information

MDPI

Article

Attention-Based Inception-Residual CNN: Skin Cancer Diagnosis with Attention-Based Inception-Residual CNN Model

Sara Younas [1], Allah Bux Sargano [1], Lihua You [2] and Zulfiqar Habib [1,*]

[1] Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Lahore 54000, Pakistan; sarayounas395@gmail.com (S.Y.); allahbux@cuilahore.edu.pk (A.B.S.)

[2] National Centre for Computer Animation, Bournemouth University, Bournemouth BH12 5BB, UK; lyou@bournemouth.ac.uk

* Correspondence: drzhabib@cuilahore.edu.pk

Abstract: Skin cancer poses a significant global health concern, demanding early diagnosis to enhance patient outcomes and alleviate healthcare burdens. Despite advancements in automated diagnosis systems, most existing approaches primarily address binary classification, with limited focus on distinguishing among multiple skin cancer classes. Multiclass classification poses significant challenges due to intra-class variations and inter-class similarities, often leading to misclassification. These issues stem from subtle differences between skin cancer types and shared features across various classes. This paper proposes an attention-based Inception-Residual CNN (AIR-CNN) model specially designed to tackle the challenges related to multiclass skin cancer classification. Incorporating the attention mechanism model effectively focuses on the most relevant features, enhancing its ability to distinguish between visually similar classes and those with intra-class variations. The attention mechanism also facilitates effective training with limited samples. The inclusion of Inception-Residual (IR) blocks mitigates vanishing gradients, improves multi-scale feature extraction, and reduces parameters, creating a lightweight yet accurate model. The experimental evaluation of the ISIC 2019 dataset demonstrates superior performance with 91.63% accuracy and fewer parameters than state-of-the-art methods, which makes it suitable for practical applications, thus contributing to the advancement of automated skin cancer diagnosis systems.

Keywords: skin cancer diagnosis; attention-based CNN; multi-scale feature extraction; inception-residual blocks; dermoscopy image analysis

check for updates

Academic Editors: Shuohong Wang, Francesco Fontanella and Birgitta Dresp-Langley

Received: 11 December 2024
Revised: 30 January 2025
Accepted: 4 February 2025
Published: 8 February 2025

Citation: Younas, S.; Sargano, A.B.; You, L.; Habib, Z. Attention-Based Inception-Residual CNN: Skin Cancer Diagnosis with Attention-Based Inception-Residual CNN Model. *Information* **2025**, *16*, 120. https://doi.org/10.3390/info16020120

1. Introduction

Skin is considered the broadest human body organ, which includes different layers. The epidermis layer is the superficial surface of the skin, which secures the human body from the outer environment and acts as a shield against injuries and infections, intercepting moisture loss and sustaining internal body temperature. One of the most crucial tasks of the skin is protection from detrimental rays coming from the sun. Sometimes, overexposure to sunlight can create problems and act as a hurdle to the smooth functioning of the skin, which leads to early aging signs, certain infections, and skin cancer. One of the most frequent disorders related to the skin is skin cancer, which is considered the deadliest and most mortal among all other types of cancers. Skin cancer is the unusual and irregular growth of cells located at the epidermis layer of skin that is stimulated due to the presence of destructive DNA in skin cells [1–3]. This disease could also be caused by some artificial sources of light and heat like sunlamps. In the USA, around 192,310 new cases of skin cancer were reported in 2019 years [4]. New Zealand and Australia have lost 55,000 people due to skin cancer in the last few years [5]. Recent research has revealed a 55% increase in skin cancer patients during the last ten years [6,7]. Skin cancer is classified into numerous types according to its severity and occurrence level. All these types of skin cancer could grow and transmit to various organs and locations of the human body [8,9]. Most of the skin cancer types look similar by appearance, which makes it difficult to correctly identify and classify them. These cancer types mostly appear on the body parts which are directly exposed to sunlight, like the arms, hands, and neck.

Among all these types, melanoma is the most maleficent type of lesion and spreads rapidly. It initiates from the pigmented section of the skin, which could penetrate the profound layer of skin, influencing the entire body. This can be colorless or visible in various colors like a pink rose, dark brown, or azure color. The recovery rate is still not satisfactory; the catastrophe caused by skin cancer, 70%, is due to this class of cancer [10–12]. A few sample images of melanoma are shown in Figure 1a. Benign is another familiar type, which grows slowly and is not fatal if it remains at the upper layer of the skin. This class often develops at locations frequently exposed to sunlight and can grow in different shapes and sizes [13]. A few images of the benign class are shown in Figure 1b.

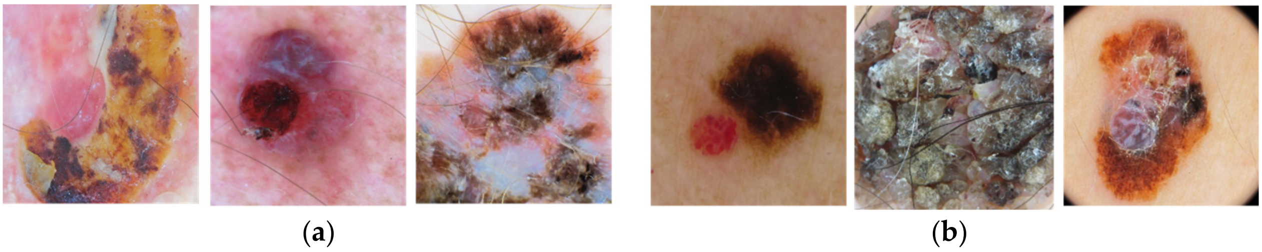


Figure 1. (**a**) Illustration of malignant melanoma; (**b**) illustration of benign melanoma.

Treatment of this disease is possible and can be fully cured if detected in the initial stage, and there are many types of treatments present to cure this disease. Some of the common and best treatments include immunotherapy, chemical peeling, chemotherapy, cryosurgery, and radiation therapy. Since the efficient cure of skin cancer is mainly based on which stage of cancer is diagnosed, identification at the initial stage would highly increase the chances of recovery [14]. The conventional method of skin cancer diagnosis is the manual examination of the skin directly through the naked eye of a dermatologist, but this method could take more time, while an automated system can greatly assist the experts in the diagnosis process [15]. For assisting the automated diagnosis of skin cancer classification, different medical imaging techniques are available. Medical imaging has revolutionized healthcare departments by providing insights and empowering medical practitioners to gather more details and deeply study the human body. Medical imaging can support timely decisions regarding medication and treatment [16]. For skin cancer, dermoscopy is preferred because this modality provides additional information on pattern, color, and other useful information that could help to categorize and compare skin cancer types. Dermoscopy is actually surface microscopy, which gives good results, especially for locating pigmented types of cancer.

Several medical imaging techniques such as CT Scan, MRI, and dermoscopy are utilized to visualize skin cancer intensely. CT scans are commonly used for fetching cross-sectional structural insights beneath the skin, but skin cancer mainly affects the upper layer of skin, and its characteristics and features are better evaluated by opting for surface imaging techniques. MRI is another medical imaging technique that is commonly applied for internal imaging, but these techniques are not used as primary imaging techniques for skin cancer classification due to their focus on deeper tissues and high cost. For most cases of skin cancer, dermoscopy images are preferred for skin cancer assessment due to their focus on surface features. Dermoscopy devices can capture high-resolution magnified images, which are used to classify cancer into different types [17,18]. The classification of skin cancer depends upon multiple factors like geometric features, color, texture, and shape. Diagnosis by visually examining the cancerous skin is challenging and not accurate due to the presence of high similarities between distinct classes. Various studies also present automatic techniques for the identification and classification of skin cancer, but those automatic identification methods also struggle to correctly classify images due to the presence of artifacts, noise, and some other irrelevant information in the images. Also, most of the solutions focused on the binary classification of skin cancer, and most of them relied on using pre-trained models by the selective utilization of network layers [1,19–21]. The development of a robust classification solution for multiclass classification was also affected due to highly imbalanced data. Furthermore, these solutions were not suitable for practical applications due to the high number of parameters, which also increases the computational complexity, leading to requirements of high computational resources. High computational complexity also effects the timely diagnosis of cancer, which is very crucial in clinical settings. The development of an advanced and practical solution was encouraged to overcome these challenges and to develop a robust and lightweight solution that is suitable for deployment in clinical settings. This paper introduces a novel framework for the multiclass classification of skin cancer to address these limitations.

The major contributions of this study are as follows:

- This study introduces an innovative architecture for multiclass skin cancer classification. The inclusion of an attention unit empowers the model to accentuate crucial features, enhancing accuracy by mitigating the impact of less significant ones during the learning process, which also helps to deal with inter-class similarity and intra-class dissimilarities.
- The proposed architecture incorporates Inception-Residual (IR) blocks, leveraging the strengths of both inception and residual networks simultaneously. These blocks address the vanishing gradient problem, facilitating the extraction of multi-scale features. This augmentation significantly boosts the model's ability to discern complex patterns across diverse skin cancer categories.
- The introduced framework achieves computational efficiency with a substantial reduction in parameters. This not only improves the overall computational performance but also ensures robust operation in resource-constrained environments. The result is an efficient and lightweight architecture that maintains high classification accuracy. Moreover, this method demonstrates robustness and practical applicability for deployment across diverse healthcare settings.

2. Related Research Work

Skin cancer is considered a lethal disease that must be diagnosed in its early stages. This is very challenging and time-consuming because different skin cancer types have a high correlation with each other due to their color, texture, or shape. Some environmental factors, like illumination, veins, hairs, etc., could also affect the classification process [22]. Initially, traditional machine learning-based techniques, such as support vector machines (SVMs), were used for skin cancer classification tasks [23]. However, in recent years, deep learning-based techniques have been in demand due to their ability to automatically learn relevant features and complex patterns. These techniques can also handle huge and diverse datasets, allowing real-time diagnosis and improved results.

In medical imaging analysis, especially in skin cancer classification, the most common issue is the non-availability of a sufficient amount of labeled datasets to develop an efficient classification model. In this direction, Hosny et al. [9] used a transfer learning-based approach to train the model with a small dataset. In this study, experiments were carried out for seven types of skin cancer categories using AlexNet as a base model [10]. A similar approach was used for the classification of eight classes of skin cancer by using different pre-trained models on the ISIC 2019 dataset. They reported that different pre-trained models produced diverse results for the same problem. In another study, Thurnhofer-Hemsi et al. [24] used five pre-trained CNN models to make simple and hierarchical classifiers to differentiate between the seven classes of skin cancer from the HAM1000 dataset. Furthermore, Arora et al. [25] compared the performance of fourteen different pre-trained models after fine-tuning on the ISIC 2018 dataset and reported the best results achieved with DenseNet201. This study by X. Chen et al. [26] used a hierarchical pre-training strategy to address challenges in sonar image classification, such as domain gaps, low resolution, and class imbalance problems. The integration of KPS Loss improves knowledge transfer, feature extraction, and classification accuracy. Some other techniques employ ensemble learning to combine two or more models for better results [27]. Chaturvedi et al. [28] compared the performance of five different pre-trained models and four types of ensembles by training them on the same dataset to classify skin cancer.

The skin cancer datasets used for the classification problem are highly imbalanced and contain images of various resolutions. Gessert et al. [29] tried to address the problem of class imbalance by using the loss balance approach, along with an ensemble of different deep learning-based models for skin cancer classification. Rahman et al. [15] introduced the concept of average ensemble learning by making an ensemble of five different models and taking an average of the final output from all the models, which significantly improved the classification results. Raza et al. [30] introduced an ensemble methodology for the classification of melanoma using four pre-trained models. This study utilized extensive data augmentation techniques and a transfer learning approach by fine-tuning each pre-trained model for classifying the acral melanoma and benign nevi. As discussed, the major portion of the research for skin cancer classification is devoted to transfer learning- and ensemble learning-based techniques, which use previously trained models as the base model. In contrast, Iqbal et al. [13] designed a novel model with fewer parameters for skin cancer classification without using any pre-trained model. They used data augmentation techniques on the ISIC 2019 dataset to overcome the class-imbalanced problem.

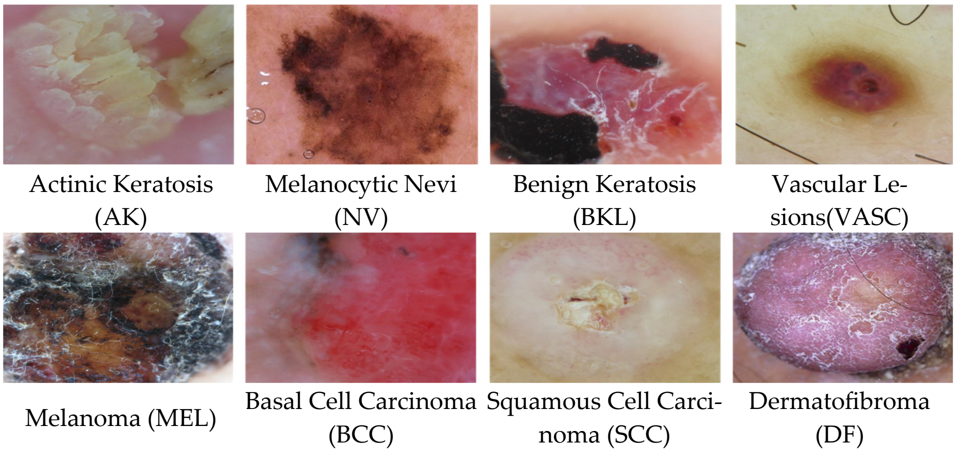
The literature suggests that skin cancer classification is challenging due to visual similarities between cancerous and normal skin images. Moreover, the inter-class similarity between the different types of skin lesions are also a major challenge for accurate classification. To cope with these challenges, Kaur et al. [31] presented a CNN for the automated classification of malignant melanoma from benign images taken from ISIC 2016, 2017, and 2020 datasets. They designed the novel lightweight and less complex DCNN by carefully adding deep layers in the model that helped capture low- to high-level features. The lack of a suitable amount of labeled data is a common problem in medical imaging because the process of labeling data is expensive, time-consuming, and requires lots of human effort. To overcome this challenge, Alzubaidi et al. [32] introduced a CNN-based model, which was trained on an excessive amount of unlabeled medical imaging datasets and was then optimized on a small amount of labeled data. Another study by Datta et al. [33] tried to cope with the problem of noise by introducing the concept of soft attention in different pre-trained models, which enhanced the performance of base networks by learning less from the noise-containing features. To improve the overall performance of multiclass classification, Hsu et al. [34] presented a novel method called HAC-LF, in which a new loss function was designed to decrease the influence of misclassification. That loss function enhances the classification efficiency by decreasing the major-type error rate. One of the crucial aspects acting as a hurdle in the performance of the skin cancer classification model is inter-class similarity and intra-reader variability, for which Wang et al. [35] introduced a new approach by adopting the technique of multimodel classification and fusing it with the attention-based mechanism. This method extracted features by using adversarial learning to obtain complementary and correlated information from both modalities. Upon the evaluation of multimodel datasets, this approach achieves superior results. Along with achieving high performance, reducing computational time is also a very crucial aspect, for which Ajmal et al. [36] launched a new algorithm based on a fuzzy entropy slime module along with the concept of deep learning for disregarding a large number of irrelevant features. The first step includes fine-tuning two deep learning models to obtain two feature vectors from fine-tuned models. In the next step, the fuzzy entropy slime mold algorithm was applied to dismiss useless features, followed by a fusion of the remaining optimal features. Then, a machine learning classifier was opted for the classification. With the evolution of deep learning-based approaches, significant progress has been made for skin cancer; still, there are many challenges, such as class imbalance problems, high computational costs, and low accuracy.

3. Material and Methods

This study introduced two novel deep learning-based models. The proposed Inception-Residual CNN (IR-CNN) is designed to deal with all the common issues that models face during training, including vanishing gradients and overfitting. In the proposed AIR-CNN, the attention unit is also present along with several IR blocks in order to correctly classify skin cancer. Thus, both models are designed specifically for the task of correctly classifying multiple classes of skin cancer. The subsequent section presents all the details related to the dataset, pre-processing steps, and their architecture.

3.1. Dataset

The International Skin Image Collaboration (ISIC) is making effort globally to improve the diagnosis of skin cancer by providing access to large dermoscopy datasets by the support of the International Society for Digital Imaging of the Skin (ISDIS). Experiments of this research are carried out on the ISIC 2019 dataset, which includes around 25,331 dermoscopic images of 8 types of skin cancer, which are Melanocytic Nevi (NV), Actinic Keratosis (AK), Basal Cell Carcinoma (BCC), Vascular cancer (VASC), Dermatofibroma (DF), melanoma (MEL), Squamous Cell Carcinoma (SCC), and Benign Keratosis (BKL). Three distinct datasets are merged to make this dataset, which also leads to a highly imbalanced no of images in various classes. A sample image for each class is shown in Figure 2.



Actinic Keratosis (AK)

Melanocytic Nevi (NV)

Benign Keratosis (BKL)

Vascular Lesions(VASC)

Melanoma (MEL)

Basal Cell Carcinoma (BCC)

Squamous Cell Carcinoma (SCC)

Dermatofibroma (DF)

Figure 2. Representative images from eight classes of skin cancer.

3.2. Pre-Processing

To enhance the performance of the model, data were passed from a pre-processing pipeline. The first step in this process was the standardization, performed by resizing all the images to the same dimension.

The ISIC 2019 dataset utilized in this study presents varying dimensions due to the amalgamation of three distinct datasets: HAM10000 with images sized at 450 × 600, BCN_20000 with images sized at 1024 × 1024, and the MSk dataset with images of various dimensions [37–39]. To ensure uniformity, all dermoscopic images were resized to 148 × 148 for one of our methods, IR-CNN, and to 224 × 224 for the other, AIR-CNN. This crucial step not only eradicated the challenge of varied image sizes, but also established a consistent input format, stimulating optimal learning across the spectrum of skin cancer. In the next step, all dermoscopic images were cropped, eliminating extraneous information and centering the cancerous regions. This deliberate action escalates the focus of the model on the pertinent features of images, placing the skin cancer area at the center of the image and significantly enriching the feature extraction capabilities, thereby elevating the validity of the model in the classification task. A visual representation of the original and cropped images can be found in Figure 3.

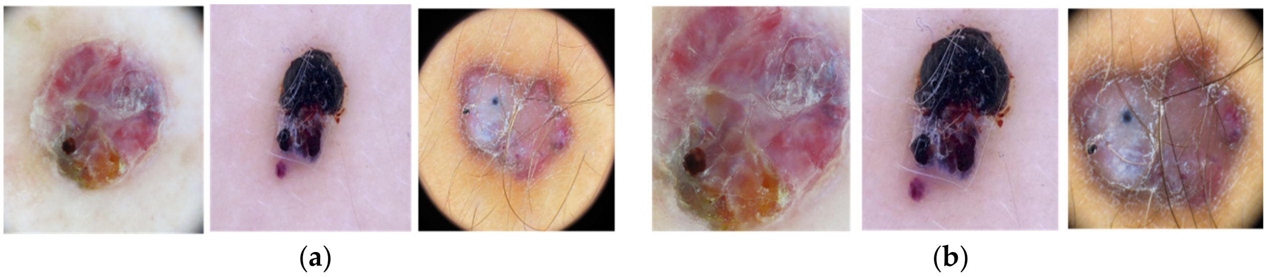


Figure 3. (**a**) original image samples (**b**) cropped image samples.

Notably, this dataset exhibited a significant class imbalance, which could lead to overfitting and bias in model predictions, particularly giving favor to those classes having more samples. To address this, data augmentation techniques were employed in the classes that have a small no of images during the training phase [40]. These methods included shifting the images by 11% in all directions (upward, rightward, leftward, and downward), introducing positional diversity. Rotations ranging from 20 degrees clockwise and counterclockwise to 60 degrees were applied in both directions, enhancing angular diversity. Horizontal and vertical flipping was also employed. These augmentation techniques not only augmented the data size but also collectively enriched the dataset's diversity, which is essential for the robust training of the model. Samples were reduced from the classes having a large no of images because, after applying a generous amount of augmentation techniques to the classes with a small no of images, the class imbalance problem still existed. Following these steps, experiments were carried out on 25,172 images from 8 skin cancer classes. From these, 70% (18,123) of images were utilized for training purposes, facilitating the model learning process; additionally, 10% (2014) of samples were allocated for validation and 20% (5035) of datasets were separated for testing the model's generalization ability. The pre-processing steps culminated with normalization. Z-Score normalization was applied to all images, bringing their pixel values to a mean of 0 and a standard deviation of 1. This final step proved essential in enhancing model convergence during training by standardizing the pixels' value. This action eliminated potential biases and reinforced its adaptability across diverse datasets. The original and augmented images are shown in Figure 4.

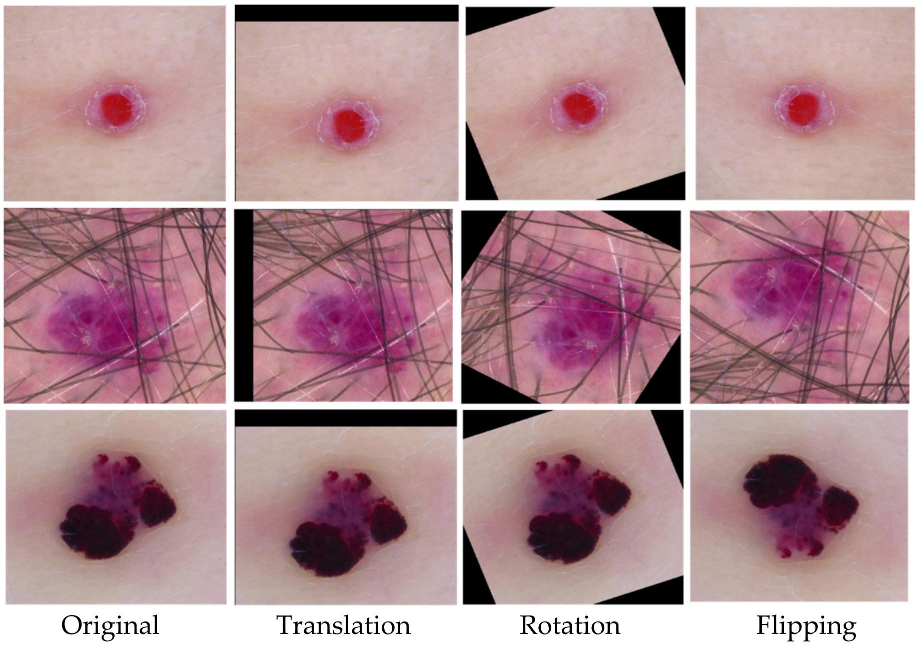


Figure 4. The leftmost column shows the original images, and the other columns show the corresponding augmented images using translation, rotation, and flipping.

3.3. Inception-Residual CNN (IR-CNN)

The proposed architecture was specially designed and developed by keeping in mind all the common and frequent challenges that frequently occur during the diagnosis and classification of skin cancer and directly affect the performance and accuracy of the model. The design of the presented architecture was finalized after multiple refinements, which were performed on the basis of the results of several experiments. The final proposed network is constructed by incorporating six Inception-Residual (IR) blocks to create a deep layered architecture. The structure and no of IR blocks were determined through a series of experiments and by considering some important factors. One factor while designing the architecture of the model was to achieve a balance between the model's complexity and performance. Integrating more than six IR blocks does not increase significant accuracy, and reducing the IR blocks from six leads to a notable decline in performance. Including six IR blocks improves the capability of the model to effectively capture and learn complex features and structures from input data. Memory and computation resources are other major factors deciding the no of blocks; the addition of six IR blocks allows us to achieve the balance between computational efficiency and predictive performance by ensuring the efficient utilization of memory and computational resources without compromising diagnostic accuracy. The decision to include six IR blocks was based on performance optimization, generalization capability, computational efficiency, and model complexity. These blocks combine elements from both the Inception and Residual architectures by applying the concept of residual connections within Inception blocks. The motivation behind introducing IR blocks into this architecture is to harness the benefits of both techniques simultaneously. Inception blocks enable the use of filters of different sizes in their convolutional layers, allowing the learning of features at various scales while minimizing the computational cost by reducing the parameter count. However, the dataset was relatively small, and the issue of vanishing gradients can emerge. To address this, residual connections are incorporated within these blocks to mitigate the vanishing gradient problem. The visual representation of the IR-CNN architecture is shown in Figure 5.

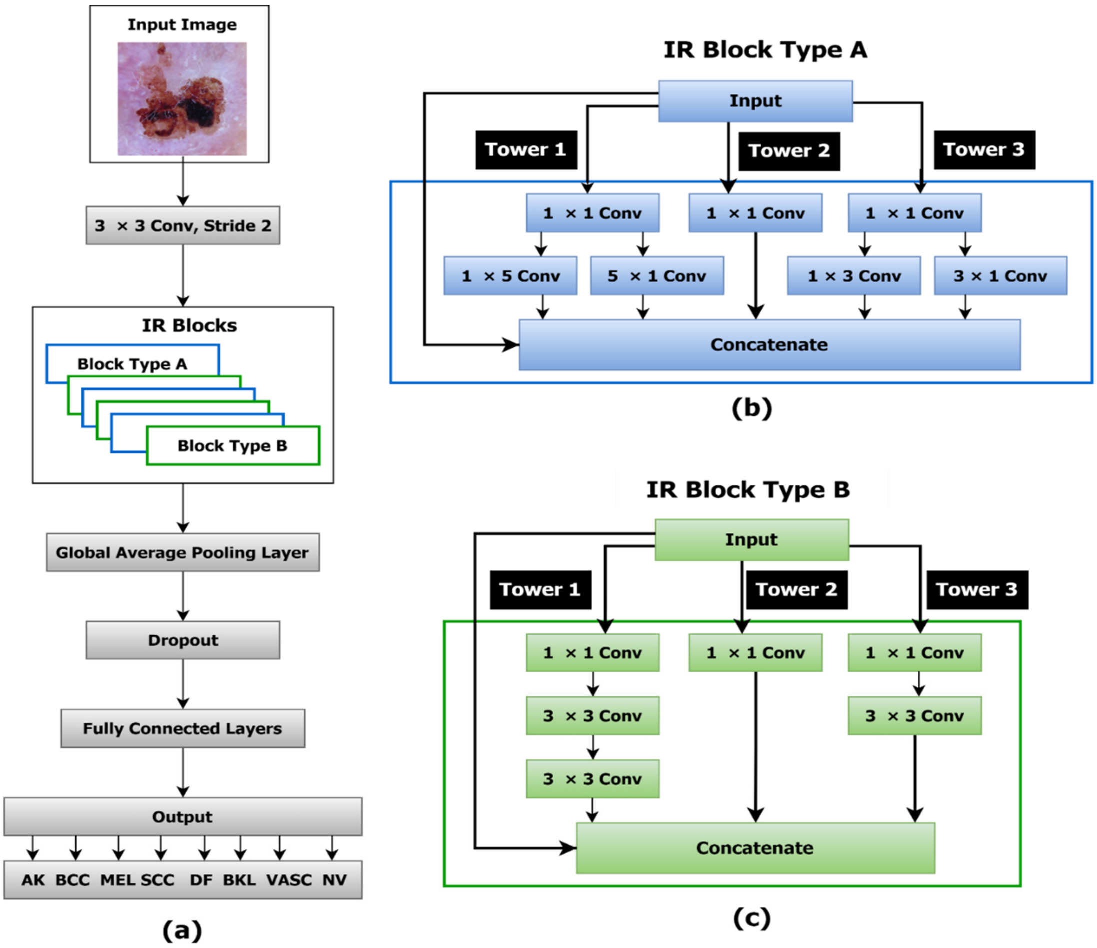


Figure 5. (**a**) Comprehensive structure of proposed IR-CNN. (**b**) Architectural details of IR block Type A. (**c**) Architectural details of IR block Type B.

Two variations of IR blocks, namely Type A and Type B, were incorporated in this model. The reason behind including two distinct types of IR blocks is to increase the model's flexibility and feature extraction diversity. Each type of IR block has different characteristics and the ability to extract features and learn from input data, leading to a more comprehensive understanding of the complex patterns hidden in input data. This architecture is also designed to handle the problem of highly similar features of different classes and distinguish features of the same classes of cancer. These blocks excel in capturing various types of information, including local spatial patterns, and extracting broader contextual information. The incorporation of two types of IR blocks increases the capacity of the IR-CNN model by capturing a broader range of features and semantic representations.

The network architecture begins with the input image, measuring 148 × 148 × 3, which is processed through the initial convolution layer, which serves as the model's input layer. This layer employs 32 filters, each with a size of 3 × 3 and a stride of 1. The resulting output is then directed to the first Inception-Residual (IR) block, a fundamental component crafted to capture both low-level and high-level features crucial for effective image analysis.

The first IR block consists of three towers, each contributing distinct convolutional operations. The first tower comprises a single 1 × 1 convolutional layer, while the second tower involves a 1 × 1 convolutional layer followed by two parallel convolutional layers with filter sizes of 1 × 3 and 3 × 1. The third tower integrates a 1 × 1 convolutional layer, succeeded by two parallel convolutional layers with filter sizes of 1 × 5 and 5 × 1. All convolutional layers within this block maintain a stride of one, and the "same" padding is applied. The outputs from these three towers are concatenated using the concatenation operator, creating the block's output. Notably, a residual connection is established in this block by combining the input and output. The second type of IR block is characterized by three towers as well. The first tower features a 1 × 1 convolutional layer followed by a 3 × 3 convolutional layer. The second tower incorporates a 1 × 1 convolutional layer followed by parallel 3 × 3 and another 3 × 3 convolutional layers. The third tower consists of a 1 × 1 convolutional layer. Stride and padding are kept consistent across all layers within this block. Similarly to the previous block, the outputs from the three towers are concatenated at the end of the block, and a residual connection is formed by concatenating the input and output. This detailed architectural arrangement ensures effective feature extraction and maintains a streamlined and efficient model structure.

The Leaky ReLU activation function is employed in each convolutional layer, providing a slight extension from zero for the negative side. The output from the last inception block is then directed to the GlobalAveragePooling layer, which serves to flatten the data. To prevent overfitting, a dropout of 50% is applied immediately after the GlobalAveragePooling layer. Subsequently, a fully connected layer with 512 neurons is added. Finally, the Softmax function is used to classify the output into eight classes. The summary of this model is given in Table 1.

Table 1. Comprehensive detail of the implemented IR-CNN.

Name of Layers	Filters Sizes (FS) and Stride (S)	Activation
Input layer		148 × 148 × 3
conv_1	FS = 3 × 3, S = 2	74 × 74 × 32
(conv_A1_1) × 2	FS = 1 × 1, S = 1	74 × 74 × 16
conv_A1_11	FS = 1 × 5, S = 1	74 × 74 × 32
conv_A1_12	FS = 5 × 1, S = 1	74 × 74 × 32
conv_A1_21	FS = 1 × 3, S = 1	74 × 74 × 32
conv_A1_22	FS = 3 × 1, S = 1	74 × 74 × 32
conv_A1_3	FS = 1 × 1, S = 1	74 × 74 × 32
concatenate_1	5 inputs	74 × 74 × 160
skip_conection	2 inputs	74 × 74 × 192
(conv_A2_1) × 2	FS = 1 × 1, S = 1	74 × 74 ×16
(conv_A2_2) × 3	FS = 3 × 3, S = 1	74 × 74 × 32
conv_A2_3	FS = 1 × 1, S = 1	74 × 74 × 32
Concatenate_2	3 inputs	74 × 74 × 96
skip_conection	2 inputs	74 × 74 × 288
(conv_A3_1) × 2	FS = 1 × 1, S = 1	74 × 74 × 16
conv_A3_11	FS = 1 × 5, S = 1	74 × 74 × 64
conv_A3_12	FS = 5 × 1, S = 1	74 × 74 × 64
conv_A3_21	FS = 1 × 3, S = 1	74 × 74 × 64
conv_A3_22	FS = 3 × 1, S = 1	74 × 74 × 64
conv_A3_3	FS = 1 × 1, S = 1	74 × 74 × 64
Concatenate_3	5 inputs	74 × 74 × 320
skip_conection	2 inputs	74 × 74 × 608
(conv_A4_1) × 2	FS = 1 × 1, S = 1	74 × 74 × 32
(conv_A4_2) × 3	FS = 3 × 3, S = 1	74 × 74 × 64
conv_A4_3	FS = 1 × 1, S = 1	74 × 74 × 64
concatenate_4	3 inputs	56 × 56 × 192
skip_conection	2 inputs	56 × 56 × 800
(conv_A5_1) × 2	FS = 1 × 1, S = 1	74 × 74 × 128
conv_A5_11	FS = 1 × 5, S = 1	74 × 74 × 128
conv_A5_12	FS = 5 × 1, S = 1	74 × 74 × 128
conv_A5_21	FS = 1 × 3, S = 1	74 × 74 × 128
conv_A5_22	FS = 3 × 1, S = 1	74 × 74 × 128

Table 1. *Cont.*

Name of Layers	Filters Sizes (FS) and Stride (S)	Activation
conv_A5_3	FS = 1 × 1, S = 1	74 × 74 × 128
concatenate_5	5 inputs	74 × 74 × 640
skip_conection	2 inputs	74 × 74 × 1440
(conv_A6_1) × 2	FS = 1 × 1, S = 1	74 × 74 × 64
(conv_A6_2) × 3	FS = 3 × 3, S = 1	74 × 74 × 128
conv_A6_3	FS = 1 × 1, S = 1	74 × 74 × 128
Concatenate_6	3 inputs	74 × 74 × 384
skip_conection	2 inputs	74 × 74 × 1824
G1	Global Average Pooling	1 × 1 × 1824
D	0.5% Dropout	1 × 1 × 1825
FC	Fully Connected	1 × 1 × 512
Output (Softmax function)	NV, BCC, DF, VASC, SCC, BKL, MEL, AK	1 × 1 ×8

3.4. Attention-Based Inception-Residual CNN (AIR-CNN)

In the novel AIR-CNN architecture, an attention mechanism is introduced. The use of attention in deep neural networks is gaining popularity due to the significant benefits it offers. In the field of medical imaging, attention modules are particularly recognized as they allow the network to focus on pertinent parts of the image. The primary aim of incorporating attention in the network is to assign higher importance to the most relevant features during training [41,42]. In the current dataset, the cancerous region could be located anywhere within the image. As only a small number of pixels contain disease-related information, while the rest may contain noise and artifacts, the attention module is integrated to guide the network on where to focus during the feature learning process. The visual representation of the proposed architecture is shown in Figure 6.

In the AIR-CNN architecture, alongside the attention module, seven similar Inception-Residual (IR) blocks are incorporated. These IR blocks are designed to strike a balance between computational efficiency and mitigating the problems of exploding and vanishing gradients, drawing inspiration from both residual and inception concepts. The selection of seven IR blocks was made to provide a higher level of adaptability and flexibility to the model to dynamically adjust its feature extraction and representation strategy based on the input data. As the skin cancer dataset consists of complex patterns and structures, a more expressive model architecture was required to effectively capture the full spectrum of relevant features. The finalization of seven IR blocks was performed after iterative experimentation and performance optimization. After varying configurations of IR blocks across different experiments, model performance was systematically evaluated, which depicted superior results by the addition of seven IR blocks.

The Stem block, which acts as the initial layer of the network, features a convolutional layer with eight 3 × 3 filters and a stride of two. This reduces the image dimensions from 224 × 224 to 112 × 112. Subsequently, the output is directed to a max-pooling layer to reduce the dimensions further to 56 × 56. The resulting feature map is then passed to the first IR block. Each IR block includes three inception towers:

- Tower 1: A convolutional layer with a 1 × 1 filter size, a stride of one, and "same" padding.
- Tower 2: A convolutional layer with a 1 × 1 filter size, followed by another convolutional layer with a 3 × 3 filter size, both with a stride of one and "same" padding.

- Tower 3: A convolutional layer with a 1 × 1 filter size and a stride of one, followed by a 3 × 3 convolutional layer with the same stride, and subsequently, another convolutional layer with a 3 × 3 filter size, "same" padding, and a stride of one.

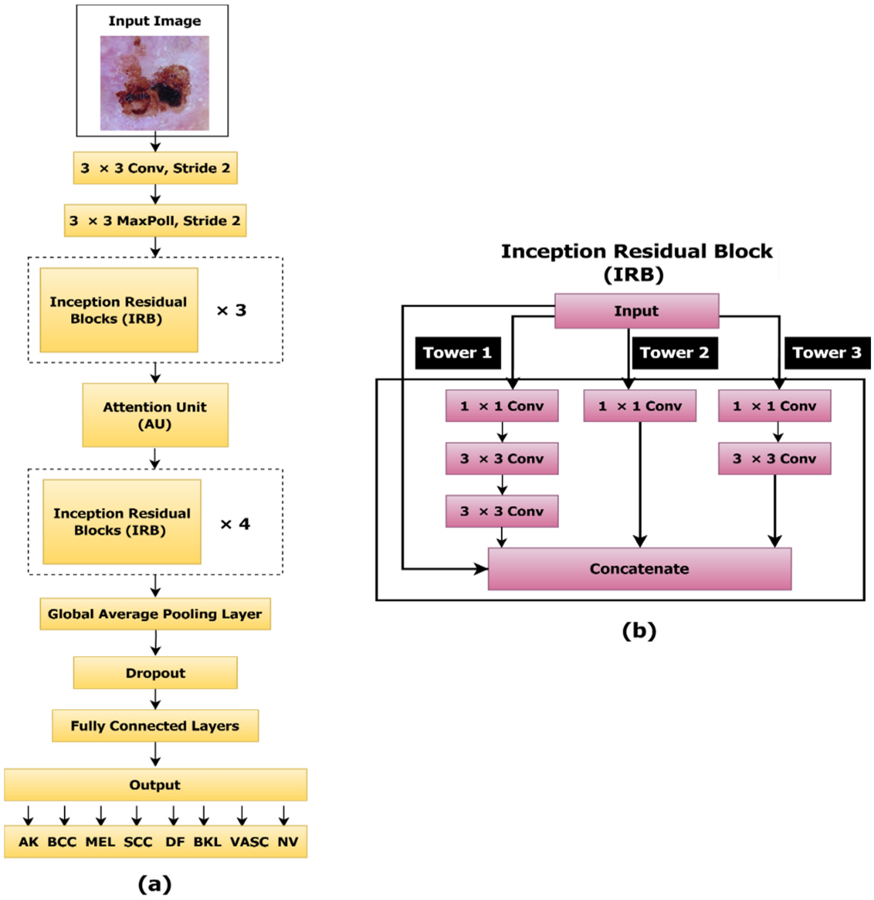


Figure 6. (**a**) Comprehensive structure of proposed architecture of AIR-CNN. (**b**) Architectural details of IR blocks.

The outputs of all three towers are concatenated using the concatenation operator, and each block establishes a residual connection by combining the input and output of the block.

Stride is an important component of the network, which is used to describe the steps to move the filter over the image. The same block is repeated six more times to construct a deep neural network for fetching complete cancer details like shape, color, edges, and complex cancer features. The no of filters increases gradually as the network becomes deeper. The first two pairs of blocks utilize a combination of 8 and 16 filters, which are doubled to 16 and 32 in the next two blocks, and the last three blocks picked 32 and 64 filters. Table 2 describes comprehensive details of the implemented model.

The module of soft attention is placed after the third IR block. The decision to select soft attention over other types of attention was taken due to its compatibility with cnn networks, interpretability, robustness to noise, and differentiability. The attention unit enhances feature extraction, improves interpretability, and effectively discriminates between supreme and noisy features within dermoscopic-based images. The soft attention mechanism, by assigning continuous weights to input elements, empowers the model to pay attention to multiple regions simultaneously. This provides facilitation for keenly understanding input data, allowing the model to dynamically allocate more focus to crucial features while minimizing the impact of noise.

Table 2. Comprehensive detail of the implemented AIR-CNN.

Name of Layers	Filters Sizes (FS) and Stride (S)	Activations
Input layer	-	224 × 224 × 3
conv_1	FS = 3 × 3, S = 2	112 × 112 × 8
max_1	FS = 3 × 3, S = 2	56 × 56 × 8
(conv_A1_1) × 2	FS = 1 × 1, S = 1	56 × 56 × 8
(conv_A1_2) × 3	FS = 3 × 3, S = 1	56 × 56 × 16
conv_A1_3	FS = 1 × 1, S = 1	56 × 56 × 16
concatenate_1	3 input	56 × 56 × 48
skip_conection	2 input	56 × 56 × 56
(conv_A2_1) × 2	FS = 1 × 1, S = 1	56 × 56 × 8
(conv_A2_2) × 3	FS = 3 × 3, S = 1	56 × 56 × 16
conv_A2_3	FS = 1 × 1, S = 1	56 × 56 × 16
Concatenate_2	3 input	56 × 56 × 48
skip_conection	2 input	56 × 56 × 104
(conv_A3_1) × 2	FS = 1 × 1, S = 1	56 × 56 × 16
(conv_A3_1) × 3	FS = 3 × 3, S = 1	56 × 56 × 32
conv_A3_3	FS = 1 × 1, S = 1	56 × 56 × 32
Concatenate_3	3 input	56 × 56 × 96
skip_conection	2 input	56 × 56 × 200
conv_Attention_1	Attention Module	56 × 56 × 128
activation_1		56 × 56 × 128
conv_Attention_2		56 × 56 × 128
conv_Transpose		56 × 56 × 128
conv_Attention_3		56 × 56 × 128
concatenate		56 × 56 × 128
activation_2		56 × 56 × 128
conv_Attention_4		56 × 56 × 1
activation_2		56 × 56 × 1
up_sampling		56 × 56 × 1
lambda		56 × 56 × 200
Multiply		56 × 56 ×200
(conv_A4_1) × 2	FS = 1 × 1, S = 1	56 × 56 × 16
(conv_A4_2) × 3	FS = 3 × 3, S = 1	56 × 56 × 32
conv_A4_3	FS = 1 × 1, S = 1	56 × 56 × 32
concatenate_4	3 input	56 × 56 × 96
skip_conection	2 input	56 × 56 × 296
(conv_A5_1) × 2	FS = 1 × 1, S = 1	56 × 56 × 32
(conv_A5_2) × 3	FS = 3 × 3, S = 1	56 × 56 × 64
conv_A5_3	FS = 1 × 1, S = 1	56 × 56 × 64
concatenate_5	3 input	56 × 56 × 160
skip_conection	2 input	56 × 56 × 456
(conv_A6_1) × 2	FS = 1 × 1, S = 1	56 × 56 × 32
(conv_A6_2) × 3	FS = 3 × 3, S = 1	56 × 56 × 64
conv_A6_3	FS = 1 × 1, S = 1	56 × 56 × 64
Concatenate_6	3 input	56 × 56 × 192
skip_conection	2 input	56 × 56 × 648
(conv_A7_1) × 2	FS = 1 × 1, S = 1	56 × 56 × 32
(conv_A7_2) × 3	FS = 3 × 3, S = 1	56 × 56 × 64
conv_A7_3	FS = 1 × 1, S = 1	56 × 56 × 64
concatenate_7	3 input	56 × 56 × 192
skip_conection	2 input	56 × 56 × 840
G1	Global Average Pooling	1× 1 × 840
D	0.5 Dropout	1 × 1 × 840
FC	Fully Connected	1 × 1 × 256
Output (Softmax function)	NV, BCC, DF, VASC, SCC, BKL, MEL, AK	1 × 1 × 8

Soft attention, particularly within the context of the suggested attention residual-based CNN, empowers the model to focus on pertinent aspects of skin cancer areas during the process of classification. The discrete examination of various sub-regions of dermoscopic images ensures adaptability to gradient descent and backpropagation, aligning seamlessly with the training of the presented CNN model.

Furthermore, soft attention is differentiable, and its weights can be updated through back-and-forth propagation. These crucial features not only facilitate optimization during training but also contribute to the flexibility of the attention mechanism within the CNN framework.

An important advantage of using the soft attention mechanism is its capability to prioritize important features and dismiss the influence of noise-containing factors. This selective attention system is particularly preferable in the context of skin cancer classification, where distinguishing between clinically important patterns and irrelevant noisy factors is necessary. By emphasizing essential features and de-emphasizing noise-containing features, the AIR-CNN achieved high accuracy and robustness in its predictions.

In a typical CNN, filters use only local and surrounding information to compute the output pixel's value, but the strategy behind soft attention is to raise or enervate the value of each pixel according to its similitude global features. In simple words, attention is the process of assigning a high value to similar features and a low value to diverse features. The location of the attention module was selected empirically after several experiments, where it was placed at various architectural locations.

After IR blocks, a layer of GlobalAveragePooling follows the last inception block, followed by dense layers containing 256 neurons. This leads to a Softmax layer acting as an output layer to classify the skin cancer. The total number of filters is a major contributing factor in deciding the parameters of a network. Models with a suitable number of parameters can boost performance, but huge numbers of parameters could drop performance and slow down the learning speed. Other state-of-the-art studies contain ~45.6 M (Liu et al. [43]), ~267.5 M (Mahbod et al. [44]), ~3.3 M (Kaur et al. [31]), ~4.8 M (Harangi. [45]), and ~256.7 M (Iqbal et al. [13]) parameters. As compared to these studies, the proposed attention CNN model has ~1.1 M parameters without compromising the network performances, which is a great achievement.

4. Experimental Setup

Both presented networks are implemented using Python 3.8 language with TensorFlow and Keras libraries. Experiments were carried out on GPU NVIDIA GeForce RTX 2080 Ti with 32 GB RAM. The model was trained with a batch size of 32; the minimum value of the batch size was selected due to the limited available resources. The dataset picked for training purposes is benchmark ISIC 2019 skin cancer dermoscopic data, which is divided into three splits with a percentage of 70, 10, and 20 for training, validation, and testing. Hyperparameters like optimization algorithms, learning rates, and activation functions play a vital role in the performance of models [46,47]. To achieve the best configuration, hyperparameters were tuned and selected after multiple experiments using the grid search method. The learning rate of 0.0001 was selected with the ADAM optimizer, and LeakyRELU was finalized as the activation function. Table 3 provides complete details of the hyperparameters setting chosen for the training of the proposed model.

Table 3. Hyperparameters of the AIR-CNN and IR-CNN.

Hyperparameter	Optimizer	Learning Rate	Batch Size	Activation Function	Dropout	No. of Epochs
Value	ADAM	0.0001	32	LeakyRELU	0.5	500 (AIR-CNN) 350 (IR-CNN)

Extensive experiments were performed with original and augmented data to evaluate the performance of the model with and without any augmentation and with different parameters and hyperparameters settings. Additionally, the performance was also evaluated by introducing batch normalization layers in each IR block, but the performance of the model drops as batch normalization is highly dependent on the batch size and its performance fluctuates by increasing the batch size. By introducing Relu as the activation function in the network, performance significantly decreases as it sometimes suffers the dying Relu Problem during the training process, due to which some neurons of hidden layers start dying by outputting only zero values. The model's performance was also examined by fine-tuning the value of the learning rate to 0.0075, 0.0005, 0.0025, 0.001, and 0.075. The overall performance of the model was evaluated based on five evaluation measures, including accuracy, precision, sensitivity, specificity, and F1 score.

5. Results and Discussion

The proposed novel CNN models for the task of skin cancer classification were designed by deeply examining all the crucial aspects of architectural design, the nature of the problem, and hyperparameter choices. The addition of inception-residual blocks in the network represents a deliberate contribution to tackling two frequent issues that arise while working with deep learning techniques: vanishing gradients and overfitting. The problem of vanishing gradients occurs during backpropagation when the gradient becomes extremely small, obstructing crucial weight updates and impeding the process of learning. For facilitating the smooth flow of gradients through residual connections, the proposed networks integrated IR blocks in the network.

Despite training on a small amount of data, the efficient architectural design of both networks alleviates the challenge of overfitting. Overfitting occurs when a model learns data too well during training, including noise and specific patterns, but is unable to generalize well on unseen data. One of the main reasons behind designing simple and lightweight networks was to avoid the risk of overfitting, generally caused by complex architectures trained on small amounts of datasets. To enable the model for better feature extraction and to discern relevant patterns without succumbing to overfitting, the attention mechanism was fused in the AIR-CNN network. Dropout was strategically applied at the rate of 50% to further guard against overfitting and to enhance the generalization capabilities of the model. This regularization technique prevents the model from being overly specialized to the training data, ensuring its adaptability to diversified skin cancer data.

The selection of hyperparameters was a delicate task, which was finalized by exploring various configurations of hyperparameters, including the optimizer, learning rate, batch size, no of epochs, and activation function. Various optimizers were explored, but the Adam optimizer outperformed SGD, Adamax, and other optimizers in terms of convergence, speed, and model accuracy due to its adaptive learning rate and moment estimation property.

The role of the activation function could not be neglected in shaping the model's performance. After experimenting with other alternative activation functions, such as Relu, LeakyRELU emerged as the most optimal choice for this classification problem. Increasing the model's overall efficiency by capturing complex features and the capability of mitigating the vanishing gradient problem makes the LeakyRELU function the best choice for this task.

Initially, the batch normalization layer was included to improve the stability of the model, but empirical results indicated a notable drop in performance when this layer was added, causing its exclusion from the final architectures. Various experiments showed that computational complexity is directly connected with batch size, as an increase in batch size also raises computational complexity without a proportional gain in performance. Therefore, the batch size of 32 was opted in both models to achieve a balance between training efficiency and model effectiveness. In conclusion, the selection and finalization of architectural design and hyperparameters were decided after numerous experiments and comparative analyses.

The design and architecture of both models were selected specially for classifying multiple skin cancer types. The evaluation of the model carried out on ISIC 2019 data and models performed well in terms of precision, F1-score, sensitivity, specificity, and accuracy. The AIR-CNN model achieved 91.63 accuracy in classifying multi classes of skin cancer. The performance of the model was also satisfactory during class-wise evaluation through precision, sensitivity, specificity, and F1-score. Another achievement of the IR-CNN model is the reduction in the number of trainable parameters to ~2.2 M, presenting architecture that was customized for efficient skin cancer classification. Despite the reduction in trainable parameters, this model shows good performance in efficiently classifying multiple types of skin cancer. The overall and class-wise results of the suggested IR-CNN model are presented in Tables 4 and 5, respectively. Tables 6 and 7 shows the overall and class-wise results of the suggested AIR-CNN.

Table 4. Results of the proposed IR-CNN.

Dataset Type	Test Data				
Metrics	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**	**F1-Score**
Results	91.53	91.6	91.5	91.5	91.60

Table 5. Class-wise results of the IR-CNN.

Classes Names	Accuracy	Precision	Sensitivity	Specificity	F1-Score
AK	93.9	93.0	94.0	92.8	93.0
BCC	89.7	90.0	90.0	89.1	90.0
BKL	88.5	86.0	92.0	87.9	89.0
DF	95.2	96.0	95.0	94.7	96.0
SCC	86.8	87.0	84.0	86.3	86.0
VASC	87.7	89.0	88.0	87.2	88.0
MEL	96.3	96.0	92.0	94.0	94.0
NV	98.6	99.0	100	98.5	99.0

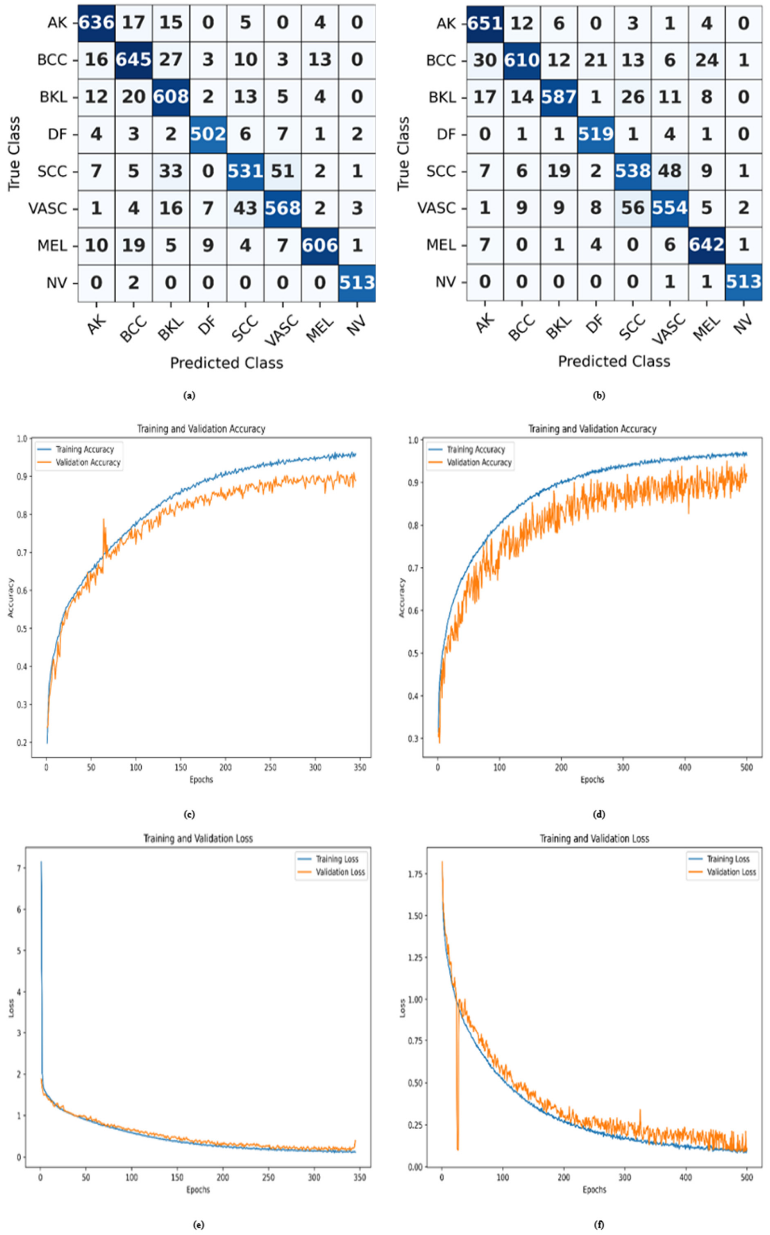
Table 6. Overall results of the proposed AIR-CNN.

Dataset Type	Test Data				
Metrics	**Accuracy**	**Precision**	**Sensitivity**	**Specificity**	**F1-Score**
Results	91.63	91.60	91.60	91.52	91.60

To verify performance and robustness, the model was also evaluated on other evaluation measures like loss curves, confusion matrix, training, and testing accuracy. The confusion matrix is an important measure specifically in classification tasks because it provides complete details of predictions and actual outcomes. The diagonal cells with dark blue color (from top-left to bottom-right) shows the number of correct classifications for each class (true positive for each class), while off-diagonal elements with light blue color represent misclassifications. The confusion matrix of the test results achieved from the IR-CNN and AIR-CNN is shown in Figure 7a and Figure 7b, respectively.

Table 7. Class-wise results of the AIR-CNN.

Classes Names	Accuracy	Precision	Sensitivity	Specificity	F1-Score
AK	90.1	91.0	96.0	89.9	94.0
BCC	92.0	94.0	85.0	91.2	89.0
BKL	89.2	92.0	88.0	88.0	90.0
DF	96.4	94.0	98.0	95.7	96.0
SCC	84.8	84.0	85.0	83.3	85.0
VASC	88.0	88.0	86.0	86.2	87.0
MEL	94.3	93.0	97.0	92.1	95.0
NV	99.5	99.0	100	99.0	99.0



True Class	AK	BCC	BKL	DF	SCC	VASC	MEL	NV
AK	636	17	15	0	5	0	4	0
BCC	16	645	27	3	10	3	13	0
BKL	12	20	608	2	13	5	4	0
DF	4	3	2	502	6	7	1	2
SCC	7	5	33	0	531	51	2	1
VASC	1	4	16	7	43	568	2	3
MEL	10	19	5	9	4	7	606	1
NV	0	2	0	0	0	0	0	513

Predicted Class

(a)

True Class	AK	BCC	BKL	DF	SCC	VASC	MEL	NV
AK	651	12	6	0	3	1	4	0
BCC	30	610	12	21	13	6	24	1
BKL	17	14	587	1	26	11	8	0
DF	0	1	1	519	1	4	1	0
SCC	7	6	19	2	538	48	9	1
VASC	1	9	9	8	56	554	5	2
MEL	7	0	1	4	0	6	642	1
NV	0	0	0	0	0	1	1	513

Predicted Class

(b)

(c)

(d)

(e)

(f)

Figure 7. (**a**) Confusion matrix of IR-CNN. (**b**) Confusion matrix of AIR-CNN. (**c**) Training and validation accuracy curve of IR-CNN. (**d**) Training and validation accuracy curve of AIR-CNN. (**e**) Training and validation loss curve of IR-CNN. (**f**) Training and validation loss of AIR-CNN.

The training and validation accuracy curves serve as valuable visual representations to understand the performance of the proposed network during the training phase. As training progresses, the accuracy curves start going upward, indicating the improvement in the model's generalization ability on unseen data. On the other side, the training and validation loss curves illustrate that error starts decreasing with time. Figure 7c,e represent the accuracy and loss curves, respectively, for the IR-CNN, and Figure 7d,f represent the accuracy and loss curves, respectively, for the AIR-CNN. In Figure 7c, validation accuracy temporarily exceeds the training accuracy because this dataset has inter-class similarities and intra-class differences, which makes it difficult to learn from some classes. At this point during training, the model might generalize better to the validation set than the training data due to stochastic effects, like batch sampling. During those epochs, the presence of more distinguishable examples in the validation set could also result in higher validation accuracy than training accuracy. The loss curve of the AIR-CNN in Figure 7f faces a significant drop in validation loss that could be due to improved feature learning at that phase of training. The model might experience a breakthrough in learning discriminative features that remarkably reduce the error rate on validation data. This non-linear performance could be the result of the sensitivity of the model in correctly distinguishing cancerous images, such as patterns, colors, and texture.

Together, these curves demonstrate learning dynamics with the enhancement of accuracy and reduction in losses with an increasing number of iterations.

5.1. K-Fold Validation

Both presented architectures, the IR-CNN and AIR-CNN, were also trained by applying the K-fold validation technique, in which the data are segmented into a 'K' number of folds. The most optimum value of 'K' selected in this work is 10, which means that the dataset is divided into 10 equal parts for a 10-fold cross-validation technique, for which images are randomly selected for each fold. During training, for each iteration, nine folds were used for training, and one fold was reserved for testing. In the end, the arithmetic mean of all folds is calculated. The results of K-fold cross-validation are shown in Table 8 and models with superior performance are shown bold.

Table 8. Results using K-fold validation with ISIC 2019 dataset.

Authors	Accuracy
Gessert et al. [29]	63.6
Proposed (IR-CNN)	**88.28**
Proposed (AIR-CNN)	**93.39**

5.2. Comparison with State-of-the-Art Studies

In this section, both suggested models are compared with the cutting-edge studies. The IR-CNN model and AIR-CNN model are compared with other studies for the task of skin cancer classification. Presented models were compared against advanced metrics such as accuracy, precision, F1-score, specificity, and sensitivity. The Mijwil et al. [2] study achieved an accuracy of 86.90% for the binary classification of skin cancer; however, the method relies on transfer learning and lacked generalizability for multiclass skin cancer classification, which is a more challenging problem in medical imaging. The Rahman et al. [15] study opted for an ensemble learning approach by combining multiple pretrained models, achieving a high sensitivity of 93.0%, but this approach increases the computational complexity and resource demand due to the integration of multiple deep learning models. Reis et al. [48] introduced the inSinEt model, which was tested on multiple datasets but performed best on a binary dataset and was not optimized for dealing with multiclass imbalanced data. Another study, by Iqbal et al. [13], presented the custom DCNN architecture for multiclass skin cancer classification, but this lacked an attention unit, which is good for highlighting relevant regions of cancer and to suppress irrelevant features like artifacts or noise present in the dataset. Without an attention unit, the model also faces challenges especially in subtle inter-class differences. In comparison, the proposed AIR-CNN deals efficiently with the intra-class differences and inter-class similarities by focusing more on the important and relevant features using the attention unit.

By achieving a superior accuracy of 91.63, the AIR-CNN proves its robustness for the multiclass classification of skin cancer. Precision and F1-score evaluation measures provide a detailed perspective on the model's ability to identify true positives. The introduced models perform excellently compared to state of the art because these models are keenly designed to deal with the challenges mentioned above. Table 9 presents the results of the suggested models with other studies and results of both proposed models are shown bold due to their superior performance.

Table 9. Comparison with state-of-the-art techniques using ISIC 2019 dataset.

Author	Accuracy	Precision	Sensitivity	Specificity	F1-Score	Parameters
Mijwil et al., 2021 [2]	86.90	87.47	86.14	87.66	-	**-**
Rahman et al., 2021 [15]	88.0	87.0	93.0	-	89.0	-
Reis et al., 2022 [48]	91.0	-	90.0	94.12	-	**-**
Iqbal et al., 2020 [13]	89.58	90.66	89.58	97.57	89.75	~256.7 M
Proposed (IR-CNN)	**91.53**	**91.6**	**91.5**	**91.5**	**91.6**	**~2.2 M**
Proposed (AIR-CNN)	**91.63**	**91.6**	**91.6**	**91.52**	**91.6**	**~1.1 M**

6. Conclusions and Future Directions

Skin cancer poses a significant worldwide health challenge, and the accurate classification of skin cancer is crucial for its initial detection. However, this task becomes particularly challenging due to the presence of similar features among different classes of skin cancer. Recent studies have introduced a range of automated deep learning-based techniques aimed at assisting dermatologists in the identification and classification of skin cancer, but most of them were primarily developed for binary-class skin cancer classification and faced challenges adapting to real-world clinical applications. The proposed solution has successfully tackled the challenges associated with multiclass skin cancer classification. This research developed a novel attention-based CNN network explicitly tailored for this complex task; within the model, the attention unit prioritizes pertinent features while effectively filtering out less significant ones. The incorporation of IR blocks in the network plays a vital role in addressing crucial aspects of skin cancer classification. These blocks excel in extracting multilevel features, enriching the model's capability to learn intricate patterns within the data and additionally enhancing the model's stability by effectively mitigating the vanishing gradient issue. Experimental results demonstrate the superior performance of the suggested model as compared to state-of-the-art methods. This superiority is evident in terms of both accuracy, with an impressive achievement of 91.63%, and efficiency, with a significantly reduced number of parameters to a lean ~1.1 M. In the future, the proposed method can also be used for other medical imaging problems such as lung cancer, pneumonia disease, and COVID-19 classification. The performance of the introduced model can be further improved by using a larger dataset or by synthetic image generation techniques such as GANS. Furthermore, the impact of using more than one attention unit in a deep learning model can also be investigated.

Author Contributions: S.Y., A.B.S., L.Y. and Z.H. collaborated in conceptualizing and designing the research. S.Y. led the methodology, executed the experiments, and prepared the manuscript. A.B.S.

and Z.H. contributed by analyzing the results and refining the architecture. 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: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

References

1. Adegun, A.; Viriri, S. Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art. *Artif. Intell. Rev.* **2021**, *54*, 811–841. [CrossRef]
2. Mijwil, M.M. Skin cancer disease images classification using deep learning solutions. *Multimed. Tools Appl.* **2021**, *80*, 26255–26271. [CrossRef]
3. Codella, N.C.F.; Nguyen, Q.B.; Pankanti, S.; Gutman, D.A.; Helba, B.; Halpern, A.C.; Smith, J.R. Deep Learning Ensembles for Melanoma Recognition in Dermoscopy Images. *IBM J. Res. Dev.* **2016**, *61*, 5:1–5:15. [CrossRef]
4. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef]
5. Joinpoint Trends in Cancer Incidence Rates for Selected Sites in Two Age Groups, US, 1995–2015 35 Figure S6. In *Trends in Cancer Death Rates for Selected Sites*; American Cancer Society: Atlanta, GA, USA, 2019.
6. Tan, T.Y.; Zhang, L.; Jiang, M. An intelligent decision support system for skin cancer detection from dermoscopic images. In Proceedings of the 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, China, 13–15 August 2016; pp. 2194–2199. [CrossRef]
7. Vijayalakshmi, M.M. Melanoma Skin Cancer Detection using Image Processing and Machine Learning. *Int. J. Trend Sci. Res. Dev.* **2019**, *3*, 780–784. [CrossRef]
8. Manne, R.; Kantheti, S.; Kantheti, S. Classification of Skin cancer using deep learning, Convolutional Neural Networks—Opportunities and vulnerabilities- A systematic Review. *Int. J. Mod. Trends Sci. Technol.* **2020**, *6*, 101–108. [CrossRef]
9. Hosny, K.M.; Kassem, M.A.; Fouad, M.M. Classification of Skin Lesions into Seven Classes Using Transfer Learning with AlexNet. *J. Digit. Imaging* **2020**, *33*, 1325–1334. [CrossRef] [PubMed]
10. Kassem, M.A.; Hosny, K.M.; Fouad, M.M. Skin Lesions Classification into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning. *IEEE Access* **2020**, *8*, 114822–114832. [CrossRef]
11. Waheed, Z.; Waheed, A.; Zafar, M.; Riaz, F. An efficient machine learning approach for the detection of melanoma using dermoscopic images. In Proceedings of the 2017 International Conference on Communication, Computing and Digital Systems (C-CODE), Islamabad, Pakistan, 8–9 March 2017; pp. 316–319. [CrossRef]
12. 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. [CrossRef]
13. Iqbal, I.; Younus, M.; Walayat, K.; Kakar, M.U.; Ma, J. Automated multi-class classification of skin lesions through deep convolutional neural network with dermoscopic images. *Comput. Med. Imaging Graph.* **2021**, *88*, 101843. [CrossRef]
14. Maron, R.C.; Schlager, J.G.; Haggenmüller, S.; von Kalle, C.; Utikal, J.S.; Meier, F.; Gellrich, F.F.; Hobelsberger, S.; Hauschild, A.; French, L.; et al. A benchmark for neural network robustness in skin cancer classification. *Eur. J. Cancer* **2021**, *155*, 191–199. [CrossRef] [PubMed]
15. Rahman, Z.; Hossain, M.S.; Islam, M.R.; Hasan, M.M.; Hridhee, R.A. An approach for multiclass skin lesion classification based on ensemble learning. *Inform. Med. Unlocked* **2021**, *25*, 100659. [CrossRef]
16. Ladd, M.E.; Bachert, P.; Meyerspeer, M.; Moser, E.; Nagel, A.M.; Norris, D.G.; Schmitter, S.; Speck, O.; Straub, S.; Zaiss, M. Pros and cons of ultra-high-field MRI/MRS for human application. *Prog. Nucl. Magn. Reson. Spectrosc.* **2018**, *109*, 1–50. [CrossRef] [PubMed]
17. Goyal, M.; Knackstedt, T.; Yan, S.; Hassanpour, S. Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities. *Comput. Biol. Med.* **2020**, *127*, 104065. [CrossRef]
18. Luu, T.N.; Phan, Q.H.; Le, T.H.; Pham, T.T.H. Classification of human skin cancer using Stokes-Mueller decomposition method and artificial intelligence models. *Optik* **2022**, *249*, 168239. [CrossRef]
19. Wang, J.; Zhu, H.; Wang, S.H.; Zhang, Y.D. A Review of Deep Learning on Medical Image Analysis. *Mob. Netw. Appl.* **2021**, *26*, 351–380. [CrossRef]

20. Alom, M.Z.; Aspiras, T.; Taha, T.M.; Asari, V.K. Skin Cancer Segmentation and Classification with NABLA-N and Inception Recurrent Residual Convolutional Networks. *arXiv* **2019**, arXiv:1904.11126.
21. Lakhani, P.; Gray, D.L.; Pett, C.R.; Nagy, P.; Shih, G. Hello World Deep Learning in Medical Imaging. *J. Digit. Imaging* **2018**, *31*, 283–289. [CrossRef]
22. Barata, C.; Celebi, M.E.; Marques, J.S. A Survey of Feature Extraction in Dermoscopy Image Analysis of Skin Cancer. *IEEE J. Biomed. Health Inform.* **2019**, *23*, 1096–1109. [CrossRef] [PubMed]
23. Hameed, N.; Shabut, A.M.; Ghosh, M.K.; Hossain, M.A. Multi-class multi-level classification algorithm for skin lesions classification using machine learning techniques. *Expert Syst. Appl.* **2020**, *141*, 112961. [CrossRef]
24. Thurnhofer-Hemsi, K.; Domínguez, E. A Convolutional Neural Network Framework for Accurate Skin Cancer Detection. *Neural Process. Lett.* **2021**, *53*, 3073–3093. [CrossRef]
25. Arora, G.; Dubey, A.K.; Jaffery, Z.A.; Rocha, A. A comparative study of fourteen deep learning networks for multi skin lesion classification (MSLC) on unbalanced data. *Neural Comput. Appl.* **2022**, *35*, 7989–8015. [CrossRef]
26. Chen, X.; Tao, H.; Zhou, H.; Zhou, P.; Deng, Y. Hierarchical and progressive learning with key point sensitive loss for sonar image classification. *Multimed. Syst.* **2024**, *30*, 380. [CrossRef]
27. Ju, C.; Bibaut, A.; van der Laan, M. The relative performance of ensemble methods with deep convolutional neural networks for image classification. *J. Appl. Stat.* **2018**, *45*, 2800–2818. [CrossRef]
28. Chaturvedi, S.S.; Tembhurne, J.V.; Diwan, T. A multi-class skin Cancer classification using deep convolutional neural networks. *Multimed. Tools Appl.* **2020**, *79*, 28477–28498. [CrossRef]
29. Gessert, N.; Nielsen, M.; Shaikh, M.; Werner, R.; Schlaefer, A. Skin lesion classification using ensembles of multi-resolution EfficientNets with meta data. *MethodsX* **2020**, *7*, 100864. [CrossRef]
30. Raza, R.; Zulfiqar, F.; Tariq, S.; Anwar, G.B.; Sargano, A.B.; Habib, Z. Melanoma classification from dermoscopy images using ensemble of convolutional neural networks. *Mathematics* **2022**, *10*, 26. [CrossRef]
31. Kaur, R.; Gholamhosseini, H.; Sinha, R.; Lindén, M. Melanoma Classification Using a Novel Deep Convolutional Neural Network with Dermoscopic Images. *Sensors* **2022**, *22*, 1134. [CrossRef]
32. Alzubaidi, L.; Al-Amidie, M.; Al-Asadi, A.; Humaidi, A.J.; Al-Shamma, O.; Fadhel, M.A.; Zhang, J.; Santamaría, J.; Duan, Y. Novel transfer learning approach for medical imaging with limited labeled data. *Cancers* **2021**, *13*, 1590. [CrossRef]
33. Datta, S.K.; Shaikh, M.A.; Srihari, S.N.; Gao, M. 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*; Springer: Cham, Switzerland, 2021; pp. 13–23. [CrossRef]
34. Hsu, B.W.Y.; Tseng, V.S. Hierarchy-aware contrastive learning with late fusion for skin lesion classification. *Comput. Methods Programs Biomed.* **2022**, *216*, 106666. [CrossRef]
35. Wang, Y.; Feng, Y.; Zhang, L.; Zhou, J.T.; Liu, Y.; Goh, R.S.M.; Zhen, L. Adversarial multimodal fusion with attention mechanism for skin lesion classification using clinical and dermoscopic images. *Med. Image Anal.* **2022**, *81*, 102535. [CrossRef] [PubMed]
36. Ajmal, M.; Khan, M.A.; Akram, T.; Alqahtani, A.; Alhaisoni, M.; Armghan, A.; Althubiti, S.A.; Alenezi, F. BF2SkNet: Best deep learning features fusion-assisted framework for multiclass skin lesion classification. *Neural Comput. Appl.* **2023**, *35*, 22115–22131. [CrossRef]
37. Tschandl, P.; Rosendahl, C.; Kittler, H. Data descriptor: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **2018**, *5*, 180161. [CrossRef]
38. 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). In Proceedings of the IEEE 15th International Symposium on Biomedical Imaging (ISBI), Washington, DC, USA, 4–7 April 2018; pp. 168–172. [CrossRef]
39. Attaran, M.; Deb, P. Machine Learning: The New 'Big Thing' for Competitive Advantage. *Int. J. Knowl. Eng. Data Min.* **2018**, *5*, 277–305. [CrossRef]
40. Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPhDW), Świnouście, Poland, 9–12 May 2018; pp. 117–122. [CrossRef]
41. Yang, X. An Overview of the Attention Mechanisms in Computer Vision. *J. Phys. Conf. Ser.* **2020**, *1693*, 012173. [CrossRef]
42. Alche, M.N.; Acevedo, D.; Mejail, M. EfficientARL: Improving skin cancer diagnoses by combining lightweight attention on EfficientNet. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 11–17 October 2021; pp. 3347–3353. [CrossRef]
43. Liu, L.; Mou, L.; Zhu, X.X.; Mandal, M. Automatic skin lesion classification based on mid-level feature learning. *Comput. Med. Imaging Graph.* **2020**, *84*, 101765. [CrossRef]
44. Mahbod, A.; Schaefer, G.; Ellinger, I.; Ecker, R.; Pitiot, A.; Wang, C. Fusing fine-tuned deep features for skin lesion classification. *Comput. Med. Imaging Graph.* **2019**, *71*, 19–29. [CrossRef] [PubMed]

45. Harangi, B. Skin lesion classification with ensembles of deep convolutional neural networks. *J. Biomed. Inform.* **2018**, *86*, 25–32. [CrossRef]
46. Kingma, D.P.; Ba, J.L. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015. [CrossRef]
47. Xu, B.; Wang, N.; Kong, H.; Chen, T.; Li, M. Empirical Evaluation of Rectified Activations in Convolutional Network. *arXiv* **2015**, arXiv:1505.00853. [CrossRef]
48. Reis, H.C.; Turk, V.; Khoshelham, K.; Kaya, S. InSiNet: A deep convolutional approach to skin cancer detection and segmentation. *Med. Biol. Eng. Comput.* **2022**, *60*, 643–662. [CrossRef]

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.