

Deep Learning Based Melanoma Diagnosis Identification

Gaole Duan*

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: DuanLuka@163.com
*corresponding author

Changyuan Wang

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: Cyw901@163.com

Abstract—Malignant melanoma is considered to be one of the deadliest types of skin cancer, and it is responsible for the death of a large number of people worldwide. However, distinguishing whether melanoma is benign or malignant has been a challenging task. Many Computer Aided Diagnosis and Detection Systems have been developed in the past for this task. This paper presents a deep learning framework based approach for melanoma diagnosis and recognition. In the proposed method, the original skin mirror image is first preprocessed and then passed to the VGG16 convolutional neural network for tumor property classification. VGG16 uses smaller convolutional kernels instead of a larger convolutional kernel to achieve a reduction in network parameters and thus improve network performance. The system is trained using segmented RGB images generated from ground truth images of the ISIC2016 dataset, and finally a softmax classifier is used for pixel-level classification of melanoma lesions. In this study, a new method to become a lesion classifier was designed to classify melanoma lesion regions into benign and malignant tumors based on the results of pixel-level classification, and experiments were conducted on two well-established public test datasets, ISIC2016 and ISIC2017, with a final accuracy of 96.1%. The results indicate that convolutional neural networks are suitable for melanoma diagnosis identification. This study is of great relevance for advanced cancer caused by malignant melanoma.

Keywords-Melanoma; Convolutional Neural Network; Convolutional Neural Network; Lesion Area; Pixel-Level Classification

I. INTRODUCTION

When the body is exposed to UV radiation for a long period of time, the skin barrier is damaged and cells in the base of the human epidermis synthesize melanin to protect against this damage and transport it to the surface of the skin to fill the skin barrier. The cells that produce melanin are

called melanocytes, but these cells that repair the skin barrier are not always beneficial, and once the growth of melanocytes gets out of control, they can become a highly malignant tumor called melanoma [1]. It is the fastest growing type of skin cancer in terms of mortality, The American Cancer Society [2] estimates that about 6930 people are expected to die of melanoma and about 92680 new melanomas is diagnosed in the United States in the year 2023. According to the statistics [2], the lifetime risk of developing melanoma is about 2.6% for whites, 0.1% for blacks, and 0.6% for Hispanics. Cutaneous melanoma is the most dangerous type of skin tumor and it contributes to 90% of skin cancer mortality, Melanoma can however be cured with prompt excision[3] [4]if diagnosed and detected early, where the depth of infiltration is an indicator of the degree of melanoma development, and melanoma with an infiltration depth of less than one millimetre can be completely treated with a minor surgery , while when the infiltration depth reaches four millimetre, there is a great possibility of metastasis even after surgery. Therefore, it is an important issue to study the moles on the patient's body and diagnose whether they are cancerous or not. Even for experienced Surgeon[5] [6], the identification of melanoma from skin lesions using dermoscopic analysis, visual inspection, clinical screening, and histopathological examination can be laborious and inaccurate, and the task of diagnosing moles on the patient's body surface is inherently difficult. Many factors have led to an urgent need for a technology that can automatically read and identify melanoma based on segmented lesion areas.

Because of the high similarity between the lesion area and the background pixels of dermoscopic images, and the diverse shapes of lesions, blurred edges, and artificial or hairy occlusions, it is necessary to first segment the dermoscopic images in order to perform automatic diagnostic classification of melanoma. To address these difficulties, scholars have proposed various semantic segmentation networks [7] [8]. Among them, MK Hasan [9] et al. present the algorithm obtained an mIoU of 87.0% on the ISIC-2017 dataset, which is 1.0% better than the winner of the ISIC-2017 challenge. From the available experimental results it is clear that the technique for segmentation of lesion regions on dermoscopic images is quite mature and there are new advances every year, which provides great help for automatic classification of segmented lesion regions.

In recent years, there have been significant improvements in the research of computational algorithms and techniques for the analysis of skin lesions. Some popular techniques use rules based on asymmetry, boundary structures, variegated color and dermatoscopical structures, which are based on rules commonly used by dermatologists to diagnose skin cancer [10]. These rules help to distinguish benign from malignant melanomas. The variegated color is always just one color in the case of Benign while the malignant always possess two or more colors. This rule has always been applied by many hand crafted methods for the analysis of skin lesions images towards the melanoma detection as shown in Figure 1.

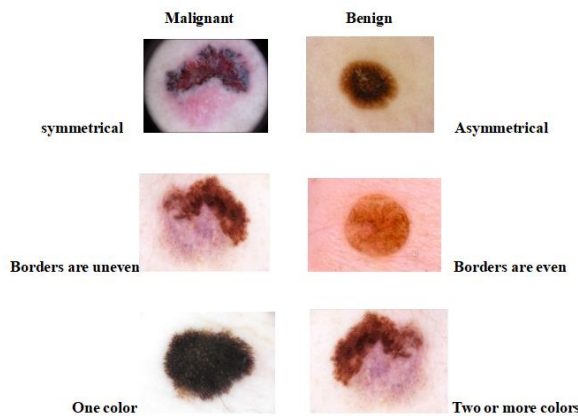


Figure 1. Image showing comparison between benign and malignant skin lesion

These methods, known as hand-crafted methods, are limited by the noise on the skin lesions and the irregular boundary features of the skin lesions. The lack of deep supervision in these methods leads to loss of information during training, which makes it difficult to analyze the complex visual features of skin lesions. This study proposes an intelligent system based on deep learning techniques to differentiate the nature of melanoma using a single VGG16.

II. METHODS AND MATERIAL

A. VGG16 network model

VGGNet is a convolutional neural network model proposed by Simonyan [11] and Zisserman. VGGNet explores the relationship between the depth of a convolutional neural network and its performance, and demonstrates that increasing the depth of the network can affect the final performance of the network to some extent. VGGNet contains a total of six network models, which are similar in structure, with the difference lies in the number of sub-layers in each convolutional layer, and the total network depth ranges from 11 to 19 layers. Under a single test scale [11], the size of the test image is set as shown in equation (1):

$$Q = S, \text{ for fixed } S \quad (1)$$

The size when the test image is jittered set as shown in the equation (2):

$$Q = \frac{1}{2}(S_{\min} + S_{\max}), \text{ for } S \in [S_{\min}, S_{\max}] \quad (2)$$

Where Q is the test set image, S is the training set image. The top-1 error and the top-5 error are 27.0% and 8.8% respectively for VGG16 with smallest image side=256. The top-1 error and the top-5 error for smallest image side = [256; 512] are 25.6% and 8.1%, respectively. The error rate decreases the most significantly and the combined error rate is the lowest, which shows that VGG16 is the best model in VGGNet.

VGG16 consists of 5 convolutional layers, 3 fully connected layers and softmax output layers. Each convolutional layer is followed by one max-

pooling layer, and the ReLU function is used for the activation units of all hidden layers. The ReLU function is shown in Figure 2.

The expression of the ReLU function is shown in equation (3) (4):

$$\text{ReLU}(x) = \max(x, 0) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } x > 0 \end{cases} \quad (3)$$

$$\begin{cases} \text{if } x < 0, f(x) = 0, f'(x) = 0 \\ \text{if } x > 0, f(x) \approx x, f'(x) = 1 \end{cases} \quad (4)$$

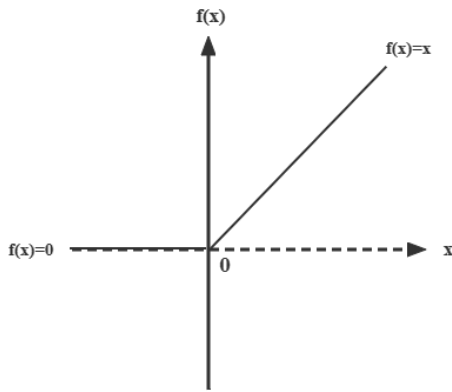


Figure 2. The Relu function in two dimensions

The ReLU function adjusts any number less than zero to zero, which makes the network partly sparse, thus reducing the overdependence between

parameters and alleviating the problem of network overflow.

VGG16 uses multiple convolutional layers with smaller convolutional kernels instead of one convolutional layer with a larger convolutional kernel. The perceptual field size obtained from a stack of two 3×3 convolutions is equivalent to a 5×5 convolution, As shown in Figure 3. while the perceptual field size obtained from a stack of three 3×3 convolutions is equivalent to a 7×7 convolution. We illustrate the principle of this substitutability with an example, and the results are shown in Table 1.

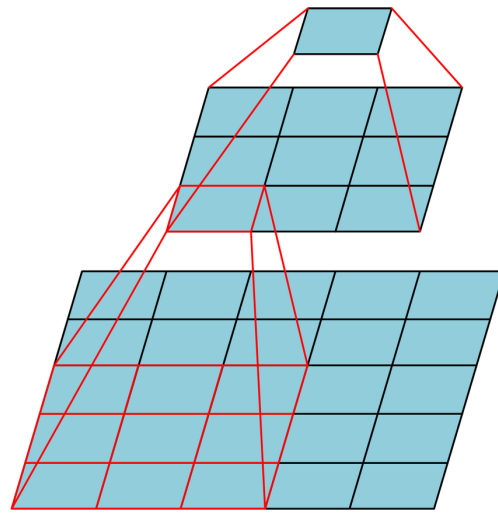


Figure 3. Mapping relationship between 3*3 convolution kernel and 5*5 convolution kernel

TABLE I. FEASIBILITY OF 3*3 CONVOLUTION KERNELS REPLACE 5*5 CONVOLUTION KERNELS

Assuming: feature_map = 28*28	Convolution step = 1	Padding = 0
1-Layer 5×5 convolutional kernel	2-Layer 3×3 convolutional kernel	
Layer1: (28-5) / 1 + 1 = 24	Layer1:(28-3) / 1 + 1 = 26	
Output: Feature map = 24×24	Layer2:(26-3) / 1 + 1 = 24	Output:Feature map = 24×24

The three Fully-Connected (FC) layers: the first two have 4096 channels each; the third performs 1000-way ILSVRC classification and thus contains 1000 channels. And Max-pooling is performed over a 2×2 pixel window; with stride 2(The effect is to halve the image size). In

summary, we can obtain the network structure diagram of VGG16, as shown in Figure 4. From it, we can see that the input of VGG16 network is a fixed size 224×224 RGB image. The only pre-processing is subtracting the mean RGB value, computed on the training set, from each pixel.

B. Datasets and hardware

The VGG16 network was trained on the publicly available International Skin Imaging Collaboration (ISIC-2016) training dataset[12], which was selected for testing since the organizers provided real-world labels in both the ISIC-2016 and ISIC-2017 test datasets. The images in the ISIC-2016 dataset were 8-bit RGB with the resolution is $540 \times 722 \sim 4499 \times 6748$ pixels, and since the training and test datasets of ISIC-2016 contain 900 and 379 images, respectively, and the proportion of malignant images in the training and test sets is 19.2% and 19.8%, respectively, considering that malignant images account for a relatively small percentage and will have an

impact on the network classification effect, choosing 374 malignant tumors images from the ISIC-2017 dataset were added to the training and test datasets of ISIC-2016, in which 300 images were added to the training set and 74 images were added to the test set.

The network was implemented in the keras framework using the python programming language with a Tensorflow2-GPU backend, and the experiments were conducted on a windows 11 operating system with the following hardware configurations: AMD RYZEN5 4000 series CPU @ 3.60 GHz \times 16 processor, GeForce GTX1660TI GPU with 8GB GDDR5 memory.

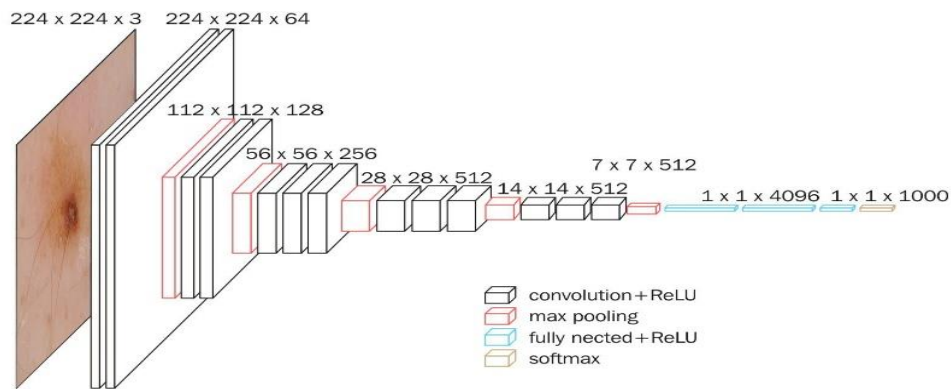


Figure 4. Overall network structure of VGG16

C. Calculation metrics

This study is a dichotomous classification of images, in the dichotomous case; the model finally needs to predict the outcome in only two cases. For each category our predictions are obtained with probabilities P and $1-P$, when the expression of the cross-entropy loss function as shown as equation (5).

$$L = \frac{1}{N} \sum_i -[y_i \log p_i + (1 - y_i) \log(1 - p_i)] \quad (5)$$

Where y_i denotes the label of sample i , positive class is 1, negative class is 0. p_i denotes the probability that sample i is predicted to be a positive class.

Comparisons were made using the most common skin lesion segmentation evaluation metrics, including Accuracy, Precision, Sensitivity, Recall, and Specificity. These metrics were used in the evaluation of the mode. They are illustrated below:

Accuracy: It measures the proportion of true results (both true positives and true negatives) among the total number of cases examined. Accuracy is expressed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Precision: It measures the accuracy and represents the proportion of examples classified as positive that are actually positive. Precision is expressed as:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

Sensitivity: It measures the proportion of those with positive values among those who are actually positive. Sensitivity is expressed as:

$$Sensitivity = Recall = \frac{TP}{TP + FN} \quad (8)$$

Specificity: This is the proportion of those that are negative among those who actually tested negative. Specificity is expressed as:

$$Specificity = \frac{TN}{TN + FP} \quad (9)$$

Eval_Top1: This is the label of the network prediction that takes the largest one inside the final probability vector as the prediction result, and if the one with the largest probability in the prediction result is correctly classified, the prediction is correct. Eval_Top1 is expressed as:

$$Eval_Top1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (10)$$

Eval_Top5 is the top five with the largest probability vector at the end, as long as the correct probability occurs, the prediction is correct. Eval_Top5 is expressed as:

$$Eval_Top5 = \frac{(1 + 5^2 \cdot Precision \cdot Recall)}{5^2 \cdot Precision + Recall} \quad (11)$$

Where FP is the number of false positive pixels, FN is the number of false negative pixels, TP is the number of true positive pixels and TN is the number of true negative pixels.

In this study, the network is evaluated using a confusion matrix, where each column of the confusion matrix expresses the category prediction of the classifier for the sample, and each row of the matrix expresses the true category to which the sample belongs. As shown in Table 2.

TABLE II. CONFUSION MATRIX UNDER BINARY CLASSIFICATION

Confusion Matrix		Predict	
		0	1
Real	0	a	b
	1	c	d

The following three equations can be obtained from the Table 2:

$$Precision = \frac{a}{a + c} \quad (12)$$

$$Recall = \frac{a}{a + b} \quad (13)$$

$$Accuracy = \frac{a + d}{a + b + c + d} \quad (14)$$

Precision, Recall and other parameters calculate the characteristics of a certain classification, while Accuracy is a criterion to determine the overall classification model.

III. RESULTS AND DISCUSSION

During the classification of melanoma lesion regions, this system was evaluated on two publicly available databases. First, the model was trained on the ISIC 2016 dermoscopy dataset using 1200 training skin lesion images. Then it was tested on 453 skin lesion images. As shown in Figure 5, the accuracy of classification of melanoma properties reached 96% at a training step of 100 epochs, and Figure 6 indicates that the loss value of the network was reduced to below 0.2 at the end of training.

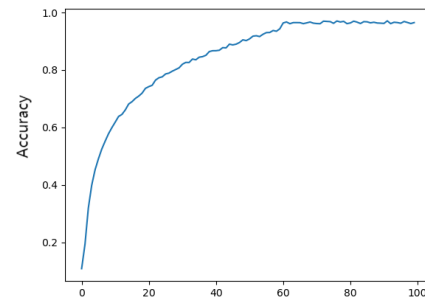


Figure 5. Training Accuracy Curves of the proposed method on the ISIC-2016 datasets

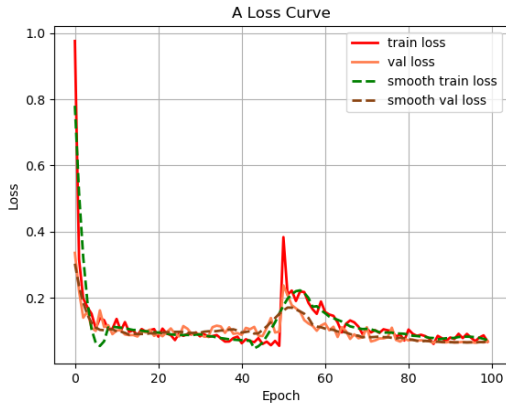


Figure 6. Training Loss Curves of the proposed method on the ISIC-2016 datasets

This study is a binary classification problem to classify melanoma images into benign and malignant tumors, and the coefficients of Eval_Top1 and Eval_Top5 calculated using the confusion matrix reached 96% and 99%, respectively, as shown in Figure 7 and Figure 8. Calculating the classification accuracy of two classes, benign and malignant tumors, and then finding the overall average accuracy is a very common evaluation metric for classification problems. After we calculate the confusion matrix, we need to quantitatively analyze the confusion matrix, and one of the most obvious metrics is to calculate the classification accuracy.

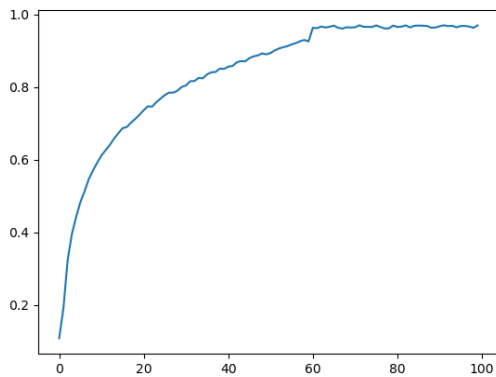


Figure 7. Training Eval_Top1 Curves of the proposed method on the ISIC-2016 datasets

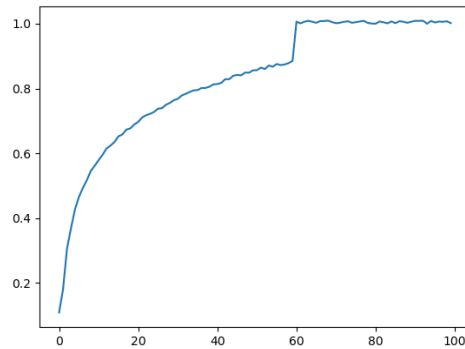


Figure 8. Training Eval_Top5 Curves of the proposed method on the ISIC-2016 datasets

Figure 9 shows the results of the method proposed in this study for predicting four malignant melanoma images, with an average accuracy of 94.9%

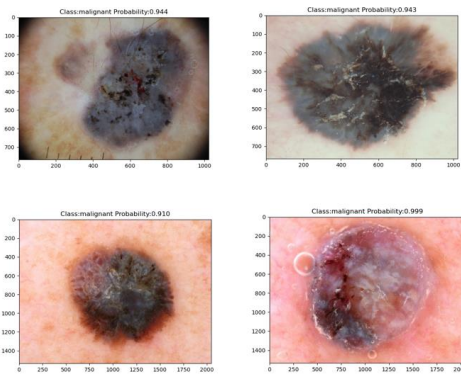


Figure 9. Prediction results of the proposed method for malignant melanoma in this study

Figure 10 shows the results of the method proposed in this study for predicting four benign melanoma images, with an average accuracy of 97.3%

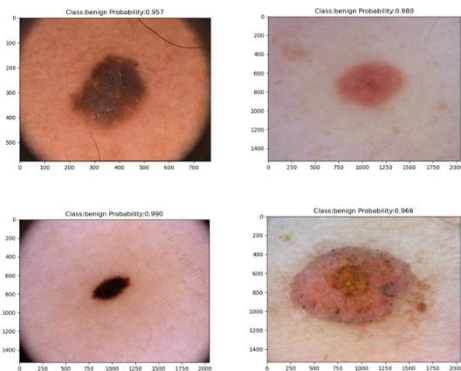


Figure 10. Prediction results of the proposed method for benign melanoma in this study

As can be seen from Figure 5, the results produced by the network trained on the ISIC 2016 dataset, the accuracy and the values of Eval_Top1 and Eval_Top5 can still be improved with the increase of the training steps and the dataset. The learning ability of the proposed model is evaluated by experiments on the improved ISIC-2016 dataset, and the accuracy curves are shown in Figure 5. The results of the curves clearly show that the ISIC 2016 dataset with a larger dataset achieves an accuracy of 96%. This improvement is due to the adoption of the cross-entropy loss function in the softmax classifier.

IV. CONCLUSIONS

In this paper, we propose a deep convolutional network-based architecture for robust detection and classification of melanoma lesion regions. The architecture uses the best model from the VGGNet model-VGG16, which uses multiple convolutional layers with smaller convolutional kernels instead of convolutional layers with larger convolutional kernels, aiming to reduce the network parameters and increase the fitting and expression capability of the network, and finally a new method is designed to classify benign and malignant melanomas based on the results of the softmax classifier. It was shown that the network depth facilitates the classification accuracy and enables state-of-the-art performance on ImageNet challenge datasets using the traditional ConvNet architecture[13]. The network with 16 weight layers used in this study boasts excellent performance in terms of final classification results. VGG16 also generalizes well to a wide range of tasks and datasets, matching or outperforming more complex recognition pipelines built around image representations with shallower depth [11]. The method proposed in this paper is feasible in medical practice, with an average processing of 5 seconds per melanoma image. The system was evaluated on one publicly available dataset of dermatological lesion images, and the overall accuracy and Eval_Top1 to Eval_Top5 coefficients of the system on the ISIC-2016 dataset were 96%, 96%, and 99%, respectively. In conclusion, this study has some reference value for

the classification of dermatological lesion images [14].

REFERENCES

- [1] National Cancer Institute. PDQ Melanoma Treatment. Bethesda, MD, USA. (Nov. 4, 2019). PDQ Adult Treatment Editorial Board. Accessed: Dec. 9, 2019.
- [2] Cancer Statistics Center. (2019). American Cancer Society.
- [3] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss, "A methodological approach to the classification of dermoscopy images," *Computerized Med. Imag. Graph.*, vol. 31, no. 6, pp. 362–373, Sep. 2007.
- [4] G. Capdehourat, A. Corez, A. Bazzano, R. Alonso, and P. Musé "Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions," *Pattern Recognit. Lett.*, vol. 32, no. 16, pp. 2187–2196, Dec. 2011.
- [5] M. E. Celebi, H. Iyatomi, W. V. Stoecker, R. H. Moss, H. S. Rabinovitz, G. Argenziano, and H. P. Soyer, "Automatic detection of blue-white veiled related structures in dermoscopy images," *Computerized Med. Imag. Graph.*, vol. 32, no. 8, pp. 670–677, Dec. 2008.
- [6] Q. Abbas, M. Celebi, C. Serrano, I. F. N. Garcia, and G. Ma, "Pattern classification of dermoscopy images: A perceptually uniform model," *Pattern Recognit.*, vol. 46, no. 1, pp. 86–97, Jan. 2013.
- [7] Bi, L, J Kim, E Ahn, A Kumar, M Fulham, D Feng "Dermoscopic Image Segmentation via Multistage Fully Convolutional Networks." *IEEE Trans Biomed Eng*, 2017. PP(9): p. 1-1.
- [8] Bi, J. Kim, E. Ahn, D. Feng, and M. Fulham, "Automatic melanoma detection via multi-scale lesion-biased representation and joint reverse classification," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 1055–1058.
- [9] Hasan, M.K, L Dahal, PN Samarakoon, FI Tushar, RM Marly. "DSNet: Automatic dermoscopic skin lesion segmentation". *Pergamon*, 2020. DOI:10.1016/J.COMBIOMED.2020.103738.
- [10] Revathi and A. Chithra, "A review on segmentation techniques in skin lesion images," *Int. Res. J. Engg Tech.*, vol. 2, no. 9, 2015.
- [11] Simonyan, K., and A. Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *Computer Science* (2014).
- [12] Codella, Ncf, Gutman D, Celebi, M.E, Helba, B, Marchetti, Dusza, S. W, Kalloo, A, Liopyris, K, Mishra, N, Kittler, H. *Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the 2017 International Symposium on Biomedical Imaging (ISBI)*, Hosted by the International Skin Imaging Collaboration (ISIC). 2017.
- [13] Hinton, G.E., Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. 2012. p. 212-223.
- [14] Version, P., *Plasma Cell Neoplasms (Including Multiple Myeloma) Treatment (PDQ®) - National Library of Medicine PubMed Health*. National Cancer Institute, 2014.