



Skin lesion classification on dermatoscopic images using effective data augmentation and pre-trained deep learning approach

Ferhat Bozkurt¹ 

Received: 11 May 2022 / Revised: 18 August 2022 / Accepted: 21 October 2022 /

Published online: 28 November 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Skin cancer is a severe disease that is common and causes death if left untreated. When skin cancer is detected early through dermatoscopic imaging, the possibility of definitive treatment is very high. Although melanoma is one of the fatal types of skin cancer, early detection dramatically increases the chances of survival. There is a low morbidity rate and limited actual data to study this deadly disease. This is a significant handicap in the application of machine learning techniques. Accurate diagnosis is essential because of the similarity of some types of lesions. The accuracy of the diagnosis is related to the professional experience of the specialist. The development of rapid and successful computerized diagnostic systems for the diagnosis and classification of skin cancer has become increasingly important. Deep learning-based applications are especially new trend in the detection of diseases from medical images. In this study, an effective data augmentation and a pre-trained deep learning approach are proposed for skin lesion classification. A hybrid network model called the Inception-Resnet-v2 is proposed to classify skin cancer images. The main aim of this study is to increase the number of images in the dataset by applying the affine transformation technique (data augmentation) and analyzing its effect on the skin cancer classification system. The highest reported accuracy in this study with an augmented dataset is 95.09% for the Inception-Resnet-v2 model while the same model achieved 83.59% with the original dataset.

Keywords Skin cancer · Dermatoscopic images · Classification · Deep learning · Data augmentation

1 Introduction

Skin cancer occurs due to the uncontrolled proliferation of DNA structures of damaged skin cells. It is among the most pervasive cancers in the world [40]. Skin cancer is the most widespread malignant disease among white people. However, the percentage of skin cancer

✉ Ferhat Bozkurt
fbozkurt@atauni.edu.tr

¹ Department of Computer Engineering, Faculty of Engineering, Ataturk University, 25240, Erzurum, Turkey

is increasing on a global scale [32]. Malignant skin cancer is one of the types of skin cancer that has a high mortality rate. About 55,000 deaths from melanoma are reported worldwide each year. This rate is equal to 0.7% of all cancer deaths. However, death rates vary widely from country to country. Due to excessive exposure to ultraviolet rays, the annual number of events of melanoma increased by 53% from 2008 to 2018 [14].

Although it is rare, it is the most malignant type of skin cancer and the number of deaths is increasing. The American Cancer Society reported that 100,350 new cases were detected in 2020. Also, estimated 6850 people died due to melanoma [44]. Although melanoma is one of the fatal kinds of skin cancer, early detection greatly increases the chances of survival. The low rate of disease results in limited real image data on this deadly disease. This is major handicap in the application of image processing and machine learning techniques. The first step in diagnosing a malignant lesion by a dermatologist is a visual inspection of the doubtful skin area. Accurate diagnosis is crucial because of the similarity of some lesion types, and diagnostic correctness correlates with the expert's professional experience [20]. When skin cancer is detected early, definitive treatment is highly likely. Optical methods are available for skin cancer scanning. These methods are superficial and give a quick response. Among the non-superficial methods, the most widely used is dermoscopic scanning. Dermoscopy is an imaging technique to obtain a magnified and illuminated image of the relevant area for accurate diagnosis of the stained area on the skin. Removing the surface reflection of the skin can improve the visual effect of deeper skin levels and give a more detailed view of skin lesions. In doubtful cases, visual inspection is assisted by dermoscopic images taken via magnifying and high-resolution cameras. In order to see the deeper skin layers, the illumination is controlled with a filter to minimize reflections on the skin during recording. Dermoscopy evaluation gives much higher accuracy than natural eye evaluation. Dermoscopy images are mostly analyzed by visual inspection. The correctness of skin lesion diagnosis can be improved thanks to this technical support [7, 8, 33, 47, 52]. The use of traditional methods like visual inspection, clinical scanning, biopsy, histopathological examination, and dermoscopic analysis of skin lesion need a high degree of skill, concentration, and time [1, 8, 52]. Even when the diagnosis of skin cancer is made by expert dermatologists, it can be erroneous due to factors such as different shapes, indistinct borders, low contrast, skin hairs, oils, and air bubbles in skin lesions. Under these circumstances, the development of rapid and high success rate computer-aided diagnostic systems for skin cancer detection and classification is becoming increasingly important. However, diagnostic accuracy can vary widely among professionals with different experiences. Consequently, there is a great interest in screening programs and the development of semi or fully automated computer-aided diagnostic systems that can be used as a second stand-alone opinion. Artificial intelligence models are the most used approaches in such computer-aided diagnosis systems [11, 12, 15]. Especially in 2012, the use of deep learning approaches in medical image classification has increased after the success of the model named AlexNet by Krizhevsky et al. [27] in the ImageNet 2012 competition.

Deep learning models have been frequently used in the classification of skin cancer images. Brinker et al. [6] split skin cancer images into two classes (Melanoma and Nevus). In the ResNet-50 model, instead of a constant learning rate, a distinct learning rate is used for each layer of the model. In addition, new methods based on the cosine function have been used to reduce learning rates. A sensitivity rate of 82.3% was achieved with this method. Hosny et al. [24] augmented the data set using the data augmentation technique for each image in the data set and used the transfer learning approach. With this approach, a classification accuracy of 95.91% was achieved. Esteva et al. [13] divided the dataset into two classes. In the preprocessing step, the images in the data set were processed

with Gaussian filtering. With the deep learning model called AdNet, 87.81% accuracy was achieved. Nugroho et al. [34] divided the HAM10000 skin cancer dataset into seven classes. The designed and trained Convolutional Neural Networks (CNN) model reached 78% classification accuracy. Alqudah et al. [5] used pre-trained AlexNet and GoogLeNet models to recognize three classes of skin cancer images. The data set is divided into two formats, the unsegmented data set and the segmented data set. The classification accuracy was 89.8% for the unsegmented dataset and 92.2% for the segmented dataset. Moataz et al. [30] proposed to use a pre-trained model with fine-tuning on Xception to recognize seven classes of skin cancer images. They executed data augmentation to improve the model performance. They used 30,294 images for training, 7,574 images for validation, and 7,714 images for testing. They obtained 96% of the average accuracy of classification for the augmented and balanced HAM10000 dataset. Chaturvedi et al. [10] proposed to use fine-tuning on Xception architecture for HAM10000 dataset (total 10,015 images, 8,912 for training and 1,103 for validation). They modified the Xception model that has a dense layer with 'Relu' activation, softmax layer (for seven classes), and Adam optimizer. The proposed method detected cancer with 91.47% accuracy. Aldwgeri and Abubacker [3] used and modified deep learning models (VGG16, VGG19, ResNet50, DenseNet121, InceptionV3, and Xception) to classify skin lesions for HAM 10,000 dataset (for unbalanced balanced and datasets). The highest reported accuracy was 80% for this ensemble model. Kassani et al. [26] studied different deep learning models to detect melanoma on the augmented HAM 10,000 dataset. They reported in this study that the highest accuracy was 92% with ResNet50 and 90% with the Xception model. Cengil et al. [9] used Alexnet and Resnet architectures and created hybrid architectures with these two models. Instead of Softmax classifier in the last layer, they used decision tree, kNN and SVM for classification with Alexnet and Resnet. The highest reported accuracy was 77.8% with Alexnet+SVM on HAM10000 dataset.

Literature review denoted that although there are different diagnosis and classification methods for skin cancer, there are still many gaps that need to be addressed, for instance, complex configuration, higher complexity of some studies, and less accuracy. It can be said that most of the skin lesion diagnosis systems in the literature give reasonable classification results to distinguish malignant melanoma from benign lesions. The performance of most machine learning techniques depends on the selected features characterizing the cancerous region and requires high computation time. Most of these studies were trained on a set of handcrafted features from images and used simple classifiers. With deep learning techniques and CNNs, impressive results have been achieved in image classification to perform skin lesion analysis and automatic diagnosis of cancer types. In skin lesion classification, transfer learning techniques were used to reduce computational and memory requirements, and data augmentation techniques were used to overcome the lack of data. Rather than training a CNN from scratch, which requires large amounts of data and cost of high computation, it is computationally efficient to use a pre-trained CNN architecture (e.g. AlexNet, DenseNet, Inception, and ResNet) and fine-tune its performance to speed up the process.

The scope and contributions of this study could be summarized as follows:

- In this study, effective data augmentation and a pre-trained deep learning approach are proposed for skin lesion classification.
- A hybrid network model called the Inception-Resnet-v2 is proposed to classify skin cancer images.
- By applying the affine transformation technique, the number of images in the dataset has increased to analyze its effect on the accuracy of skin cancer classification.

- Performance comparison of the proposed method with other pre-trained methods is performed on an augmented skin cancer dataset.

The rest of this study is carried out as follows. In Section 2, the material and method are presented. In this section, the studied original and augmented datasets, and pre-trained Inception-Resnet-v2 architecture are explained in detail. Experimental results are given in Section 3, and the discussion and conclusion part of this study is given in Section 4.

2 Material and method

Deep learning-based models have recently been performing above human-level accuracy in classification tasks [49]. There exist a great impact of hyper-parameter on the performance of these models. Furthermore, the size of the dataset on which deep models are trained has a great impact on performance.

In this study, effective data augmentation and a pre-trained deep learning approach are proposed for skin lesion classification. Figure 1 shows the general flowchart of system design. In this study, a hybrid network model called the Inception-Resnet-v2 is proposed to classify skin cancer images. We have increased the number of images in the dataset by applying the affine transformation technique and analyzing its effect on the skin cancer

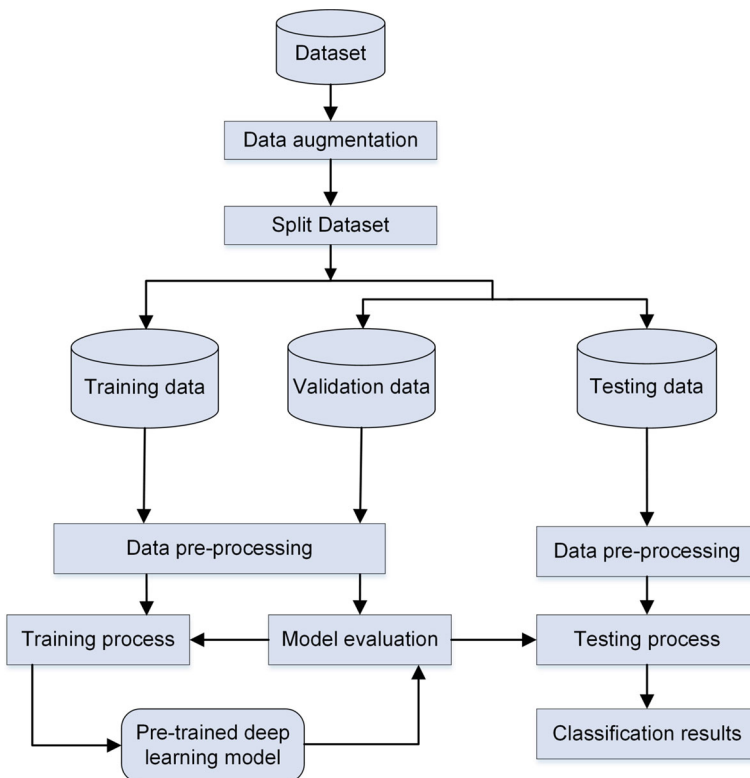


Fig. 1 General flowchart of system design

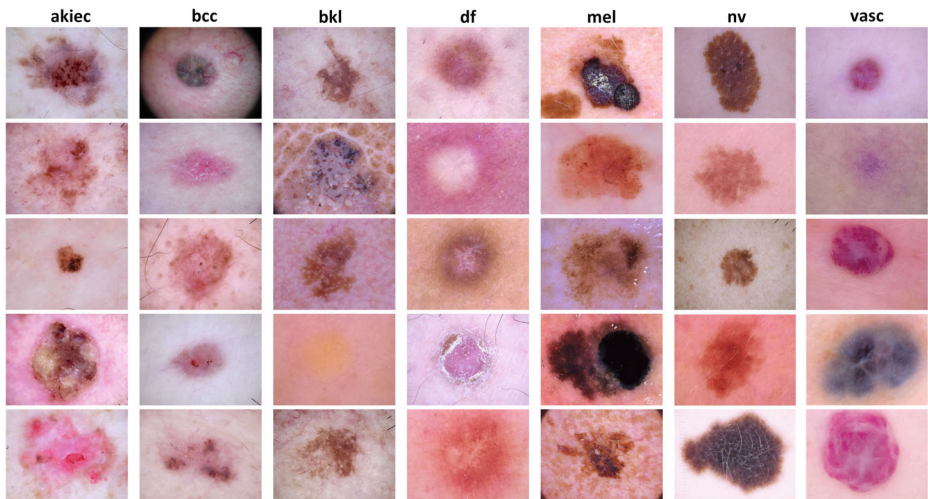


Fig. 2 Some example images from skin cancer dataset

classification system. The datasets used in this study are defined below as the original and augmented datasets. The original dataset contains images that have not applied any preprocessing. The augmented dataset consists of the images in the original dataset and the new images obtained by applying the affine transform to the related images.

2.1 Original dataset

In this study, the public skin cancer MNIST HAM10000 dataset [48] was used to classify skin cancer. The dataset consists of samples of pigmented lesions from distinct populations, as shown in Fig. 2. The classes and their types of data contained in the dataset are given in Table 1. As shown in Table 2, there exist images belonging to seven distinct classes (as shown in Table 1) in the dataset. It contains an extensive catalog of multi-source dermatoscopic images of pigmented injuries. In the original dataset, there are 10015 skin-threatening dermatoscopic images from different classes gathered from different sources. Furthermore, these images have a size of 600x450 in the RGB format. Then they are rescaled to 224x224 pixels for model.

Table 1 The classes of the skin cancer MNIST HAM10000 dataset

Class	Abbreviation	Type
akiec	Actinic keratoses	Benign or malignant
bcc	Basal cell carcinoma	Malignant
bkl	Benign keratosis-like lesions	Benign
df	Dermatofibroma	Benign
mel	Melanoma	Malignant
nv	Melanocytic nevi	Benign
vasc	Vascular lesions	Benign or malignant

Table 2 The number of data before and after data augmentation process

Class	Original images	Augmented images
akiec	327	3,269
bcc	514	5,135
bkl	1,099	10,987
df	115	1,150
mel	1,113	11,122
nv	6,705	6,705
vasc	142	1,419
Total	10,015	39,787

2.2 Augmented dataset

The biggest problem that machine and deep learning algorithms face is that there is not enough data to train the model. The lack of sufficient data creates an overfitting problem. This is a big problem that often occurs in these algorithms. This event causes the network to memorize the training data, and it fails when it encounters an input other than the training data. One of the most important methods of getting rid of this problem is data augmentation. This method is applied to the training set, and many images are obtained artificially by changing the properties of the available data [4].

The size of datasets affects deep learning and classification models. Creating a skin cancer dataset from scratch is a difficult and time-consuming problem. Also, there are more images in the dataset in some classes than in others, as shown in Table 2. Dealing with unbalanced data can lead to a lower performance of the minority class. This situation can lead to data misclassification in the most machine and deep learning approaches. The aim of our study is to generate new images from existing images with the affine transformation technique and analyze the effect of these images on skin cancer classification. At this stage, we have performed data augmentation to improve the performance of the Inception-ResNet model. The affine transformation could be brightness, rotation, shift, flip, and zoom for image data augmentation [19, 50]. In this section, random rotation augmentation technique is used to enlarge the size of the dataset.

Rotation augmentation is obtained by randomly rotating the image clockwise, a certain number of degrees from 0 to 360. Rotation returns pixels from the image frame, leaving the areas of the frame without pixel data that needs to be filled in. The figure shows random rotations applied to the image between 0 and 90 degrees. In this stage, each image in each class except *nv* class is randomly rotated nine times, as shown in Fig. 3. Therefore, when the original data was added to the image, the data of each class increased ten times, as shown in Table 2. At the end of this process, the number of images in the data set increased from 10,015 to 39,787 as shown in Table 2. Dataset image distribution before and after augmentation are also given in Tables 3 and 4. The dataset is split as 70% testing, 10% validation, and 20% testing as shown Tables 3 and 4.

2.3 Method

In this study, a hybrid network called the Inception-Resnet-v2 is proposed for skin lesion classification, which is composed of the Inception and the Residual modules. A pre-trained

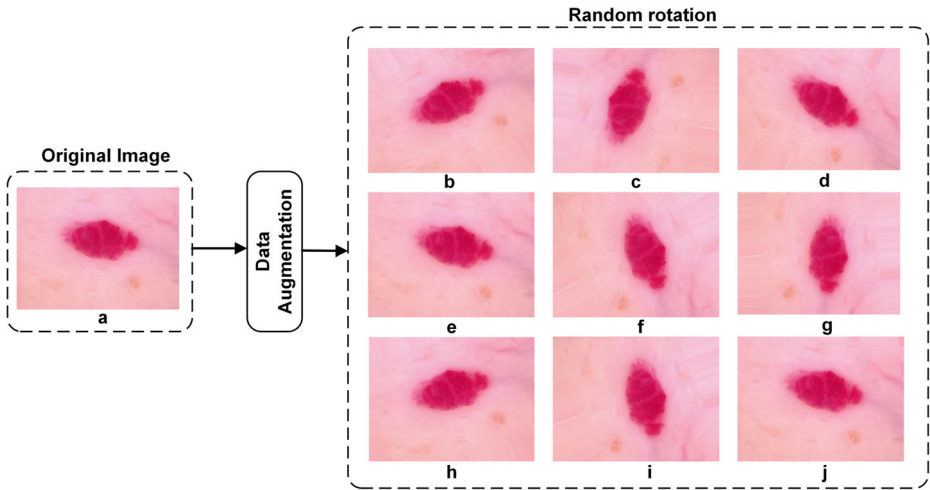


Fig. 3 A sample original image (a) from vasc class and randomly rotated nine images, (b-j) augmented images

model has already been trained on a dataset and includes biases and weights. The model represents the features of the dataset on which it was trained.

As shown in Fig. 4(a), the Inception network has many convolution kernels at different sizes to improve the adaptability of the network and extract many features of representations. Through the Inception network structure, the parameters of the model are reduced. Thus, the network does not lose the model feature representation. As a result, the number of convolution kernels is reduced as much as possible. Figure 4(b) shows a residual network structure. In order to make network training and parameter optimization quickly, signals of different units and layers can be transmitted directly forward and backward to any layer. The residuals are necessary for deep network to prevent the problem of degradation. Now we can say that using residual networks aids the network learn both depths and weights at the same time. We also provide that the new layer ($l+1$) learns something new by ensuring the output of the previous layer (l) without making any changes to the output of the current layer ($l+1$). Thus, this technique overcomes both degradation and vanishing problems in very deep networks. The number of feature maps of x_i may differ from that of the feature map in the residual convolution network, so it is required to use 1×1 convolution to increase or decrease the dimension. Meanwhile, the residual operation is stated by (1), (2), (3) as follows [36, 51]:

$$F(X_i) = X_i * w + \alpha \tag{1}$$

$$Y_i = R(F) + h(X_i) \tag{2}$$

Table 3 Data split and dataset image distribution before augmentation (original dataset)

	akiec	bcc	bkl	df	mel	nv	vasc
Training	228	359	769	80	779	4,693	99
Validation	33	52	110	12	111	671	14
Testing	66	103	220	23	223	1,341	29
Total	327	514	1,099	115	1,113	6,705	142

Table 4 Data split and dataset image distribution after augmentation (augmented dataset)

	akiec	bcc	bkl	df	mel	nv	vasc
Training	2,288	3,594	7,690	805	7,785	4,693	993
Validation	327	514	1,099	115	1,112	671	142
Testing	654	1,027	2,198	230	2,225	1,341	284
Total	3,269	5,135	10,987	1,150	11,122	6,705	1,419

$$X_{i+1} = R(Y_i) \tag{3}$$

$$R(z) = \max(0, z) \tag{4}$$

$$R(z) = \begin{cases} 0, & z < 0 \\ z, & z \geq 0 \end{cases} \tag{5}$$

In (1); X_i is the input; w is the weight; α is the offset; $F(X_i)$ points to the convolution operation. In (2); R is the ReLU function; $h(X_i)$ is a basic transformation for the X_i input; Y_i is the sum of two branches. In (3); X_{i+1} is the final output of the residual module. There exist many different activation functions that could be used. The three most common activation functions are the tanh, sigmoid, and rectified linear unit (ReLU) function. In this study, ReLU activation function is used as (4). Because, ReLU is simple and increases the nonlinearity and prevents network saturation. Especially, ReLU has a good effect because it removes vanishing gradients and is utilized in hidden layers. But, the weak point here is dead neurons. ReLU thresholds all negative values to zero, and its positive side has a fixed gradient of 1 as (5). While $z \geq 0$, $R(z) = z$, and its lead is 1; while $z < 0$, $R(z) = 0$, with a lead of 0. As a result, The ReLU will not be saturated on the positive side. However, the gradient of ReLU with respect to the input is zero on the negative side. This means that the gradient flow to the neurons will always be zero because the ReLU neuron starts producing a negative output. Therefore, due to the zero gradient, the weight of the neuron can never be optimized [51].

$$\frac{\partial X_n}{\partial X_i} = \frac{\partial X_i + F(X_i, \omega_i, \alpha_i)}{\partial X_i} = 1 + \frac{\partial F(X_n, \omega_n, \alpha_n)}{\partial X_n} \tag{6}$$

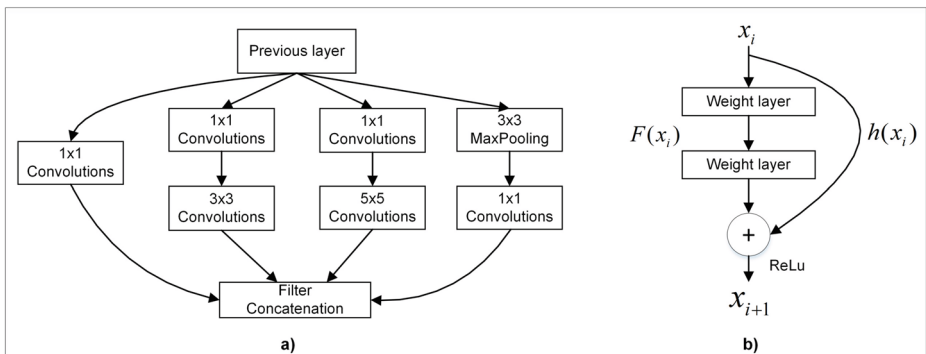


Fig. 4 a) An Inception building block [46] b) A residual building block [22]

The aim of utilizing a residual network learning unit is to prevent the problem of the gradient disappearing entirely while training the Inception network model. When the performance of the network model achieves an exact saturation, the residual network layer could be mapped in the same way. This enables the training network to converge faster and easier. From shallow i layer to deep n layer for the learning characteristics with (6), we understand that no matter how deep the network layers are, the gradient will never reach zero. In (6); X_i points the input of the i th residual unit, and X_n points the input of the n th unit, and $F(\cdot)$ is the residual function [51].

In this study, it is proposed to classify skin cancer images with Inception-ResNet-v2 architecture as shown in Fig. 5. Inception-ResNet-v2 is a model created by combining Inception and ResNet architectures with improved recognition and classification performance. The Inception and Residual modules use benefits of each other to enhance the detection accuracy and decrease the number of computations. Inception-ResNet-v2 has been trained on the ImageNet database with too many images and is a convolutional neural network (CNN). This CNN network has 164 layers and is able to split images into a thousand object categories like keyboard, pencil, mouse, and many animals [53]. Consequently, the network learned about the plentiful feature representations for a wide variety of images. ResNet and Inception give boosting performance in image recognition with low computational cost when it is compared to other models. ResNet architecture is about growing deep, while Inception is about growing wide. Therefore, with the Inception-ResNet-v2 architecture, we can achieve the optimum result in going both deep and wide. Inception-ResNet-v2 is a CNN algorithm based on the Inception architecture and includes residual connections. The connections now allow for shortcuts while the model is being trained. Thus researchers are able to set up deeper neural networks for better performance. This also provides a significant simplification of the initial blocks. This structure allows optimization of the residual layer by changing the size of the first convolution operation to 1×1 . The process of

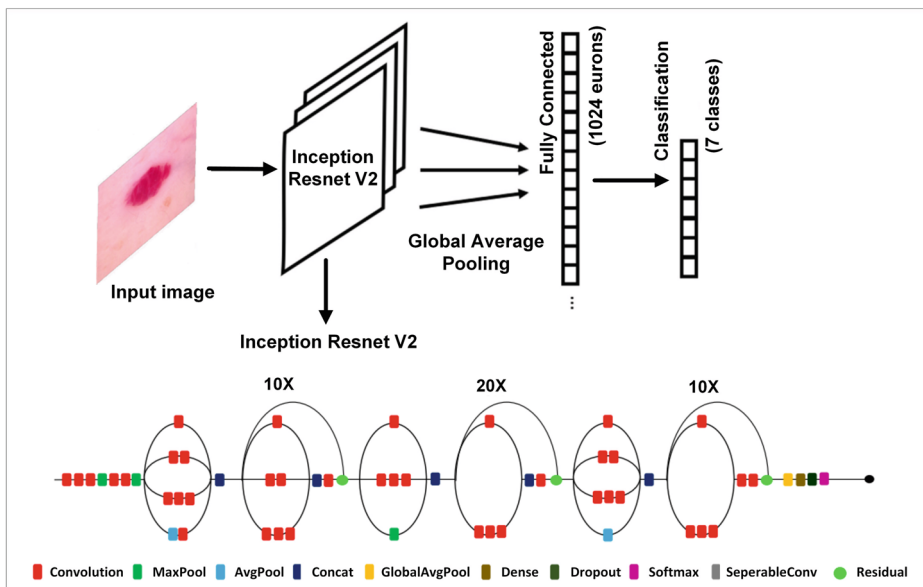


Fig. 5 Proposed Inception-Resnet-v2 architecture for skin lesion classification

transferring the previous activation value to the output also continues even if learning stops [17, 22].

$$f(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, \text{ for } i = 1 \dots k \text{ and } z = z_1 \dots z_k \quad (7)$$

In this study, Inception-ResNet-V2 model has 54,339,810 total parameters which are 54,279,266 of trainable and 60,544 of non-trainable. As shown in Fig. 5, the top layers of method contain a global average pooling layer, a FCL (fully-connected layer) that has 1024 neurons with ReLU activation function, and the end of the layer the neurons that provide classification in each of the seven classes with Softmax activation function. When there are multiple classes to be predicted, the Softmax activation function is usually used. For k classes, Softmax is calculated by (7). In this equation, \vec{z} is input vector to the Softmax function, all the z_i values are elements of the input vector, e^{z_i} standard exponential function of each element in the input vector, and $\sum_{j=1}^K e^{z_j}$ normalization term at the bottom of the formula [31].

Global average pooling is a process that calculates the average output of the feature map in the previous layer. There is a global average pooling layer that precedes the fully connected layer at the end of the network to obtain the deep features, as shown in Fig. 5. This fairly simple operation significantly reduces the data, preparing the model for the final classification layer. In the global average pooling process, overfitting is prevented by taking the average value of the feature map. A dropout layer was used to reduce overfitting during the training process. It is the elimination of some memorizing nodes in the network to prevent the memorization of the network. Thus, the memorization of the network is tried to be eliminated. The dropout layer is a flattening layer for fully connected layers. The dropout operation increases the neural network's ability to be flattened [28]. With the dropout process, a random zero weight value was assigned to the neurons in the network. The dropout ratio for this process was determined as 0.5. Thus, it is ensured that the model becomes resistant to small changes in the input and achieves a higher accuracy rate.

3 Experimental results

We compared the proposed Inception-Resnet-v2 method with the others, trying to find the best accuracy value according to different deep networks. In this study, we ran each method ten times on different training and test sets. Each method was run for 100 iterations in training. We performed all comparisons on the same machine and on the same data set. We recorded and compared the runtime of all methods in the experimental results. We used the classification report tool in the Python-Sklearn library to evaluate classification performance. In order to obtain the best accuracy in classification, we observed and compared the methods on the same machine by recording execution time under the same conditions. All experiments conducted in this study were developed using the Python 3.10.0 Jupyter Notebook development environment on a computer with an CPU with i7 (8700U) @ 3.20 GHz processor, 4 GB graphics card, and 16 GB primary memory hardware. In addition, the Keras and Tensorflow libraries are used.

In preprocessing stage, all images are uniformly rescaled to 224×224 to reduce the computational load. In experimental studies, the data is split into 70% training, 10% validation and 20% testing. We settled the best parameters through the quantitative experiments, and then skin lesion classification have been carried out according to those parameters. Our

main aim is to demonstrate the functionality of the proposed model on large skin cancer data set by comparing it with well-known pre-trained deep learning models. Model evaluations are carried out using a running average of the parameters calculated over time. In this study, in order to make an acceptable contrast between the various approaches to application configurations, we decided to adjust the basic parameters throughout all the studies. There are many hyperparameters that help to adjust the accuracy of the approximation. In this section, we have performed experimental setup throughout with model hyperparameters. We have tuned hyperparameters according to the accuracy of the model’s experimental results. We utilized ‘Adam’ as an adaptive optimizer. We can say that Adam is one of the time-efficient and important optimizers for deep networks. With learning rate = 0.0001, this optimizer uses ‘categorical cross-entropy’ as a loss function. We trained the models for 100 epochs with a 128 batch size. We used dropout (0.5) to generalize the network. Furthermore, tuning training parameters are; learning rate $\eta = e^{-5}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = e^{-8}$ dropout rate (0.5) and batch size (128) are set respectively. The momentum rate (0.9) and the weight decay parameters (e^{-5}) are set respectively. We set regularization parameter to be 0.0001 to prevent overfitting.

At this stage, the performance of the method was evaluated with different evaluation criteria. The performance of the method is measured using four performance criteria that are recall, precision, F1-score, and accuracy (Acc). The confusion matrix gives these values for each class (6=vasc, 5=nv, 4=mel, 3=df, 2=bkl, 1=bcc, 0=akiec). The performance of the method was evaluated according to the accuracy value calculated over the confusion matrices for the original and augmented datasets as shown in Fig. 6. For example, performance records of the Inception-Resnet-v2 model are given in Table 5 for the original dataset. The calculated scores for recall, F1-score, precision for each class, and obtained average results are presented in this table. According to this experimental study, the accuracy value obtained with the Inception-Resnet-v2 model is 83.59% for the original dataset. Similarly, performance scores of the Inception-Resnet-v2 model are given in Table 6 for the augmented dataset. Obtained accuracy value with the Inception-Resnet-v2 model is 95.09% for the augmented dataset. A comparison of overall accuracy rates from original and augmented datasets with Inception-Resnet-v2 model is also shown in Table 7.

In Fig. 7(a), graph of training/test accuracy and graph of training/test loss for 100 iterations of the InceptionResNetV2 model are given for the original dataset. Similarly in Fig. 7(b), the graph of training/test accuracy and the graph of training/test loss for 100

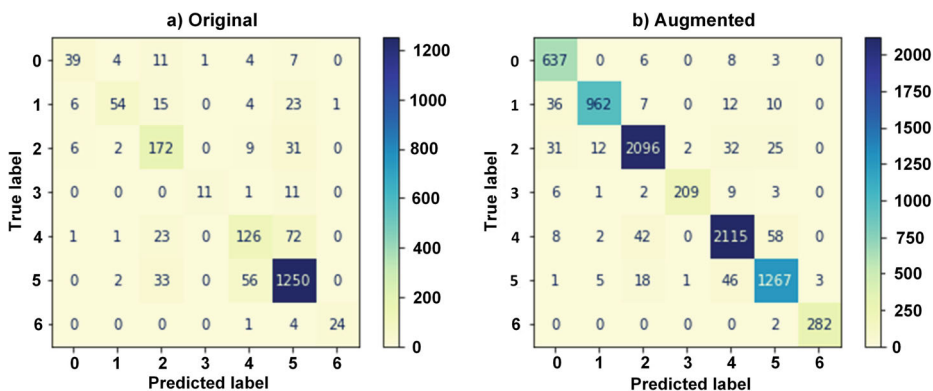


Fig. 6 Confusion matrices of the Inception-ResNet-v2 model from original and augmented datasets

Table 5 Performance scores of the Inception-Resnet-v2 model for original dataset

Accuracy=83.59%	Precision	Recall	F1-Score
akiec	0.7500	0.5909	0.6610
bcc	0.8571	0.5243	0.6506
bkl	0.6772	0.7818	0.7257
df	0.9167	0.4783	0.6286
mel	0.6269	0.5650	0.5943
nv	0.8941	0.9321	0.9127
vasc	0.9600	0.8276	0.8889
Average:	0.8117	0.6714	0.7231

iterations of the Inception-Resnet-v2 model are also given for the augmented dataset. Both the test and training accuracy curves indicate that when iteration number increases, learning occurs, and test accuracy gives successful results as shown in Fig. 7(a) and (b). Both the test and training loss curves indicate that when iteration number increases, learning occurs, and test error rate decreases as shown in Fig. 7 (a) and (b).

In addition, the performance comparison of the proposed InceptionResNetV2 model with other pre-trained models in terms of accuracy was performed in this study. Table 8 shows the performance comparison of the different pre-trained methods on the augmented skin cancer dataset. These pre-trained methods are VGG16, VGG19, SqueezeNet, LeNet-5, AlexNet, and an established deep CNN model. The deep CNN model consists of four sequential convolution pooling layers, one flatten layer, three fully connected layers, and a softmax classifier. We compared the proposed InceptionResNetV2 method in this study with the others, trying to find the best accuracy value according to different deep networks and trainable parameters. Other studied existing pre-trained models in experimental studies with low accuracy are not included in this table. The proposed InceptionResNetV2 model achieved the highest accuracy with 95.09% among all other methods as shown in Table 8 and the boxplot graph in Fig. 8. Performance comparison of these methods is also shown by using boxplots, as shown in Fig. 8. In this study, we ran each method ten times on different training and test sets. Each method was run for 100 iterations in training. The average accuracy value of each method was calculated. The boxplot graph was formed with the min-max and average values of each method.

In addition, the execution times of different pre-trained methods are analyzed by taking into account the training time over approximately 30k images for 100 iterations.

Table 6 Performance scores of the Inception-Resnet-v2 model for augmented dataset

Accuracy=95.09%	Precision	Recall	F1-Score
akiec	0.8860	0.9740	0.9279
bcc	0.9796	0.9367	0.9577
bkl	0.9655	0.9536	0.9595
df	0.9858	0.9087	0.9457
mel	0.9518	0.9506	0.9512
nv	0.9262	0.9448	0.9354
vasc	0.9895	0.9930	0.9912
Average:	0.9549	0.9516	0.9527

Table 7 Comparison of overall accuracy rates from original and augmented datasets with Inception-Resnet-v2 model

Dataset	# of images	Accuracy (%)
Original	10,015	83,59
Augmented	39,787	95,09

Depending on the complex configuration of the methods, the depth of the network and the number of trainable parameters; high computation times can be obtained. The execution time of the InceptionResNetV2 model is approximately recorded as 1h 50 min 4 sec as shown in Table 8. According to the running analysis, the lowest execution time was obtained during the training period with the InceptionResNetV2 model, and the lowest system response in the test evaluations was obtained with this model. Furthermore, Table 9 shows the performance comparison of similar studies on the MNIST HAM10000 dataset recently. By researching the best practice methods in the literature, we applied an effective data augmentation and a pre-trained deep learning approach to get a high classification accuracy. The classification accuracy for the augmented dataset is achieved to 95.09% by effective data augmentation and pre-trained deep learning approach. In comparison, the InceptionResNetV2 model gives 83.59% accuracy with the original dataset, as shown in Table 9.

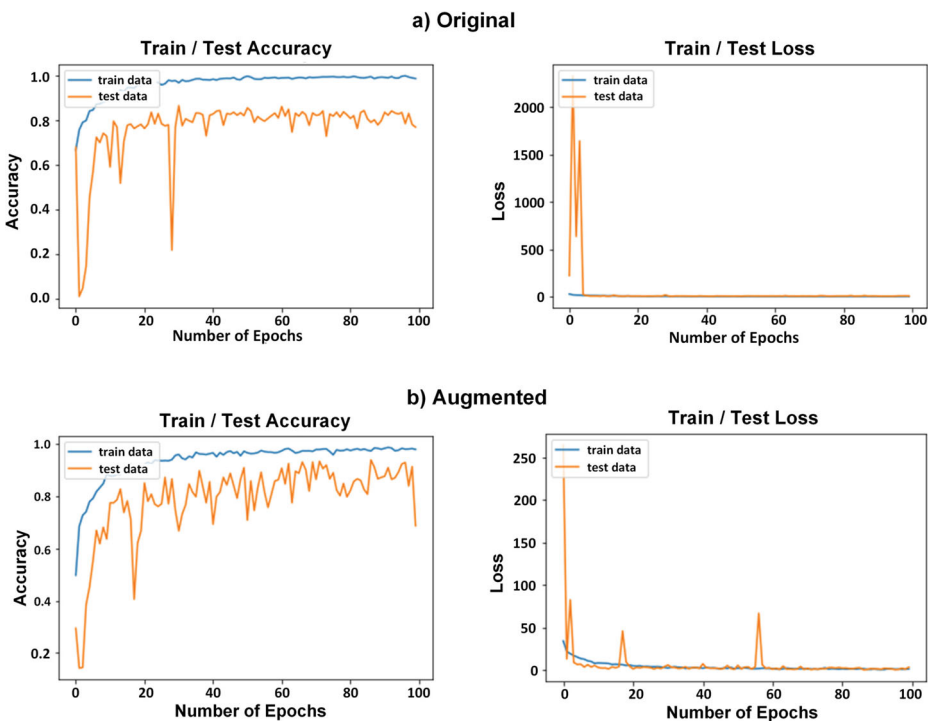


Fig. 7 Accuracy and loss graphs of the Inception-Resnet-v2 model for (a) original dataset and (b) augmented dataset

Table 8 Performance comparison of the different pre-trained methods on augmented skin cancer dataset

Method	# of Trainable params	Precision	Recall	F1-score	Acc (%)	Training time
VGG19	66,567	0.6354	0.5417	0.5638	61,55	6h 10min 16s
VGG16	66,567	0.6420	0.5824	0.5912	62,96	6h 29min 35s
SqueezeNet	119,399	0.6737	0.6332	0.6141	67,41	8h 54min 40s
LeNet-5	5,407,031	0.6963	0.6970	0.6867	69,33	2h 52min 52s
AlexNet	50,851,015	0.7331	0.7227	0.7253	72,43	1h 58min 20s
Deep CNN	274,258,951	0.8559	0.8585	0.8565	85,51	7h 49min 42s
InceptionResNetV2	55,857,255	0.9549	0.9516	0.9527	95,09	1h 50min 04s

4 Discussion and conclusion

Skin cancer is a serious disease that is common and causes death if left untreated. If skin cancer is not diagnosed early, it can lead to fatal cases. Dermatoscopic images are of great importance in the early diagnosis of skin cancer. When skin cancer is detected early from dermatoscopic images, definitive treatment is highly likely. The low rate of disease results in limited real image data on this deadly disease. This is a significant handicap in the application of deep learning techniques. Thus, we have increased total number of images in the dataset using by the data augmentation technique. Deep learning-based models have recently been performing above human-level accuracy in classification tasks. There exist a significant impact of hyper-parameter on the performance of models. Furthermore, the size of the dataset on which deep models are trained has a significant impact on performance. The biggest problem encountered in the machine and deep learning algorithms is that there is not enough data to train the model. The lack of enough data creates the overfitting problem, a big problem that frequently occurs in these algorithms. This event causes the network to memorize the training data, and fails when it encounters an input other than the training data.

In this study, data augmentation is carried out to the training set, and more images are obtained artificially by changing the properties of the available data. In this context, effective data augmentation and pre-trained deep learning approach are proposed for skin lesion classification. A hybrid network model called the Inception-Resnet-v2 is proposed

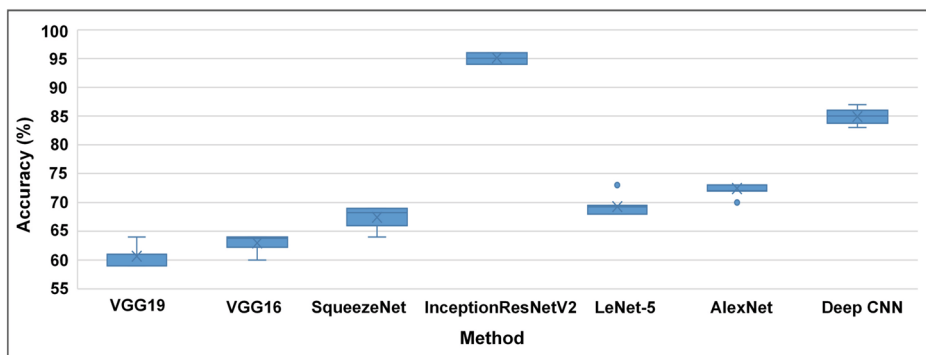
**Fig. 8** Boxplot graph of studied pre-trained methods for accuracy

Table 9 Performance comparison of similar studies in the literature on MNIST HAM10000 dataset

Study	Year	Method	# of images	Acc (%)
Purnama et al. [37]	2019	Inception V3	10,015	72
Nugroho et al. [34]	2019	CNN	10,015	78
Pai & Giridharan [35]	2019	CNN	10,015	79.56
Kassani & Kassani [26]	2019	Augmentation + ResNet50	39,060	92.08
Garg et al. [18]	2020	ResNet	10,015	90.51
Chaturvedi et al. [10]	2020	InceptionResNetV2 + ResNeXt101	10,015	92.83
Ratul et al. [39]	2020	InceptionV3	10,015	89.81
Moataz et al. [30]	2021	Fine-tuned xception model	30,294	96
Cengil et al. [9]	2021	Alexnet+SVM	10,015	77.80
Salamaa & Aly [42]	2021	ResNet50+ Preprocessing	10,015	93.98
Salamaa & Aly [42]	2021	ResNet50+ Preprocessing+SVM	10,015	98.43
Hameed et al. [21]	2021	Stacked CNN	36,800	92.5
Ramachandro et al. [38]	2021	Dense net	10,015	95
Kalaivani et al. [25]	2021	Fine-Tuned VGG-CNN	10,015	92
Mehra et al. [29]	2021	ResNet-50	10,015	84.87
Srinivasu et al. [45]	2021	MobileNet V2-LSTM	10,015	85.34
Salma & Eltrass [43]	2022	CNN	1400	98.79
Hoang et al. [23]	2022	Lightweight deep learning	10,015	86.33
This study	2022	InceptionResNetV2	10,015	83.59
This study	2022	Augmentation+InceptionResNetV2	39,787	95.09

to classify skin cancer images. The aim is to increase the number of images in the dataset by applying the affine transformation technique and analyzing its effect on the skin cancer classification system. Creating a skin cancer dataset from scratch is a complex and time-consuming problem and many datasets are unbalanced. Dealing with unbalanced data can cause lower performance of the minority class. This situation can lead to data misclassification in most deep and machine learning approaches. Our study aims to generate new images from existing images with the affine transformation technique and analyze the effect of these images on skin cancer classification. Thus, we have performed data augmentation to improve the performance of the Inception-ResNet model. ResNet and Inception give boosting performance in image recognition with low computational cost when it is compared to other models. ResNet architecture is about growing deep, while Inception is about growing wide. Therefore, with the Inception-ResNet-v2 architecture, we can achieve the optimum result in going both deep and wide. The highest reported accuracy in this study with an augmented dataset is 95.09% for the Inception-Resnet-v2 model while the same model achieved 83.59% with the original dataset.

Although these pre-trained deep learning models can be utilized to solve many important problems, their usage is still seriously criticized. Since it is extremely difficult to determine which data descriptors are the most sufficient to represent a particular phenomenon of special interest. Classifiers could be unsuccessful when there are too many variables and there exists a high correlation relationship between these variables. At this stage, vectors belonging to the representation set are kept at lower dimensions and the number of random variables is reduced by dimension reduction techniques. Dimension reduction is a

preprocessing step in machine learning to eliminate unwanted features and improve learning accuracy. There are methods of data representation, each of which has its own advantages to reduce redundancy characteristics. In addition, imbalance data and high dimensionality common problems in pattern recognition and machine learning. Imbalance is a major problem in classification, and this process becomes more complex when the dataset has numerous features. For attribute selection, the traditional classification generally prefers the majority class. This situation leads to poor performance for parameter setting or the selection of attributes that better define the majority class [2, 16, 41]. In order to solve the imbalance problem, in data-driven methods, the expected balance is tried to be achieved by reducing the majority class data. Another data-driven method is to generate data from the minority class distribution [54]. In this context, data from the minority class (for example akiec, bcc, df, vasc) has been augmented by using the data-driven method in this study. When the studies on this topic are examined, Rocchetti et al. [41] modified the training strategy by re-evaluating categorical data in the light of the Pareto analysis approach. They have developed a tool that gives a new shape to the dataset based on the Pareto rule. In this way, they used these categorical descriptors as a tool, not as an input, to train their deep learning model. With this data arrangement, they developed a more efficient deep learning model. Akram et al. [2] proposed a new framework for classification of skin lesion that incorporates in-depth feature information to generate the best distinguishing feature vectors while maintaining the original feature domain. To select distinctive features and reduce dimensionality, they used entropy-controlled neighbor component analysis. To test the success of the method, they examined the accuracy success with different classifiers. Fattahi et al. [16] proposed a hybrid method that performs the process of feature extracting and selecting concurrently to reduce the data dimensionality in the shape of a cost-sensitive optimization problem.

As a result, our main aim is to demonstrate the functionality of the proposed model on a large skin cancer data set by comparing it with well-known pre-trained deep learning models. This model assembles the advantages of Inception and Residual module, expanding the network width and lightening the training problem of the deep network. Since, these modules can benefit from each other to increase accuracy of detection and reduce the total number of calculations. Residual connections have been observed to significantly increase the training speed of the Inception architecture. We have achieved high accuracy by building both deep and wide networks through the proposed model on the augmented dataset. To the best of our knowledge, there exist limited studies that combine effective data augmentation and Inception-Resnet-v2 model to increase the accuracy of skin lesion classification and compare it with other pre-trained deep learning models. In further studies, we will try to construct a cost-sensitive model of misclassification of minority class data by proposing a new cost-sensitive or model-based method. We will also expand this work on effective dimension reduction on high-dimensional data. In addition, other deep learning models and hybrid methods will be studied, and comparisons will be made on the augmented datasets, which include more affine transformation techniques.

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

Declarations

Conflict of Interests The author has declared that no conflict of interests or competing interests exist.

References

1. Abbas Q, Celebi ME, Serrano C, Garcia IF, Ma G (2013) Pattern classification of dermoscopy images: a perceptually uniform model. *Pattern Recognit* 46(1):86–97. <https://doi.org/10.1016/j.patcog.2012.07.027>
2. Akram T, Lodhi HMJ, Naqvi SR, Naeem S, Alhaisoni M, Ali M, Qadri NN (2020) A multilevel features selection framework for skin lesion classification. *HCIS* 10(1):1–26. <https://doi.org/10.1186/s13673-020-00216-y>
3. Aldwgeri A, Abubacker NF (2019) Ensemble of deep convolutional neural network for skin lesion classification in dermoscopy images. In: *International visual informatics conference*. Springer, Cham, pp 214–226. https://doi.org/10.1007/978-3-030-34032-2_20
4. Ali MS, Miah MS, Haque J, Rahman M, Islam MK (2021) An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models. *Mach Learn Appl* 5:100036. <https://doi.org/10.1016/j.mlwa.2021.100036>
5. Alqudah AM, Alquraan H, Qasmieh IA (2019) Segmented and non-segmented skin lesions classification using transfer learning and adaptive moment learning rate technique using pretrained convolutional neural network. In: *Journal of biomimetics, biomaterials and biomedical engineering*, vol 42. Trans Tech Publications Ltd, pp 67–78
6. Brinker TJ, Hekler A, Enk AH, Berking C, Haferkamp S, Hauschild A, Utikal JS (2019) Deep neural networks are superior to dermatologists in melanoma image classification. *Eur J Cancer* 119:11–17. <https://doi.org/10.1016/j.ejca.2019.05.023>
7. Capdehourat G, Corez A, Bazzano A, Alonso R, Musé P (2011) Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions. *Pattern Recogn Lett* 32(16):2187–2196. <https://doi.org/10.1016/j.patrec.2011.06.015>
8. Celebi ME, Iyatomi H, Stoecker WV, Moss RH, Rabinovitz HS, Argenziano G, Soyer HP (2008) Automatic detection of blue-white veil and related structures in dermoscopy images. *Comput Med Imaging Graph* 32(8):670–677. <https://doi.org/10.1016/j.compmedimag.2008.08.003>
9. Cengil E, Çınar A, Yıldırım M (2021) Hybrid convolutional neural network architectures for skin cancer classification. *Avrupa Bilim ve Teknoloji Dergisi, Ejosat Special Issue 2021 (ICAENS)*, pp 694–701. <https://doi.org/10.31590/ejosat.1010266>
10. Chaturvedi S, Tembhurne JV, Diwan T (2020) A multi-class skin Cancer classification using deep convolutional neural networks. *Multimed Tools Appl* 79(39):28477–28498. <https://doi.org/10.1007/s11042-020-09388-2>
11. Demir F (2021) Derin Öğrenme Tabanlı Yaklaşımla Kötü Huyulu Deri Kanserinin Dermatoskopik Görüntülerden Saptanması. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi* 33(2):617–624. <https://doi.org/10.35234/fumbd.900170>
12. Ergün E, Kılıç K (2021) Derin Öğrenme ile Artırılmış Görüntü Seti üzerinden Cilt Kanseri Tespiti. *Black Sea J Eng Sci* 4(4):192–200. <https://doi.org/10.34248/bsengineering.938520>
13. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542(7639):115–118. <https://doi.org/10.1038/nature21056>
14. Fabbrocini G, Triassi M, Mauriello MC, Torre G, Annunziata MC, De Vita V, Monfrecola G (2010) Epidemiology of skin cancer: role of some environmental factors. *Cancers* 2(4):1980–1989. <https://doi.org/10.3390/cancers2041980>
15. Fabbrocini G, De Vita V, Pastore F, D'Arco V, Mazzella C, Annunziata MC, Monfrecola A (2011) Teledermatology: from prevention to diagnosis of nonmelanoma and melanoma skin cancer. *Int J Telemedicine Appl* 2011. <https://doi.org/10.1155/2011/125762>
16. Fattahi M, Moattar MH, Forghani Y (2022) Improved cost-sensitive representation of data for solving the imbalanced big data classification problem. *Journal of Big Data* 9(1):1–24. <https://doi.org/10.1186/s40537-022-00617-z>
17. Ferreira CA et al (2018) Classification of breast cancer histology images through transfer learning using a pre-trained inception Resnet V2. In: *Campilho A, Karray F, ter Haar Romeny B (eds) Image analysis and recognition. ICIAR 2018. Lecture notes in computer science()*, vol 10882. Springer, Cham. https://doi.org/10.1007/978-3-319-93000-8_86
18. Garg R, Maheshwari S, Shukla A (2021) Decision support system for detection and classification of skin cancer using CNN. In: *Innovations in computational intelligence and computer vision*. Springer, Singapore, pp 578–586. https://doi.org/10.1007/978-981-15-6067-5_65
19. Gocer E (2020) Image augmentation for deep learning based lesion classification from skin images. In: *2020 IEEE 4th international conference on image processing, applications and systems (IPAS)*. IEEE, pp 144–148. <https://doi.org/10.1109/IPAS50080.2020.9334937>

20. Haenssle HA, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Zalaudek I (2018) Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Ann Oncol* 29(8):1836–1842. <https://doi.org/10.1093/annonc/mdy166>
21. Hameed A, Umer M, Hafeez U et al (2021) Skin lesion classification in dermoscopic images using stacked convolutional neural network. *J Ambient Intell Human Comput*. <https://doi.org/10.1007/s12652-021-03485-2>
22. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 770–778
23. Hoang L, Lee SH, Lee EJ, Kwon KR (2022) Multiclass skin lesion classification using a novel lightweight deep learning framework for smart healthcare. *Appl Sci* 12(5):2677. <https://doi.org/10.3390/app12052677>
24. Hosny KM, Kassem MA, Foad MM (2019) Classification of skin lesions using transfer learning and augmentation with alex-net. *PLoS one* 14(5):e0217293. <https://doi.org/10.1371/journal.pone.0217293>
25. Kalaivani A, Karpagavalli S, Bibi MJ (2021) A deep learning approach for real-time defect classification in skin disease. *New Arch-International Journal of Contemporary Architecture* 8(2):443–451
26. Kassani SH, Kassani PH (2019) A comparative study of deep learning architectures on melanoma detection. *Tissue Cell* 58:76–83. <https://doi.org/10.1016/j.tice.2019.04.009>
27. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* 25
28. Li Y, Wang K (2020) Modified convolutional neural network with global average pooling for intelligent fault diagnosis of industrial gearbox. *Eksploracja i Niezawodność* 22(1). <https://doi.org/10.17531/ein.2020.1.8>
29. Mehra A, Bhati A, Kumar A, Malhotra R (2021) Skin cancer classification through transfer learning using ResNet-50. In: *Emerging technologies in data mining and information security*. Springer, Singapore, pp 55–62. https://doi.org/10.1007/978-981-33-4367-2_6
30. Moataz L, Salama GI, Abd Elazeem MH (2021) Skin cancer diseases classification using deep convolutional neural network with transfer learning model. In: *Journal of physics: conference series*, vol 2128, no 1. IOP Publishing, p 012013
31. Mohbey K (2020) Multi-class approach for user behavior prediction using deep learning framework on twitter election dataset. *J Data Inf Manag* 2(1):1–14. <https://doi.org/10.1007/s42488-019-00013-y>
32. Nami N, Giannini E, Burroni M, Fimiani M, Rubegni P (2012) Teledermatology: state-of-the-art and future perspectives. *Expert Rev Dermatol* 7(1):1–3. <https://doi.org/10.1586/edm.11.79>
33. Narayanamurthy V, Padmapriya P, Noorasafirin A, Pooja B, Hema K, Nithyakalyani K, Samhuri F (2018) Skin cancer detection using non-invasive techniques. *RSC Adv* 8(49):28095–28130. <https://doi.org/10.1039/C8RA04164D>
34. Nugroho AA, Slamet I, Sugiyanto (2019) Skin cancer identification system of HAM10000 skin cancer dataset using convolutional neural network. In: *AIP conference proceedings*, vol 2202, no 1. AIP Publishing LLC, p 020039. <https://doi.org/10.1063/1.5141652>
35. Pai K, Giridharan A (2019) Convolutional neural networks for classifying skin lesions. In: *TENCON 2019-2019 IEEE region 10 conference (TENCON)*. IEEE, pp 1794–1796. <https://doi.org/10.1109/TENCON.2019.8929461>
36. Pouyanfar S, Chen SC, Shyu ML (2017) An efficient deep residual-inception network for multimedia classification. In: *2017 IEEE international conference on multimedia and expo (ICME)*. IEEE, pp 373–378. <https://doi.org/10.1109/ICME.2017.8019447>
37. Purnama IKE et al (2019) Disease classification based on dermoscopic skin images using convolutional neural network in teledermatology system. In: *2019 international conference on computer engineering, network, and intelligent multimedia (CENIM)*, pp 1–5. <https://doi.org/10.1109/CENIM48368.2019.8973303>
38. Ramachandro M, Daniya T, Saritha B (2021) Skin cancer detection using machine learning algorithms. In: *2021 innovations in power and advanced computing technologies (i-PACT)*. IEEE, pp 1–7. <https://doi.org/10.1109/i-PACT52855.2021.9696874>
39. Ratul MAR, Mozaffari MH, Lee WS, Parimbelli E (2020) Skin lesions classification using deep learning based on dilated convolution. *BioRxiv*, p 860700. <https://doi.org/10.1101/860700>
40. Rey-Barroso L, Peña-Gutiérrez S, Yáñez C, Burgos-Fernández FJ, Vilaseca M, Royo S (2021) Optical technologies for the improvement of skin cancer diagnosis: a review. *Sensors* 21(1):252. <https://doi.org/10.3390/s21010252>
41. Rocchetti M, Delnevo G, Casini L, Mirri S (2021) An alternative approach to dimension reduction for pareto distributed data: a case study. *Journal of Big Data* 8(1):1–23. <https://doi.org/10.1186/s40537-021-00428-8>

42. Salamaa WM, Aly MH (2021) Deep learning design for benign and malignant classification of skin lesions: a new approach. *Multimed Tools Appl* 80:26795–26811. <https://doi.org/10.1007/s11042-021-11000-0>
43. Salma W, Eltrass AS (2022) Automated deep learning approach for classification of malignant melanoma and benign skin lesions. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-022-13081-x>
44. Siegel RL, Miller KD, Fuchs HE, Jemal A (2022) Cancer statistics, 2022. *CA: a cancer journal for clinicians*. <https://doi.org/10.3322/caac.21708>
45. Srinivasu PN, SivaSai JG, Ijaz MF, Bhoi AK, Kim W, Kang JJ (2021) Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors* 21(8):2852. <https://doi.org/10.3390/s21082852>
46. Szegegy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Rabinovich A (2015) Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1–9
47. Thomas L, Puig S (2017) Dermoscopy, digital dermoscopy and other diagnostic tools in the early detection of melanoma and follow-up of high-risk skin cancer patients. *Acta Derm Venereol* 97. <https://doi.org/10.2340/00015555-2719>
48. Tschandl P, Rosendahl C, Kittler H (2018) The HAM 10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci Data* 5:180161. <https://doi.org/10.1038/sdata.2018.161>
49. Ural A, Kilimci Z.H (2021) The prediction of chiral metamaterial resonance using convolutional neural networks and conventional machine learning algorithms. *Int J Comput Exp Sci Eng* 7(3):156–163. <https://doi.org/10.22399/ijcesen.973726>
50. Verma R, Kumar N, Patil A, Kurian NC, Rane S, Graham S, Sethi A (2021) MoNuSAC2020: a multi-organ nuclei segmentation and classification challenge. *IEEE Trans Med Imaging* 40(12):3413–3423. <https://doi.org/10.1109/TMI.2021.3085712>
51. Wang J, He X, Faming S, Lu G, Cong H, Jiang Q (2021) A real-time bridge crack detection method based on an improved inception-resnet-v2 structure. *IEEE Access* 9:93209–93223. <https://doi.org/10.1109/ACCESS.2021.3093210>
52. Xie Y, Zhang J, Xia Y, Shen C (2020) A mutual bootstrapping model for automated skin lesion segmentation and classification. *IEEE Trans Med Imaging* 39(7):2482–2493. <https://doi.org/10.1109/TMI.2020.2972964>
53. Zhang Y, Davison BD (2019) Modified distribution alignment for domain adaptation with pre-trained inception ResNet. <https://doi.org/10.48550/arXiv:1904.02322>
54. Zunair H, Hamza AB (2020) Melanoma detection using adversarial training and deep transfer learning. *Phys Med Biol* 65(13):135005. <https://doi.org/10.1088/1361-6560/ab86d3>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.