

Article

Not peer-reviewed version

Skin Cancer Detection Using Transfer Learning and Deep Attention Mechanisms

[Areej Alotaibi](#) * and [Duaa AlSaeed](#)

Posted Date: 26 November 2024

doi: 10.20944/preprints202411.1898.v1

Keywords: skin cancer; deep learning; transfer learning; attention mechanism; computer vision; pretrained models; Xception; dermoscopic images; medical imaging



Preprints.org is a free multidisciplinary 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 open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Skin Cancer Detection Using Transfer Learning and Deep Attention Mechanisms

Areej Alotaibi * and Duaa AlSaeed

College of Computer and Information Sciences, King Saud University, Riyadh 11451, Saudi Arabia

* Correspondence: areej.400400a@gmail.com

Abstract: Early and accurate diagnosis of skin cancer improves survival rates; however, dermatologists often struggle with lesion detection due to similar pigmentation. Deep learning and transfer learning models have shown promise in diagnosing skin cancers through image processing. Integrating Attention Mechanisms (AMs) with deep learning have further enhanced the accuracy of medical image classification. While significant progress has been made, further research is needed to improve detection accuracy. Previous studies have not explored the integration of attention mechanisms with the pre-trained Xception transfer learning model for binary classification of skin cancer. This study investigates the impact of various attention mechanisms on the Xception model's performance in detecting benign and malignant skin lesions. Using the HAM10000 dermatoscopic image dataset, four experiments were conducted. Three models incorporated self-attention (SL), hard-attention (HD), and soft-attention (SF), respectively, while the fourth model used standard Xception without AMs. Results demonstrated the effectiveness of AMs, with models incorporating self, soft, and hard attention mechanisms achieving accuracies of 94.11%, 93.29%, and 92.97%, respectively, compared to 91.05% for the baseline model, representing a 3% improvement. Both self-attention and soft-attention models outperformed previous studies on recall metrics, which are crucial for medical investigations. These findings suggest that AMs can enhance performance on complex medical imaging tasks, potentially supporting earlier diagnosis and improving treatment outcomes.

Keywords: skin cancer; deep learning; transfer learning; attention mechanism; computer vision; pretrained models; Xception; dermoscopic images; medical imaging

1. Introduction

Cancer, a collective term referring to a group of related diseases, occurs when certain body cells begin to divide uncontrollably and invade nearby tissues [1]. Cancer commonly affects the skin, prostate, lungs, breasts and pancreas, and is a primary cause of death worldwide [2]. Skin cancer is the most prevalent, accounting for 75% of all cancer cases globally [3].

Skin diseases involve abnormal changes in the outer layer of the skin. According to the World Health Organization (WHO), skin cancer accounts for one-third of global cancer cases. In Saudi Arabia, it is the ninth most common malignancy [4]. Skin cancer primarily develops on sun-exposed areas such as the head, face, lips, ears, neck, chest, arms, and hands, as well as on the legs in women. Several risk factors may cause skin cell damage, including genetics, lighter natural skin color, older age, and certain medical conditions. Most cases result from overexposure to ultraviolet (UV) light, with risk increasing with exposure duration [5].

Skin cancer includes two primary categories: malignant melanoma (MM) and non-melanoma skin cancer (NMSC), each exhibiting distinct clinical outcomes. Malignant melanoma is the least common form of skin cancer, accounting for only 5% of cases, and early detection plays a crucial role in determining the patient's prognosis [6]. While NMSC does not significantly contribute to overall cancer mortality, its incidence increases with age. A study presented at the Conference (EADV) 2023 revealed that NMSC is responsible for more global deaths than melanoma skin cancer [7]. The two

primary types of NMSC are basal cell carcinoma (BCC) and squamous cell carcinoma (SCC), with other types less common. BCC accounts for approximately 80% to 85% of NMSC cases, while SCC comprises 15% to 20%. SCC demonstrates a higher propensity for metastasis compared to BCC [6]. The contrast in the propensity for metastasis between MM and NMSC positions MM as the primary contributor to mortality among individuals with skin cancer.

Early diagnosis significantly increases the likelihood of recovery. Detecting skin cancer in its initial stages not only facilitates easier treatment but also enhances overall prognosis. According to Cancer Research UK, the chances of successful treatment are higher when cancer is discovered before it has advanced or spread [8].

Currently, dermatologists rely on biopsy, a method involving sample removal from suspicious lesions, which is inconvenient and potentially painful for patients [9].

As a result, various anatomical and molecular imaging techniques have been developed and utilized to detect different types of skin cancer. However, there are over 2000 dermatological diseases, and similar-looking skin lesions from different conditions can complicate visual examination and lead to misdiagnosis [10].

The use of artificial intelligence (AI) and deep learning (DL) in medical diagnostics significantly enhances skin cancer diagnostic accuracy beyond visual examination alone. However, optimal deep learning performance requires extensive medical image data. Transfer learning is proposed in this study as a solution to this challenge.

Transfer learning leverages knowledge from a pre-trained model to learn new tasks in a target domain. This approach is particularly effective when target data is scarce, leading to extensive research in deep transfer learning methods for skin cancer diagnosis and classification.

Deep attention mechanisms (AMs) have also gained increased attention in the classification of medical images [11,12] because of the promising results obtained when DL algorithms are combined with deep attention mechanisms. Deep attention mechanisms give special attention to specific areas of the image instead of according to equal importance to all batches of the image. This special attention increases the possibility of enhancing the performance of the classification technique.

Despite advances in skin cancer detection, further research is needed to explore techniques that enhance accuracy. No previous research has explored attention mechanisms with pre-trained Xception transfer learning for binary skin cancer classification. This study investigates Xception-based deep transfer learning, both with and without attention mechanisms, in detecting benign and malignant skin lesions. This model could help in the early detection of skin cancer, which, in turn, could enhance the chances of successful treatment.

The major contributions of this study are presented below:

- Proposal of a novel model based on the Xception architecture that incorporates various AMs for binary classification of skin lesions as benign or malignant.
- A thorough investigation of how different AMs impact the Xception model's performance.
- Comparison of the proposed models with recent state-of-the-art skin cancer detection methods in binary classification, using the same dataset.

This study is divided into the following sections: Section 2 presents the related works; materials and methods are discussed in Section 3; experimental results and discussion are presented in Section 4; and the study draws conclusions in Section 5.

2. Related Work

In recent years, numerous studies have employed deep learning-based approaches to diagnose and classify skin cancer. These approaches have demonstrated improved performance compared to traditional machine learning methods [13,14]. Most recent literature uses convolutional neural networks (CNNs) and has shown competitive performance in diagnosing skin lesions, as demonstrated in studies [15–19], to name several examples. One recent study [16] used a CNN to classify seven skin cancer types from the HAM10000 dataset. Using Enhanced Super Resolution Generative Adversarial Network (ESRGAN) for image enhancement, the model achieved accuracies of 98.77%, 98.36%, and 98.89% for protocols I, II, and III respectively.

CNN was also proposed in [20] for skin lesion detection and classification, using systematic meta-heuristic optimization and CNN-based image detection techniques. The study compared various Keras classifiers, including MobileNet, on HAM10000 and ISIC-2017 datasets. The proposed system outperformed others, achieving 78% accuracy.

Deep convolutional networks have excelled in image segmentation and localization. Nawaz et al., [21] proposed a novel approach to locating and segmenting melanoma cells using faster region-based convolutional neural networks (RCNN) with fuzzy k-means clustering. They applied this method to three datasets: ISBI-2016, ISIC-2017, and PH2. Their approach outperformed existing methods, achieving average accuracies of 95.40%, 93.1%, and 95.6% on the respective datasets.

Another study [22] detected melanoma using a novel lightweight convolutional neural network (LWCNN). The study utilized the HAM10000 dataset, with images labeled as either melanoma or non-melanoma. The proposed LWCNN model outperformed pre-trained models like GoogLeNet, ResNet-18, and MobileNet-v2. The model achieved 91.05% accuracy, with only 22.54 minutes of total processing time. On the same dataset, Renith and Senthilselvi [23] proposed a novel approach for classifying skin lesions as benign or malignant. Their model combined an Improved Adaboost algorithm with Aphid-Ant Mutualism optimization, built on the AlexNet architecture. This approach outperformed existing methods, achieving 95.7% accuracy, 95% specificity, 94.8% sensitivity, 95.4% precision, and a 95% F-measure.

Recent research has focused on exploring transfer learning approaches. In [24], a MobileNetV2-based deep transfer learning model was introduced for melanoma classification using 33,126 images extracted from the SIIM-ISIC 2020 dataset. The model achieved 98.2% accuracy, outperforming other techniques in both accuracy and computational efficiency.

Authors of a different study [25] introduced SCDNet, a novel method based on VGG16 and CNN architectures, to classify 25,331 images from ISIC 2019 into four skin cancer categories. This model achieved 96.91% accuracy, surpassing the performance of ResNet50 at 95.21%, AlexNet at 93.14%, VGG19 at 94.25%, and Inception V3 at 92.54%.

The pre-trained ResNet and InceptionV3 models were used in [26] to extract features from 14,033 dermoscopic images from HAM 10000, ISBI 2016, and ISBI 2017 datasets. After applying data augmentation techniques, a CNN was used for classification. The method achieved accuracies of 89.30%, 97.0%, and 94.89% on the datasets.

Imran et al., [27] developed an ensemble CNN model that combined VGG, Caps-Net, and ResNet to detect cancerous and non-cancerous conditions. They used 25,000 images from the ISIC dataset. The results demonstrated that this ensemble approach of pre-trained models achieved 93.5% accuracy, with a training time of 106 seconds, outperforming individual models across various performance metrics.

Accurate detection and diagnosis of skin lesions can be challenging due to the similarity in appearance of various types of lesions, such as melanoma and nevi, especially in color images. Ashraf et al., proposed an automatic classification approach using a pre-trained AlexNet CNN model, with Region of Interest (ROI) for accurately extracting discriminative features of melanoma [28]. The model, trained on 3,738 images from DermIS and DermQuest with extensive augmentation, achieved 97.9% and 97.4% accuracy on the datasets.

There has been growing interest in the comparative analysis of different transfer learning approaches. One study [29] evaluated six networks (VGG19, InceptionV3, InceptionResNetV2, ResNet50, Xception, and MobileNet) on the HAM10000 skin cancer dataset. Xception outperformed others, achieving 90.48% accuracy, 89.57% recall, 88.76% precision, and an 89.02% F-measure.

Transfer learning models have encouraged research into their effectiveness for classifying skin cancer images in both binary and multiple categories. In [30], the authors evaluated modified EfficientNet V2-M and B4 models for both multiclass and binary classification of skin lesions, using 58,032 dermoscopic images from ISIC datasets. EfficientNetV2-M outperformed in both tasks. For multiclass classification, it achieved accuracies of 97.62%, 95.49%, and 94.8% on ISIC 2020, 2019, and HAM datasets, respectively. In binary classification, it reached 99.23%, 97.06%, and 95.95% accuracy on the same datasets.

Another study [31] performed binary (cancer vs. non-cancerous) and multiclass (six lesion types) classification on 2,298 dermoscopy images from the PAD-UFES-20 dataset, using a CNN-based transfer learning model. The pre-trained CNN model improved accuracy rates by over 20% compared to conventional models. The proposed model achieved a mean accuracy and an F1 score of 0.86 for both classification types.

Many studies have used pre-trained models on the HAM10000 dataset for binary melanoma classification. These studies typically focus on melanoma cases, which comprise 1,113 instances in the dataset, while classifying the remaining skin lesion types as non-melanoma. One such study in [32] tested four pre-trained models (ResNet50, InceptionV3, Xception, and VGG16) with three segmentation techniques (SegNet, BCDU-Net, and U-Net). The results showed that Xception performed best, achieving accuracies of 95.2% and 95.2% with BCDU-Net and SegNet, respectively. The study in [33] used augmented HAM and PH2 datasets, totaling 18,004 images, to classify melanoma. Combining handcrafted features, EfficientNet-B0 and hair removal techniques, they achieved 94.9% and 98% accuracy on the HAM and PH2 datasets, respectively. Another study [34] compared a non-pretrained CNN with three pretrained models (MobileNetV2, EfficientNetV2, and DenseNet121) for melanoma classification on the HAM dataset. The pretrained models achieved better overall accuracies of 93.77%, 93.60%, and 93.34%, respectively, compared to the CNN, which achieved 91.35%.

Previous studies focused on exploring deep learning and transfer learning techniques, treating all image patches equally. A shift in research introduced deep attention mechanisms to highlight regions of interest and extract optimal features, potentially enhancing skin cancer detection accuracy. In [35], the authors proposed a Soft Attention-Based Convolutional Neural Network (SAB-CNN) for classifying HAM dataset images. (SMOTE) was used to address dataset imbalance. The model achieved 95.94% accuracy, 95.30% Matthews Correlation Coefficient, and 95.97% Balanced Accuracy Score. This study highlighted the importance of attention mechanisms and data balancing in improving deep neural network performance.

The use of AMs has encouraged research investigating the effectiveness of CNNs with AMs [12,36,37]. One study [12] examined soft attention's impact on five deep neural networks (ResNet34, ResNet50, Inception ResNet v2, DenseNet201, and VGG16) for skin cancer image classification. Soft attention aims to enhance crucial elements while reducing noise. Results showed a 4.7% improvement over the baseline, achieving 93.7% precision on the HAM10000 dataset. On the ISIC-2017 dataset, soft-attention coupling improved sensitivity by 3.8%, reaching 91.6% compared to other approaches.

Another study [11] proposed a dual-track deep learning model for skin cancer classification. The first track used a modified DenseNet-169 with a Coordinate Attention Module (CoAM) for local features, while the second employed a custom CNN with a feature pyramid and global context networks for multiscale and global features. By combining these features, the model achieved 93.2% accuracy, 95.3% precision, 91.4% recall, and a 93.3% F1-Score on the HAM10000 dataset.

Table 1 provides a concise overview of related works, revealing that few studies have combined attention mechanisms with pre-trained models for binary skin cancer classification. Notably, no previous research has explored the use of deep attention mechanisms with the pre-trained Xception model for diagnosing skin cancer in a binary classification context. Studies reviewed in the previous section demonstrate that integrating AMs has shown promising accuracy in extracting spatial information, and highlighting regions of interest in images.

This encourages further investigation into the effects of integrating Xception-based deep transfer learning with AMs for detecting benign and malignant skin lesions. Our contribution in this work is to explore and compare the use of different types of deep attention mechanisms with Xception in the detection of skin cancer. Based on a review of related works, this approach has not been previously employed.

Table 1. Skin cancer detection and classification studies.

Ref	Approaches	Dataset	Classification Type	Evaluation metrics			
				Precision	Recall	Accuracy	F1-Score
[20]	CNN	HAM1-ISIC 2017	multi-class	NA	NA	78%	NA
[16]	CNN	HAM	multi-class	NA	NA	98.89%	NA
[21]	RCNN-FKM	ISIC-2016 ISIC-2017 PH2	Binary	NA	97.2%	96.1%	NA
[23]	Adaboost + IAB-AAM + AlexNet	HAM	Binary	95.4%	94.8%	95.7%	95%
[22]	LWCNN	HAM	Binary	NA	NA	91.05%	NA
[25]	CNN-VGG16	ISIC 2019	Multi-class	92.19%	92.18%	96.91%,	92.18%
[28]	Pre-trained CNN model AlexNet + ROI	DermIS DermQue t	Binary	NA	NA	97.9%	NA
[24]	MobileNetV2	ISIC-2020	Binary	98.3%	98.1%	98.20%	98.1%
[31]	Pre-trained CNN	PAD-UFES-20	Binary/ multi-class	B=88% M=90%	B=81% M=83%	B=86% M =NA	B=NA M =86%
[38]	Modified VGG16 architecture	Kaggle	Binary	NA	NA	89.09%	93.0%
[27]	CNN-VGGNet, CapsNet, and ResNet	ISIC	multi-class	94%	NA	93.5%	92.0%
[26]	CNN-ResNet, InceptionV3	ISBI 2016 ISBI 2017 Ham	Multi-class	95.30%	NA	95.89%	94.90%
[33]	HC + ResNet50V2 and EfficientNet	HAM and PH2	Binary	92.8%	97.5%	98%	95%
[30]	EfficientNet V2-M and EfficientNet-B4	ISIC 2020, ISIC 2019 HAM	Multi-class / Binary	B=96% M=96%	B=95% M=95%	B= 97.06% M=95%	B=95% M=95%
Ref	Approaches	Dataset	Classification Type	Evaluation metrics			
				Precision	Precision	Precision	Precision
[29]	Six transfer learning networks	HAM	Multi-class	88.76%	89.57%	90.48%,	89.02%
[32]	Four pretrained + image	Ham	Binary	NA	94.16%	96.10%	96.02%

segmentation							
n							
[34]	MobileNetV2,	Ham	Binary	93.77%	89.78%	93.77%	93.51%
	EfficientNetV2+						
	DenseNet121 + CNN						
	CNNs +Soft attention						
[35]	Six pre-trained models+ Soft attention	HAM and ISIC 2017	Multi-class	93.7%	NA	93.4%	NA
[12]	Densenet-169 with CoAM + customized CNN	HAM	Binary	95.3%	91.4%	93.2%	93.3%
[11]		HAM	Binary	95.3%	91.4%	93.2%	93.3%

3. Materials and Methods

As previously stated, this study investigates the impact of integrating Xception deep transfer learning methods with different attention mechanisms (SL, SF, and HD) for detecting skin cancer in dermoscopy images. Figure 1 illustrates the architecture of the proposed model.

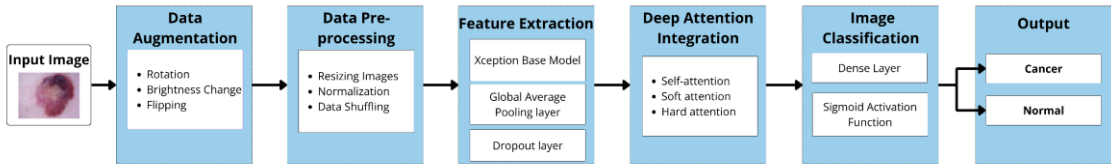


Figure 1. Model architecture

In the following sections, the implementation of the main components of our proposed models are described. This includes the dataset description, data augmentation, data pre-processing, Xception-based models for feature extraction, deep attention integration, image classification, and model evaluation. The specifications related to each model are identified, both with and without attention mechanisms (AMs), highlighting any differences between them.

3.1. Dataset

In this study, the HAM10000 ('Human Against Machine') dataset [39,40]], a collection of pigmented skin lesion images publicly available on the ISIC archive, was used [41]. The dataset comprises 10,015 images representing seven types of pigmented skin lesions. These types include actinic keratosis (AKIEC), basal cell carcinoma (BCC), benign keratosis (BKL), dermatofibroma (DF), melanocytic nevi (NV), melanoma (Mel), and vascular skin lesions (VASC). Figure 2 illustrates examples of these seven lesion types, while Figure 3 displays their class distribution. The x-axis represents the lesion types, and the y-axis shows their corresponding counts.

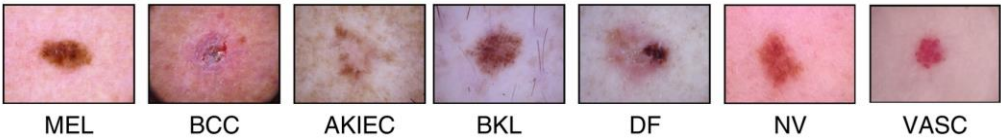


Figure 2. HAM10000 dataset.

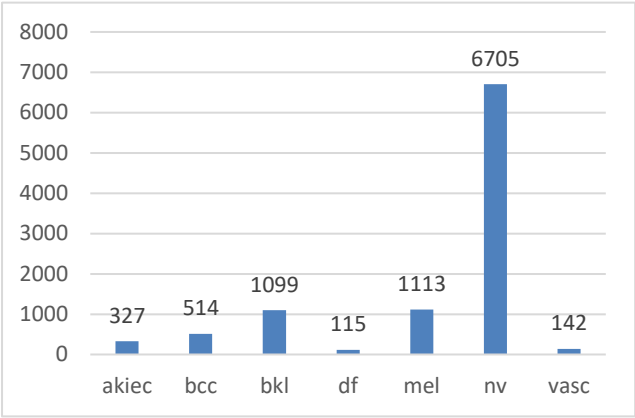


Figure 3. Class Distribution in the Ham10000 Dataset.

As the HAM10000 dataset is multi-class, the study focused on binary classification. The seven classes were classified into either malignant (cancerous) or benign (normal) groups. MEL and BCC were grouped as cancer, while DF, BKL, NV, VASC, and AKIEC were identified as normal. The dataset thus comprised two binary classes: cancer and normal, as illustrated in Table 2.

Table 2. Image Statistics of the Ham10000 Dataset.

Cancer			Total	Normal				Total
MEL	BCC	AKIEC		DF	BKL	NV	VASC	
1113	514	327	1954 (19.56%)	115	1099	6705	142	8,061 (80.49%)

The normal class comprised 80% of the entire image dataset, leading to an imbalanced database that significantly impacted the training process. The cancer class represented only 19.56% of the images, with the AKIEC labels constituting less than 3% of the total. To ensure a balanced dataset, data augmentation techniques were implemented.

3.2. Data Augmentation

Various data augmentation techniques were applied to the cancer class, including rotation, brightness adjustment and flipping, using the real-time image data generator function from the Keras library in Python. These techniques increased the sample size of the cancer class and enhanced the diversity of the training data.

The original images were rotated by up to 40°, applying a random rotation angle between -40° and 40°. The brightness of the images was rotated to between 1.0 and 1.3 times the original brightness, simulating different lighting conditions. We also flipped the images randomly, both vertically and horizontally. The parameters and their selected values are provided in Table 3.

Table 3. Parameters of data augmentation.

Parameters	Values	Description
Rotation range	40	Randomly rotate images within a range of 40 degrees
Brightness range	[1.0,1.3]	Adjust brightness 1.0-1.3 times original

Horizontal flip	True	Flipping the Image horizontally
Vertical flip	True	Flipping the Image Vertically

After augmentation, three augmented versions were generated for each original cancer image, increasing the size of the cancer class from 1,954 to 7,816 images. This brought the cancer class size much closer to the normal class size, resulting in a total dataset of 15,877 dermoscopy images. The statistics in Figure 4 illustrate the differences in the dataset before and after augmentation.

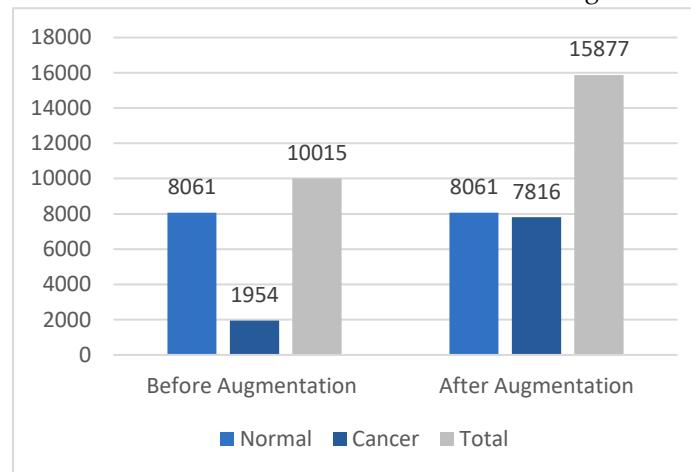


Figure 4. Difference of HAM10000 Dataset Size Before and After Data Augmentation.

3.3. Data Preprocessing

After augmenting the images, preprocessing and preparation procedures were applied to the dermatoscopic images from the HAM10000 dataset. These procedures encompassed resizing, normalization, and data shuffling. The dermatoscopic images were resized from 450 × 600 pixels to 299 × 299 pixels to match the Xception model's default input size. Pixel values from the 0-255 range to the 0-1 range were then standardized, which is optimal for neural network models. Finally, to prevent bias during training, data shuffling was applied, ensuring randomness in batch selection and preventing the model from learning patterns based on the order of the data.

3.4. Feature Extraction with Pre-Trained Xception Models

The Xception base model serves as a powerful feature extractor due to its pre-training on the extensive ImageNet dataset. When loading the Xception model, the "include top" parameter was set to "false". This meant that the fully connected top layers originally designed for the ImageNet classification task were not loaded. Instead, these layers were replaced by custom fine-tuning layers suited for the specific classification task at hand. This approach leveraged pre-learned knowledge, provided flexibility for customization, improved accuracy, enhanced computational efficiency, and substantially reduced the number of unnecessary parameters.

Following the base Xception model, a GlobalAveragePooling2D layer was added to reduce spatial dimensions by extracting global features from the feature maps generated by the Xception base model. This was followed by a dropout layer to prevent overfitting, enhancing the model's generalization capabilities.

3.5. Deep Attention Integration

An attention layer was integrated into Xception architecture, which was one of three mechanisms: hard, soft, or self-attention. Figure 5 shows an illustration of the Xception model with AMs.

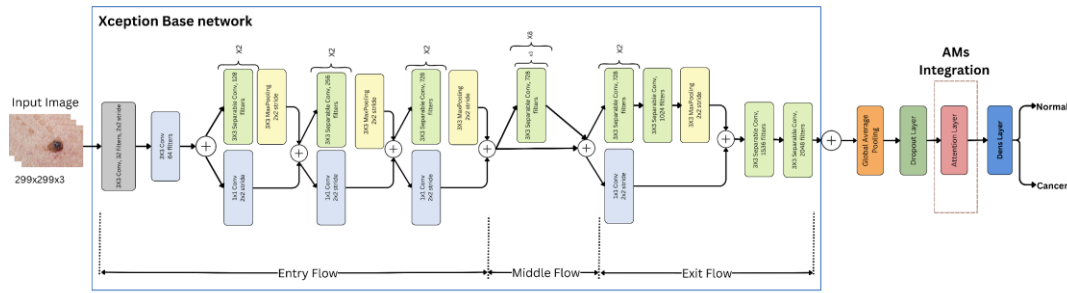


Figure 5. Xception model with AMs.

Each of these mechanisms was integrated separately into the overall architecture, utilizing the same preceding and subsequent layers and procedures. Each attention mechanism (AM) layer has its own method for analyzing the features extracted from the base Xception model. These methods determined which parts of the image were most important for the classification task, as follows:

- 1 SL layer: This layer transformed the input into query (Q), key (K), and value (V) vectors through linear transformations. Attention scores were computed as the dot product of the query with all keys divided by $\sqrt{d_k}$. These scores were then normalized using Softmax to obtain attention weights for the values [42]. In this project, self-attention was internally implemented using Keras's built-in attention layer [43], following this equation:

$$\text{SelfAttention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

- 2 SF layer: This layer discredited irrelevant areas of the image by multiplying the corresponding feature maps with low weights. The low attention areas had weights closer to 0, allowing the model to focus on the most relevant information, which enhanced performance [12]. A dense layer was used with softmax activation to compute attention weights α_i for each feature x_i , where Softmax ensured that these weights sum to 1, as shown in the following equation [44]:

$$\alpha_i = \frac{\exp(w_i \cdot x_i)}{\sum_{j=1}^n \exp(w_j \cdot x_j)} \text{ for } i = 1, 2, \dots, n \quad (2)$$

These attention weights were then applied to the feature map x . using a dot product operation

$$y = \sum_{i=1}^n \alpha_i x_i \quad (3)$$

- 3 HD layer: This layer compelled the model to focus exclusively on crucial elements, disregarding all others. The weight assigned was either 0 or 1 for each input component. This applied a binary mask to the attention scores between queries Q and keys K . The mechanism assigned a value of 1 to the top k highest-scoring elements (selected by TopK), and 0 to the rest [45]. This forced the model to focus only on the most important elements, disregarding others, without involving gradients in the selection process. The process is represented by the following equation:

$$A_{hard}(Q, K) = 1[\text{score}(Q, K) \in \text{TopK}(\text{score}(Q, K), k)] \quad (4)$$

Based on the attention mechanism's type and analysis, a weighted feature map was created, assigning higher weights to the more relevant features and lower weights to the less important ones.

3.6. Image Classification

This is the last layer in the Xception model's architecture. At this stage, input images were classified into two main categories - normal and cancer - using the following process:

3.6.1. Dense (Fully Connected) Layer

The output from the AM's layers was flattened and fed into the dense layer. This process combined all the information gathered from the network's previous layers to classify the input image.

3.6.2. Sigmoid Layer

A sigmoid function was utilized to transform the output of the fully connected layer into binary (0 or 1), which was interpreted as classification probabilities.

3.6.3. Classification Layer

The dense layer with a sigmoid activation function served as the final classification layer. In the case of integrated AMs, it took the output of the AM layers as input and generated the final classification predictions into one of two classes (normal or cancer). To mitigate overfitting, L2 regularization techniques were applied to the weights of this dense layer. In the original Xception-based model without AMs, the dense layer with sigmoid activation took the learned features from the Xception base model and the Global Average Pooling layer to make final class predictions. L2 regularization was not used for this model as it was unnecessary.

3.7. Model Evaluation

After the training process, the proposed models were tested on the testing dataset. The architecture's performance was evaluated using accuracy, F1 score, precision, and recall. These performance metrics are explained in detail below, along with their definitions and equations. In these equations, TP stands for true positives, TN for true negatives, FN for false negatives, and FP for false positives.

3.7.1. Classification Accuracy

The model's ability to correctly classify samples compared to the total number of samples in the evaluation dataset is known as accuracy. It is calculated as follows:

$$\text{Accuracy} = \frac{\# \text{ correctly classified samples}}{\# \text{ all samples}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (4)$$

3.7.2. Recall

It is also referred to as sensitivity or True Positive Rate (TPR), indicating the rate of correctly classified positive samples. This metric is considered a crucial factor in medical research, as the aim is to miss as few positive cases as possible, leading to high recall [46].

$$\text{REC} = \frac{\# \text{ true positive samples}}{\# \text{ samples classified positive}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (5)$$

3.7.3. Precision

This represents the proportion of retrieved samples that are pertinent and is computed as the ratio of correctly classified samples to all samples assigned to a specific class.

$$\text{PREC} = \frac{\# \text{ samples correctly classified}}{\# \text{ samples assigned to class}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

3.7.4. F1 Score

As a widely used metric in binary and multi-class classification, the F1 score combines precision and recall through their harmonic mean. It balances these metrics, making it especially valuable for imbalanced datasets.

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{7}$$

3.7.5. False Alarm Rate (FAR)

Incorrectly classifying negative instances as positive occurs at a rate measured by the false positive rate. This metric quantifies how often a model raises erroneous alerts for cases that are actually negative.

$$FAR = \frac{\text{Incorrectly classified actual negatives}}{\text{All actual positives}} = \frac{FP}{FP + TN} \tag{8}$$

3.7.6. Cohen’s kappa

This quantitative measure assesses the level of agreement between two raters evaluating the same subject, while accounting for the possibility of chance agreement. This metric is widely adopted across various fields, including statistics, psychology, biology, and medicine.

$$k = \frac{P_A - P_E}{1 - P_E} \tag{9}$$

3.7.7. AUC Score and ROC Curve

The receiver operating characteristic (ROC) curve graphically illustrates the relationship between the false positive rate and the true positive rate of a classifier. The probability curve's area under curve (AUC) indicates how well the model distinguishes between classes. A higher AUC value reflects better class separation by the classifier.

4. Results and Discussion

In this section, the results and findings of the Xception architecture incorporating different AMs are discussed, beginning with an explanation of the experimental settings and then analyzing the results of the four developed models, comparing them to those of previous studies.

3.1. Experimental Settings

The proposed models were trained for skin cancer diagnosis on dermoscopic images from the HAM10000 dataset. Augmentation techniques were applied to expand the sample size and enhance the diversity of the training data. The dataset was split into 20% for testing and 80% for training. A 10-fold cross-validation technique was used for training and testing the models. Table 4 provides a detailed breakdown of the dataset split. To prevent bias, the order of the images was shuffled randomly using a data shuffling technique.

Table 4. Data Breakdown.

Dataset Size	Training Sets	Testing Sets
15,877	12,702	3,175

All four proposed models were trained using the Adam Optimizer with a learning rate of 0.001, a batch size of 32, and a total of 50 epochs across all folds. The binary cross-entropy loss function was used, with a default probability threshold of 0.5. The activation function in the model with AMs includes both sigmoid and Softmax, while the model without AMs uses only sigmoid. Table 5 provides an overview of the optimization hyperparameters used in the experiments.

Table 5. Hyperparameter Settings: Xception-Based Models with and without AMs.

Parameter	With AMs	Without AMs
Epochs	50	50
Dropout	0.7	Not Used
Shuffle	True	True
Activation function	Sigmoid/Softmax	Sigmoid
L2 Regularization	0.001	Not Used
Loss-Function	binary-cross-entropy	binary-cross-entropy
Probability Threshold	0.5	0.5
Optimizer	Adam	Adam
Learning rate	0.001	0.001
Batch size	32	32

Early stopping was applied, with a patience of 5 epochs, to prevent overfitting. Additionally, an error-handling technique was used that ensured robust training and maintained optimization integrity by handling errors. The best model weights were saved, based on decreased validation loss and higher accuracy.

3.2. Classification Results

Four experiments were conducted to evaluate the performance of the Xception model, both with and without three attention mechanisms. These experiments compared the base Xception model to versions incorporating different types of AMs, resulting in four distinct models: Xception (base), Xception-SL, Xception-SF, and Xception-HD. The performance of the proposed models was evaluated by measuring the accuracy, recall, precision, F1 score, Cohen's kappa, and AUC scores for each model.

Table 6 shows the results of the four models.

Table 6. Model classification results.

Models	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)	AUC	Cohen's Kappa
Xception (Base)	91.05%	91.68%	90.78%	91.23%	0.972	0.821
Xception-SL	94.11%	95.47%	93.10%	94.27%	0.987	0.882
Xception-SF	93.29%	95.28%	91.81%	93.51%	0.983	0.865
Xception-HD	92.97%	93.98%	92.32%	93.14%	0.983	0.859

The results of the experiments revealed a significant effect of AMs on the Xception model's performance. Incorporating AMs into the Xception model led to improvements across all metrics. In contrast, the base Xception alone exhibited the lowest performance on all metrics. Nevertheless, the results were good.

The results also showed a convergence between the three AMs models with Xception integration. Incorporating self-attention (Xception-SL) into the Xception architecture yielded the highest performance across all metrics. This suggests that the self-attention mechanism significantly improved the Xception architecture by capturing relationships between distant elements in a sequence.

The Xception-SF model demonstrated the second-best performance across all metrics, except for precision when compared to Xception-HD. Both models yielded similar and promising results. Notably, Xception-SF achieved a recall of 95.28%, almost matching that of Xception-SL.

The results strongly suggest that incorporating deep attention mechanisms improved Xception's performance, particularly in terms of recall—a critical metric in medical applications. All AM-enhanced Xception models achieved promising results. These recall performances indicate that each variant showed promise in effectively identifying skin cancer cases.

After examining the performance of the models through recall, accuracy, precision and F1 score, their agreement was evaluated using Cohen's kappa, which provided insights into the agreement between the models' predictions and the true labels.

Table 6 presents the aggregate Cohen's kappa scores obtained from 10-fold cross-validation. The Xception models, both with and without the three AMs, demonstrated strong performance for this metric. All models achieved scores between 0.821 and 0.882, indicating substantial agreement between their predictions and the ground truth.

Having discussed the performance of the models using Cohen's kappa, their performance was evaluated, as reflected in the AUC scores. Figure 6 illustrates the ROC curves and AUC results of four models. The AUC score reflects each model's ability to differentiate between classes—in this case, normal and cancer. The Xception models with the three AMs achieved convergence in their results, which were 0.98, while the Xception performed slightly lower, with a score of 0.97.

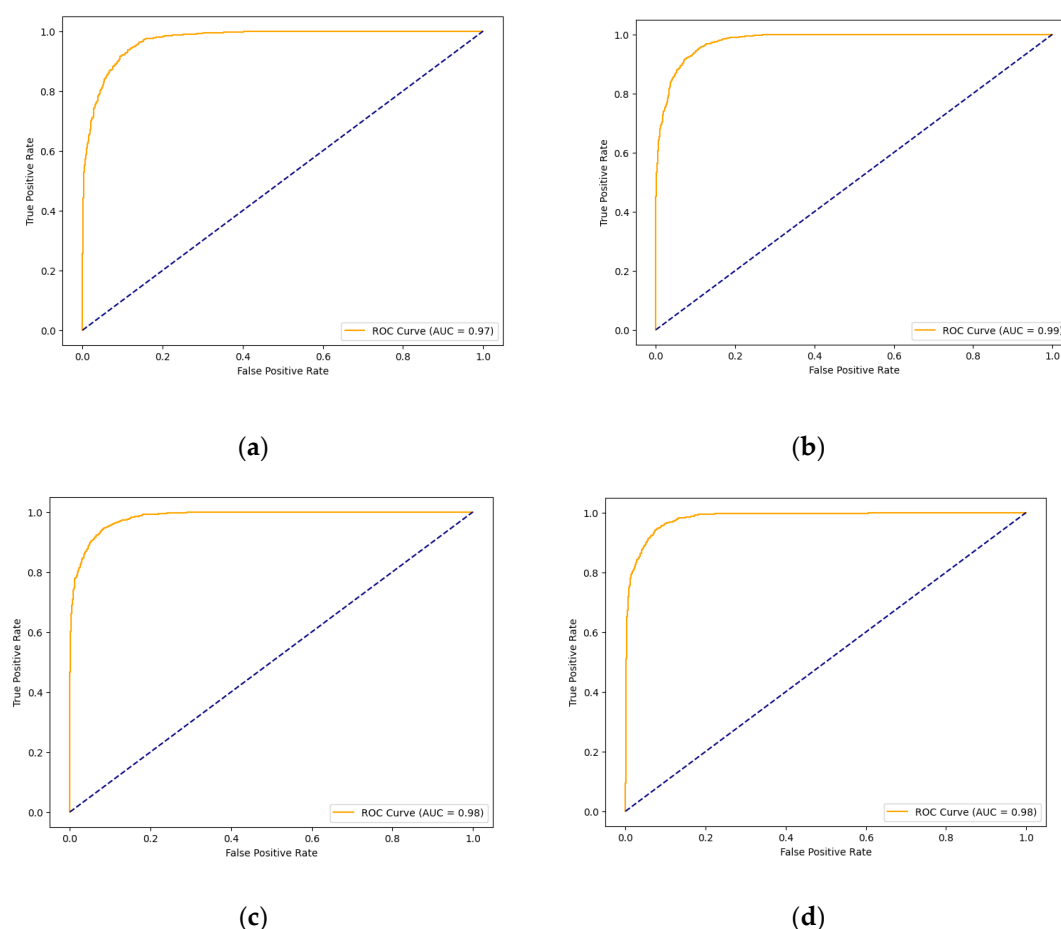


Figure 6. ROC curves of models for skin cancer: The ROC curve of the Xception model (a); the ROC curve of Xception-SL model (b); the ROC curve of Xception-HD model (c); the ROC curve of Xception-SF model (d).

Lastly, the confusion matrix results present a detailed view of the models' classification performance. Figure 7 shows the confusion matrices for four Xception models (the base Xception, Xception-SL, Xception-SF, and Xception-HD).

As shown in Figure 7, the Xception models with the three AMs exhibited similar performance patterns. The Xception-SL model demonstrated the best performance, correctly classifying 1,539 normal pigmented skin images and 1,449 cancerous images, with 114 false positives and 73 false negatives. The Xception-SF model accurately classified 1,536 normal pigmented skin images and 1,426 cancerous images, with 137 false positives and 76 false negatives. The Xception-HD model exhibited a slightly different pattern, correctly classifying 1,515 normal pigmented skin images and 1,437 cancerous images, but with a higher number of 126 false positives and 97 false negatives.

In contrast, the base Xception model correctly identified 1,478 normal pigmented skin images and 1,413 cancerous images, with the highest number of 150 false positives and 134 false negatives. These results highlight the varying impacts of different AMs on the models' classification accuracy and error distribution.

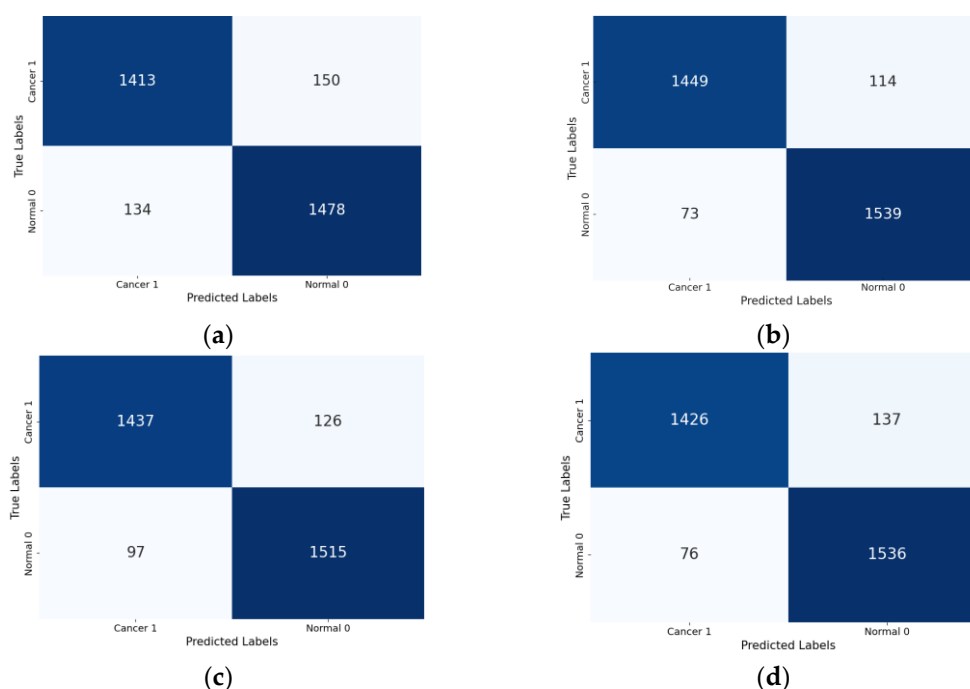


Figure 7. Confusion Matrices for Xception Models with/without AMs: Xception model (a); Xception-SL model (b); Xception-HD model (c); Xception-SF model (d).

From the confusion matrix, the false alarm rate was determined for each model. The Xception-SL model achieved the best results, with the highest classification rate of 94.11% and the lowest false alarm rate among all models at 6.90%. The Xception-HD and Xception-SF models demonstrated similar performance. The Xception-HD model achieved a classification rate of 92.97% with a false alarm rate of 7.68%, while the Xception-SF model slightly outperformed it with a classification rate of 93.29% but had a slightly higher false alarm rate of 8.19%. The Xception alone had the highest false alarm rate of 9.21%, which negatively impacted its overall classification rate of 91.05%. These results highlight the significant impact of different AMs on the models' ability to accurately classify skin lesions while minimizing false positives.

Our extensive performance evaluation and analysis indicates that the Xception model, incorporating three attention mechanisms (self, hard, and soft), consistently outperformed the standard Xception model across all metrics. These findings strongly suggest that integrating attention mechanisms into the Xception architecture effectively enhances the detection of pigmented skin lesions in dermatoscopic images. This highlights the potential of attention mechanisms to improve performance on complex medical imaging tasks, particularly in oncological applications.

Despite the promising results achieved by our proposed models, several factors affected the models' ability to achieve higher performances. One factor is the limited size and diversity of the HAM1000 dataset, which led to overfitting. While Xception-based models with/without AMs performed exceptionally well on the training set, their test set performance was lower than expected, indicating generalization challenges. Although image augmentation and anti-overfitting techniques including L2 regularization, early stopping, and dropout layers were employed, overfitting remains a concern. Future research could explore generative AI techniques, particularly generative adversarial networks (GANs), to create diverse synthetic images. To mitigate overfitting and enhance model robustness, additional strategies could be investigated, such as implementing weight decay and expanding the training dataset.

Data quality is another crucial factor that may have affected the performance of the models. The HAM10000 dataset suffered from image noise caused by varying lighting conditions, different device types used for capture, inconsistent image resolution, and variable clarity of lesion boundaries. This might have introduced inconsistencies into the models' learning process, potentially hindering their ability to identify and learn relevant features. Continuous efforts in noise filtering techniques could improve image quality and readability, potentially enhancing the models' classification accuracy.

The high computational resources required for training and evaluating the four developed models should also be considered, especially as the complexity of the fine-tuning architecture increased. Despite these challenges, the proposed models showed promising enhancements in skin cancer detection and classification.

3.3. Comparison with Other Models

Diverse transfer learning approaches applied to the HAM10000 dataset were investigated. Two recent and closely related studies were identified for comparison with our proposed method. One study was published in 2023 and the other in 2024, after our experiments were completed.

These two studies were selected for comparison because they share several key characteristics with our research. Both studies divided the HAM10000 dataset into malignant and benign categories, maintained a similar distribution of samples, and employed transfer learning models. One of the studies also incorporated deep attention mechanisms. Table 7 (below) compares the performance of our proposed models on the HAM10000 dataset with the results from the two studies. Both studies reported accuracy and provided additional metrics, including recall, precision, and F1-score.

The study in [30] reported the highest accuracy of 95.95%, using modified versions of EfficientNet V2-M and EfficientNet-B4 for classifying malignant and benign skin lesions. While this accuracy was slightly higher than that of our best-performing Xception-SL model (94.11%), our proposed models outperformed theirs on all other key metrics, including recall, precision, and F1-Score. Specifically, our Xception-SL model achieved recall, precision, and F1-Score values of 95.47%, 93.10%, and 94.27%, respectively, while our Xception-SF model achieved 95.28%, 91.81%, and 93.51%. Additionally, our Xception models, incorporating attention mechanisms, recorded higher AUC scores than the EfficientNet models, with an improvement of more than 0.03.

The other study in [11] demonstrated that integrating a modified DenseNet-169 network with a coordinate attention mechanism (CoAM) and a customized CNN improves precise localization and modeling of long-range dependencies in dermoscopic images. This approach achieved the highest precision of 95.3%, an accuracy of 93.2%, a recall of 91.4%, and an F1 score of 93.3%.

Our proposed approach of integrating three different AMs (SF, HD, and SL) into the Xception architecture significantly enhanced network performance. This integration selectively focused on the most relevant areas of skin lesion images, improving accuracy and capturing long-range dependencies.

Our results were competitive in terms of both accuracy and F1-score. The Xception-SL and Xception-SF models achieved higher accuracy than the approach in [11], with Xception-SL being the most accurate at 94.11%, followed by Xception-SF at 93.29%. Both models also outperformed [11] in F1-score, with an improvement of more than 0.48 points. Although the precision of our four developed models was slightly lower, recall is often a more critical measure than precision in medical applications, as minimizing false negatives is essential to ensure that as few actual cases as possible are missed. All our proposed models (Xception-SL, Xception-SF, Xception-HD, and Xception) outperformed the approach of these studies in terms of recall, with scores of 95.47%, 95.28%, 93.98%, and 91.68% respectively.

In summary, our proposed models demonstrate promising advancements compared to recent studies in classifying and detecting skin cancer. Notably, both the self-attention and soft-attention models outperformed the previous studies in the recall metric, a critical measure for medical investigations.

These results shows that integrating different AMs into the Xception architecture improves the classification of malignant and benign skin lesion, potentially enhancing medical diagnostics and patient care.

Table 7. Comparison with state-of-the-art models.

Ref/Year	Dataset Relabeling Method	Approach	Precision	Recall	Accuracy	F1-Score	AUC
[30] 2023	Benign= 8,388 Malignant= 1,627	EfficientNetV2-M and EfficientNet-B4	95.95%	94%	83%	88%	0.980
[11] 2024	Benign= 8,061 Malignant= 1,954	Modified DenseNet-169 with CoAM+ Customized CNN	93.2%	91.4%	95.3%	93.3%	-
Our Proposed Models	Normal= 8,061 Cancer= 1,954	Xception (Base)	91.05%	91.68%	90.78%	91.23%	0.972
		Xception-SL	94.11%	95.47%	93.10%	94.27%	0.987
		Xception-SF	93.29%	95.28%	91.81%	93.51%	0.983
		Xception-HD	92.97%	93.98%	92.32%	93.14%	0.983

5. Conclusions

In this study, a novel model based on Xception architecture was proposed, incorporating three attentional mechanisms (self, hard, and soft) to classify skin cancer as benign or malignant. The impact of these AMs on model performance was thoroughly investigated. The results demonstrate that integrating AMs into the Xception architecture effectively enhances its performance. The accuracy of Xception alone was 91.05%. With AMs, the accuracy increased to 94.11% with self-attention, 93.29% with soft attention, and 92.97% with hard attention. Notably, both the self-attention and soft-attention models outperformed previous studies on the recall metric, which is crucial for medical investigations. To our knowledge, this is the first study to investigate the impact of attention mechanisms in Xception-based deep transfer learning for binary skin cancer classification. The findings suggest that attention mechanisms can enhance pre-trained models, with potential applications in aiding dermatologists in early skin cancer diagnosis, potentially improving treatment outcomes and survival rates.

A limitation of our study was the limited size and diversity of the HAM10000 dataset, which led to overfitting. Although image augmentation and anti-overfitting techniques were employed, overfitting remains a concern. Additionally, image noise may have impeded the ability of the models to learn relevant features effectively, reducing their overall accuracy and performance. Future work will focus on experimenting with larger combined datasets and using GANs to synthesize realistic skin lesion images, implementing noise filtering techniques and exploring various attention mechanisms to improve model performance. Transfer learning approaches will be evaluated using EfficientNet and ResNet, along with ensemble methods.

Author Contributions: Conceptualization, D.A.; methodology, D.A.; software, A.A.; investigation, A.A.; writing—original draft preparation, A.A.; writing—review and editing, D.A.; supervision, D.A.; project administration, D.A. 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.

Data Availability Statement: The HAM10000 dataset is available at: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T> (accessed on 18 November 2024).

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

References

1. Brown, J.S.; Amend, S.R.; Austin, R.H.; Gatenby, R.A.; Hammarlund, E.U.; Pienta, K.J. Updating the Definition of Cancer. *Mol Cancer Res* **2023**, *21*, 1142–1147, doi:10.1158/1541-7786.MCR-23-0411.
2. Rovenstine, R. Skin Cancer Statistics 2023. *The Checkup* 2022.
3. Razmjoooy, N.; Ashourian, M.; Karimifard, M.; Estrela, V.V.; Loschi, H.J.; Nascimento, D. do; França, R.P.; Vishnevski, M. Computer-Aided Diagnosis of Skin Cancer: A Review. *Current Medical Imaging* **16**, 781–793.
4. Al-Dawsari, N.A.; Amra, N. Pattern of Skin Cancer among Saudi Patients Attending a Tertiary Care Center in Dhahran, Eastern Province of Saudi Arabia. A 20-Year Retrospective Study. *International Journal of Dermatology* **2016**, *55*, 1396–1401, doi:10.1111/ijd.13320.
5. Khan, N.H.; Mir, M.; Qian, L.; Baloch, M.; Ali Khan, M.F.; Rehman, A.-; Ngowi, E.E.; Wu, D.-D.; Ji, X.-Y. Skin Cancer Biology and Barriers to Treatment: Recent Applications of Polymeric Micro/Nanostructures. *Journal of Advanced Research* **2022**, *36*, 223–247, doi:10.1016/j.jare.2021.06.014.
6. Zambrano-Román, M.; Padilla-Gutiérrez, J.R.; Valle, Y.; Muñoz-Valle, J.F.; Valdés-Alvarado, E. Non-Melanoma Skin Cancer: A Genetic Update and Future Perspectives. *Cancers* **2022**, *14*, 2371, doi:10.3390/cancers14102371.
7. PhD, J.N. Non-Melanoma Skin Cancer Deaths Exceed Melanoma Deaths Globally Available online: <https://www.cancertherapyadvisor.com/home/cancer-topics/skin-cancer/non-melanoma-skin-cancer-deaths-exceed-melanoma-deaths-globally/> (accessed on 30 November 2023).
8. Why Is Early Cancer Diagnosis Important? Available online: <https://www.cancerresearchuk.org/https%3A/www.cancerresearchuk.org/about-cancer/spot-cancer-early/why-is-early-diagnosis-important> (accessed on 10 July 2023).
9. Kato, J.; Horimoto, K.; Sato, S.; Minowa, T.; Uhara, H. Dermoscopy of Melanoma and Non-Melanoma Skin Cancers. *Front. Med.* **2019**, *6*, doi:10.3389/fmed.2019.00180.
10. Li, Z.; Koban, K.C.; Schenck, T.L.; Giunta, R.E.; Li, Q.; Sun, Y. Artificial Intelligence in Dermatology Image Analysis: Current Developments and Future Trends. *Journal of Clinical Medicine* **2022**, *11*, 6826, doi:10.3390/jcm11226826.
11. Ramamurthy, K.; Thayumanaswamy, I.; Radhakrishnan, M.; Won, D.; Lingaswamy, S. Integration of Localized, Contextual, and Hierarchical Features in Deep Learning for Improved Skin Lesion Classification. *Diagnostics* **2024**, *14*, 1338, doi:10.3390/diagnostics14131338.
12. Datta, S.K.; Shaikh, M.A.; Srihari, S.N.; Gao, M. Soft-Attention Improves Skin Cancer Classification Performance 2021.
13. Jones, O.T.; Matin, R.N.; van der Schaar, M.; Prathivadi Bhayankaram, K.; Ranmuthu, C.K.I.; Islam, M.S.; Behiyat, D.; Boscott, R.; Calanzani, N.; Emery, J.; et al. Artificial Intelligence and Machine Learning Algorithms for Early Detection of Skin Cancer in Community and Primary Care Settings: A Systematic Review. *Lancet Digit Health* **2022**, *4*, e466–e476, doi:10.1016/S2589-7500(22)00023-1.
14. Ravi, V. Attention Cost-Sensitive Deep Learning-Based Approach for Skin Cancer Detection and Classification. *Cancers* **2022**, *14*, 5872, doi:10.3390/cancers14235872.
15. Arshed, M.A.; Mumtaz, S.; Ibrahim, M.; Ahmed, S.; Tahir, M.; Shafi, M. Multi-Class Skin Cancer Classification Using Vision Transformer Networks and Convolutional Neural Network-Based Pre-Trained Models. *Information* **2023**, *14*, 415, doi:10.3390/info14070415.
16. Mukadam, S.B.; Patil, H.Y. Skin Cancer Classification Framework Using Enhanced Super Resolution Generative Adversarial Network and Custom Convolutional Neural Network. *Applied Sciences* **2023**, *13*, 1210, doi:10.3390/app13021210.
17. Mridha, K.; Uddin, Md.M.; Shin, J.; Khadka, S.; Mridha, M.F. An Interpretable Skin Cancer Classification Using Optimized Convolutional Neural Network for a Smart Healthcare System. *IEEE Access* **2023**, *11*, 41003–41018, doi:10.1109/ACCESS.2023.3269694.
18. Shapna Akter, M.; Shahriar, H.; Sneha, S.; Cuzzocrea, A. Multi-Class Skin Cancer Classification Architecture Based on Deep Convolutional Neural Network. *arXiv e-prints* 2023.
19. Kekal, H.P.; Saputri, D.U.E. Optimization of Melanoma Skin Cancer Detection with the Convolutional Neural Network. *Journal Medical Informatics Technology* **2023**, 23–28, doi:10.37034/medinftech.v1i1.5.
20. Nour, A.; Boufama, B. Convolutional Neural Network Strategy for Skin Cancer Lesions Classifications and Detections. In Proceedings of the Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics; Association for Computing Machinery: New York, NY, USA, November 24 2020; pp. 1–9.

21. Nawaz, M.; Mehmood, Z.; Nazir, T.; Naqvi, R.A.; Rehman, A.; Iqbal, M.; Saba, T. Skin Cancer Detection from Dermoscopic Images Using Deep Learning and Fuzzy K-Means Clustering. *Microscopy Research and Technique* **2022**, *85*, 339–351, doi:10.1002/jemt.23908.
22. Alabdulkreem, E.; Elmannai, H.; Saad, A.; Kamil, I.; Elaraby, A. Deep Learning-Based Classification of Melanoma and Non-Melanoma Skin Cancer. *Traitement du Signal* **2024**, *41*, 213–223, doi:10.18280/ts.410117.
23. G, R.; Ayothi, S. An Efficient Skin Cancer Detection and Classification Using Improved Adaboost Aphid–Ant Mutualism Model. *International Journal of Imaging Systems and Technology* **2023**, *33*, n/a–n/a, doi:10.1002/ima.22932.
24. Rashid, J.; Ishfaq, M.; Ali, G.; Saeed, M.R.; Hussain, M.; Alkhalifah, T.; Alturise, F.; Samand, N. Skin Cancer Disease Detection Using Transfer Learning Technique. *Applied Sciences* **2022**, *12*, 5714, doi:10.3390/app12115714.
25. Naeem, A.; Anees, T.; Fiza, M.; Naqvi, R.A.; Lee, S.-W. SCDNet: A Deep Learning-Based Framework for the Multiclassification of Skin Cancer Using Dermoscopy Images. *Sensors* **2022**, *22*, 5652, doi:10.3390/s22155652.
26. Alabduljabbar, R.; Alshamlan, H. Intelligent Multiclass Skin Cancer Detection Using Convolution Neural Networks. *Computers, Materials & Continua* **2021**, *69*, 831–847, doi:10.32604/cmc.2021.018402.
27. Imran, A.; Nasir, A.; Bilal, M.; Sun, G.; Alzahrani, A.; Almuhaimeed, A. Skin Cancer Detection Using Combined Decision of Deep Learners. *IEEE Access* **2022**, *10*, 118198–118212, doi:10.1109/ACCESS.2022.3220329.
28. Ashraf, R.; Afzal, S.; Rehman, A.U.; Gul, S.; Baber, J.; Bakhtyar, M.; Mehmood, I.; Song, O.-Y.; Maqsood, M. Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection. *IEEE Access* **2020**, *8*, 147858–147871, doi:10.1109/ACCESS.2020.3014701.
29. Jain, S.; Singhania, U.; Tripathy, B.; Nasr, E.A.; Aboudaif, M.K.; Kamrani, A.K. Deep Learning-Based Transfer Learning for Classification of Skin Cancer. *Sensors* **2021**, *21*, 8142, doi:10.3390/s21238142.
30. Venugopal, V.; Raj, N.I.; Nath, M.K.; Stephen, N. A Deep Neural Network Using Modified EfficientNet for Skin Cancer Detection in Dermoscopic Images. *Decision Analytics Journal* **2023**, *8*, 100278, doi:10.1016/j.dajour.2023.100278.
31. Dīmīlīler, K.; Sekeroglu, B. Skin Lesion Classification Using CNN-Based Transfer Learning Model. *Gazi University Journal of Science* **2023**, *36*, 660–673, doi:10.35378/gujs.1063289.
32. A Comparative Study of Neural Network Architectures for Lesion Segmentation and Melanoma Detection Available online: <https://ieeexplore.ieee.org/document/9230969> (accessed on 17 September 2024).
33. Bansal, P.; Garg, R.; Soni, P. Detection of Melanoma in Dermoscopic Images by Integrating Features Extracted Using Handcrafted and Deep Learning Models. *Computers & Industrial Engineering* **2022**, *168*, 108060, doi:10.1016/j.cie.2022.108060.
34. Parmonangan, I.H.; Marsella, M.; Pardede, D.F.R.; Rijanto, K.P.; Stephanie, S.; Kesuma, K.A.C.; Cahyaningtyas, V.T.; Anggreainy, M.S. Training CNN-Based Model on Low Resource Hardware and Small Dataset for Early Prediction of Melanoma from Skin Lesion Images. *Engineering, MAThematics and Computer Science Journal (EMACS)* **2023**, *5*, 41–46, doi:10.21512/emacsjournal.v5i2.9904.
35. Alhudhaif, A.; Almaslukh, B.; Aseeri, A.O.; Guler, O.; Polat, K. A Novel Nonlinear Automated Multi-Class Skin Lesion Detection System Using Soft-Attention Based Convolutional Neural Networks. *Chaos, Solitons & Fractals* **2023**, *170*, 113409, doi:10.1016/j.chaos.2023.113409.
36. Alshehri, A.; AlSaeed, D. Breast Cancer Detection in Thermography Using Convolutional Neural Networks (CNNs) with Deep Attention Mechanisms. *Applied Sciences* **2022**, *12*, 12922, doi:10.3390/app122412922.
37. Liu, J.; Zhang, K.; Wu, S.; Shi, H.; Zhao, Y.; Sun, Y.; Zhuang, H.; Fu, E. An Investigation of a Multidimensional CNN Combined with an Attention Mechanism Model to Resolve Small-Sample Problems in Hyperspectral Image Classification. *Remote Sensing* **2022**, *14*, 785, doi:10.3390/rs14030785.
38. Anand, V.; Gupta, S.; Altameem, A.; Nayak, S.R.; Poonia, R.C.; Saudagar, A.K.J. An Enhanced Transfer Learning Based Classification for Diagnosis of Skin Cancer. *Diagnostics* **2022**, *12*, 1628, doi:10.3390/diagnostics12071628.
39. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. *Sci Data* **2018**, *5*, 180161, doi:10.1038/sdata.2018.161.
40. Codella, N.C.F.; Gutman, D.; Celebi, M.E.; Helba, B.; Marchetti, M.A.; Dusza, S.W.; Kalloo, A.; Liopyris, K.; Mishra, N.; Kittler, H.; et al. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI), Hosted by the International Skin Imaging Collaboration (ISIC) 2018.
41. ISIC | International Skin Imaging Collaboration Available online: <https://www.isic-archive.com> (accessed on 21 October 2023).
42. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need 2023.

43. Tf.Keras.Layers.Attention | TensorFlow v2.16.1 Available online: https://www.tensorflow.org/api_docs/python/tf/keras/layers/Attention (accessed on 13 October 2024).
44. “Soft & Hard Attention” Available online: <https://jhui.github.io/2017/03/15/Soft-and-hard-attention/> (accessed on 13 October 2024).
45. Papadopoulos, A.; Korus, P.; Memon, N. Hard-Attention for Scalable Image Classification 2021.
46. Hicks, S.A.; Strümke, I.; Thambawita, V.; Hammou, M.; Riegler, M.A.; Halvorsen, P.; Parasa, S. On Evaluation Metrics for Medical Applications of Artificial Intelligence. *Sci Rep* **2022**, *12*, 5979, doi:10.1038/s41598-022-09954-8.

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.