NTU Journal of Engineering and Technology (2023) 2 (2) : 1-6
DOI: https://doi.org/10.56286/ntujet.v2i2.532

IRAQI Academic Scientific Journals

P-ISSN: 2788-9971 E-ISSN: 2788-998X
NTU Journal of Engineering and Technology
Available online at: https://journals.ntu.edu.iq/index.php/NTU-JET/index

Skin Cancer Classification: A Transfer Learning Approach Using Inception-v3

1st Yaarob Younus Al Badrani[1], 2nd Abbas Mgharbel[2]
1. Computer Systems Department, Northern Technical University, Mosul, Iraq, E-mail: yaarobalbdrany@ntu.edu.iq
2. Lebanese University · Faculty Of Science,E-mail: mgharbel.abbas@gmail.com

Article Informations

Received: 13-06- 2023,
Revised :20-09-2023
Accepted: 01-10-2023,
Published Online: 17-10-2023

Corresponding author:
Name:yaarob younus albdrany
Affiliation : Northern Technical University
Email: yaarobalbdrany@ntu.edu.iq

Key Words:
Skin cancer classification,
deep learning,
Inception-v3,
Transfer learning,
skin lesions.

A B S T R A C T

In the human body, the skin serves as the primary layer of defense for essential organs. However, as a result of ozone layer degradation, exposure to UV radiation, fungal and viral infections. Skin cancer is becoming more common .
This study proposes a novel deep learning-based framework for the multi-classification of eight different types of skin cancer. The suggested framework is divided into several steps. The initial phase is the data augmentation of images. In the second step, deep models are fine-tuned. The model is opted, for Inception-v3, and updated their layers. In the third step, The suggested model has been applied to train both fine-tuned on augmented datasets.
After optimization, the pre-trained model performs well for classifying skin tumors, with Inception-v3 having accuracy and an F-score of 81% and 81%, respectively.

THIS IS AN OPEN ACCESS ARTICLE UNDER THE CC BY LICENSE:
https://creativecommons.org/licenses/by/4.0/

1. Introduction

There are various unique organs in the human body. One of which, is the skin. The biggest organ, composed of water, protein, lipids, and minerals, covers the whole human body. The skin controls body temperature and guards against pathogens. Skin nerves aid in the perception of hot and cold feelings [1]. The stratified, cellular epidermis and underlying dermis make up human skin. Skin cancer in the last ten years is the most prevalent kind of cancer in people, with melanoma being the most dangerous and fatal form[2].

Diagnosing skin cancer with the naked eye is highly subjective and rarely generalizable. Besides, Most of the prior works are based on the classification of a single disease or a few types of skin diseases. In this study, the researcher tries to provide a deep learning-based algorithm that classifies a variety of benign and malignant skin lesions into different categories and produces more accurate findings than earlier work in the field of skin cancer classification [3]. In order to help patients receive the right medical care, this thesis intends to quickly detect and categorize skin diseases and to define the type of the disease through medical picture analysis.

1.1. The Development of Skin Cancer's Spread

According to a World Health Organization (WHO) report, cancer is one of the major causes of mortality worldwide. According to the forecast, there would be a doubling of cancer diagnoses during the following 20 years [4, 5]. According to the World Health Organization (WHO), about 49 persons in Iraq died from skin cancer in 2020.

10,000 people lose their lives to melanoma every year in the US alone, and the total number of deaths is projected to increase by 6.5 percentage points in 2022. In Europe, almost 100,000 new cases of melanoma are discovered each year, and melanoma is detected in 15,229 Australians per year [6].

1.2. The proposed Types for classification of Skin Cancer

In our study, we classified eight types of benign and malignant skin cancer [7], shown in (figure 1), it is as follows:

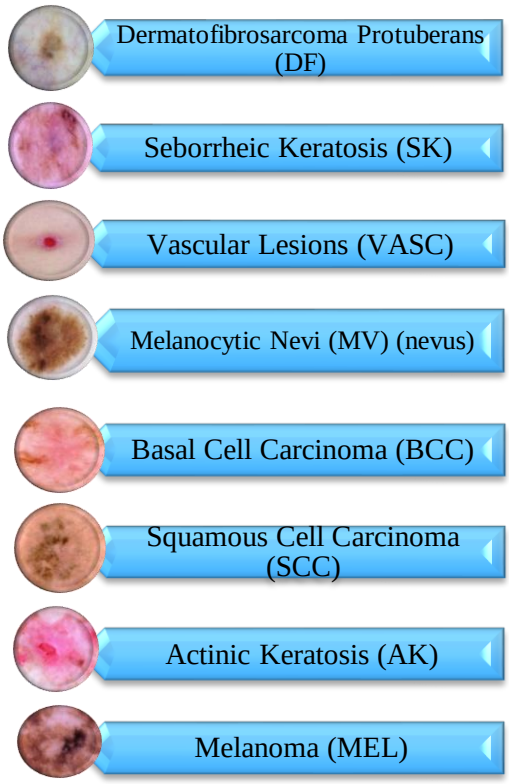

Figure 1 : Eight different types of diseases caused by skin lesions

2. Methodology

2.1. Enhancement Inception-v3 Model Architecture (modified)

A pre-trained convolutional neural network architecture from the Inception family called Inception-v3 is 48 layers deep and has been trained on more than a million photos from the imagenet 2012 database. The network accepts images of a 299 by 299 resolution. In the first section, the model extracts general features from input photos and then classifies them in the second section using those features [8].

The structure of the model is retrained and fine-tuned so that the last three layers are removed and the results of the bottleneck layer are used as the feature results in order to make the model more appropriate for our experiment, which significantly reduces the size of the dataset needed and training time. Retraining the final layer enables to preserve the knowledge that the model had acquired throughout its first training and apply it to your smaller dataset, producing accurate classifications without requiring intensive training and processing resources.

For training, validation, and testing, respectively, 70%, 20%, and 10% of the dataset were used. The sizes kernel-passed activation function fully-connected the three final Dense layers (256, 256,128). Two 0.3 dropout layers are placed after the activation unit to stabilize the focus layer's output. Used a layer called globalaveragepooling2d. The final layer uses a softmax activation function after a relu activation. The model makes considerable use of batch normalization, which is also applied to the activation inputs. The model is optimized using the Adam gradient descent method with a learning rate of Adam (lr = 0.001). Additionally, callback is used, and lr is reduced by a factor of 0.5 if accuracy doesn't increase after three epochs. After training on 40 epochs and 32 sample sizes and fine-tuning on those same numbers, the network is assessed. The structural schematic diagram of the Inception-v3 model in this work is shown in Figure 2.

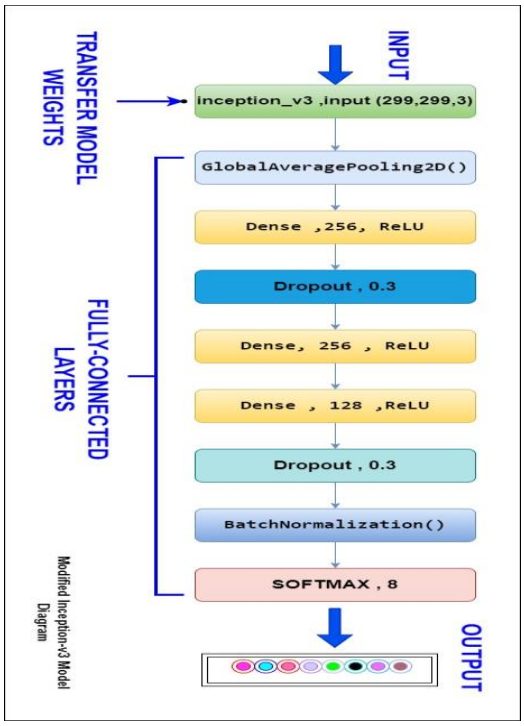

Figure 2: Modified Inception-v3 Model diagram

2.2. Performance evaluation metrics

Precision: is the ratio of the percentage of correctly classified images to the total number of images the model correctly assigned to a positive class [9]. The formula is as follows:

$$Precision = \frac{True\ Positive(TP)}{True\ Positive(TP) + False\ Positive(FP)} \times 100\% \quad (1)$$

Recall: The recall is the percentage of all correctly classed positive samples that have been appropriately identified [9].

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative(FN)} \times 100\% \quad (2)$$

F1- score: In statistical analysis of binary classification, a test's efficiency is measured by the F1-score, also referred to as the F measure [10]. We calculated the F1-Score over all classes using the following expression to get a thorough evaluation of the model's performance in the context of a multiclass classification problem.

$$F1 - score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (3)$$

3. Results and discussions

3.1. Descriptions of utilized datasets

Two sets of data on melanoma were utilized in this work. Where the researcher gathered a big number of pictures of various skin cancers. By merging the "Skin Cancer Dataset" from the first set [11] and the "Skin Cancer Detection using CNN" from the second set [12]. Whereas A second group, which included nine different forms of benign and malignant skin lesions, was joined with the previous group, which only contained five different types of skin cancer. In order to maximize the amount of datasets regarding benign and malignant melanoma that could be gathered, the two groups were mixed and increased their number of images totaling 7,534 images to improve the effectiveness of the CNN algorithms. With rising data volumes and the application of data augmentation techniques, the network can be trained to achieve the highest classification accuracy. The data set comprises a number of high-quality, vivid, and high-resolution (JPG) photographs (the majority of which are 450 × 600 pixels) of various skin cancer types in the training and test groups. The number of photos for each category that were produced after combining the first data set and the second data set to achieve the study's objective of classifying eight different types of skin lesions is shown in Table 1.

Table 1 : Dataset collection

Dataset	*Dataset for Each Type of Skin Cancer*							
	AK	**SCC**	**DF**	**MV**	**MEL**	**SK**	**BCC**	**VASC**
First Dataset	988 Image	1015 Image	-	1318 Image	1200 Image	1100 Image	-	-
Second Dataset	130 Image	197 Image	246 Image	-	438 Image	80 Image	569 Image	253 Image
The total number of images				**7,534 images**				

3.2. Training and Validation Results and Discussions

In order to validate the categorical accuracy, training accuracy curves in (Figure 3 a) and training validation loss curves have been provided in (Figure 3 b) in order to evaluate the model learning and to show how accurate the learning is.

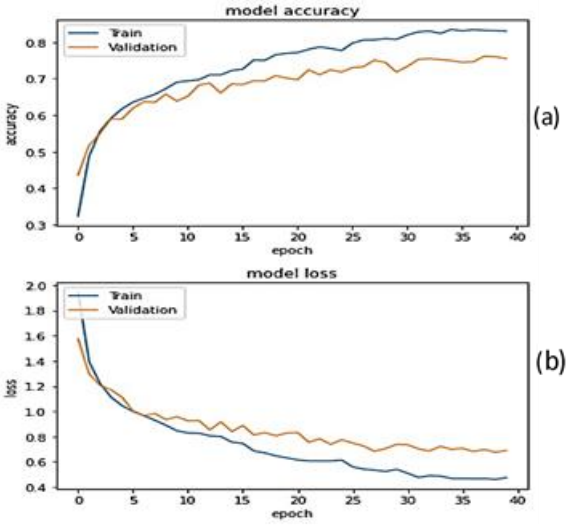

Figure 3 : Training-validation accuracy curves (a) and training-validation loss curves (b) for Inception-v3 model

The loss curve drops symmetrically downward as the number of iterations grows, and the training accuracy index rises as a result, showing that the model is picking up information at an acceptable pace. The ability of the model to generalize to new images is demonstrated by the fact that the training and validation curves barely deviate from each other and that there is overfitting. The accuracy of the test was indicated, and it was 76.36% before fine-tuning.

Figure 1.3 depicts the results in visual form. With a total of 40 epochs used and a loss of 1.49 over the course of data training, the accuracy started out at 32.33%. During each training step, the accuracy of the training and verification data increased and the loss fell. After epoch 16, the learning rate is lowered to allow for the most accurate training of the data. The validation set is only used to evaluate the model's performance after the model has been trained on the training set. Given that the level of accuracy in training and verification tends to increase with the number of epochs, it can be deduced that epoch 40 is the highest possible score for training the data in the context of the modified Inception-v3 method.

3.3. Testing Results and Discussions

In order to assess the findings acquired from unselected samples during the testing phase and to assess the efficacy of the suggested approach, we developed test accuracy curves in (Figure 4 a) and test validation loss curves in (Figure 4 b) to validate categorical accuracy.

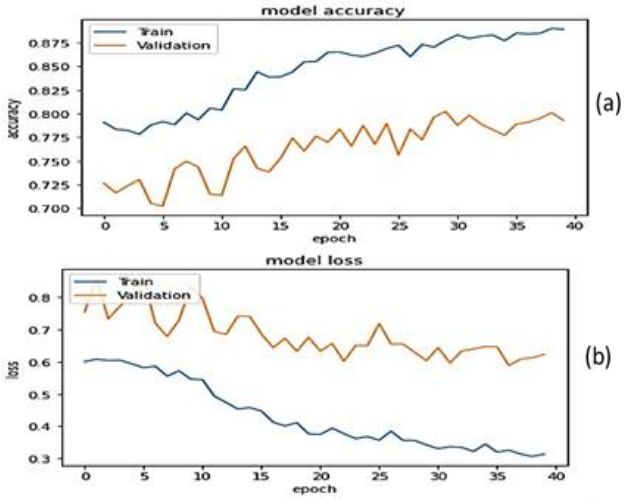

Figure 4 : (a) Testing accuracy curves an (b)Testing loss curves for Inception-v3 model

The training loss curve slopes down symmetrically, the test accuracy rises with the number of repetitions, and after fine tuning, the accuracy averages 81.08%. This shows that the model learned well throughout the training phase, has learned new information, and is clearly capable of recognizing the majority of the new photos. This finding suggests that the suggested algorithm accurately categorizes the majority of the eight forms of skin cancer. As testing data for the model, 762 photos from the validity test set of an unlabelled sample were employed. 40 epochs in all were employed, with a test loss of 0.60 and an accuracy starting point of 79.07%. Given that the Inception-v3 algorithm has been pre-trained on images similar to skin images, which improves the algorithm's ability to recognize different types of skin cancer, we observe an improvement in the accuracy of the test data and a decrease in the loss rate and an increase training and testing speed when compared to the related worked algorithm. Epoch 40 achieves the greatest data test accuracy, The learning rate is slowed down from epoch 11 in order to lower mistake rates and attain the best level of accuracy while evaluating data.

3.4. Confusion Matrix and Classification Report

762 unselected sample images from the validation set were used to validate the model. Figure 5 displays the confusion matrices for the proposed model. For the eight classes, the diagonal

value of the confusion matrix is higher. This shows that it can categorize those numbers of test samples from our test dataset accurately.

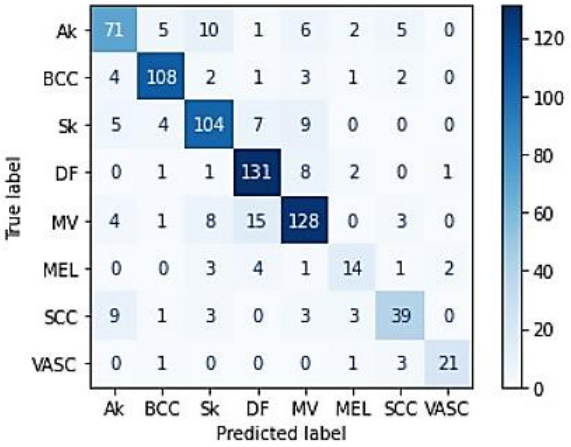

Figure 5: Confusion Matrix for Inception-v3 Model

In order to evaluate how well the Inception-v3 model performed on unselected images from the validation set, we measured accuracy, recall, and f1 score using partial mean and weight. For accuracy, recall, and F1 score, the weighted average was 81%, 81%, 81%, while the partial average was 79%, 77%, 78%. For basal cell carcinoma, the Inception-v3 model has the best accuracy, retrieval and f1 score values and as shown in (Figure 1.5). The Multiclass Rating report containing the partial average, accuracy, weighted average, F1 score, and, recall is shown in Table 2.

Table 2: Classification Report

	Precision	Recall	F1-score	Support
AK	0.76	0.71	0.74	100
BCC	0.89	0.89	0.89	121
SK	0.79	0.81	0.80	129
DF	0.82	0.91	0.86	144
MV	0.81	0.81	0.81	159
MEL	0.61	0.56	0.58	25
SCC	0.74	0.67	0.70	58
VASC	0.88	0.81	0.84	26
accuracy			0.81	762
macro avg.	0.79	0.77	0.78	762
weighted avg.	0.81	0.81	0.81	762

3.5. Comparisons

State-of-the-art research is used in comparisons with various CNN network models. For the dataset being used, several network models are simulated and used. These include the proposed transfer learning model, the VGG-16 model [71], the VGG-19 model [73], and CNN's model [13]. The outcomes of the applied comparisons for various image sizes are displayed in Table 3. (input shapes).

Table 3: Comparisons between different CNN network models

DL Network Models	Accuracy Before Fine-Tuning	Accuracy After Fine-Tuning	Size of Image
VGG-16 [11]	**72.06 %**	**77%**	**224×224 pixel**
VGG-19 [14]	**71.51 %**	**80%**	**224×224 pixel**
CNN [13]	**70%**	**79%**	**100×100 pixel**
Modified Inception-v3 Model	**76.36 %**	**81 %**	**299×299 pixel**

As previously stated, 10% of the total photos in the same dataset are chosen at random to be used in the testing stage. In the testing phase, the proposed model is assessed. Additionally, it provides comparisons with a number of cutting-edge CNN network architectures after being tested and simulated using the ISIC dataset's used images.

Table 3 compares the testing accuracies of several CNN network structures and shows that the networks have varying degrees of accuracy. The amount of trained data, the kind of algorithm, the size of the input image, the RGB image, and the quantity of layers employed during training all affect accuracy. Comparing the proposed Inception-v3 network to the others reveals that it is well suited to the ISIC dataset. However, it is obvious that the proposed Inception-v3 model achieves higher accuracy and faster training and testing speeds than the comparison models. This is as a result of its properly thought-out architecture and ISIC adaptation.

In the research described in the table, a network's depth in the VGG-16 and VGG-19 reveals its negative impact on identifying features of skin cancer application, and adding additional layers (stack of layers) results in a large increase in training error. They had poor accuracy, and Slow training speed when compared to the Proposed Inception-v3 model.

The modified network Inception-v3 model is distinguished by specific layers with various parameters. As a result, they can improve the outcomes in comparison to the CNN, VGG-16 and

VGG-19 and find some skin cancer traits.To categorize skin cancer into eight categories and obtain the greatest accuracy for all used skin lesion kinds, this model can successfully identify the characteristics of the ISIC pictures used.

4. Conclusions

Skin cancer is a fatal condition, and chances of survival are increased when it is detected early. Nevertheless, skin cancer risk may be increased by viral, fungal, or exposure to UV exposure diseases. This thesis focused on examining the use of deep learning techniques to categorize ISIC photos of skin cancer. A model has been developed for transfer learning which was Inception v3. The designed general structure of the model involved many processing stages.

Firstly, collecting skin cancer images where collected 7,534 images of malignant and benign oncological diseases, which were formed by The International Skin Imaging Collaboration (ISIC).

Second, the images for each suggested model were divided into three phases: training, validation, and testing, with respective allocations of 70%, 20%, and 10%.

As a result, measures like the F-score and accuracy show that the proposed Inception v3 model performs remarkably well for the multiclassification of skin lesions after fine-tuning shown up to 81% performance.

References

[1] C. (n.d.).@ClevelandClinic. "Skin: Layers, Structure and Function." https://my.clevelandclinic.org/health/articles/10978-skin (accessed 28 December 2022.

[2] M. Vestita, P. Tedeschi, and D. Bonamonte, "Anatomy and Physiology of the Skin," in *Textbook of Plastic and Reconstructive Surgery: Basic Principles and New Perspectives*, M. Maruccia and G. Giudice Eds. Cham: Springer International Publishing, 2022, pp. 3-13.

[3] M. Dildar *et al.*, "Skin cancer detection: a review using deep learning techniques," *International journal of environmental research and public health,* vol. 18, no. 10, p. 5479, 2021.

[4] A. Naeem, M. S. Farooq, A. Khelifi, and A. Abid, "Malignant melanoma classification using deep learning: datasets, performance measurements, challenges and opportunities," *IEEE Access*, vol. 8, pp. 110575-110597, 2020.

[5] A. V. Demidova, D. S. Kulyabov, and E. Y. Shchetinin, "Skin cancer classification computer system development with deep learning," in *ITTMM*, 2020, pp. 57-69.

[6] A. Naeem, T. Anees, M. Fiza, R. A. Naqvi, and S.-W. Lee, "SCDNet: A Deep Learning-Based Framework for the Multiclassification of Skin Cancer Using Dermoscopy Images," *Sensors*, vol. 22, no. 15, p. 5652, 2022.

[7] a. A. o. D. association. "Types of skin cancer" https://www.aad.org/public/diseases/skin-cancer/types/common (accessed.

[8] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.

[9] A. Bassel, A. B. Abdulkareem, Z. A. A. Alyasseri, N. S. Sani, and H. J. Mohammed, "Automatic malignant and benign skin cancer classification using a hybrid deep learning approach," *Diagnostics*, vol. 12, no. 10, p. 2472, 2022.

[10] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using EfficientNets–a first step towards preventing skin cancer," *Neuroscience Informatics*, vol. 2, no. 4, p. 100034, 2022.

[11] A. W. KHAN. "Skin Cancer dataset." https://www.kaggle.com/datasets/abdulwadoodkhan199/skin-cancer-dataset (accessed.

[12] J. AHUJA. "Skin Cancer Detection using CNN." https://www.kaggle.com/datasets/jaiahuja/skin-cancer-detection(accessed.

[13] N. Rezaoana, M. S. Hossain, and K. Andersson, "Detection and classification of skin cancer by using a parallel CNN model," in *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, 2020: IEEE, pp. 380-386.

[14] A. W. KHAN. "Skin Cancer dataset-vgg19." https://www.kaggle.com/code/abdulwadoodkhan199/skin-cancer-using-transfer-learning-vgg19(accessed.