Skin Lesions Segmentation and Classification for Medical Diagnosis

Merve Gun
Department of Computer Engineering
Istanbul Sabahattin Zaim University
Istanbul, Turkey
0000-0001-5190-9466

Alaa Ali Hameed
Department of Computer Engineering
Istanbul Sabahattin Zaim University
Istanbul, Turkey
0000-0002-8514-9255

Mirsat Yesiltepe
Department of Computer Engineering
Yildiz Technical University
Istanbul, Turkey
0000-0003-4433-5606

Akhtar Jamil
Department of Computer Engineering
Istanbul Sabahattin Zaim University
Istanbul, Turkey
0000-0002-2592-1039

***Abstract*— Classification and segmentation of various skin lesions play a very important role in the field of dermoscopy. Using computer-aided applications to detect cancerous cells and predict the lesion as benign and malignant can yield better results. Automatic estimation of skin disease from skin lesion images help practitioners to perform rapid diagnosis, provide early treatment and quick decision making. In this paper, Convolution Neural Network (CNN) is used to identify cancer-prone skin lesions from dermoscopy images. Experiments were performed on ISIC 2016 data set with two lesion classes (Malignant and Benign). The training was carried out with the Multiple Residual Neural Network (ResNet) architecture, where the data is pre-processed with different methods. Finally, the comparative analysis with other methods was also performed. The results indicated that the performance of our proposed method is also in line with state of the art methods.**

***Keywords*— lesion classification, Resnet, convolutional neural network, medical image analysis.**

I. Introduction

Advances in deep learning methods in recent years have also led to the development of intelligent medical imaging-based diagnostic systems that can help the analysis of medical images and help professionals make better decisions for patient health.

Some lesions on the skin can be the harbinger of serious cancer types. Especially early detection of melanoma is very important for the success of the treatment process. In this project, skin lesion classification is made. A deep learning-based approach is presented to classify a dermoscopic image containing a skin lesion as malignant or benign.

Malignant is a type of skin cancer that occurs in the skin tissue and can cause death. The disease can be treated with early diagnosis. Benign is a more common but not dangerous type of skin lesion. The dermoscopic images of these two species are similar and may not be distinguished by human vision. For this reason, Computer Aided Diagnosis (CAD) systems, which are frequently used in the detection of diseases, can help patients and physicians in the detection of skin cancer [1]. Even if expert dermatologists use dermatology images for diagnosis, the accuracy rate of the expert diagnosis is estimated at 75–84% [2].

In the study conducted by Hameed et al. [3] in 2019, they used AlexNet CNN model for feature extraction on 9144 skin images. For classification, they reached an accuracy of 86.21% with SVM classifier. They showed that the features obtained from CNN models by using a 10-fold cross-validation technique to prevent overcompliance increased the classification performance of multiple skin lesions. In the study conducted by Ahmad et al. [4] in 2020, they proposed a new frame by fine-tuning the layers of ResNet152 and InceptionResNet-V2 models to the skin lesion images they received from Wuhan Hospital in China and achieved an accuracy of 87.42%.

Studies conducted with Deep Learning methods have shown that better prediction success has been achieved in many areas such as image classification compared to machine learning methods [1]. In the study of Rodrigues et al. [5] for the year 2020, it was seen that using CNN and classical machine learning methods together gave better predictions according to the results obtained by comparing them.

Deep Learning models do not need extra preprocessing to extract features that represent image data. CNN is the most widely used architecture among deep learning models. In 2015, He et al. [6] proposed a new CNN model called ResNet with a low computational cost. ResNet achieved a great success in the 2015 ImageNet competition with an error rate of 3.57% [1]. However, very deep neural networks are difficult to train as the number of parameters needed to training exponential increase with increasing number of layers increase. In addition, highly powerful computational resources are also required to train the algorithm faster. In theory, training error is expected to decrease as the depth increases. However, in reality, the training error increases with adding more layers to CNN. ResNet architecture has brought a solution to the problem of training error that gets worse / disappears as its depth increases.

Fig. 1 represents the training error (%) and test error (%) of a 20-layer and 56-layer flat network. It was observed that as the depth increased, that is, the number of iterations, training and test data errors increased.

In this paper we investigate the power of deep neural network using CNN for classification of skin lesion from images. Specifically, we employed ResNet on ISIC 2016 data set with for classification of lesion into malignant or benign.

The rest of the paper is organized as follows. The data set used in this study is described in section II. The proposed methodology for classification and segmentation is described in Section II. Section IV summarized the results obtained and paper was completed with concluding remarks.

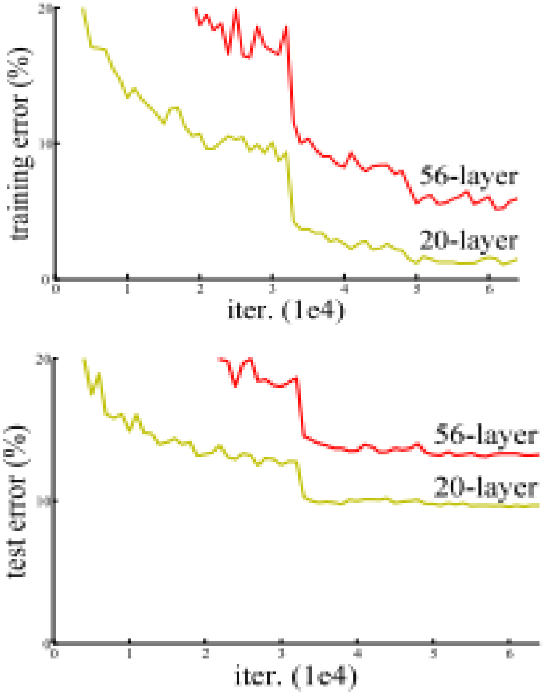


Fig. 1. Training error (top) and test error (down) [6]

II. DATA SET

Dermoscopic images are commonly used by dermatologists for various disease identification. In this study, data was obtained from International Skin Imaging Collaboration (ISIC) [7] which is freely available online. This data set includes total 1279 skin lesion images in RGB format which hare further divided into training (900) and testing (379) subsets. There are two classes in this data set: Malignant and Benign. Fig. 2 shows some samples for malignant tissues (a) and benign tissues (b).

(a)

(b)

Fig. 2. Sample images a) Malignant Tissue Image, b) Benign Tissue Image

TABLE I. DATA SET USED IN THIS STUDY

Class	Training	Test	Total
Malignant	173	75	248
Benign	727	304	1031
Total	**900**	**379**	**1279**

III. PROPOSE METHOD

A. *Deep Learning (DL)*

Deep Learning is a new machine learning method based on artificial neural networks that can automatically extract features from data and generally produces higher accuracy. Unlike classical machine learning methods, it does not require manual feature engineering. CNN is one of the most successful deep learning models for image data which is described in following section.

B. *Convolutional Neural Network (CNN)*

There are many CNN models such as ResNet, GoogleNet, VGGNet and AlexNet. CNN has a multiple neurons arranged layer structure [1]. These include Input Layer, Convolution Layer, Pooling Layer, Fully-Connected Layer and Output Layer.

In Convolution Neural Network, neurons are expressed in 3 dimensions; width, height and depth. Depth is equal to 3 due to the RGB image format, for example if the image data is color.

The basic architecture of Convolutional Neural Network is shown in Fig. 3 while fig.4 shows the architecture of ResNet18. In addition to these basic layers, the layers of many different CNN architectures vary.

1. *Convolutional layer:* They are the most important elements of deep learning architectures. Each layer consists of a certain size and fixed number of filters selected before the training process. Each filter detects a specific feature and produces a feature map in response [9].

   Filters create convolutional layers and are trainable feature extractors that are learned from data. Filters have the ability to learn properties of images such as color, size, edge during training.

2. *Pooling layer:* The Pooling layer, a subsampling layer, reduces the size of feature maps. As a result, only the activated properties are transmitted to the next layer [1].

3. *Fully connected layer:* It acts as a classifier. It usually consists of 1-2 fully bonded layers. Next comes the classification layer for multi-class data.

C. *Residual Network (ResNet)*

Residual blocks form the basis of the restoration architecture. One of the problems of deep learning is the problem of overfitting due to increasing depth. Overfitting means good training accuracy but poor test accuracy. Blocks are now suggested to solve this problem. Now in the block the input x is directly added to the output of the network i.e. it becomes f(x) + x.

In Figure 5, the output generated when a block is added to the CNN architecture is now summarized. In this case, output is given as in (1);

$$H(x) = F(x)+x \qquad (1)$$

When the input of the network is equal to its output, F (x) = 0 ‘. So H (x) = x. The aim is to skip certain layers using jump links or leftover blocks to improve performance in the deeper layers of deep learning. Thus, improvement is provided to the problem of optimization and distortion in a deep network.

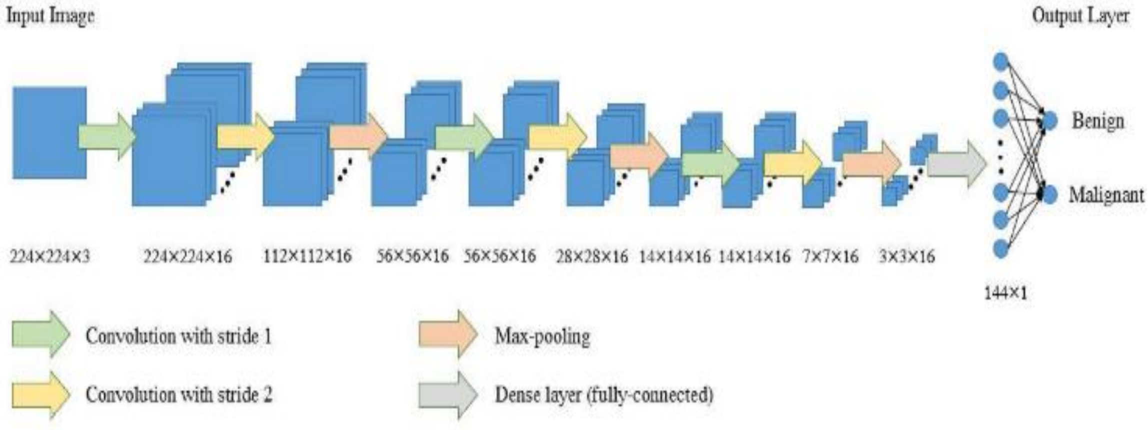


Fig. 3 Architecture of convolutional neural network [8].

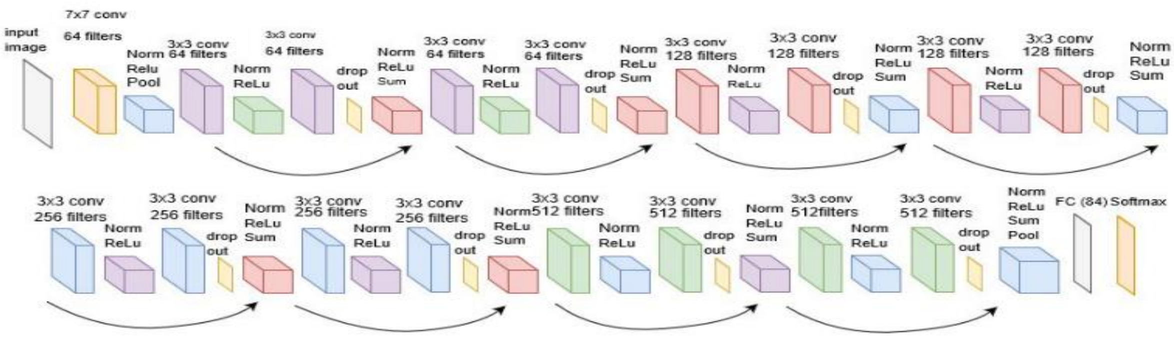
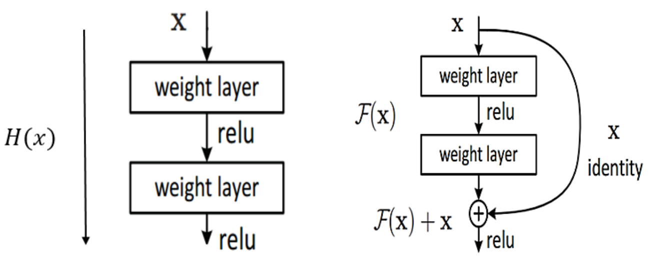


Fig. 4. ResNet18 Layers Architecture [10]



a. b.

Fig. 5 (a). Normal CNN architecture, b) Residual connection CNN [6]

IV. EXPERIMENTAL RESULTS

In this study, 900 images were used for training and 379 images for testing. After training the data set using CNN, the classification and accuracy of skin lesions were calculated by learning with the ResNet18 model. We trained our model with 30 epochs. Accuracy and loss values are compared by using different optimizers to analyze their performance.

The images of our data set are divided into 2 classes. To measure the performance of the network, sensitivity, specificity, precision and F1 values were calculated and compared with other research results.

A. *ResNet Classification*

In our study, we applied a classification to our data set with the ResNet18 architecture. We resized the images we used in the data set to 224x224x3. ResNet18 architecture consists of 5 Convolution Layers, 2 Max Pooling Layers and 1 Fully Connection Layer (Figure 4). Since we have 2 classes in the data set, the "Output Size" value is arranged as "2" in the Fully Connection Layer. Accordingly, the "Classes and Output Size" values in the classification layer have been updated as "auto".

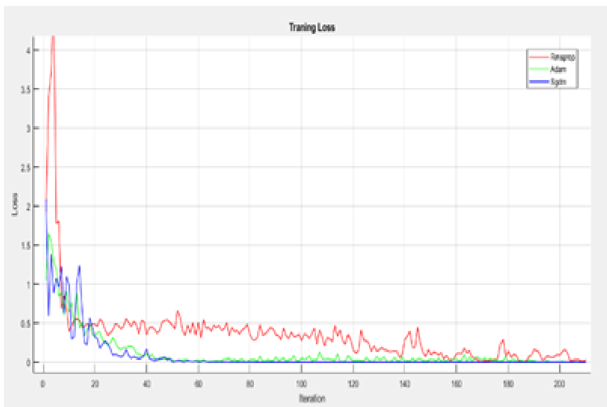


Fig. 6. Training accuracy according to different optimizers

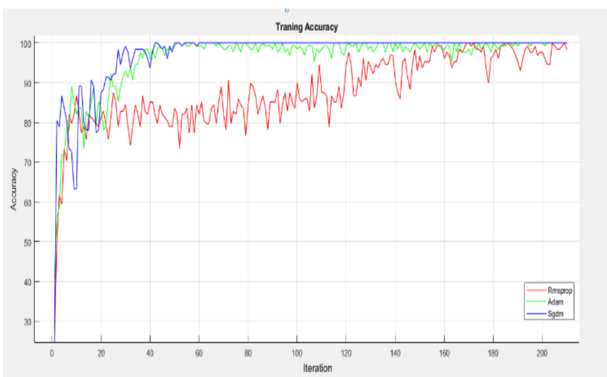


Fig. 7. Loss values according to different optimizer

B. Training

The image data was resized to 224x224x3. ReLU activation function has been selected as the activation function. Training data and test data were fed into classifier by loading into the image data store in the Matlab environment.

Table II summarizes the results obtained for each class using confusion matrix while Table III summarizes the results obtained for each classifier. The results show that Sgdm produced highest classification accuracy with 84,96%, then Adam which was 81,53%, and Rmsprop optimizer was 77,57%. It is also seen that the optimizer with the lowest training and test loss value is again Sgdm. For each optimizer, parameters were empirically obtained during the training. The accuracy and loss obtained for each classifier are shown in Fig. 6. and Fig. 7. Moreover, the quantitative results are presented in Table IV.

C. Evaluation

Confusion matrix is a special table layout that gives us the opportunity to visualize the performance of the algorithm. Each row of the matrix represents instances in a real class, and each column of the matrix represents instances in a predicted class [11].

TABLE II. CONFUSION MATRIX

		Predicted Values		
		Positive (Benign)	**Negative (Malignant)**	**Actual Totals**
Actual Values	**Positive (Benign)**	**290** 76,5%	**43** 11,3%	87,1% 12,9%
	Negative (Malignant)	**14** 3,7%	**32** 8,4%	69,6% 30,4%
	Predicted Totals	95,4% 4,6%	42,7% 57,3%	85,0% 15,0%

True Positive (TP) and False Negative (FN) values correspond to the number of lesions in a particular class that are true and false classified, respectively. True Negative (TN) refers to the number of lesions that do not belong to a particular class that are classified as not belonging to this class. False Positive (FP) values are the number of lesions incorrectly classified as belonging to a particular species [5].

In order to better evaluate the system performance, sensitivity (TPR - true positive rate) and specificity (TNR - true negative rate) measurements were calculated. In this case, sensitivity (3) means the ratio of appropriately classified Malignant cases to the number of Malignant cases, while specificity (4) is the ratio of appropriately classified Benign cases to the number of all Benign cases [9].

Accuracy measures the number of data correctly classified (2). High accuracy, sensitivity and specificity value indicate an efficient classification method [8].

$$Accuracy = \frac{\mathrm{TP + TN}}{\mathrm{TP + FP + TN + FN}} \quad (2)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{TN + FP} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$F1\_Score = \frac{(2 * Precision * Sensitivity)}{Precision + Sensitivity} \quad (6)$$

TP = True Positive TN = True Negative

FN = False Negative FP = False Positive

P - number of positive samples (Benign), N - number of negative samples (Malignant), TP-True Positive: Benign is correctly defined as Benign; FP-False Positive: Benign is incorrectly defined as Malignant; TN-True Negative: Malignant is correctly identified as Malignant; FN-False Negative: Represents Malignant image data incorrectly identified as Benign.

TABLE III. CLASSIFICATION ACCURACY (%)

Optimizer	**ACC.**	**SEN.**	**SPE.**	**PRE.**	**F1 Score**
Sgdm	**84,96**	**87,09**	**69,57**	**95,39**	**91,05**
Adam	81,53	84,82	55,81	93,75	89,06
Rmsprop	77,57	85,21	42,65	87,17	86,18

In the training process, optimizing the training with different parameters is important for different performance results. Accuracy, sensitivity, specificity, precision and F1 results that we obtained with different optimizers calculated according to the experiments performed Table III. As shown in ‘. According to the experimental results, it is seen that he is the optimizer Adam with high accuracy values.

TABLE IV. COMPARISON WITH THE MODELS (%)

Authors	**Method**	**Accuracy**	**Sensitivity**	**Specificity**
Ahmad et al. [4]	ResNet152	87,42	97,04	97,23
Kwasigroch et al. [9]	ResNet50	75,5	90	61
Vesal et al. [12]	SkinNet	93	93	90
Al-Masni et al. [13]	Inception-ResNet-v2	81,79	81,80	71,40
Pham et al. [14]	Inception-v3	87	70	91,2
Pham et al. [14]	ResNet50	79,5	80	96,2
Jayalakshmi et al. [15]	BN_CNN	83,05	83,06	83,23
Cabas et al. [8]	CNN	84,76	91,97	78,71
Wu et al. [16]	ResNet50	82	84	85
Our Model	**ResNet18**	**84,96**	**87,09**	**69,57**

In medical applications, the highest sensitivity coefficient is required because a malignant lesion that is misdiagnosed can seriously affect patient health.

The model Table IV, which we proposed with the methods, Accuracy, Sensitivity and Specificity values of previous studies using different deep learning methods. It has been compared in.

V. CONCLUSION

In this study, we investigated a deep convolution network, ResNet18, for classification of skin lesions from RGB images. The investigated model can automatically classify the given lesion image without need of traditional segmentation and feature extraction process. According to the experimental results, ResNet was effective for skin lesion classification. Moreover, Sgdm, Adam and Rmsprop optimizers were also investigated. Performance results were compared with previous research using various deep learning methods and the proposed model was found to be promising. We hope that the model presented in the study can be further developed into real applications where it can help experts for better diagnosis and treatment.

ACKNOWLEDGMENT

I would like to thank all the authors who contributed to the emergence of this study, the International Skin Imaging Collaboration (ISIC), who published the data set we used in the study, and Istanbul Sabahattin Zaim University.

REFERENCES

[1] O. Yıldız, "Melanoma detection from dermoscopy images with deep learning methods: A comprehensive study," *J. Fac. Eng. Archit. Gazi Univ.*, vol. 34, no. 4, pp. 2241–2260, 2019, doi: 10.17341/gazimmfd.435217.

[2] S. Jain, V. Jagtap, and N. Pise, "Computer aided melanoma skin cancer detection using image processing," *Procedia Comput. Sci.*, vol. 48, no. C, pp. 735–740, 2015, doi: 10.1016/j.procs.2015.04.209.

[3] N. Hameed, A. M. Shabut, and M. A. Hossain, "Multi-Class Skin Diseases Classification Using Deep Convolutional Neural Network and Support Vector Machine," *Int. Conf. Software, Knowl. Information, Ind. Manag. Appl. Ski.*, vol. 2018-Decem, pp. 1–7, 2019, doi: 10.1109/SKIMA.2018.8631525.

[4] B. Ahmad, M. Usama, C. M. Huang, K. Hwang, M. S. Hossain, and G. Muhammad, "Discriminative Feature Learning for Skin Disease Classification Using Deep Convolutional Neural Network," *IEEE Access*, vol. 8, pp. 39025–39033, 2020, doi: 10.1109/ACCESS.2020.2975198.

[5] D. de A. Rodrigues, R. F. Ivo, S. C. Satapathy, S. Wang, J. Hemanth, and P. P. R. Filho, "A new approach for classification skin lesion based on transfer learning, deep learning, and IoT system," *Pattern Recognit. Lett.*, vol. 136, pp. 8–15, 2020, doi: 10.1016/j.patrec.2020.05.019.

[6] K. He and J. Sun, "Deep Residual Learning for Image Recognition," 2016, doi: 10.1109/CVPR.2016.90.

[7] ISIC, "Challenge Isic Archive." https://challenge.isic-archive.com/data#2016.

[8] A. P *et al.*, "Deep Convolutional Neural Network for Melanoma Image Classification," no. September, p. 92027, 2020, [Online]. Available: http://repositorio.unan.edu.ni/2986/1/5624.pdf.

[9] A. Kwasigroch, A. Mikołajczyk, and M. Grochowski, "Deep neural networks approach to skin lesions classification - A comparative analysis," *2017 22nd Int. Conf. Methods Model. Autom. Robot. MMAR 2017*, pp. 1069–1074, 2017, doi: 10.1109/MMAR.2017.8046978.

[10] M. A. R. Alif, S. Ahmed, and M. A. Hasan, "Isolated Bangla handwritten character recognition with convolutional neural network," *20th Int. Conf. Comput. Inf. Technol. ICCIT 2017*, vol. 2018-Janua, no. March 2018, pp. 1–6, 2018, doi: 10.1109/ICCITECHN.2017.8281823.

[11] H. Zhou, F. Xie, Z. Jiang, J. Liu, S. Wang, and C. Zhu, "Multi-classification of skin diseases for dermoscopy images using deep learning," *IST 2017 - IEEE Int. Conf. Imaging Syst. Tech. Proc.*, vol. 2018-Janua, pp. 1–5, 2017, doi: 10.1109/IST.2017.8261543.

[12] S. Vesal, N. Ravikumar, and A. Maier, "SkinNet: A deep learning framework for skin lesion segmentation," *arXiv*, pp. 1–3, 2018.

[13] M. A. Al-masni, D. H. Kim, and T. S. Kim, "Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification," *Comput. Methods Programs Biomed.*, vol. 190, p. 105351, 2020, doi: 10.1016/j.cmpb.2020.105351.

[14] T. C. Pham *et al.*, "Improving binary skin cancer classification based on best model selection method combined with optimizing full connected layers of Deep CNN," *2020 Int. Conf. Multimed. Anal. Pattern Recognition, MAPR 2020*, pp. 1–6, 2020, doi: 10.1109/MAPR49794.2020.9237778.

[15] G. S. Jayalakshmi and V. S. Kumar, "Performance analysis of convolutional neural network (CNN) based cancerous skin lesion detection system," *ICCIDS 2019 - 2nd Int. Conf. Comput. Intell. Data Sci. Proc.*, pp. 1–6, 2019, doi: 10.1109/ICCIDS.2019.8862143.

[16] X. Li, J. Wu, E. Z. Chen, H. Jiang, and C. V Feb, "What evidence does deep learning model use to classify Skin Lesions ?"