# Skin Cancer Classification using Deep Learning and Transfer Learning

**Conference Paper** · December 2018

DOI: 10.1109/CIBEC.2018.8641762



**3 authors:**

Khalid Hosny
Zagazig University
**105** PUBLICATIONS **1,359** CITATIONS
SEE PROFILE

Mohamed A. Kassem
Kafr El-Sheikh University
**7** PUBLICATIONS **128** CITATIONS
SEE PROFILE

Mohamed Fouad
Zagazig University
**1** PUBLICATION **19** CITATIONS
SEE PROFILE

# Skin Cancer Classification using Deep Learning and Transfer Learning

Khalid M. Hosny, Mohamed A. Kassem, and Mohamed M. Foaud

*Abstract*— **Skin cancer, specially melanoma is one of most deadly diseases. In the color images of skin, there is a high similarity between different skin lesion like melanoma and nevus, which increase the difficulty of the detection and diagnosis. A reliable automated system for skin lesion classification is essential for early detection to save effort, time and human life. In this paper, an automated skin lesion classification method is proposed. In this method, a pre-trained deep learning network and transfer learning are utilized. In addition to fine-tuning and data augmentation, the transfer learning is applied to AlexNet by replacing the last layer by a softmax to classify three different lesions (melanoma, common nevus and atypical nevus). The proposed model is trained and tested using the ph2 dataset. The well-known quantative measures, accuracy, sensitivity, specificity, and precision are used in evaluating the performance of the proposed method where the obtained values of these measures are 98.61%, 98.33%, 98.93%, and 97.73%, respectively. The performance of the proposed method is compared with the existing methods where the classification rate of the proposed method outperformed the performance of the existing methods.**

## I. Introduction

Skin cancer is one of the most world-wide diseases that cause death [1]. There are two types of skin cancer called melanoma and non-melanoma. Early detection of these lesions may increase the curing rate to 90% [2]. The high similarity between different types of skin lesions makes the visual examination hard and may lead to wrong investigation [3]. Therefore, an automated system is required for skin lesion classification [4]. This classification system utilized the techniques of image processing and artificial intelligence.

The earlier computer-aided methods for dermatological image classification suffered from two major problems. First, insufficient data [5-6]. Imaging process is the second challenging problem where skin images acquired by a special device, dermoscopy, while other medical images such as histology images were acquired by using the microscopy and biopsy [7]. The earlier methods [8, 9], required extensive pre-processing, segmentation and feature extraction processes in order to classify the skin images.

Recently, researchers around the world successfully utilized deep learning in visual task [10,11] and object recognition [12]. Codella et al. [3] proposed a hybrid system for melanoma classification. In this system, support vector machine (SVM), sparse coding, and a deep learning method were combined where the classification accuracy was 93.1%. In [13], Ozkan and Koklu developed a skin lesion classification system. They applied four machine learning methods to classify the lesion into melanoma, abnormal, and normal where the highest accuracy percentage, 92.50%, was achieved by the artificial neural networks (ANN). Bi et al. [14] proposed a melanoma detection system using the multi-scale lesion-biased representation (MLR) and joint reverse classification (JRC). In spite of the lesions classified into two classes melanoma and non-melanoma, the accuracy was 92%. Chakravorty et al. [15] proposed a skin lesion classification for irregular distribution of colors and structures using Kullback-Leibler divergence of color histogram and Structural Similarity metric. Their method achieved a classification rate, 83%. Waheed et al. [16] proposed a machine learning model based on discriminating properties such different types of color and texture features of skin lesion to classify melanoma. Their model achieved a classification rate of, 96%. Premaladha et al. [17] proposed a melanoma classification system that enhances images by contrast limited adaptive histogram equalization technique, and then, the filtered grey-scale image was segmented by normalized method of Otsu. The classification accuracy using deep learning was 92.89%.

In this paper, a deep convolutional neural network (DCNN) is applied to classify the color images of skin cancer into three types: Melanoma, atypical nevus, and common nevus. The proposed method has two main advantages over the earlier computer-aided methods for skin cancer. First, the proposed method has the ability to work with any type of image (dermoscopic and photographic). Second, the proposed method does not require any pre-processing. It works directly with the acquired color images of skin.

The rest of this paper consists of four sections. In section II, a brief description of DCNN is presented. In section III, the proposed method and the utilized tools are presented. In section IV, description of the performed experiments and the obtained results are presented. The conclusion is presented in section V.

## II. Background

A number of neural network layers were used to comprise a deep convolutional neural network (DCNN). These layers are divided into convolutional and pooling layers [18]. These layers were used to extract features from the input color images of skin. A fully connected layer is the last stage-layer which used to get the predicted classes by the computed class

K. M. Hosny is with Information Technology Department, Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt (e-mail: k_hosny@yahoo.com).

M. A. Kassem is with Information Systems department, Modern Academy, Cairo 11435, Egypt (e-mail: cs.engineer.mohamed.1987@gmail.com).

M. M. Foaud is with: 1- Electronics and Communication Department, Faculty of Engineering, Zagazig University, Zagazig 44519, Egypt. 2- Alexandria Higher Institute of Engineering and Technology, Alexandria 21311, Egypt.

score [19]. The DCNN could be used with new labeled data that differ from the previous data which used to train the DCNN for the first time [20]. The insufficient number of labeled-data, noise, and aberrations [21] limited the utilization of DCNN in the classification task of skin lesions [22]. The same feature may have a large variation in dermoscopic images while lesions of different kinds may have visual similarity [23]. All of these challenges motivated the authors to present the proposed method for efficient classification of skin lesions.

A DCNN has been used in a number of application to improve performance such visual tasks, natural language processing, action recognition [24, 25]. VGGNet, ZFNet, ResNet, GoogLeNet, AlexNet, and LeNet [26] are different DCNN architectures that can be used in different applications. The AlexNet is the utilized DCNN architecture in the proposed skin cancer classification model. A brief description of these net is presented in the next subsection.

### *A. AlexNet*

AlexNet was developed by Krizhevsky et al. [24] and used in the visual recognition of imageNet [27]. With 8 compromise layers, AlexNet is divided into 5 convolutional layers and 3 fully connected layers. The dimension of the input layer cannot exceed the predicated width (W) and height (H) equal to 224×224 where the depth (D) of this layer is 3. The depth of the input image refers to red, green and blue (color space). The input image filtered through the first convolutional layer. A number of kernels (K) equal 96 with a filter (F) of size 11×11 and a stride of 4 pixels are used in first layer. The stride (S) is the distance between the centers of responsive field to neighboring neurons in the kernel map. A number of padding (P) may be added to some of the convolutional layers. The output size of the convolutional layers can be computed using the mathematical expression (W−F+2P)/S + 1. From this mathematical form, the size of the convolutional output will be (224–11+0) / 4 + 1 ≈ 55. The input of the second convolutional layer will be 55×55×96. By using 2 graphical processing units (GPUs), the load of the work will be 55×55×48 for each of the GPUs.

Generally, the process of dimensional reduction must be applied with each feature map. In AlexNet, the pooling layer which follows the convolutional layer used to reduce the dimension of the feature map. The pooling may be Average, Max, Sum, etc. A max pooling layer is used in AlexNet, where the size of this layer is $\left((55/2) \times (55/2) \times 96\right) \approx 27 \times 27 \times 96$. This pooling layer is associated with 256 filters of size 5×5×48 and a stride of 2 pixels. As we mentioned before, there are 2 GPUs, therefore, the load will be 27×27×128 for each GPU.

The output of the second layer which is connected with 384 kernels constitute the input of the third layer. The size of each kernel in this layer is 3×3×192 for 2 GPUs. The fourth and fifth layers has 384 and 256 kernels respectively. In the fourth layer, the size of each kernel will be 3×3×192 while the kernel size is 3×3×128 for the fifth layer. The third and fourth layers has been created without any normalization or pooling layer, while the fifth layer has a max pooling layer. Finally, two fully connected layers are created. The input of these created layers come from the third, fourth and fifth convolutional layers. Each fully connected layer has 4096 neurons. An illustration of the AlexNet architecture is displayed in Fig. (1).

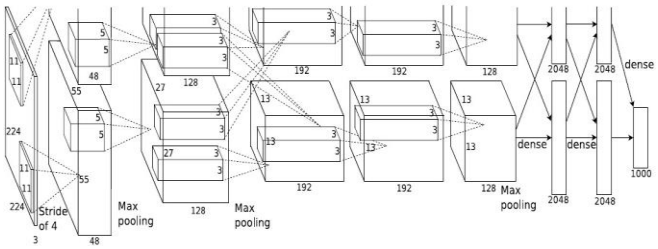


Figure 1. Architecture of Alex-Net [24]

## III. METHODOLOGY

The description of the proposed method is presented in the next two subsections. The first subsection is devoted to describe the datasets of skin lesion and the augmentation process for the color images of skin lesion. The different ways used in the pre-trained DCNN are presented in the second subsection.

### *A. Augmentation Process*

The DCNN requires a massive number of images for training and testing to achieve high classification rates. This is a big challenge especially with skin cancer datasets where the number of available labeled images for training and testing is very limited. For example, the ph2 dataset [28] consists of 200 RGB color images. This dataset was divided into three classes: melanoma, common nevus, and atypical nevus where the number of images in these classes are 40, 80, and 80 respectively.

The augmentation process is a successful way to overcome this problem. All images in each class are rotated with different angels from $0^0$ to $270^0$ with a constant step $5^0$. The rotation process results in 55 rotated images for each color image. The total number of melanoma images becomes $40 \times 55 = 2200$ while the total number of common nevus and atypical nevus images are $80 \times 55 = 4400$.

### *B. Transfer Learning*

Training new DCNN model required a massive number of images. Unfortunately, no available dataset of skin lesion with thousands of labelled images. The theory of transfer learning is the key solution for this challenging problem where a big size medical dataset such as the ImageNet could be used in pertained AlexNet. Three different ways have been applied in the proposed model to AlexNet: First, the last layer that used for classification has been replaced with new softmax layer to classify the input images into three classes instead of 1000 classes as in the ImageNet. Second, the back-propagation have been used to fine-tune the weights to gain new weights to classify skin lesion in a better way. A small rate of learning has been used, so the convolutional layers weights will not dramatically change while the weights of the fully connected layers are randomly initialized. The weights are updated using stochastic gradient descent (SGD) algorithm based on the dataset images. Third, an augmentation step has been done to dataset to overcome the limitation of labelled images required to train the DCNN to gain the optimal weights.

In this work, the classification layer has been replaced with a softmax layer to classify three different skin lesion, melanoma, common nevus, and atypical nevus. Fig (2) describe the modified model of AlexNet.

## IV. Experiments, Results and Discussion

The experiments were performed using an IBM-computer equipped with a core i5 processor, 8 GB DDRAM and NVIDIA GeForce 920m graphical processing unit. The proposed method is codded using MATLAB 2017x 64-bit, where the executed version is coded using CUDA to be run using the GPU.

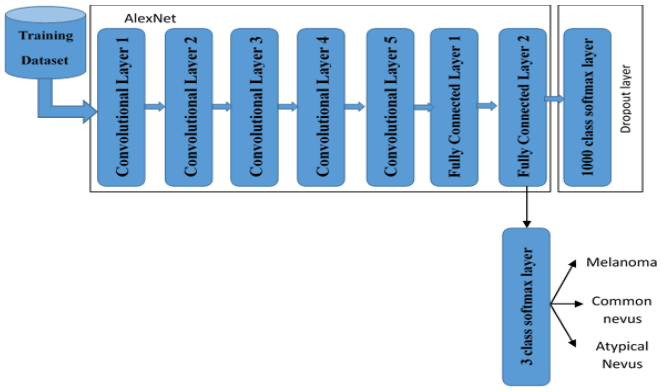


Figure 2. Modified Alex-Net

GPU enables us to train a huge number of images which have a good impact to minimize the error rate. Based on the required task, the pre-trained DCNN-based models change the classification layer.

We performed two types of experiments. The first type of experiments was designed to evaluate the proposed method with original dataset images without any augmentation. The second type of experiments aimed to evaluate the proposed model with the augmented images for the same dataset.

For creditable results, the same values of initial learning rate, batch size and number of training epochs are used for all experiments. These values are 10, 32, and 0.001 for batch size, the number of training epochs and initial learning rate respectively. Using MATLAB, the dataset is randomly divided to training and testing where 80% of the dataset was used for training and 20% for testing.

The pre-trained AlexNet cannot process color images with size greater than 227 ×227 ×3. To overcome this problem, we perform a segmentation process [29] of the region of interest (ROI) to reduce the size and ignore unnecessary parts of the input images. Then, a check step has been run over all images to ensure that the width and height do not exceed 227 ×227 while the depth must be 3 because of using images in the color space R, G, and B.

The performance of the proposed model has been evaluated using four quantitative measures: Accuracy, sensitivity, specificity, and precision [30]. These measures are computed using the following forms:

$$\text{Accuracy(ACC)} = \frac{t_p+t_n}{t_p+f_p+f_n+t_n} \quad (1)$$

$$\text{Sensitivity(TPR)} = \frac{t_p}{t_p+f_n} \quad (2)$$

$$\text{Specificity(TNR)} = \frac{t_n}{f_p+t_n} \quad (3)$$

$$\text{Precision(PPV)} = \frac{t_p}{t_p+f_p} \quad (4)$$

where $t_p$, $f_p$, $f_n$, $and\ t_n$ refer to true positive, false positive, false negative, and true negative. TPR in (2) means true positive rate, TNR in (3) refers to true negative rate while PPV in (4) is an abbreviation for positive prediction value. The receiver operating characteristic (ROC) [31] is additional qualitative measure.

Two experiments are performed using the same dataset. The first experiment is performed with the original dataset images without augmentation. The second experiment is performed with augmented images. The color images are rotated with angles from $0^0$ to $270^0$ with fixed step $5^0$. The computed performance measures are shown in table I.

The performance of the proposed method is compared with the existing methods [13-16] using the same dataset of skin lesions. Some of the existing methods classify lesions into two classes as discussed in the previous section and compute the accuracy for these classes and list the highest accuracy. The proposed method classifies lesions into three classes, melanoma, common nevus, and atypical nevus where the average of performance is measure. Results of the comparison process are presented in Table II where the proposed method achieves the highest classification rate. The comparison of obtained classification rates is visualized and displayed in Fig (3). The ROC curves for the proposed and the existing methods [13-16] are displayed in Fig (4).

All measures prove the superiority of the proposed DCNN-based classification method over the other existing methods.

TABLE I. PERFORMANCE MEASURE OF PROPOSED MODEL

| Ph2 dataset | Average accuracy (%) | Average sensitivity (%) | Average specificity (%) | Average Precision (%) |
|---|---|---|---|---|
| Original images | 80 | 72.92 | 83.33 | 75.81 |
| Augmented images | **98.61** | **98.33** | **98.93** | **97.73** |

TABLE II. COMPARATIVE STUDY

| Methods | Classification Method | TPR (%) | TNR (%) | PPV (%) | ACC (%) |
|---|---|---|---|---|---|
| R. Chakravorty [15] | Symmetry type | 67 | 89 | 69 | 83 |
| L. Bi [14] | Joint Reverse | 87.50 | 93.13 | ---- | 92 |
| I. A. Ozkan [13] | ANN | 90.86 | 96.11 | 92.38 | 92.5 |
| Z. Waheed [16] | SVM | 97 | 84 | ---- | 96 |
| **Proposed method** | DCNN | **98.33** | **98.93** | **97.73** | **98.61** |

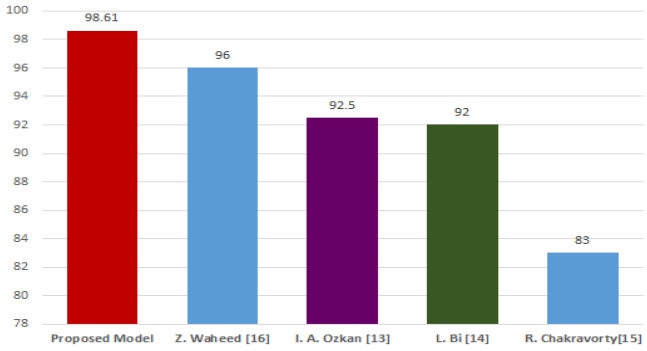


Figure 3. Visualized classification rates for the proposed and the esisting methods [13-16].

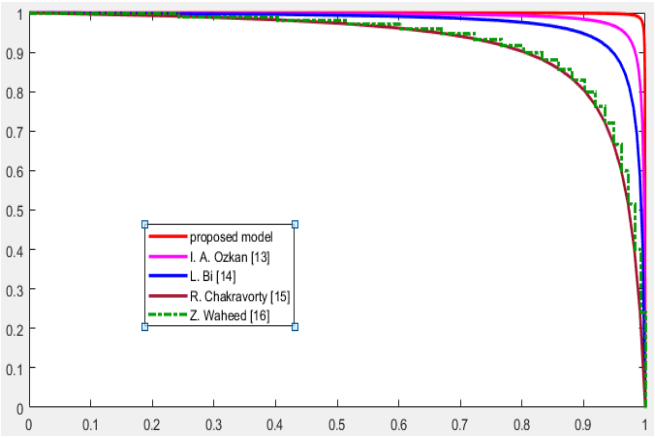


Figure 4. ROC curves for the different methods

## V. CONCLUSION

A huge number of labeled images is needed to build a successful deep neural network. The transfer learning and image augmentation are applied to a pre-trained AlexNet to overcome this major challenge. The proposed method has the ability to classify three different lesions by replacing the last layer to softmax with three classes only. According to the transfer learning, the weights of the modified model have been fine-tuned in addition to the augmentation of the dataset images. Four performance measures have been computed for the proposed to compare with existing methods where the obtained results prove that the proposed method outperformed the existing methods. The achieved rates are 98.61%, 98.33%, 98.93% and 97.73% for accuracy, sensitivity, specificity, and precision respectively.

## REERENCES

[1] American Cancer Society: Cancer facts and figures 2018. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf , Accessed: 15 Aug 2018.

[2] N. C. F. Codella et al., "Deep learning ensembles for melanoma recognition in dermoscopy images" IBM Journal of Research and Development, Vol.61, pp.5:1- 5:15, 2017.

[3] N. Codella et al., "Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images", LNCS, Springer, vol. 9352, pp. 118–126, 2015. Available: https://doi.org/10.1007/978-3-319-24888-2_15 , Accessed: 15 Aug 2018

[4] N. K. Mishra and M. E. Celebi. "An overview of melanoma detection in dermoscopy images using image processing and machine learning", 2016. Available: https://arxiv.org/abs/1601.07843 , Accessed: 15 Aug 2018.

[5] Masood, A. & Al-Jumaily, A. A. Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms. Int. J. Biomed. Imaging 2013, vol. 323268, 2013.

[6] Gutman, D. et al. Skin lesion analysis toward melanoma detection. International Symposium on Biomedical Imaging (ISBI), (International Skin Imaging Collaboration (ISIC), 2016.

[7] Binder, M. et al. Epiluminescence microscopy-based classification of pigmented skin lesions using computerized image analysis and an artificial neural network. Melanoma Res., Vol. 8, pp. 261–266, 1998.

[8] Burroni, M. et al. Melanoma computer-aided diagnosis: reliability and feasibility study. Clin. Cancer Res., Vol. 10, pp. 1881–1886, 2004.

[9] Schindewolf, T. et al. Classification of melanocytic lesions with color and texture analysis using digital image processing. Anal. Quant. Cytol. Histol., Vol. 15, pp. 1–11, 1993.

[10] Mnih, V. et al. Human-level control through deep reinforcement learning. Nature, vol. 518, pp. 529–533, 2015.

[11] Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. Nature, vol. 529, pp. 484–489, 2016.

[12] Russakovsky, O. et al. Imagenet large scale visual recognition challenge. Int. J. Comput. Vis., vol. 115, pp. 211–252, 2015.

[13] I. A. Ozkan, and M. Koklu, "Skin Lesion Classification using Machine Learning Algorithms", Intelligent Systems and Applications in Engineering, vol. 5, no. 4, pp. 285-289, 2017.

[14] L. Bi, J. Kim, E. Ahn, D. Feng, and M. Fulham, "Automatic melanoma detection via multi-scale lesion-biased representation and joint reverse classification", 2016 IEEE 13th Int. Symposium on Biomedical Imaging (ISBI), pp. 1055-1058, 2016.

[15] R. Chakravorty, S. Liang, M. Abedini and R. Garnavi, "Dermatologist-like feature extraction from skin lesion for improved asymmetry classification in PH2 database", 38th Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3855-3858, 2016.

[16] Z. Waheed, A. Waheed, M. Zafar, and F. Riaz, "An Efficient Machine Learning Approach for the Detection of Melanoma using Dermoscopic Images", Int. Conf. on Communication, Computing and Digital Systems (C-CODE), IEEE, pp. 316-319, 2017.

[17] Premaladha, J., and K. S. Ravichandran. "Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms." Journal of medical systems 40.4 (2016): 96

[18] Y. LeCun, Y. Bengio, G.Hinton, "Deep learning", Nature, vol. 521, PP.436–444, 2015.

[19] Convolutional Neural Networks (CNNs / ConvNets), the Stanford CS class notes, Spring 2017 Assignments, Available: http://cs231n.github.io/convolutional-networks/, Accessed: 15 Aug 2018.

[20] N. Tajbakhsh et al., "Convolutional neural networks for medical image analysis: Full training or fine tuning?" IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1299–1312, 2016.

[21] X. Yang, et al., "A Novel Multi-task Deep Learning Model for Skin Lesion Segmentation and Classification", Available: https://arxiv.org/abs/1703.01025 , Accessed: 15 Aug 2018.

[22] H. Greenspan, B. van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1153–1159, 2016.

[23] U. O. Dorj, K. K. Lee, J. Y. Choi, M. Lee, "The skin cancer classification using deep convolutional neural network", Multimedia Tools and Applications, springer, vol. 77, no. 8, pp. 9908-9924, 2018.

[24] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", In Proc. Neural Information Processing Systems (NIPS), vol. 1, pp.1097-1105, 2012.

[25] S. Suraj, et.al. , "A taxonomy of Deep Convolutional Neural Nets for Computer Vision", Frontiers in Robotics and AI, VOL.2, 2016, Available: https://doi.org/10.3389/frobt.2015.00036 , Accessed: 15 Aug 2018.

[26] Convolutional Neural Networks (CNNs / ConvNets), the Stanford CS class notes, Spring 2017 Assignments, Available: http://cs231n.github.io/convolutional-networks/, Accessed: 15 Aug 2018.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 248–255, 2009.

[28] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. S. Marcal and J. Rozeira, "PH2- A dermoscopic image database for research and benchmarking," 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, 2013, pp. 5437-5440.

[29] MATLAB Central Program or Color Image Segmentation – Athi Narayan S, K.S.R. College of Engineering, Erode, Tamil Nadu, India, Available: https://www.mathworks.com/matlabcentral/fileexchange/25257-color-image-segmentation?focused=5191437&tab=function, Accessed: 15 Aug 2018.

[30] M. Stojanovi et.al., "Understanding sensitivity, specificity, and predictive values", Vojnosanit Pregl, vol. 71, no11, pp. 1062–1065,2014.

[31] T. Fawcett, "An introduction to ROC analysis" Pattern Recognition Letter, Vol. 27, no. 8, p. 861–874, 2006.