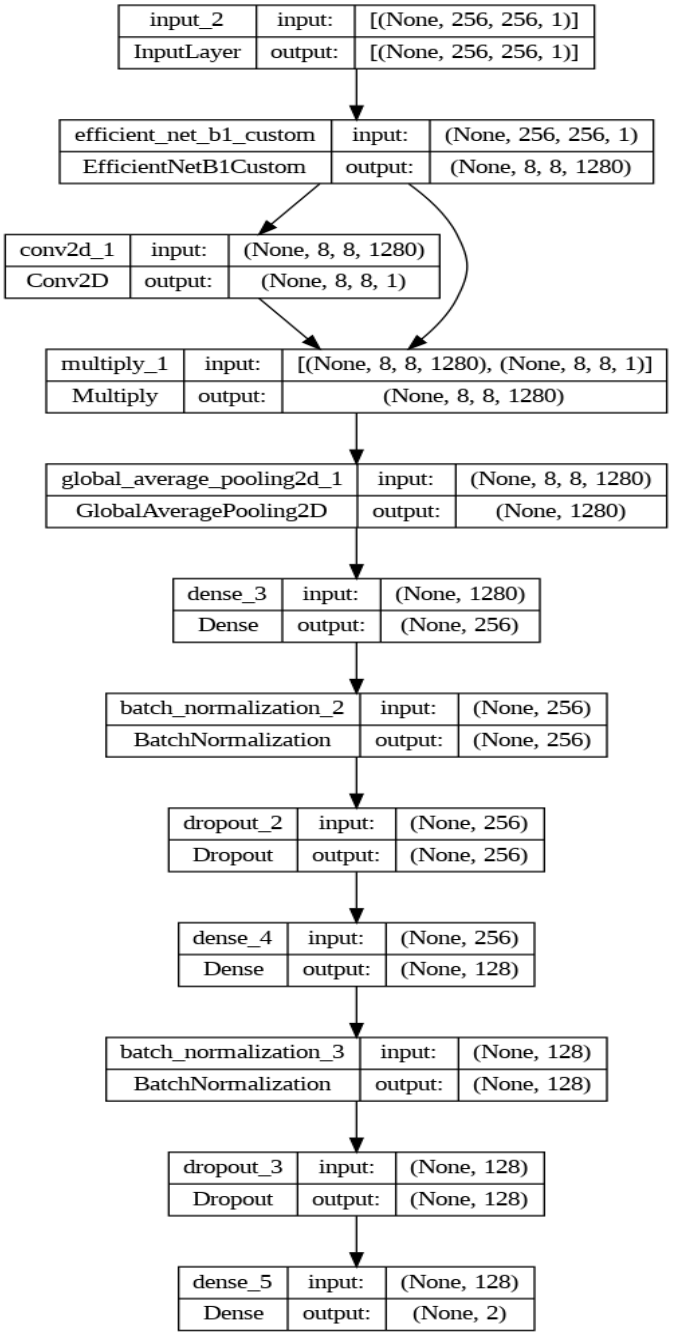

Figure 4: Mathematical Representation of the model

Preprocessing and model development

Preprocessing for Cancer Diagnosis Model: The preprocessing for training a model specifically for cancer diagnosis involved grouping the various skin lesion classes into two broader categories: cancerous and non-cancerous. This section discusses the steps taken to prepare the data for this binary classification task.

Grouping Classes into Cancerous and Non-Cancerous

The lesion types were grouped into two categories:

- Cancerous: Includes AKIEC, BCC, BKL, DF, MEL, NV.
- Non-Cancerous: Includes Normal and VASC.

This grouping was based on the nature of the lesions, with the cancerous group containing all malignant or potentially malignant types, and the non-cancerous group containing benign or normal skin images.

Loading Images and Labels: The new binary categories were used to label the photographs after they were loaded from the dataset. Reading the photos, making any necessary augmentations, and translating the labels to a binary format were all steps in this procedure.

Each image was correctly classified into either the malignant or non-cancerous group in order to prepare the dataset. The training of a model designed especially for cancer diagnosis was made easier by this binary labeling.

```python
def load_images_and_labels(base_dir, img_size=(256, 256), augment=False):
    images = []
    labels = []
    class_names = sorted(os.listdir(base_dir))  # Ensure that the order is consistent
    class_indices = {name: index for index, name in enumerate(class_names)}

    for folder in class_names:
        folder_path = os.path.join(base_dir, folder)
        for img_file in os.listdir(folder_path):
            img_path = os.path.join(folder_path, img_file)
            img = load_img(img_path, color_mode='rgb', target_size=img_size)  # Corrected color_mode to 'rgb'
            img_array = img_to_array(img)

            if augment:
                img_array = apply_random_transformations(img_array)

            images.append(img_array)
            labels.append(class_indices[folder])

    images = np.array(images, dtype='float32') / 255.0  # Normalize the images
    labels = to_categorical(labels, num_classes=len(class_names))  # Convert labels to one-hot encoding

    return images, labels, class_indices

def apply_random_transformations(image):
    image = Image.fromarray(np.uint8(image))  # Convert to PIL image for transformations
    # Random flip
    if random.choice([True, False]):
        image = image.transpose(Image.FLIP_LEFT_RIGHT)
    if random.choice([True, False]):
        image = image.transpose(Image.FLIP_TOP_BOTTOM)

    # Random rotation
    rotation = random.choice([0, 90, 180, 270])
    image = image.rotate(rotation)

    return np.array(image)
```

Figure 5: Loading Images and Augmentation

Loading Images and Labels: Given the class imbalance in the dataset, class weights were calculated to ensure that the model did not become biased towards the more frequent class. The calculated class weights were:
Class weights:
Cancerous: 0: 2.3266331658291457
Non-Cancerous: 1: 0.6368638239339752
These weights indicated a significant imbalance, with non-cancerous images being less frequent than cancerous ones. By applying these weights during training, the model was guided to pay more attention to the minority class, improving its ability to correctly classify non-cancerous images.

Training the Cancer Diagnosis Model: The preprocessed dataset and the determined class weights were used to train the cancer diagnosis model. The model's performance on the training and validation sets was tracked during several training epochs.
Training Progress and Results: During the training process, the model showed significant improvements in both accuracy and loss for the binary classification task. Key observations from the training process included:

Hyperparameters and their representation
The following table (2) summarizes the hyperparameters used in the model and their respective values:

Table 2: Hyperparameters and their Representation

Hyperparameter	**Symbol**	**Value**
Input Shape	x	(256, 256, 3)
Base Model Weights		ImageNet
Attention Map Channels		1
Attention Activation	σ	Sigmoid
Dense Layer 1 Units	$k1$	256
Dense Layer 2 Units	$k2$	128
Dropout Rate	p	0.5
Regularization Factor	λ	0.01
Learning Rate	η	0.001
Number of Classes	num_classes	Varies (8 for multi-class, 2 for binary)

The model architecture and the mathematical representation of the transformations applied to the input data highlight the structured approach to feature extraction, attention mechanism integration, and regularization, ensuring the robustness and effectiveness of the model for skin lesion classification and cancer diagnosis.

Training Process
The training process involved several important steps to ensure the model learned effectively from the data while avoiding overfitting. Key components of the training process included learning rate scheduling, the use of callbacks, and monitoring the model's performance on a validation set.

Learning Rate Scheduling
Learning rate scheduling was implemented to adjust the learning rate dynamically during training. The learning rate was initially set to 0.001, and it was reduced by a factor of 5 when the validation loss plateaued. This was achieved using the `ReduceLROnPlateau
The effect of learning rate scheduling was observed through the gradual stabilization and eventual improvement in the model's performance. Initially, a higher learning rate allowed the model to make significant updates to the weights, facilitating rapid learning. As training progressed, reducing the learning rate helped fine-tune the model by making smaller adjustments, which was crucial for minimizing validation loss and improving generalization.

Callbacks Used
Several callbacks were utilized to enhance the training process:
1. EarlyStopping:
- Monitor: `val_loss`
- Patience: 20 epochs
- Restore Best Weights: True
- The EarlyStopping callback was used to halt training when the validation loss stopped improving. This helped prevent overfitting by ensuring that the model did not continue to train beyond the point where it was no longer improving.
2. ModelCheckpoint:
- Monitor: `val_loss`
- Save Best Only: True
- Filepath: `best_model.keras` or `best_model.h5`
- The ModelCheckpoint callback saved the model weights at each epoch where the validation loss improved. This ensured that the best performing model, in terms of validation loss, was retained for further evaluation and deployment.
3. ReduceLROnPlateau:
- Monitor: `val_loss`
- Factor: 0.2
- Patience: 5 epochs
- Min LR: 0.00001

Metrics used
These metrics helps in evaluating a model's performance identifying the model's strengths and weaknesses and also guiding the model's improvement.

True Positive (TP)
True positive occurs when a positive prediction (for instance, lesion present) matches an actual positive outcome (lesion indeed present).
False Positive (FP)
A false positive occurs when a positive prediction (for instance, lesion present) does not match positive outcome (lesion absent). False positive is also known as type I error or false alarm.

False Negative (FN)
A false negative occurs when a negative prediction (for instance, lesion absent) does not match negative outcome (lesion present). False negative is also known as type II error or false prediction.

Accuracy
accuracy measures the proportion of correct predictions out of the total predictions made.
Formula: Accuracy=(TP+TN)/(TP+TN+FP+FN)
Precision: precision measures the proportion of true positives among all positive predictions made.
Formula: precision = TP/(TP+FP)
Recall: recall measures the proportion of true positive among all actual positive instances
Formula: Recall= TP/(TP+FN)
F1-Score: F1-score is the harmonic mean of precision and recall, providing a balanced measure for both.
Formula: F1-score= 2*(Precision*Recall)/(Precison+Recall)

RESULTS AND DISCUSSION
The classification report in table 3 and table4 provided a detailed breakdown of the model's performance on the validation set. The report included precision, recall, and F1-score for each class:
- ReduceLROnPlateau: This callback played a crucial role in refining the model. Initially, a higher learning rate allowed the model to converge quickly. As the learning rate was reduced upon plateauing of the validation loss, the model could make more precise adjustments, leading to better fine-tuning and improved accuracy on the validation set.

Training Results
The training results showcased the effectiveness of the training strategy. For instance, the multi-class classification model showed an initial rapid improvement in accuracy, followed by stabilization as the learning rate was reduced. The validation loss and accuracy wre closely monitored, and the model checkpoint saved the best performing model.
The binary classification model for cancer detection similarly benefitted from this strategy, achieving high precision, recall, and F1-scores on the validation set. The final classification reports for both multi-class and binary classification models highlighted the model's performance in detail.

Table 3: Key Observations and Observation

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	8.1181	30.72%	7.6832	14.63%
5	3.3551	64.74%	94.0215	6.83%
10	1.2784	79.70%	4.3213	13.17%
16	0.5782	92.55%	3.9850	20.00%
28	0.2156	99.35%	0.9506	75.12%
50	0.0738	100.00%	0.1907	98.05%

Table 4: Classification of Report

Lesion Type	Precision	Recall	F1-Score
AKIEC	0.76	0.63	0.69
BCC	0.83	0.67	0.74
BKL	0.64	0.70	0.67
DF	0.47	0.64	0.54
MEL	0.67	0.53	0.59
NV	0.58	0.70	0.64
Normal	1.00	1.00	1.00
VASC	0.78	1.00	0.88

Overall Accuracy: 72%

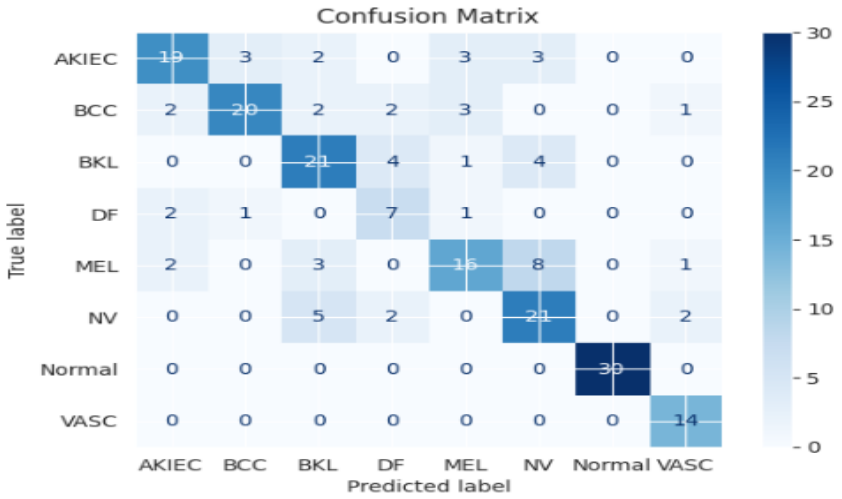

Figure 6: Confusion Matrix

These results indicated that the model performed well on distinguishing between lesions, achieving high precision, recall, and F1-scores. However, there was variability in performance across the different lesion types, with some classes achieving lower scores. This variability highlighted areas for potential improvement, such as additional data augmentation, model tuning, or exploring alternative architectures.

Table 5: Training Progress and Result

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	5.9211	72.79%	9.3217	21.46%
8	0.6336	98.33%	3.2017	20.98%
13	0.3504	99.57%	0.7971	89.27%
28	0.1510	99.84%	0.1653	99.51%
50	0.0738	100.00%	0.1907	98.05%

Figure 7: Training and Validation

The training process demonstrated the model's ability to generalize well to the validation data, achieving high accuracy and low loss values over multiple epochs.
Classification Report for Cancer Diagnosis: The classification report for the cancer diagnosis model provided detailed metrics for the cancerous and non-cancerous categories. The report included precision, recall, and F1-score for each category:
Classification Report: - Cancerous: Precision: 0.99, Recall: 0.98, F1-Score: 0.99. - Non-Cancerous: Precision: 0.93, Recall: 0.98, F1-Score: 0.96

The model showed high accuracy (98%) for binary classification of cancerous vs. non-cancerous lesions but struggled with multi-class classification, achieving only 72% accuracy. The F1 score varied, with challenges in identifying certain lesion types like Dermatofibroma (0.54). Compared to other studies, which achieved accuracies above 93% by integrating multimodal data or using advanced attention mechanisms, the current model's multi-class performance is lower. This suggests that incorporating additional patient information and exploring more sophisticated architectures could enhance classification accuracy and consistency across all lesion types.

Table 6: Classification Report

Class	Includes	Precision	Recall	F1-Score
Cancerous	AKIEC, BCC, BKL, DF, MEL, NV	1.00	0.98	0.99
Non-Cancerous	Normal, VASC	0.92	1.00	0.96

Overall Accuracy: 98%

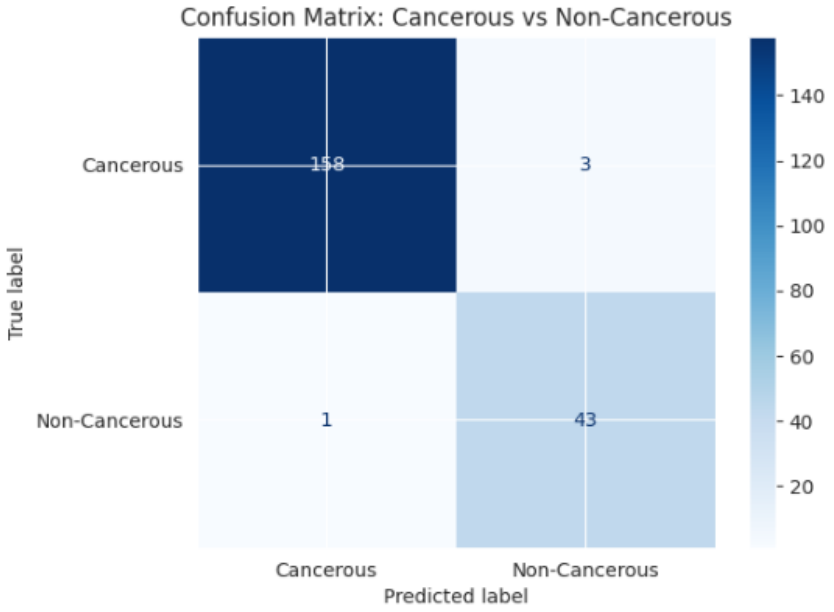

Figure 8: Confusion Matrix: Cancerous vs Non-Cancerous

The model was quite successful in differentiating between malignant and non-cancerous lesions, as seen by the excellent precision and recall values for both categories. With an overall accuracy of 98%, the model proved to be dependable and durable for the binary classification task.

CONCLUSION

An important development in medical imaging technology is the use of attention processes in the identification of skin lesions. By employing these principles, the model has proven to perform better in properly detecting and classifying skin lesions. Attention processes enhance sensitivity and specificity, two critical metrics in medical diagnostics, by concentrating on pertinent features and areas within the images. In addition to helping with early detection, this capacity lowers false positives and boosts diagnostic confidence in general.
Utilizing the HAM10000 dataset, which includes 10,015 dermatoscopic images of pigmented lesions, the implementation used in this study produced precise class labels to guarantee that the images were correctly identified and categorized, as well as directories to further organize the dataset. Class weight is used to address unbalanced datasets, while splitting is used to guarantee a balanced training process.
On a dataset of more than 10,000 image instances, the experimental results demonstrate that the model that classifies the images into different lesion types had the lowest accuracy of 72%, while the model that determines whether the lesion is cancerous or non-cancerous had the highest accuracy of 98%, demonstrating its robustness and reliability for the binary classification task. With a precision of 1.0, Normal had the highest F-1 score, while DF had the lowest at 0.54. Thus, this study's methodology for classifying skin lesions as either malignant or non-cancerous was effective. The development of a lesion detection system based on machine learning is advised.

REFERENCES

Alzubaidi L., Zhang J., Humaidi A.J., Al-Dujaili A., Duan Y., Al-Shamma O., Santamaría J., Fadhel M.A., Al-Amidie M., Farhan L. (2021). Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions.

Volume 8. Springer International Publishing; Cham, Switzerland.

Balaji, D., Dong, C., Xu, S., Yan, Q., Li, Z., Zhang, C., Luo, N., (2020). Ms red: A novel multi-scale residual encoding and decoding network for skin lesion segmentation. Medical Image Analysis 75, 102293.

Debelee W., Li M., Wu R., Du W., Zhang S., Yin S., Chen Z., Huang H. (2020). The Design and Application of an Automated Microscope Developed Based on Deep Learning for Fungal Detection in Dermatology. Mycoses. 64:245–251. https://doi.org/10.1111/myc.13209

Eze, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2023). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118.

Goyal P., Caron M., Lefaudeux B., Xu M., Wang P., Pai V., Singh M., Liptchinsky V., Misra I., Joulin A. (2021). Self-Supervised Pretraining of Visual Features in the Wild. arXiv. 20212103.01988

Grignaffini G.A., Jung N.L., Valle A.A., Bhattacharya S.D., Keel C.E. (2022). Robotic Inguinal Lymph Node Dissection for Melanoma: A Novel Approach to a Complicated Problem. J. Robot. Surg. 12:745–748. https://doi.org/10.1007/s11701-017-0776-z

Hauser, A., Ying Cao, Xuguang Wang, Peng Li, and Wei Xu. (2022). Deep recurrent models with fast-forward connections for neural machine translation. CoRR, abs/1606.04199, 2016.

Hoang L., Lee S.-H., Lee E.-J., Kwon K.-R. Multiclass skin lesion classification using a novel lightweight deep learning framework for smart healthcare. *Appl. Sci.* 2022;12:2677. https://doi.org/10.3390/app12052677

Jeong J., Y. Xie, Y. Xia, and C. Shen. (2023). Attention residual learning for skin lesion classification. IEEE transactions on medical imaging, 2019. 1, 2, 6, 7

Li H., He X., Zhou F., Yu Z., Ni D., Chen S., Wang T., Lei B. (2021). Dense deconvolutional network for skin lesion segmentation. *IEEE J. Biomed. Health Inform.* 2018;23:527–537. https://doi.org/10.1109/JBHI.2018.2859898

Meng, H., Marghoob, A. A., Argenziano, G., Carrera, C., Curiel-Lewandrowski, C., Hofmann-Wellenhof, R., & Zalaudek, I. (2021). Standardization of terminology in dermoscopy/dermatoscopy: Results of the third consensus conference of the International Society of Dermoscopy. Journal of the American Academy of Dermatology, 74(6), 1093-1106.

Mohammed, A. and Al-Tuwaijari, I. (2021). Spot-adaptive knowledge distillation. IEEE Transactions on Image Processing.

Montagna W. (2012). *The Structure and Function of Skin*. Amsterdam, Netherlands:Elsevier.

Nwoye H., Ferlay J., Siegel R.L., Laversanne M., Soerjomataram I., Jemal A., Bray F. (2021). Global Cancer Statistics: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J. Clin. 2021;71:209–249. https://doi.org/10.3322/caac.21660

Salidi, A. (2005). Machine learning and health care disparities in dermatology. JAMA Dermatology, 154(11), 1247-1248.

Sanga, P., Rosendahl, C., & Kittler, H. (2023). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5(1), 1-9

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

Wolff, K., Johnson, R. A., & Saavedra, A. P. (2017). Fitzpatrick's Color Atlas and Synopsis of Clinical Dermatology. McGraw-Hill Education.

Yang Y, Cui Z, Xu J, Zhong C, Wang R, Zheng WS. (2019). Continual Learning With Bayesian Model Based on a Fixed Pre-Trained Feature Extractor. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer)*. New York, NY, United States: Springer; (2021).

Zhang L, Yang G, Ye X. (2019). Automatic Skin Lesion Segmentation by Coupling Deep Fully Convolutional Networks and Shallow Network With Textons. *J Med Imaging* (2019)

Zia Ur Rehman S., Xie B., Li Y., Zhao X., Kuang Y., Su J., He X., Wu X., Fan W., Huang K., et al. (2020). Smart Identification of Psoriasis by Images Using Convolutional Neural Networks: A Case Study in China. J. Eur. Acad. Dermatol. Venereol. 2020;34:518–524. https://doi.org/10.1111/jdv.15965