Article

Enhancing Skin Cancer Detection and Classification: Exploring the Impact of Attention Mechanisms in Transfer Learning Models

Dana Halabi *

Posted Date: 13 December 2023

doi: 10.20944/preprints202312.0943.v1

Keywords: Skin Cancer; transfer learning; attention; computer vision

Article

Enhancing Skin Cancer Detection and Classification: Exploring the Impact of Attention Mechanisms in Transfer Learning Models

Dana Halabi

SAE Institute, Luminus Technical University College (LTUC), Amman, Jordan; d.alhalabi@saejordan.com

Abstract: Skin cancer is a global health concern with increasing prevalence, necessitating effective early detection and classification systems. Transfer learning has emerged as a powerful tool in this domain, enhancing diagnostic accuracy. Attention mechanisms, which selectively focus on pertinent image features, play a pivotal role in transfer learning. While previous studies have highlighted their effectiveness, their applicability across different models and datasets remains understudied. This empirical study explores the impact of various attention mechanisms on the performance of transfer learning models for skin cancer detection and classification. Using five transfer learning models (DenseNet121, InceptionV3, MobileNet, VGG16, and Xception) and six attention mechanisms (Channel Attention, Global Context Attention, Guided Attention, Nonlocal Attention, Positional Attention, and Spatial Attention), we conducted 105 experiments across three datasets. Traditional metrics (accuracy, precision, recall, and f1-score) were employed for empirical validation. The results reveal a nuanced relationship between attention mechanisms, transfer learning models, and datasets. Overall, attention mechanisms exhibit the potential to enhance skin cancer classification. Spatial and channel attention mechanisms consistently outperform others, offering simplicity and effectiveness. Model-specific selection of attention mechanisms is crucial, with a trade-off between model complexity and performance evident. This study provides insights into developing efficient skin cancer classification models utilizing attention mechanisms.

Keywords: skin cancer; transfer learning; attention; computer vision

1. Introduction

Increasing rates of skin cancer worldwide are a growing global health concern. The World Health Organization (WHO) estimates that approximately three million cases of non-melanoma skin cancer and more than 132,000 cases of skin cancer are diagnosed annually, which increases the annual mortality rate significantly [1]. The high prevalence of skin cancer underscores the critical need to find effective diagnosis systems that help in the early detection and accurate classification of the disease promptly to ensure early intervention and increase the chance of successful treatment and recovery from the skin cancer, which will lead to a decrease in the death rate resulting from the skin cancer [2].

In recent years, the use of transfer learning in diagnostic systems has increased due to its effectiveness in improving the accuracy of skin cancer detection and classification, especially since available labeled data are limited [3]. Transfer learning allows knowledge to be transferred from one domain to another by leveraging pre-trained models leading to improved performance in the target domain [4]. Several studies have demonstrated the effectiveness of transfer learning in various aspects of skin cancer detection [5]. These studies have utilized various pre-trained models such as MobileNetV3, Visual Geometry Group network (VGG), Inception V3, and ResNet for detecting and classifying skin cancer [6], [7].

Transfer learning has shown promise in skin cancer detection and classification. Attention mechanisms play a crucial role in focusing on relevant features during the image classification process in transfer learning [8], [9]. Several studies have investigated the effectiveness of attention

 doi:10.20944/preprints202312.0943.v1

mechanisms in skin cancer classification. For example, Song et al. found that combining squeeze-and-excitation attention with CNN architectures can accurately classify skin moles between benign and malignant without severe bias [10]. Datta et al. conducted a study to assess the effectiveness of the soft attention mechanism within deep-learning neural networks. Their findings revealed that it improved the performance of classification models specifically for skin lesions [11]. While these studies emphasize the significance of attention mechanisms and their contribution to enhancing the performance of transfer learning models in skin cancer classification, they primarily concentrate on the application of cutting-edge strategies and the presentation of results, consistently demonstrating the effectiveness of employing attention mechanisms. However, these studies typically do not address scenarios where these mechanisms may prove ineffective or computationally demanding. The objective of this study is to conduct experimental assessments to thoroughly evaluate the impact of attention mechanisms on the performance of transfer learning models, particularly in the domain of skin cancer detection and classification. By comprehensively evaluating different mechanisms of interest, this study seeks to identify the most effective approaches to improve the accuracy and reliability of skin cancer diagnostic systems.

To achieve the study's objectives, the following research questions will be addressed:

1. How do different attention mechanisms affect the performance of transfer learning models in the field of skin cancer detection and classification?
2. What is the trade-off between computational complexity and performance when incorporating attention mechanisms into transfer learning models for skin cancer detection and classification?

The organization of this study is as follows: Section 2 presents the related work and Section 3 briefly explains the involved transfer learning models and the involved attention mechanisms. Section 4 describes the research methodology used in this study and Section 5 provides details about experimental design and the obtained results are stated, analyzed, and discussed. Section 7 concludes the final outcome of the study.

2. Related Work

2.1. Transfer Learning in Skin Cancer Detection and Classification

Transfer learning has been widely used in skin cancer detection methods to improve accuracy and efficiency. Several studies have employed pre-trained deep learning models, such as AlexNet, VGG, ResNet, Inception, DenseNet, MobileNet, and Xception, for skin cancer detection. These models extract relevant image representations and learn features from skin lesion images. The extracted features are then used for the classification and detection of different types of skin cancer. Wang et al. examined the performance of the VGG model on the ISIC 2019 challenge dataset, which achieved a high accuracy of 0.9067 and an AU ROC over 0.93 [12]. Sathish et al. developed an IoT-based smartphone app by using transfer learning deep learning models, which included VGG-16 and AlexNet, for the automatic identification of skin cancer. The metrics showed that AlexNet performed better for cancer prediction [13].

Hemalatha et al. combined image processing tools with Inception V3 to enhance the structure and increase accuracy to 84% [14]. Barman et al. conducted a comparative analysis of the performance of the GoogLeNet transfer learning model with other transfer learning models such as Xception, InceptionResNetV2, and DenseNet. The evaluation was carried out on the ISIC dataset, resulting in a training accuracy of 91.16% and a testing accuracy of 89.93%. These results suggest that the GoogLeNet transfer learning model is more reliable and resilience compared to existing transfer learning models in the realm of skin cancer detection [15].

Rashid et al. proposed a deep transfer learning model using MobileNetV2 for melanoma classification and achieved better accuracy compared to other deep learning techniques [16]. Khandizod et al. also used MobileNet for a skin cancer classifier algorithm and obtained high accuracy in diagnosing skin cancer from skin lesion images [17]. Agrahari et al. employed a pre-trained MobileNet model for building a multiclass skin cancer detection system with high

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

performance comparable to that of a dermatology expert [18]. Bansal et al. analyzed the performance of the fine-tuned MobileNet model on the HAM10000 dataset and achieved a classification accuracy of 91% for seven classes [19]. Hum et al. used transfer learning on MobileNetV2 for skin lesion detection in a mobile application and achieved an evaluation accuracy of 93.9% [20].

Transfer learning using DenseNet has been explored for skin cancer detection. Khan et al. utilized a low-resolution, highly imbalanced, grayscale HAM10000 skin cancer dataset and applied transfer learning with DenseNet169, achieving a performance of 78.56% for 64 × 64 pixel images [21]. Panthakkan et al. developed a unique deep-learning model that integrates Xception and ResNet50, achieving a prediction accuracy of 97.8% [22]. Mehmood et al. introduced SBXception, a modified version of the Xception network that is shallower and broader. This architecture demonstrated significant performance improvements, achieving a reduction of 54.27% in training parameters and a decrease in training time [23]. Shaaban et al. proposed a diagnosis system for classifying between cancer and no cancer that used Xception and achieved an accuracy of 96.66% [24]. Ashim et al. compared different transfer learning models, including DenseNet and Xception, for predicting skin cancer using a Kaggle skin cancer dataset. The results showed that DenseNet achieved an accuracy of 81.94%, and Xception achieved an accuracy of 78.41% [3].

The proposed models have shown superior performance in terms of accuracy and efficiency compared to traditional machine-learning approaches. Overall, transfer learning-based methods have proven to be effective in skin cancer detection and classification by leveraging pre-trained models and extracting relevant features from skin lesion images. Table 1 summarizes the related work for utilizing Transfer Learning in Skin Cancer Detection and Classification.

Table 1. Summarization of the related work for utilizing Transfer Learning in Skin Cancer Detection and Classification.

ID	Research	Transfer Learning Model	Dataset	Achievements
1	Wang et al. [12]	VGG model	ISIC 2019	Achieved a high accuracy of 0.9067 and an AU ROC over 0.93
2	Sathish et al. [13]	VGG-16 and AlexNet	Skin cancer dataset	The metrics showed that AlexNet performed better for cancer prediction
3	Hemalatha et al. [14]	Inception V3	NA	Increase accuracy to 84%
4	Barman et al. [15]	GoogLeNet, Xception, InceptionResNetV2, and DenseNet	ISIC dataset	Increase the training accuracy to 91.16% and the testing accuracy to 89.93%.
5	Rashid et al. [16]	MobileNetV2	Melanoma dataset	Achieved better accuracy compared to other deep learning techniques
6	Khandizod et al. [17]	MobileNet	Skin lesion images	Obtained high accuracy in diagnosing skin cancer
7	Agrahari et al. [18]	MobileNet	Skin cancer dataset	high performance comparable to that of a dermatology expert
8	Bansal et al. [19]	MobileNet	HAM10000	Increase accuracy to 91%
9	Hum et al. [20]	MobileNetV2	Skin lesion dataset	Achieved an evaluation accuracy of 93.9%
10	Khan et al. [21]	DenseNet169	HAM10000	Achieving a performance of 78.56%
11	Panthakkan et al. [22]	Xception and ResNet50	NA	Achieving a prediction accuracy of 97.8%
12	Mehmood et al. [23]	SBXception	NA	Achieving a reduction of 54.27% in training parameters
13	Shaaban et al. [24]	Xception	Skin cancer dataset	Achieved an accuracy of 96.66%

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

14	Ashim et al. [3]	DenseNet and Xception	Skin cancer dataset	DenseNet achieved an accuracy of 81.94%, and Xception achieved an accuracy of 78.41%

2.2. Attention Mechanisms in Skin Cancer Detection and Classification

Attention mechanisms have been widely used in the field of skin cancer detection. These mechanisms aim to enhance the representativeness of extracted features and improve classification performance.

Several papers have proposed different attention-based models for skin cancer diagnosis. Song et al. introduced a SE-CNN model that utilized squeeze-and-excitation attention to accurately classify skin moles between benign and malignant. The SE-CNN model obtained similar performance using fewer parameters compared to other state-of-the-art models, including ResNet, DenseNet, and EfficientNet, in classifying the property of skin moles between benign and malignant [10]. La Salvia et al. introduced a hyperspectral image classification architecture that utilized Vision Transformers. This architecture demonstrated superior performance compared to the state-of-the-art methods in terms of false negative rates and processing times. Their results were showcased using a hyperspectral dataset comprising 76 skin cancer images [25].

The researchers, He et al., devised a co-attention fusion network known as CAFNet to tackle the drawbacks associated with independent feature extraction and rough fusion features in multimodal image-based approaches. The CAFNet model achieved a mean accuracy of 76.8% on the dataset comprising a seven-point checklist, outperforming the performance of state-of-the-art methods [26]. Datta et al. conducted a study investigating the efficacy of the Soft-Attention mechanism within deep neural networks, and they exhibited improved performance in the classification of skin lesions. The combination of the Soft-Attention mechanism with different neural architectures yielded a precision rate of 93.7% on the HAM10000 dataset and a sensitivity score of 91.6% on the ISIC-2017 dataset [11].

In a study by Li et al., different lightweight versions of the YOLOv4 object detection algorithm were evaluated for skin cancer detection. The YOLOv4-tiny model with the CBAM attention mechanism achieved a good balance between model size and detection accuracy, maintaining 67.3% of the mAP of the full YOLOv4 model while reducing the weight file to 9.2% of the original [27]. Aggarwal et al. introduced an attention-guided D-CNN model that enhances the precision of a conventional D-CNN architecture by around 12%. This research endeavor makes a noteworthy contribution to the realm of biomedical image processing by providing a mechanism to enhance the efficacy of D-CNNs and enable timely identification of skin cancer [28]. Ravi employed attention-cost-sensitive deep learning models that fuse features and utilize ensemble meta-classifiers to detect and classify skin cancer. The proposed methodology demonstrates a detection and classification accuracy of 99% for skin diseases, outperforming the performance of existing methods [29].

The aforementioned studies emphasize the significance of attention mechanisms in improving the performance of models used for detecting and classifying skin cancer. By incorporating these mechanisms into deep learning models, the models can concentrate on crucial regions of the image, consequently enhancing the accuracy of detection and classification. Therefore, attention mechanisms play a vital role in augmenting the effectiveness of deep learning models for the detection and classification of skin cancer. Table 2 summarizes the related work for utilizing Attention Mechanisms in Skin Cancer Detection and Classification.

Table 2. Summarization of the related work for utilizing Attention Mechanisms in Skin Cancer Detection and Classification.

ID	Research	Attention Mechanism	Dataset	Achievements
1	Song et al. [10]	squeeze-and-excitation attention	Skin moles dataset	Obtained similar performance using fewer parameters
2	La Salvia et al. [25]	Vision Transformers	76 skin cancer images	outperformed the state-of-the-art in terms of false negative

 doi:10.20944/preprints202312.0943.v1

3	He et al., [26]	co-attention fusion network	Dataset comprising a seven-point checklist	achieved a mean accuracy of 76.8%
4	Datta et al. [11]	Soft-Attention	HAM10000 and ISIC-2017	precision rate of 93.7% on the HAM10000 dataset and a sensitivity score of 91.6% on the ISIC-2017 dataset
5	Li et al. [27].	CBAM attention	Skin cancer dataset	Achieved a good balance between model size and detection accuracy
6	Aggarwal et al. [28]	guided attention	Skin cancer dataset	Enhances the precision by around 12%
7	Ravi [29]	cost-sensitive attention	Skin cancer dataset	classification accuracy of 99% for skin diseases

3. Background

This section will briefly review the set of computer vision transfer learning models and the various attention mechanisms involved in this study.

3.1. Transfer Learning in Computer Vision

In computer vision, transfer learning involves using pre-trained models that are then used to perform related or similar tasks. Harnessing the knowledge gained from these pre-trained models significantly saves time and computational resources. [30–32]. In computer vision, transfer learning involves using a pre-trained deep convolutional neural network to extract basic features from lower layers and capture complex and abstract features in higher layers [33]. The lower layers are typically frozen to retain features extracted from a pre-trained neural network, and the upper layers are typically retrained only to meet the specific task [34]. This approach has demonstrated its effectiveness in improving accuracy and reducing training duration compared to traditional techniques. It also finds applications across diverse domains within computer vision [35], [36].

Certainly, here's a concise overview of the transfer learning process using five state-of-the-art deep Convolutional Neural Network (CNN) models employed in this study:

1. **DenseNet**
2. DenseNet is a robust deep learning framework renowned for its unique dense connectivity structure, which facilitates direct connections between each layer. This design promotes efficient feature reuse, rendering feature extraction remarkably effective. DenseNet's architecture enables the development of larger networks that remain resource-efficient, making it a favored option for a wide range of computer vision tasks [37]. The efficacy of the DenseNet framework has been demonstrated in addressing various challenges, such as poor convergence, overfitting, and gradient disappearance, that may arise in comprehensive architectures [38].
3. **Inception**

   Inception, also known as GoogLeNet, is a deep convolutional neural network architecture that brought forth the notion of inception modules. These modules employ various filter sizes within the same layer to capture features at various scales. Inception accomplished high accuracy in image classification tasks while minimizing computational complexity [39].
4. **MobileNet**

   MobileNet, a deep convolution neural network, was introduced by Google in the year 2017. Its distinguishing characteristic lies in its ability to effectively utilize computational resources and model size, rendering it suitable for environments with limited resources such as mobile devices and embedded systems. The efficiency of MobileNet is achieved through the implementation of depth-wise separable convolutions, which serve to decrease the number of parameters while simultaneously enabling the extraction of meaningful features by the model [40].
5. **VGG**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

VGG (Visual Geometry Group) signifies a classical convolutional neural network (CNN) architecture from the University of Oxford celebrated for its simplicity and efficacy in image analysis, boasting a uniform structure comprising 16 or 19 layers of 3x3 convolutional layers and max-pooling layers. VGG's impact extends beyond ImageNet, excelling in various tasks and datasets. Its architecture includes 64-channel 3x3 convolutional layers with 1x1 convolution filters and ReLU units, concluding with three fully connected layers with 4096 channels and 1000 classes [41].

6. **Xception**

Xception, a deep learning framework, was introduced by Google in the year 2016. The incorporation of depth-wise separable convolutions in Xception enhances the performance and efficiency of convolutional neural networks (CNNs). Through the separation of spatial and depth-wise convolutions, Xception reduces the number of parameters, while simultaneously preserving the ability to accurately capture complex features. Consequently, this produces a more compact and computationally efficient model [42].

3.2. Attention Mechanisms

Attention mechanisms are vital in computer vision tasks as they boost the performance of deep neural networks by enabling them to concentrate on pertinent information in images. In this section, we offer a brief overview of the six attention mechanisms involved in this study.

1. Spatial Attention

Spatial Attention is a deep learning technique that improves a model's focus on specific areas (regions) of an image while reducing attention to others, allowing the model to prioritize important parts of the input data.

The mathematical equation for the Spatial Attention applied in this study can be described as follows:

Let $\boldsymbol{I}$ be the input tensor with dimensions $\boldsymbol{H}$ **x** $\boldsymbol{W}$ **x** $\boldsymbol{C}$, where $\boldsymbol{H}$, $\boldsymbol{W}$, and $\boldsymbol{C}$ represent the height, width, and number of channels in the input tensor, respectively.

1. **Average Pooling:** First, an average pooling operation is applied along the channel axis **(axis=-1)** to calculate the average value of each spatial location across all channels. This operation is represented as **avg_pool**, and the resulting tensor has dimensions $\boldsymbol{H}$ **x** $\boldsymbol{W}$ **x 1**. This operation is expressed mathematically as:

$$\text{avg\_pool} = \frac{1}{C}\sum_{c=1}^{C} I_{i,j,c}$$

where $\boldsymbol{I}_{i,j,c}$ represents the value of the input tensor at spatial position **(i, j)** and channel **(c)**.

2. **Sigmoid Convolution:** Next, a convolution operation is applied with a kernel size of (1, 1) and a single output channel. This convolution layer has a sigmoid activation function. The purpose of this convolution is to learn spatial attention weights for each spatial location. The output of this convolution, denoted as **conv_output**, has the dimensions **H x W x 1** and contains values between 0 and 1 due to the sigmoid activation:

$$\text{conv\_output}_{i,j} = \sigma\left(\sum_{c=1}^{C} W_{i,j,c} \cdot \text{avg\_pool}_{i,j}\right)$$

where **δ** represents the sigmoid activation function, $\mathbf{W}_{i,j,c}$ represents the learned convolution kernel weights, and $\mathbf{avg\_pool}_{i,j}$ represents the average-pooled value at spatial position **(i, j)**.

3. **Spatial Attention Applied to Input**: Finally, the output of the sigmoid convolution is element-wise multiplied with the original input tensor $\boldsymbol{I}_{i,j,c}$ to produce the final output of the Spatial Attention layer. This operation assigns higher weights to spatial locations that are more important based on the learned attention values:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

$$\text{output}_{i,j,c} = \text{conv\_output}_{i,j} \cdot I_{i,j,c}$$

The Spatial Attention layer computes attention weights for each spatial location in the input tensor and applies these weights to the input data, allowing the network to focus on specific regions of the input during processing. The mathematical equation of Spatial Attention is illustrated in Figure 1.

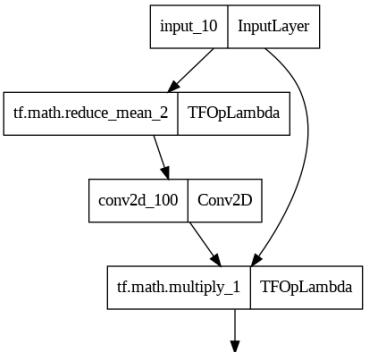

Figure 1. Spatial Attention.

2. Channel Attention

The Channel Attention mechanism aims to enhance the importance of certain channels within each feature map. This helps the model focus on relevant information across different channels and improve its ability to make accurate predictions.

The mathematical equation for the Channel Attention applied in this study can be described as follows:

Let ***I*** be the input tensor with dimensions ***H* x *W* x *C***, where ***H***, ***W***, and ***C*** represent the height, width, and number of channels in the input tensor, respectively.

1. **Average Pooling:** First, an average pooling operation is applied to the input tensor ***I*** along the spatial dimensions (height and width). This operation calculates the average value for each channel, resulting in a tensor **avg_pool** with dimensions **1 x 1 x C**. The average pooling operation is represented mathematically as:

$$\text{avg\_pool} = \frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} I_{i,j,c}$$

Where $I_{i,j,c}$ represents the value of the input tensor at spatial position **(i, j)** and channel **(c)**.

2. **Fully Connected Layers (FC1 and FC2):** Two fully connected (dense) layers are used to process the **avg_pool** tensor:
   - **fc1:** This layer reduces the dimensionality of the **avg_pool** tensor by applying a linear transformation followed by a ReLU activation. The output of **fc1**, denoted as **fc1_output**, has dimensions 1 x 1 x (C/8), where (C/8) represents a reduction of one-eighth of the original channel dimension.

$$\text{fc1\_output} = \text{ReLU}(W_1 \cdot \text{avg\_pool} + b_1)$$

   - **fc2:** This layer takes the output of **fc1** and applies another linear transformation followed by a sigmoid activation function. The output of **fc2**, denoted as **channel_attention**, has the same dimensions as **avg_pool**, which is **1 x 1 x C**.

$$\text{channel\_attention} = \sigma(W_2 \cdot \text{fc1\_output} + b_2)$$

where **δ** represents the sigmoid activation function, $\mathbf{W}_1$ and $\mathbf{W}_2$ are learnable weight matrices, and $\mathbf{b}_1$ and $\mathbf{b}_2$ are learnable bias vectors.

3. **Channel Attention Application:** Finally, the channel attention tensor **channel_attention** is element-wise multiplied with the original input tensor ***I*** to produce the final output of the Channel Attention layer:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

$$\text{output}_{i,j,c} = \text{channel\_attention}_{1,1,c} \cdot I_{i,j,c}$$

The Channel Attention layer computes attention weights for each channel in the input tensor and scales each channel's features accordingly. It allows the network to emphasize or de-emphasize specific channels based on their importance for a given task. The mathematical equation of Channel Attention is illustrated in Figure 2.

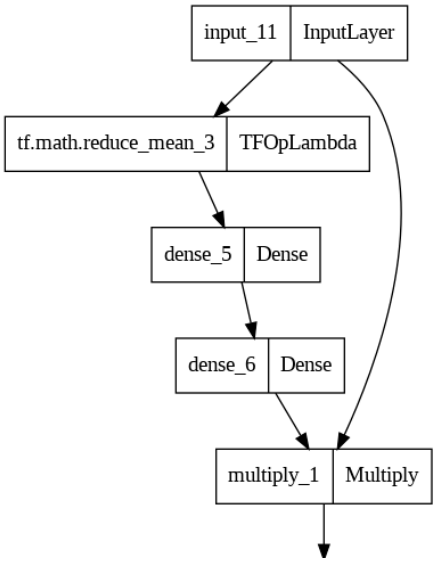

Figure 2. Channel Attention.

3. Positional Attention

The Positional Attention mechanism is designed to enhance specific spatial positions within a feature map based on their significance for the task. It does this by learning attention scores for different spatial locations and using these scores to modulate the feature map.

The mathematical equation for the Positional Attention applied in this study can be described as follows:

Let ***I*** be the input tensor with dimensions ***H* x *W* x *C***, where ***H***, ***W***, and ***C*** represent the height, width, and number of channels in the input tensor, respectively.

1. **Query Computation:** The first step is to compute a query tensor, denoted as **query**, which will determine the attention weights for each position in the input tensor. This is done using a 1x1 convolution layer with a sigmoid activation. The output of this convolution, **query**, has the same dimensions as the input tensor **H x W x 1** and contains values between 0 and 1 due to the sigmoid activation:

$$\text{query}_{i,j} = \sigma\left(\sum_{c=1}^{C} W_{i,j,c} \cdot I_{i,j,c}\right)$$

where **δ** represents the sigmoid activation function, $\mathbf{W}_{i,j,c}$ represents the learned convolution kernel weights, and $\boldsymbol{I}_{i,j,c}$ represents the value of the input tensor at spatial position **(i, j)** and channel **(c)**.

2. **Positional Attention Application:** The final output of the Positional Attention layer is obtained by element-wise multiplying the original input tensor ***I*** with the query tensor **query**:

$$\text{output}_{i,j,c} = \text{query}_{i,j} \cdot I_{i,j,c}$$

The Positional Attention layer applies attention to each position in the input tensor independently. The attention mechanism is learned through the convolutional layer, allowing the network to emphasize or de-emphasize specific positions based on their importance for the task at hand. The mathematical equation of Positional Attention is illustrated in Figure 3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

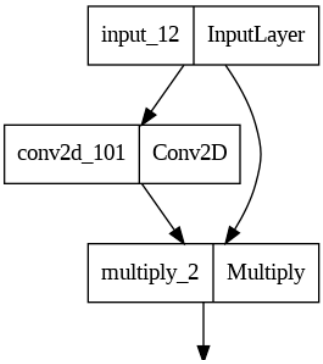

Figure 3. Positional Attention.

4. Nonlocal Attention

The Nonlocal Attention mechanism is designed to capture long-range dependencies and correlations within a sequence or feature map by considering relationships between all possible pairs of positions. It does so by computing attention scores that indicate the relevance of each position to others.

The mathematical equation for the Non-Local Attention applied in this study can be described as follows:

Let ***I*** be the input tensor with dimensions ***H* x *W* x *C***, where ***H***, ***W***, and ***C*** represent the height, width, and number of channels in the input tensor, respectively.

1. **Query, Key, and Value Computation:**
   - **Query (Q)**: Calculate query matrices by applying a 1x1 convolution operation to the input tensor ***I***. The query tensor, denoted as **queries**, has the same dimensions as the input tensor **H x W x C**.
   - **Key (K):** Calculate key matrices by applying another 1x1 convolution operation to ***I***. The key tensor, denoted as **keys**, also has dimensions **H x W x C**.
   - **Value (V):** Calculate value matrices by applying a third 1x1 convolution operation to ***I***. The value tensor, denoted as **values**, has dimensions **H x W x C**.
2. **Attention Score Calculation:** Compute attention scores **attention_scores** by taking the dot product of the queries and keys, followed by a Softmax operation along the last dimension (channel dimension):

$$\text{attention\_scores} = \text{softmax}(\text{queries} \cdot \text{keys}^{T})$$

where **(.)** represents the dot product, and the Softmax operation is applied along the channel dimension to obtain normalized attention scores for each position in the input.

3. **Applying Attention to Values:** Apply the computed attention scores to the values to obtain the attended values **attention_values**:

$$\text{attention\_values} = \text{attention\_scores} \cdot \text{values}$$

The multiplication of attention scores and values calculates the weighted sum of values based on the attention weights.

4. **Combining Attention Values with Inputs:** Multiply the attended values **attention_values** element-wise with the original input tensor ***I***:

$$\text{output} = I \cdot \text{attention\_values}$$

The output tensor represents the result of applying non-local attention to the input tensor, where each position in the output is a weighted sum of values from all positions in the input.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

The Nonlocal Attention layer allows the network to capture long-range dependencies and relationships between positions in the input tensor by computing and applying attention scores globally across spatial dimensions. The mathematical equation of Nonlocal Attention is illustrated in Figure 4.

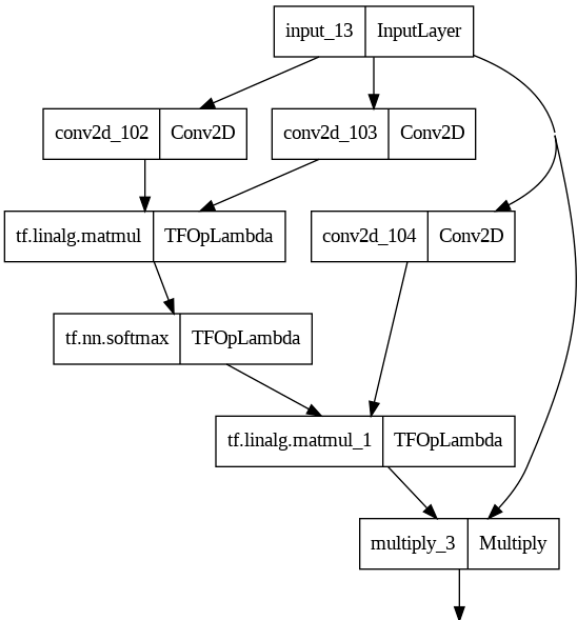

Figure 4. Non-Local Attention.

5. Global Context Attention

The Global Context Attention mechanism is designed to capture the overall context or semantic information of a feature map by considering a global representation of the entire map. It generates a context vector based on the collective information present in the feature map and uses this vector to enhance the features.

The mathematical equation for the Global Context Attention applied in this study can be described as follows:

Let I be the input tensor with dimensions H x W x C, where H, W, and C represent the height, width, and number of channels in the input tensor, respectively.

1. **Global Average Pooling:** The first step is to compute a global context vector by applying global average pooling to the input tensor I. Global average pooling computes the average value across all spatial positions, resulting in a tensor of size **1x1xC**:

$$\text{global\_context} = \text{GlobalAveragePooling2D}(I)$$

2. **Context Vector Generation:** To generate a context vector that can be applied to the input tensor, a fully connected (dense) layer is used. This layer takes the global context tensor as input and produces a context vector **context_vector** with dimensions **1x1xC** using a sigmoid activation function:

$$\text{context\_vector} = \text{sigmoid}\left(\text{Dense}(\text{global\_context})\right)$$

3. **Expanding the Context Vector**: To match the dimensions of the input tensor, the context vector **context_vector** is expanded by adding two dimensions with size **1** at the beginning. This results in a context vector with dimensions **1x1x1xC**:

$$\text{context\_vector} = \text{expand}(\text{expand}(\text{context\_vector}, 1), 1)$$

4. Applying Global Context to Input: Finally, the input tensor I is multiplied element-wise by the context vector **context_vector** to produce the final output of the Global Context Attention layer:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

$$\text{output} = I \cdot \text{context\_vector}$$

The Global Context Attention layer computes a global context vector from the input tensor and then scales the input tensor at each position using this context. The output contains information that has been enhanced by considering the global context of the entire input tensor. The mathematical equation of Global Context Attention is illustrated in Figure 5.

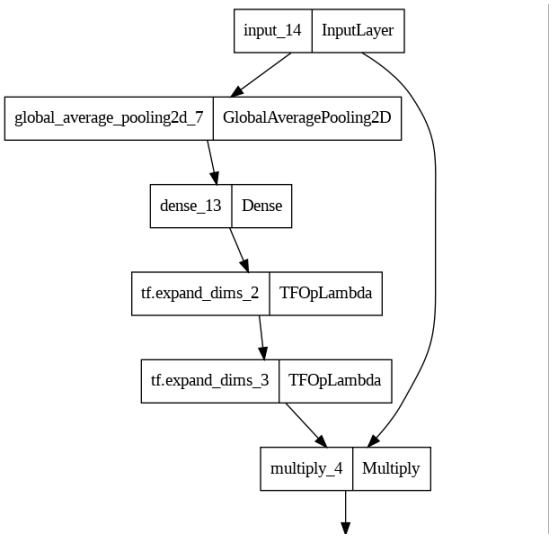

Figure 5. Global Context Attention.

6. Guided Attention

The Guided Attention mechanism is designed to selectively enhance specific regions within a feature map based on a learned attention map. It guides the model's focus to relevant areas and helps improve the interpretation of the model's decisions.

The mathematical equation for the Guided Context Attention applied in this study can be described as follows:

Let I be the input tensor with dimensions H x W x C, where H, W, and C represent the height, width, and number of channels in the input tensor, respectively.

1. **Attention Map Calculation:** The first step is to compute an attention map, denoted as **attention_map**, using a convolutional layer. The convolutional layer, with a kernel size of (3, 3) and a sigmoid activation function, calculates the attention map with the same spatial dimensions as the input tensor **(H x W)** but with a single channel:

$$\text{attention\_map} = \sigma\big(\text{Conv2D}(I)\big)$$

where **δ** represents the sigmoid activation function.

2. **Guided Attention Application:** The final output of the Guided Attention layer is obtained by element-wise multiplying the original input tensor I with the attention map **attention_map**:

$$\text{output} = I \cdot \text{attention\_map}$$

This operation applies the attention map to the input tensor, enhancing regions or features in I that are assigned higher values in the attention map while suppressing regions with lower values.

The Guided Attention layer allows the network to focus on specific spatial regions in the input tensor based on the learned attention map. It can be particularly useful in tasks where different regions of the input may have varying levels of importance or relevance to the task at hand. The mathematical equation of Guided Attention is illustrated in Figure 6.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

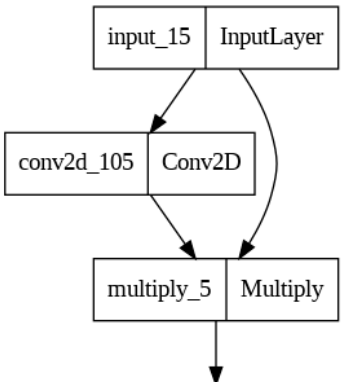

Figure 6. Guided Attention.

4. Methodology

4.1. Data collection and preprocessing

4.1.1. Data collection

The performance of deep learning techniques depends on the availability of suitable and valid datasets. This study utilizes the following datasets:

1. HAM10000 dataset

The HAM10000 dataset[1] was published in 2018 and consists 10,015 dermatoscopic images with a size of 600 X 450 (width X height) pixel resolution [43]. There are seven types of skin lesions: actinic keratosis/intraepithelial carcinoma (AKIEC), basal cell carcinoma (BCC), benign keratosis (BKL), dermatofibroma (DF), melanoma (MEL), melanocytic nevi (NV), and vascular lesions (VASC). Among them, BKL, DF, NV, and VASC are benign tumors, whereas AKIEC, BCC, and MEL are malignant tumors. Samples from the HAM10000 dataset are illustrated in Figure 7.

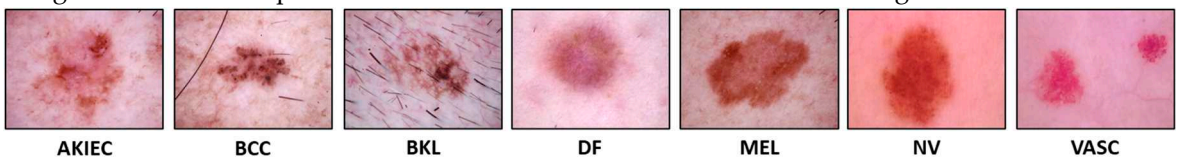

Figure 7. Samples of Skin lesion images from the HAM10000 dataset.

The images were gathered over 20 years from two distinct sources: the Department of Dermatology at the Medical University of Vienna, Austria, and the Skin Cancer Practice of Cliff Rosendahl in Queensland, Australia. The labels assigned to these images underwent validation through various methods, including histopathology, reflectance confocal microscopy, follow-up examinations, or expert consensus. Despite comprising a total of 10,015 skin lesion images, it is worth noting that this dataset displays a considerable class imbalance, as illustrated in Figure 8. To provide an example, the most extensive category (NV) encompasses 6,705 images, while the smallest category (DF) comprises just 115 images.

[1] https://www.kaggle.com/datasets/surajghuwalewala/ham1000-segmentation-and-classification

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

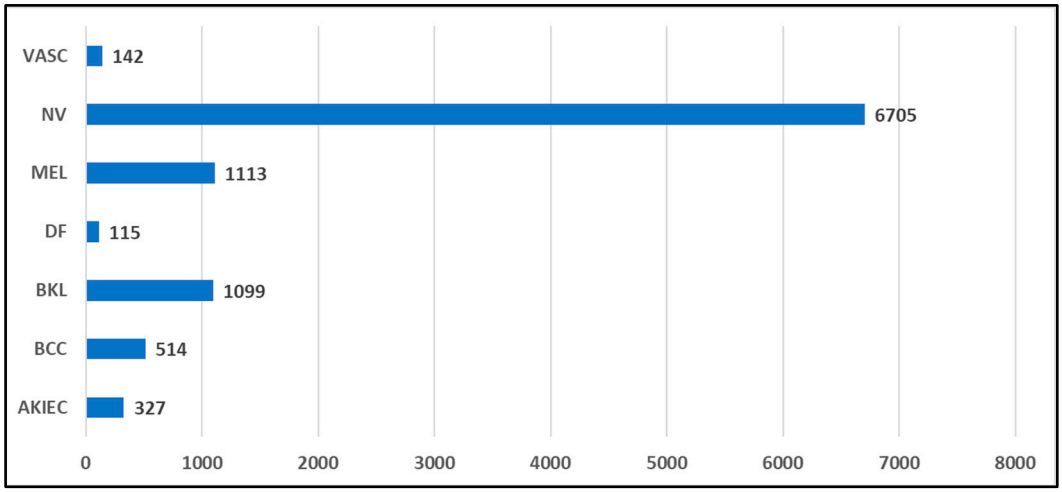

Figure 8. The distribution of 7 classes in the HAM10000 dataset.

2. ISIC2017 dataset

The International Skin Imaging Collaboration skin lesion segmentation dataset (ISIC2017)[2] is a publicly available dataset from 2017, as detailed in Codella et al. [44]. It comprises 2,750 images with varying resolutions. The ISIC dataset was released by the International Skin Imaging Collaboration (ISIC) as a comprehensive collection of dermoscopy images. This dataset originated from a challenge involving lesion segmentation, dermoscopic feature identification, and disease classification. Within this dataset, images are categorized into three types of skin lesions: melanoma, nevus, and seborrheic keratosis. Figure 9 provides visual examples of samples from the ISIC2017 dataset.

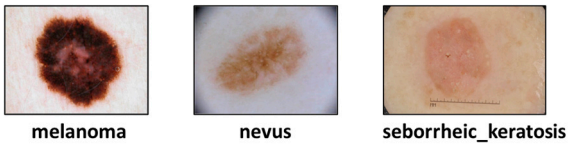

Figure 9. Samples from the ISIC2017 dataset are illustrated.

Despite comprising 2,750 skin lesion images, this dataset exhibits significant class imbalance, as evident in Figure 10. Notably, the largest class (nevus) encompasses 1,843 images, while the remaining two classes (melanoma and seborrheic keratosis) consist of only 521 and 386 images, respectively.

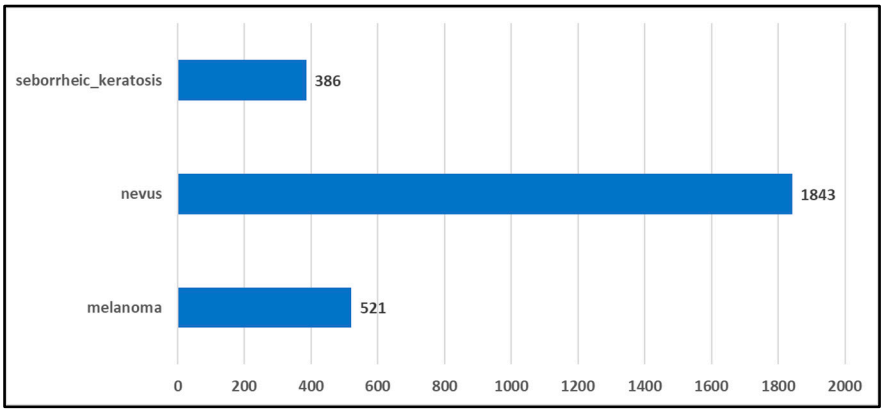

Figure 10. The distribution of 3 classes in the ISIC2017dataset.

[2] https://www.kaggle.com/datasets/wanderdust/skin-lesion-analysis-toward-melanoma-detection

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

3. **Melanoma dataset**

The Melanoma dataset is a modified version of the HAM10000 dataset. In contrast to the original dataset, which categorized skin cancer images into seven classes, the Melanoma dataset simplifies this into two categories: "Melanoma" and "Not Melanoma." The "Melanoma" group comprises 1,113 images, while the "Not Melanoma" group contains 8,902 images. To balance the dataset, data augmentation techniques were applied to the "Melanoma" group, resulting in a total of 8,903 images. This transformed dataset now represents a balanced version derived from the original HAM10000 dataset. Figure 11 provides visual samples from the Melanoma dataset, while Figure 12 illustrates the distribution of binary classes within the Melanoma dataset.

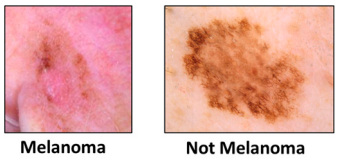

Figure 11. Samples from the Melanoma dataset.

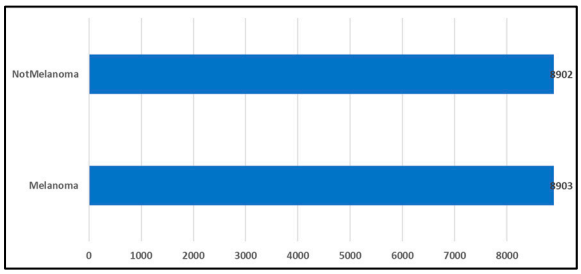

Figure 12. The distribution of the classes in the Melanoma dataset.

4.1.2. Data Preprocessing

1. Prepare Datasets

I. HAM10000 dataset

The publicly available HAM10000 dataset initially lacked divisions into training, validation, and testing datasets. To prepare this dataset for training and evaluating the proposed transfer learning models, we set aside ten percent of the data for testing, resulting in a testing dataset of 1,002 samples. Subsequently, the remaining dataset was split into a training dataset comprising 7,210 samples and a validation dataset comprising 1,803 samples, maintaining an 80:20 ratio. At each stage of this process, stratified sampling was employed to ensure that the distribution of classes was consistent across subsets and to prevent the possibility of overlooking minority classes between subsets. Figure 13 provides an overview of the entire process of preparing the HAM10000 dataset.

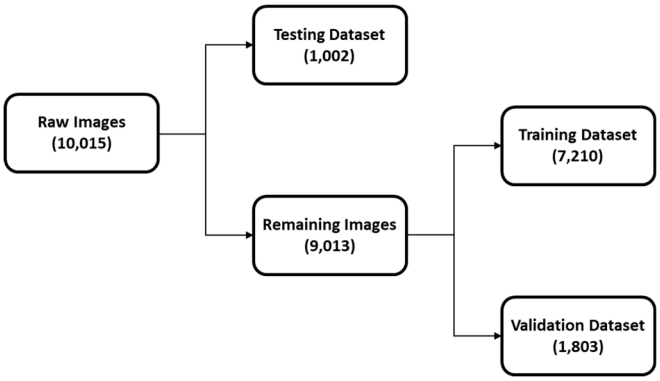

Figure 13. The whole preparation process of the HAM10000 dataset.

Table 3 shows the details of training, validation, and testing datasets for the HAM10000 dataset.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

Table 3. the details of training, validation, and testing datasets for the HAM10000 dataset.

	AKIEC	BCC	BKL	DF	MEL	NV	VASC	Total
Training	237	373	788	82	801	4825	104	7210
Validation	62	86	206	24	191	1212	22	1803
Testing	28	55	105	9	121	668	16	1002
Total	327	514	1099	115	1113	6705	142	10015

II. ISIC2017 dataset

ISIC2017 offers a dataset configuration comprising 2,000 images for training, 150 images for validation, and 600 images for testing. The specifics of these training, validation, and testing datasets are detailed in Table 4.

Table 4. the details of training, validation, and testing datasets for the ISIC2017 dataset.

Class	Melanoma	Nevus	Seborrheic Keratosis	Total
Training	374	1372	254	2000
Validation	30	78	42	150
Testing	117	393	90	600
Total	521	1843	386	2750

III. Melanoma dataset

The Melanoma dataset[3], derived from the Skin Cancer MNIST: HAM10000 dataset, is an augmented dataset comprising 17,805 images. For this dataset, 10,682 images were allocated to the training set, 3,562 to the validation set, and 3,561 to the testing set. Table 5 provides a breakdown of the training, validation, and testing datasets for the Melanoma dataset.

Table 5. the details of training, validation, and testing datasets for the Melanoma dataset.

Class	Melanoma	Not Melanoma	Total
Training	5341	5341	10682
Validation	1781	1781	3562
Testing	1781	1780	3561
Total	8903	8902	17805

2. Preprocessing data

Regarding data preprocessing, all images were resized into the dimensions of 224x224 pixels to train and test VGG, MobileNet, and DenseNet models. On the other hand, all images were resized into the dimensions of 299x299 pixels to train and test Inception, and Xception models. Then the pixel values of images were rescaled to a range between 0 and 1.0/255.0, which helps standardize the input.

3. Data augmentation

A set of data augmentation techniques were applied to images in the training dataset. Data augmentation techniques help in reducing overfitting and improving the generalization and robustness of machine learning models. These techniques include randomly shifting the image horizontally by up to 10% of its width (width_shift_range=0.1) and vertically by up to 10% of its height (height_shift_range=0.1), and it may horizontally flip the image (horizontal_flip=True).

4.2. Transfer learning framework

[3] Alexander Scarlat (2019) 'Dermoscopic pigmented skin lesions from HAM10k' Available at: https://www.kaggle.com/datasets/drscarlat/melanoma

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

4.2.1. Transfer Learning Models

In this study, five deep convolution network models which are VGG16, InceptionV3, MobileNet, DenseNet121, and Xception, are implemented with a set of attention mechanism, to classify skin cancer images. These models are all state-of-the-art feature extractors which are trained on ImageNet dataset.

4.2.2. Details of Transfer Learning Framework

In the context of skin cancer detection and classification, we employed transfer learning by making subtle adjustments to the architecture and fine-tuning the weights of pre-trained VGG, Inception, MobileNet, DenseNet, and Xception models initially trained on the ImageNet dataset. These adaptations encompassed replacing conventional average pooling with global average pooling and substituting the top layer of the pre-trained CNN models with a configuration involving a fully connected layer, a dropout layer (with a rate of 0.5), and another fully connected layer. The specifications of the last fully connected layer varied based on the dataset, with seven units and "Softmax" activation for HAM10000, three units and "Softmax" activation for ISIC2017, and seven units with "Sigmoid" activation for the Melanoma dataset. This transfer learning approach, outlined in Figure 14, entailed unfreezing all layers of the pre-trained CNN models to adapt to the unique dataset features (HAM10000, ISIC2017, Melanoma). Furthermore, global average pooling and dropout layers were employed to combat overfitting by retaining essential features while reducing excessive detail capture in the learned features.

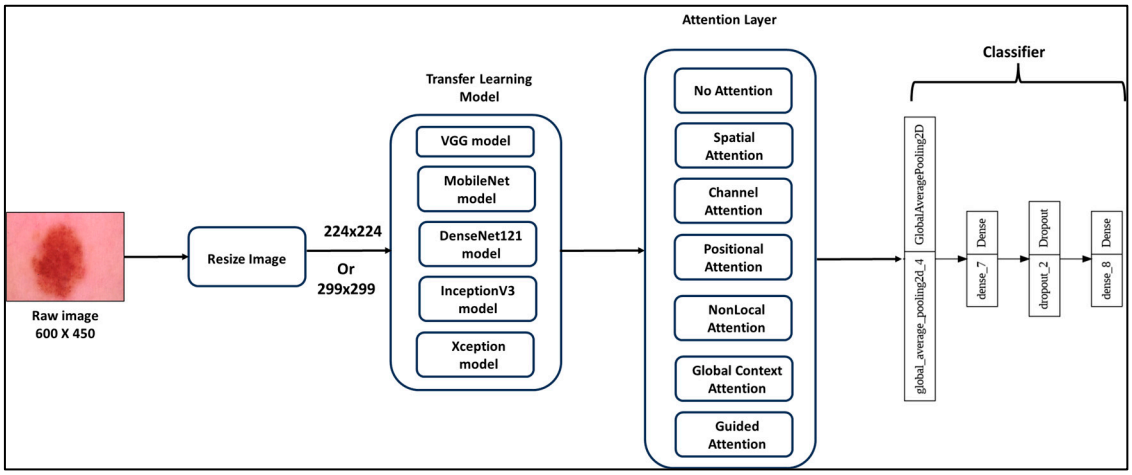

Figure 14. Transfer Learning Attention Framework.

4.3. Attention Mechanisms Integration

In this study, six attention mechanisms which are Spatial Attention, Channel Attention, Positional Attention, Nonlocal Attention, Global Context Attention, and Guided Attention, are incorporated into transfer learning models. Figure 14 illustrates the whole Transfer Learning Attention Framework.

4.4. Evaluation metrics

The performance of the model is analyzed using traditional four metrics accuracy, precision, recall, and F1 score. These metrics are developed from True positive (TP), True negative (TN), False positive (FP), and False negative (FN) predictions.

- Accuracy:

Accuracy measures the proportion of correctly classified instances out of all the instances in the dataset.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

- Precision:

Precision measures the proportion of true positive predictions (correctly predicted positive instances) out of all positive predictions made.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- Recall (Sensitivity or True Positive Rate):

Recall measures the proportion of true positive predictions (correctly predicted positive instances) out of all actual positive instances in the dataset.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- F1-Score:

The F1 score is the harmonic mean of precision and recall, providing a balance between the two metrics. It is especially useful when dealing with imbalanced datasets.

$$\text{F1 Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. Experimental Setup:

5.1. Experiment Design:

All the models were trained for the different datasets using the same optimization options. The proposed models were compiled using the Adam optimizer. However, the choice of loss function varied depending on the dataset. The "categorical cross-entropy" loss function was used for the HAM10000 and ISIC2017 datasets, while the "binary cross-entropy" loss function was employed for the Melanoma dataset. Table 6 provides an overview of the optimization hyperparameters used in our experiments.

Table 6. Set of hyper-parameters used during the training step of all the phases. Please note that the batch size is equal to 4 due to computational constraints (e.g., availability of GPU RAM).

Hyper-parameter	Value
Number of epochs	**20**
Loss function	**Cross-entropy (Binary/Categorical)**
Optimizer	**Adam**
Learning rate	**0.001**
Batch Size	**4**

To reduce overfitting and improve training efficiency, an Early stopping technique was applied during training, and the best model weights were saved according to a decrease in the validation loss.

5.2. Model Complexity

The complexity information, including the total number of parameters, the number of trainable parameters, the number of non-trainable parameters, and the attention layer parameters, for all the transfer learning models used in the experiments (DenseNet121, InceptionV3, MobileNet, VGG16, and Xception), is presented in Tables 7–11 respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

Table 7. The complicity information (i.e., the number of total parameters, the number of trainable parameters, the number of none-trainable parameters, and the attention layer parameters) for the DenseNet121 models.

DenseNet121	Total params	Trainable params	Non-trainable params	Layer params	Ranked based on Layer params
Base (No Attention)	7563843	526339	7037504	0	0
Channel Attention	7827139	789635	7037504	263296	2
Global Context Attention	8613443	1575939	7037504	1049600	5
Guided Attention	7573060	535556	7037504	9217	6
Nonlocal Attention	10712643	3675139	7037504	3148800	4
Positional Attention	7564868	527364	7037504	1025	3
Spatial Attention	7563845	526341	7037504	2	1

Table 8. The complicity information (i.e., the number of total parameters, the number of trainable parameters, the number of none-trainable parameters, and the attention layer parameters) for the InceptionV3 models.

InceptionV3	Total params	Trainable params	Non-trainable params	Layer params	Ranked based on Layer params
Base (No Attention)	22853411	1050627	21802784	0	0
Channel Attention	23904291	2101507	21802784	1050880	2
Global Context Attention	27049763	5246979	21802784	4196352	5
Guided Attention	22871844	1069060	21802784	18433	6
Nonlocal Attention	35442467	13639683	21802784	12589056	4
Positional Attention	22855460	1052676	21802784	2049	3
Spatial Attention	22853413	1050629	21802784	2	1

Table 9. The complicity information (i.e., the number of total parameters, the number of trainable parameters, the number of none-trainable parameters, and the attention layer parameters) for the MobileNet models.

MobileNet	Total params	Trainable params	Non-trainable params	Layer params	Ranked based on Layer params
Base (No Attention)	3755203	526339	3228864	0	0
Channel Attention	4018499	789635	3228864	263296	2
Global Context Attention	4804803	1575939	3228864	1049600	5
Guided Attention	3764420	535556	3228864	9217	6
Nonlocal Attention	6904003	3675139	3228864	3148800	4
Positional Attention	3756228	527364	3228864	1025	3
Spatial Attention	3755205	526341	3228864	2	1

Table 10. The complicity information (i.e., the number of total parameters, the number of trainable parameters, the number of none-trainable parameters, and the attention layer parameters) for the VGG16 models.

VGG16	Total params	Trainable params	Non-trainable params	Layer params	Ranked based on Layer params
Base (No Attention)	14978883	264195	14714688	0	0
Channel Attention	15044995	330307	14714688	66112	2
Global Context Attention	15241539	526851	14714688	262656	5
Guided Attention	14983492	268804	14714688	4609	6
Nonlocal Attention	15766851	1052163	14714688	787968	4
Positional Attention	14979396	264708	14714688	513	3

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

Spatial Attention	**14978885**	**264197**	**14714688**	**2**	**1**

Table 11. The complicity information (i.e., the number of total parameters, the number of trainable parameters, the number of none-trainable parameters, and the attention layer parameters) for the Xception models.

Xception	Total params	Trainable params	Non-trainable params	Layer params	Ranked based on Layer params
Base (No Attention)	**21912107**	**1050627**	**20861480**	**0**	**0**
Channel Attention	**22962987**	**2101507**	**20861480**	**1050880**	**2**
Global Context Attention	**26108459**	**5246979**	**20861480**	**4196352**	**5**
Guided Attention	**21930540**	**1069060**	**20861480**	**18433**	**6**
Nonlocal Attention	**34501163**	**13639683**	**20861480**	**12589056**	**4**
Positional Attention	**21914156**	**1052676**	**20861480**	**2049**	**3**
Spatial Attention	**21912109**	**1050629**	**20861480**	**2**	**1**

6. Experimental Results

To assess the impact of the six attention mechanisms on the five transfer learning models, a total of 105 experiments were conducted across three datasets: HAM10000, ISIC2017, and Melanoma. Within these 105 experiments, 35 experiments were performed on each dataset. These 35 experiments comprised five experiments aimed at establishing the base model for each transfer learning model (i.e., the transfer learning model without any attention mechanisms), and an additional 30 experiments that involved combining the five transfer learning models with the six attention mechanisms.

1. Experiments on the **HAM10000** dataset

Comparisons in accuracy, precision, recall, and f1-score between experiments associated with the **HAM10000** dataset were reported in Tables 12–15 respectively. In these tables, for each transfer model, the scores of base transfer learning models are highlighted in green, the scores that are higher than the score of the base model are highlighted in blue, and the scores that are lower than the score of the base model are highlighted in red[4].

Table 12. Values of accuracy for the 35 experiments associated with the HAM10000 dataset.

accuracy	DenseNet121	InceptionV3	MobileNet	VGG16	Xception
Base (No Attention)	**91.87**	**91.87**	**93.50**	**91.25**	**93.16**
Channel Attention	**92.76**	**92.64**	**93.70**	**91.70**	**92.64**
Global Context Attention	**92.81**	**92.07**	**93.19**	**91.93**	**93.44**
Guided Attention	**92.02**	**91.25**	**91.87**	**91.79**	**91.87**
Nonlocal Attention	**92.13**	**90.73**	**92.22**	**92.22**	**91.93**
Positional Attention	**93.33**	**91.50**	**92.96**	**91.90**	**92.99**
Spatial Attention	**92.56**	**92.53**	**93.19**	**91.65**	**93.16**

Table 13. Values of weighted precision for the 35 experiments associated with the HAM10000 dataset.

weighted precision	DenseNet121	InceptionV3	MobileNet	VGG16	Xception
Base (No Attention)	**66.44**	**67.00**	**74.97**	**67.94**	**74.55**
Channel Attention	**74.27**	**71.35**	**76.26**	**66.68**	**70.83**
Global Context Attention	**74.51**	**67.53**	**74.40**	**68.78**	**75.47**
Guided Attention	**67.73**	**64.51**	**68.48**	**69.33**	**69.47**

[4] This highlighting policy is applied also for the experiments associated with ISIC2017 dataset and Melanoma dataset.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

Nonlocal Attention	67.31	55.55	68.87	70.11	70.15
Positional Attention	75.25	65.07	74.15	69.59	75.55
Spatial Attention	72.67	72.87	75.34	65.94	74.76

Table 14. Values of weighted recall for the 35 experiments associated with the HAM10000 dataset.

weighted recall	DenseNet121	InceptionV3	MobileNet	VGG16	Xception
Base (No Attention)	71.56	71.56	77.25	69.36	76.05
Channel Attention	74.65	74.25	77.94	70.96	74.25
Global Context Attention	74.85	72.26	76.15	71.76	77.05
Guided Attention	72.06	69.36	71.56	71.26	71.56
Nonlocal Attention	72.46	67.56	72.75	72.75	71.76
Positional Attention	76.65	70.26	75.35	71.66	75.45
Spatial Attention	73.95	73.85	76.15	70.76	76.05

Table 15. Values of weighted F1-score for the 35 experiments associated with the HAM10000 dataset.

weighted F1-score	DenseNet121	InceptionV3	MobileNet	VGG16	Xception
Base (No Attention)	65.41	67.66	74.65	62.35	74.56
Channel Attention	73.63	70.92	76.43	66.51	70.97
Global Context Attention	73.58	66.78	72.38	70.00	75.37
Guided Attention	68.16	65.33	67.02	69.03	67.38
Nonlocal Attention	69.08	57.93	68.63	69.37	69.99
Positional Attention	75.57	66.68	73.60	67.40	74.34
Spatial Attention	72.51	71.52	75.14	66.16	74.03

A. Impact of Attention Mechanisms on Base Models

The base model, without any attention mechanisms, showed that MobileNet performed the best in terms of the weighted F1-score, achieving 74.65%. When Channel Attention was applied to MobileNet, it improved the F1-score to 76.43%, but it also added a significant number of parameters (263,296), indicating a trade-off between performance and model complexity[5].

B. Performance of Different Transfer Learning Models

DenseNet121 and VGG16 benefited from the addition of all six attention mechanisms, indicating their flexibility and adaptability to attention mechanisms. MobileNet and InceptionV3, on the other hand, showed enhanced performance when combined with only two attention mechanisms: Channel Attention and Spatial Attention. Xception's performance improved only when paired with Global Context Attention.

C. Influence of Attention Mechanisms

Channel Attention and Spatial Attention generally improved the performance of multiple transfer learning models, but they negatively impacted the Xception model. Global Context Attention enhanced the performance of DenseNet121, VGG16, and Xception but reduced the performance of InceptionV3 and MobileNet. Guided Attention, Nonlocal Attention, and Positional Attention primarily benefited DenseNet121 and VGG16 but had negative effects on InceptionV3, MobileNet, and Xception.

2. Experiments on the **ISIC2017** dataset

Comparisons in accuracy, precision, recall, and f1-score between experiments associated with the **ISIC2017** dataset were reported in Tables 16–19.

[5] It is worth mentioning that all the performance comparisons in this study are compared to the performance of base models.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

Table 16. Values of accuracy for the 35 experiments associated with the ISIC2017 dataset.

accuracy	DenseNet121	InceptionV3	MobileNet	VGG16	Xception
Base (No Attention)	77.48	78.04	80.57	75.61	80.91
Channel Attention	77.15	77.37	78.92	78.81	79.80
Global Context Attention	75.39	77.04	78.81	75.72	78.81
Guided Attention	73.73	74.94	77.04	77.92	74.83
Nonlocal Attention	74.39	73.40	76.82	78.37	78.81
Positional Attention	77.81	77.04	78.92	76.93	80.57
Spatial Attention	72.63	80.13	79.91	76.49	75.06

Table 17. Values of weighted precision for the 35 experiments associated with the ISIC2017 dataset.

weighted precision	DenseNet121	InceptionV3	MobileNet	VGG16	Xception
Base (No Attention)	66.52	56.01	71.65	57.11	69.88
Channel Attention	64.64	65.50	68.90	60.96	70.77
Global Context Attention	64.38	63.64	67.61	61.08	70.99
Guided Attention	52.63	51.53	42.98	52.40	62.32
Nonlocal Attention	62.00	64.38	70.24	65.34	67.76
Positional Attention	74.90	67.23	68.27	58.06	69.09
Spatial Attention	67.60	68.84	70.35	60.94	68.39

Table 18. Values of weighted recall for the 35 experiments associated with the ISIC2017 dataset.

weighted recall	DenseNet121	InceptionV3	MobileNet	VGG16	Xception
Base (No Attention)	66.23	67.05	70.86	63.41	71.36
Channel Attention	65.73	66.06	68.38	68.21	69.70
Global Context Attention	63.08	65.56	68.21	63.58	68.21
Guided Attention	60.60	62.42	65.56	66.89	62.25
Nonlocal Attention	61.59	60.10	65.23	67.55	68.21
Positional Attention	66.72	65.56	68.38	65.40	70.86
Spatial Attention	58.94	70.20	69.87	64.74	62.58

Table 19. Values of weighted F1-score for the 35 experiments associated with the ISIC2017 dataset.

weighted F1	DenseNet121	InceptionV3	MobileNet	VGG16	Xception
Base (No Attention)	60.80	60.83	67.63	58.11	69.53
Channel Attention	61.26	65.66	68.11	60.95	68.19
Global Context Attention	62.08	64.41	67.10	59.62	65.05
Guided Attention	55.93	56.40	51.93	55.83	59.06
Nonlocal Attention	60.18	60.47	66.24	65.86	61.88
Positional Attention	60.91	62.24	67.51	56.45	66.76
Spatial Attention	61.06	68.73	67.90	59.85	62.45

A. Performance of Base Transfer Learning Models

The best-performing base transfer learning model was Xception, achieving an F1 score of 69.53%. Xception's performance was not outperformed by any other model, irrespective of the transfer learning model and attention mechanism applied. This suggests that Xception is a strong choice as a base model for this dataset.

B. Performance from the Perspective of Transfer Learning Models

VGG16's performance improved when combined with certain attention mechanisms (channel, global context, nonlocal, and spatial attention). This indicates that these mechanisms helped VGG16 focus on relevant features, enhancing its performance. DenseNet121's performance also improved with channel, global context, positional, and spatial attention. However, guided and nonlocal attention had a negative impact. Different attention mechanisms affect models differently, suggesting that model architecture plays a role in compatibility with attention mechanisms. InceptionV3

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

benefited from the channel, global context, positional, and spatial attention but suffered a performance drop with guided and nonlocal attention. MobileNet had mixed results, performing well with channel and spatial attention but declining with global context, guided, nonlocal, and positional attention. Xception consistently performed worse when combined with any of the attention mechanisms, suggesting that this model might already have attention mechanisms embedded in its architecture or that the additional attention mechanisms are not suitable for it.

C. Performance from the Perspective of Attention Mechanisms

Channel attention and spatial attention consistently enhanced the performance of several transfer learning models, including DenseNet121, MobileNet, InceptionV3, and VGG16. However, they harmed the Xception model. Global context attention generally improved the performance of DenseNet121, InceptionV3, and VGG16 but decreased the performance of MobileNet and Xception. Positional attention had a positive effect on DenseNet121 and InceptionV3 but negatively impacted MobileNet, VGG16, and Xception. Nonlocal attention only improved the performance of the VGG16 model, while it adversely affected the other models. Guided attention did not improve the performance of any transfer learning model, suggesting that this mechanism may not be suitable for skin cancer classification.

3. Experiments on Melanoma dataset

Comparisons in accuracy, precision, recall, and F1-score between experiments associated with the Melanoma dataset were reported in Tables 20–23.

Table 20. Values of accuracy for the 35 experiments associated with the Melanoma dataset.

accuracy	DenseNet121	InceptionV3	MobileNet	VGG16	Xception
Base (No Attention)	93.37	92.92	93.57	84.02	93.23
Channel Attention	93.43	93.18	93.29	85.82	93.37
Global Context Attention	93.68	92.84	93.15	85.17	93.79
Guided Attention	93.01	92.73	93.18	86.44	93.34
Nonlocal Attention	93.37	92.90	93.37	85.40	93.04
Positional Attention	93.32	92.87	93.20	85.40	93.23
Spatial Attention	93.54	93.26	93.77	84.84	93.63

Table 21. Values of weighted precision for the 35 experiments associated with Melanoma dataset.

precision	DenseNet121	InceptionV3	MobileNet	VGG16	Xception
Base (No Attention)	88.76	88.51	89.55	86.59	88.26
Channel Attention	89.12	88.37	89.53	85.48	88.76
Global Context Attention	89.09	88.07	91.42	85.09	89.83
Guided Attention	87.88	87.97	88.21	85.85	88.67
Nonlocal Attention	89.07	88.46	89.15	82.14	88.49
Positional Attention	89.18	88.07	88.07	84.16	88.26
Spatial Attention	89.58	89.41	89.42	86.26	89.72

Table 22. Values of weighted recall for the 35 experiments associated with Melanoma dataset.

weighted recall	DenseNet121	InceptionV3	MobileNet	VGG16	Xception
Base (No Attention)	99.33	98.65	98.65	80.51	99.72
Channel Attention	98.93	99.44	98.03	86.29	99.33
Global Context Attention	99.55	99.10	95.22	85.28	98.76
Guided Attention	99.78	98.99	99.66	87.25	99.38
Nonlocal Attention	98.88	98.65	98.76	90.45	98.93
Positional Attention	98.60	99.16	99.94	87.19	99.72
Spatial Attention	98.54	98.15	99.27	82.87	98.54

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

Table 23. Values of weighted F1-score for the 35 experiments associated with Melanoma dataset.

weighted F1-score	DenseNet121	InceptionV3	MobileNet	VGG16	Xception
Base (No Attention)	93.74	93.30	93.88	83.44	93.64
Channel Attention	93.77	93.58	93.59	85.88	93.74
Global Context Attention	94.03	93.26	93.29	85.19	94.09
Guided Attention	93.45	93.15	93.59	86.54	93.72
Nonlocal Attention	93.72	93.28	93.71	86.10	93.42
Positional Attention	93.65	93.29	93.63	85.65	93.64
Spatial Attention	93.85	93.57	94.09	84.53	93.92

A. Base Transfer Learning Model Performance

MobileNet emerged as the best-performing base transfer learning model with an F1-score of 93.88%, outperforming other models. It had a moderate number of trainable parameters (526,339). Spatial attention slightly improved the F1-score of MobileNet, indicating that even a small attention mechanism can have a positive impact on performance.

B. Transfer Learning Model Performance with Attention Mechanisms

VGG16 exhibited performance improvement when combined with all six attention mechanisms, suggesting that it is a versatile base model that benefits from various attention strategies. Xception performed well with five out of six attention mechanisms but showed a decrease in performance with nonlocal attention. This indicates sensitivity to certain attention mechanisms. The performance of DenseNet121 improved with channel attention, global context attention, and spatial attention, but declined with guided attention, nonlocal attention, and positional attention. This model's behavior suggests that the choice of attention mechanism matters significantly. InceptionV3 performed best with channel attention and spatial attention, while its performance decreased with global context attention. This suggests that some attention mechanisms are more suitable for specific base models. The performance of MobileNet only improved with spatial attention, indicating that it is less responsive to most attention mechanisms compared to other models.

C. Attention Mechanism Performance

Spatial attention consistently improved the performance of all transfer learning models. Channel attention was effective for DenseNet121, InceptionV3, Xception, and VGG16, but had a negative impact on MobileNet. Global context attention improved the performance of DenseNet121, VGG16, and Xception but harmed InceptionV3 and MobileNet. Guided attention and positional attention were generally less effective, with improvements seen only in VGG16 and Xception, but reductions in performance for other models. Nonlocal attention was effective only for VGG16 and had a detrimental effect on the other models.

7. Discussion

- Dataset-specific results

The following are some specific observations from the results of each dataset:

- HAM10000 dataset:

The spatial attention and channel attention mechanisms were the most effective for all models on this dataset. The other attention mechanisms were generally less effective, and sometimes even decreased the performance of the models.

- ISIC Archive dataset

The spatial attention mechanism was the most effective for all models on this dataset. The channel attention mechanism was also effective for most models, but it decreased the performance of the Xception model. The other attention mechanisms were generally less effective, and sometimes even decreased the performance of the models.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 December 2023 doi:10.20944/preprints202312.0943.v1

- Melanoma dataset

The spatial attention and channel attention mechanisms were the most effective for all models on this dataset. The global context attention mechanism was also effective for the DenseNet121, VGG16, and Xception models, but it decreased the performance of the InceptionV3 and MobileNet models. The other attention mechanisms were generally less effective, and sometimes even decreased the performance of the models.

The choice of base model should be dataset-specific. Models that perform well on one dataset may not generalize to others due to differences in data distribution.

- Consideration of Model Complexity:

Each transfer learning model participated in 18 experiments, with an additional three experiments conducted to construct a base model for each dataset. Table 24 provides a summary of the model complexity, including the total and trainable parameters, for the five transfer learning models used in the study. It also presents the percentage of experiments, out of the total 18, in which improvements in performance were observed when compared to the performance of the transfer learning base model. The results reveal that there is no direct correlation between model complexity and performance enhancement with the inclusion of attention mechanisms. For instance, MobileNet, despite having fewer parameters, did not consistently benefit from attention mechanisms, whereas models with more parameters, such as VGG16, demonstrated potential improvements.

Table 24. Information of the complexity of transfer learning models vs. the percentage of number of experiments that acquire improve in the performance.

Model	Total params	Trainable params	Rank of Total params	Rank of Trainable params	Number of "Increase Performance" in 18 experiments	Percentage of "Increase Performance" (%)	Rank of Percentage of "Increase Performance"
DenseNet121	7563843	526339	2	3	13	72	2
InceptionV3	22853411	1050627	5	5	8	44	3
MobileNet	3755203	526339	1	2	5	28	5
VGG16	14978883	264195	3	1	16	89	1
Xception	21912107	1050627	4	4	6	33	4

Choosing attention mechanisms should be a deliberate process, carefully balancing performance gains with model complexity, as overly complex models may not be practical for deployment.

- Attention Mechanism Selection:

Each attention mechanism was tested in 15 experiments. Table 25 summarizes the percentage of experiments, out of the total 15, where improvements in performance were observed when compared to the performance of the transfer learning base models. The results indicate that the effectiveness of attention mechanisms varies depending on the base model and dataset used. Notably, spatial attention appears to be a reliable choice, consistently enhancing performance without introducing significant increases in model complexity.

Table 25. for each attention mechanism the percentage of the number of experiments from the total experiments (15 experiments) that acquire improvements in the performance when compared to the performance of transfer learning base models.

Attention Mechanism	Number of "Increase Performance" in 15 experiments	Percentage of Increase Performance (%)	The rank of Percentage of Increase Performance	Layer params - VGG16	Layer params - DenseNet121 and MobileNet	Layer params - InceptionV3 and Xception
Channel Attention	12	80	2	66112	263296	1050880
Global Context Attention	9	60	3	262656	1049600	4196352
Guided Attention	4	27	5	4609	9217	18433
Nonlocal Attention	4	27	5	787968	3148800	12589056
Positional Attention	6	40	4	513	1025	2049
Spatial Attention	13	87	1	2	2	2

8. Conclusion and Future Work

This study assesses the effect of utilizing attention mechanisms on the performance of transfer learning in detecting and classifying skin cancer. The study uses five transfer learning models namely DenseNet121, InceptionV3, MobileNet, VGG16, and Xception to build skin cancer detection and classification models. Also, the study focuses on six attention mechanisms: Channel Attention, Global Context Attention, Guided Attention, Nonlocal Attention, positional attention, and spatial attention. We implemented all the above-mentioned attention mechanisms to perform our analysis. We conducted 105 experiments across three datasets, and the empirical validation for the results was done using four traditional metrics: accuracy, precision, recall, and f1-score. The conclusions of the study are as follows:

The study's findings reveal that the impact of attention mechanisms on transfer learning models in skin cancer classification is complex, and influenced by the model, attention method, and dataset. Overall, the study suggests that attention mechanisms can enhance the performance of these models in skin cancer classification.

The study's findings indicate that spatial attention and channel attention mechanisms prove to be the most efficient for skin cancer classification, irrespective of the transfer learning model or dataset employed. These mechanisms are straightforward to grasp, directing the models toward crucial image details. In contrast, the other attention mechanisms, being more intricate, might pose challenges for effective learning by the models. Moreover, they may not be as pertinent to the skin cancer classification task as spatial and channel attention mechanisms.

The results indicate that selecting an attention mechanism should align with the specific model in use, as not all mechanisms suit every model. Furthermore, it becomes apparent that there is a trade-off between the complexity of the model and its performance, particularly when contemplating the incorporation of attention mechanisms.

Hence, the analysis performed in this study can be used to develop efficient skin cancer classification models that utilize attention mechanisms. The Future work will focus on more complex attention mechanisms such as transformer-based attention and hybrid attention mechanisms. Other transfer learning models and ensemble methods may also be investigated.

Data Availability Statement: Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study

Conflicts of Interest: The author declares that she has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

1. R. M. Fernandez, "SDG3 good health and well-being: integration and connection with other SDGs," *Good Health Well-Being*, pp. 629–636, 2020.
2. S. A. Gandhi and J. Kampp, "Skin cancer epidemiology, detection, and management," *Med. Clin.*, vol. 99, no. 6, pp. 1323–1335, 2015.
3. L. K. Ashim, N. Suresh, and C. V. Prasannakumar, "A comparative analysis of various transfer learning approaches skin cancer detection," in *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, IEEE, 2021, pp. 1379–1385.
4. H. K. Kondaveeti and P. Edupuganti, "Skin cancer classification using transfer learning," in *2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI)*, IEEE, 2020, pp. 1–4.
5. M. Fraiwan and E. Faouri, "On the automatic detection and classification of skin cancer using deep transfer learning," *Sensors*, vol. 22, no. 13, p. 4963, 2022.
6. R. C. Suganthe, N. Shanthi, M. Geetha, R. Manjunath, S. M. Krishna, and P. M. Balaji, "Performance Evaluation of Transfer Learning Based Models On Skin Disease Classification," in *2023 International Conference on Computer Communication and Informatics (ICCCI)*, IEEE, 2023, pp. 1–7.
7. P. P. Naik, B. Annappa, and S. Dodia, "An Efficient Deep Transfer Learning Approach for Classification of Skin Cancer Images," in *International Conference on Computer Vision and Image Processing*, Springer, 2022, pp. 524–537.
8. M. E. Haque, M. R. Ahmed, R. S. Nila, and S. Islam, "Classification of human monkeypox disease using deep learning models and attention mechanisms," *ArXiv Prepr. ArXiv221115459*, 2022.

9. Z. Cheng, G. Huo, and H. Li, "A multi-domain collaborative transfer learning method with multi-scale repeated attention mechanism for underwater side-scan sonar image classification," *Remote Sens.*, vol. 14, no. 2, p. 355, 2022.
10. Z. Song and Y. Zhou, "Skin cancer classification based on CNN model with attention mechanism," in *Second International Conference on Medical Imaging and Additive Manufacturing (ICMIAM 2022)*, SPIE, 2022, pp. 281–287.
11. S. K. Datta, M. A. Shaikh, S. N. Srihari, and M. Gao, "Soft attention improves skin cancer classification performance," in *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data: 4th International Workshop, iMIMIC 2021, and 1st International Workshop, TDA4MedicalData 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 4*, Springer, 2021, pp. 13–23.
12. X. Wang, Y. Yang, and B. Mandal, "Automatic detection of skin cancer melanoma using transfer learning in deep network," in *AIP Conference Proceedings*, AIP Publishing, 2023.
13. K. Sathish, A. Mohanraj, R. Raman, V. Sudha, A. Kumar, and V. Vijayabhaskar, "IoT based Mobile App for Skin Cancer Detection using Transfer Learning," in *2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, IEEE, 2022, pp. 16–22.
14. D. Hemalatha, K. N. Latha, and P. M. Latha, "Skin Cancer Detection Using Deep Learning Technique," in *2023 2nd International Conference for Innovation in Technology (INOCON)*, IEEE, 2023, pp. 1–5.
15. S. Barman, M. R. Biswas, S. Marjan, N. Nahar, M. S. Hossain, and K. Andersson, "Transfer Learning Based Skin Cancer Classification Using GoogLeNet," in *International Conference on Machine Intelligence and Emerging Technologies*, Springer, 2022, pp. 238–252.
16. J. Rashid *et al.*, "Skin cancer disease detection using transfer learning technique," *Appl. Sci.*, vol. 12, no. 11, p. 5714, 2022.
17. S. Khandizod, T. Patil, A. Dode, V. Banale, and C. D. Bawankar, "Deep Learning based Skin Cancer Classifier using MobileNet".
18. P. Agrahari, A. Agrawal, and N. Subhashini, "Skin Cancer Detection Using Deep Learning," in *Futuristic Communication and Network Technologies*, A. Sivasubramanian, P. N. Shastry, and P. C. Hong, Eds., in Lecture Notes in Electrical Engineering. Singapore: Springer Nature, 2022, pp. 179–190. doi: 10.1007/978-981-16-4625-6_18.
19. N. Bansal and S. Sridhar, "Skin lesion classification using ensemble transfer learning," in *Second International Conference on Image Processing and Capsule Networks: ICIPCN 2021 2*, Springer, 2022, pp. 557–566.
20. Y. C. Hum *et al.*, "The development of skin lesion detection application in smart handheld devices using deep neural networks," *Multimed. Tools Appl.*, vol. 81, no. 29, pp. 41579–41610, 2022.
21. M. R. H. Khan, A. H. Uddin, A.-A. Nahid, and A. K. Bairagi, "Skin cancer detection from low-resolution images using transfer learning," in *Intelligent Sustainable Systems: Proceedings of ICISS 2021*, Springer, 2022, pp. 317–334.
22. A. Panthakkan, S. M. Anzar, S. Jamal, and W. Mansoor, "Concatenated Xception-ResNet50—A novel hybrid approach for accurate skin cancer prediction," *Comput. Biol. Med.*, vol. 150, p. 106170, 2022.
23. A. Mehmood, Y. Gulzar, Q. M. Ilyas, A. Jabbari, M. Ahmad, and S. Iqbal, "SBXception: A Shallower and Broader Xception Architecture for Efficient Classification of Skin Lesions," *Cancers*, vol. 15, no. 14, p. 3604, 2023.
24. S. Shaaban, H. Atya, H. Mohammed, A. Sameh, K. Raafat, and A. Magdy, "Skin Cancer Detection Based on Deep Learning Methods," in *The International Conference on Artificial Intelligence and Computer Vision*, Springer, 2023, pp. 58–67.
25. M. La Salvia *et al.*, "Attention-based Skin Cancer Classification Through Hyperspectral Imaging," in *2022 25th Euromicro Conference on Digital System Design (DSD)*, IEEE, 2022, pp. 871–876.
26. X. He, Y. Wang, S. Zhao, and X. Chen, "Co-attention fusion network for multimodal skin cancer diagnosis," *Pattern Recognit.*, vol. 133, p. 108990, 2023.
27. P. Li, T. Han, Y. Ren, P. Xu, and H. Yu, "Improved YOLOv4-tiny based on attention mechanism for skin detection," *PeerJ Comput. Sci.*, vol. 9, p. e1288, 2023.
28. A. Aggarwal, N. Das, and I. Sreedevi, "Attention-guided deep convolutional neural networks for skin cancer classification," in *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, IEEE, 2019, pp. 1–6.
29. V. Ravi, "Attention cost-sensitive deep learning-based approach for skin cancer detection and classification," *Cancers*, vol. 14, no. 23, p. 5872, 2022.
30. A. Nehvi, R. Dar, and A. Assad, "Visual Recognition of Local Kashmiri Objects with Limited Image Data using Transfer Learning," in *2021 International Conference on Emerging Techniques in Computational Intelligence (ICETCI)*, IEEE, 2021, pp. 49–52.
31. X. Liu, W. Yu, F. Liang, D. Griffith, and N. Golmie, "Toward deep transfer learning in industrial internet of things," *IEEE Internet Things J.*, vol. 8, no. 15, pp. 12163–12175, 2021.

32. E. Antonio, C. Rael, and E. Buenavides, "Changing Input Shape Dimension Using VGG16 Network Model," in *2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS)*, IEEE, 2021, pp. 36–40.
33. S. Roy and S. Saravana Kumar, "Feature Construction Through Inductive Transfer Learning in Computer Vision," in *Cybernetics, Cognition and Machine Learning Applications: Proceedings of ICCCMLA 2020*, Springer, 2021, pp. 95–107.
34. B. Maschler, D. Braun, N. Jazdi, and M. Weyrich, "Transfer learning as an enabler of the intelligent digital twin," *Procedia CIRP*, vol. 100, pp. 127–132, 2021.
35. A. Brodzicki, M. Piekarski, D. Kucharski, J. Jaworek-Korjakowska, and M. Gorgon, "Transfer learning methods as a new approach in computer vision tasks with small datasets," *Found. Comput. Decis. Sci.*, vol. 45, no. 3, pp. 179–193, 2020.
36. A. Liang, "Effectiveness of Transfer Learning, Convolutional Neural Network and Standard Machine Learning in Computer Vision Assisted Bee Health Assessment," in *2022 International Communication Engineering and Cloud Computing Conference (CECCC)*, IEEE, 2022, pp. 7–11.
37. B. Chen, T. Zhao, J. Liu, and L. Lin, "Multipath feature recalibration DenseNet for image classification," *Int. J. Mach. Learn. Cybern.*, vol. 12, pp. 651–660, 2021.
38. Soniya, S. Paul, and L. Singh, "Sparsely Connected DenseNet for Malaria Parasite Detection," in *Advances in Systems Engineering: Select Proceedings of NSC 2019*, Springer, 2021, pp. 801–807.
39. M. Long, S. Long, F. Peng, and X. Hu, "Identifying natural images and computer-generated graphics based on convolutional neural network," *Int. J. Auton. Adapt. Commun. Syst.*, vol. 14, no. 1–2, pp. 151–162, 2021.
40. A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." arXiv, Apr. 16, 2017. doi: 10.48550/arXiv.1704.04861.
41. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." arXiv, Apr. 10, 2015. Accessed: Sep. 12, 2023. [Online]. Available: http://arxiv.org/abs/1409.1556
42. S. Sharma and S. Kumar, "The Xception model: A potential feature extractor in breast cancer histology images classification," *ICT Express*, vol. 8, no. 1, pp. 101–108, 2022.
43. P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, no. 1, pp. 1–9, 2018.
44. N. C. Codella *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)," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, IEEE, 2018, pp. 168–172.