Lesion Detection on Skin Images Using Improved U-Net

*Elif Işılay Ünlü**, *Ahmet Çınar*
Fırat University, Department of Computer Engineering, Elazığ, 23119, Türkiye
**Corresponding author: eiunlu@firat.edu.tr*

Abstract

One of the most prevalent cancers in humans is skin cancer. The deadliest form of skin cancer is malignant melanoma and the incidence rate has increased rapidly in recent years. In the treatment of melanoma, early diagnosis is very critical. It is difficult and time consuming to automatically detect melanoma from images taken from dermoscopy devices. Computer-aided systems are needed, therefore. In this paper, a deep learning-based method for melanoma segmentation and classification with color images taken from dermoscopy devices is proposed. This technique uses ISIC 2017 International Skin Imaging Collaboration.

In this paper, for segmentation and classification measures, 1317 skin images taken from the ISIC archive were used. The approach is based on the architecture of Preprocessing, U-Net and VGGNet. Operations such as mean subtraction, image normalization, image cropping, and scaling are implemented in the preprocessing phase. It is intended to make pictures of the skin more convenient before segmentation. The training precision rate and jaccard similarity coefficient reached 93% as a result of segmentation with these results, and the dice coefficient reached 79%. The accuracy rate is 85.5% as a result of the classification in the two-class dataset in the pre-trained VGG16 network. The accuracy rate of dataset classification obtained with cross-validation is 95.86%.

Keywords: deep learning, image segmentation, image classification, melanoma, U-Net, VGGNet

Introduction

The need for more computer-aided systems has contributed to increased health concerns and forms of illnesses. Cancer cases are among the most significant of these diseases. Cancer is one of the leading causes of death in many countries in recent years. In the treatment of cancer disorders, early diagnosis is of high significance. Skin cancer is another significant kind of cancer that needs early diagnosis. Skin cancer is more common than breast, prostate, and colon cancer cases [1]. Melanoma is a deadly type of skin cancer. According to studies conducted in 2015, almost 160,000 people are diagnosed with melanoma every year and 48,000 people die from the disease [2].

In the detection of melanoma, examination of color images from dermoscopy equipments plays a significant role. However, changes in the appearance of melanoma and similarities with other benign tumors may cause wrong decisions among expert dermatologists. Using deep learning methods, structures that human perception cannot recognize can be detected and early diagnosis can be made in patient treatment [3].

Deep learning is an algorithm that is inspired by the brain and named artificial neural networks, one of the sub-branches of machine learning. Deep learning techniques are rooted in neural networks, where 'deep' refers to multiple layers layered on top of each other. A deep learning architecture is comprised of stack layers, each of which is expected to subtract from higher-level representations, ie features. For example, an image consists of pixels, and a deep model learns the edges of a picture, object parts, and objects from layers. Deep learning approaches are used in classification tasks, natural language processing, image processing, speech recognition, and are much more powerful than shallow architectures [4].

Deep learning algorithms have become a rapidly preferred methodology to analyze medical images, especially in convolution networks. Deep learning methods in the medical field in general; Its use for image classification, object detection, segmentation, recording and other tasks has become increasingly common. Image segmentation is the classification of each pixel on the image, creating a map of all the object areas on the image. Segmentation divides an image into various meaningful regions containing identical attributes for each pixel. Similar brightness may be present in the image, and this brightness may reflect objects in different areas of the image. Segmentation can be used to identify, classify, and diagnose object particles with the same brightness within the image [5].

As convolutional neural networks improve, classification and segmentation processes produce more efficient outcomes. Studies using different models and architectures in the literature are discussed.

Many studies have been conducted on the detection of melanoma from images since the 2000s. In the first studies, k-nearest neighbor (kNN), support vector machines (SVM) etc. classical machine learning methods were used. Divided 5393 different images into three classes: healthy, diseased, and diseased [6]. As a result of this research, they were able to reach 87% sensitivity and 92% specificity with the kNN classifier.

Celebi et al. performed binary classification as healthy and diseased using support vector machines (Support Vector Machine-SVM) after operations such as border detection and color removal for lesion segmentation. In this study, which was carried out using classical methods, an accuracy value of approximately 92% was reached [7].

Kawahara et al. have applied deep learning to 10 different skin diseases. The proposed solution achieved a favorable precision of 85.8% for 5 classes, with major increases of 60% precision for low-represented classes. The system achieved an accuracy of 81.8% in all 10 datasets of 1300 images collected from a regular (non-dermoscopic) camera [8].

Codella et al. performed a melanoma classification by extracting features in dermoscopic images with deep learning methods and classical machine learning methods, and an accuracy value of 93% was achieved [9].

Esteva et al. suggested a system using a single CNN to identify skin lesions by directly inputting pixels and disease tags. A dataset of 12932 clinical images and 2,032 different diseases has been used to train CNN. The proposed approach has been tested against at least 21 dermatologist information. In the 3-classification task, CNN reached 72.1% total accuracy and 65.56% and 66.0% precision in a subgroup of two dermatologist validation sets. In the second task, the classification with 9 classes, CNN reached 55.4% accuracy, while dermatologists achieved 53.3% and 55.0% accuracy [10].

Dandi et al. suggested a system for lesion quality detection. For these two tasks, they used a U-Net model consisting of three architectures. For the lesion segmentation task, they used 192 U-Net models with different architectures. Lesion feature detection was evaluated as five separate binary classification problems. Each feature has been identified by a classifier that integrates 60 U-Net models. In the given verification set, they were able to reach the threshold of 0.820 in task 1, Jaccard index in task 2, and 0.432 in task 2 [11].

Nasr-Esfahania et al. proposed a new class of fully convolutional networks with new dense pool layers for segmentation of lesion sites on non-dermoscopic images. Based on the Dermquest dataset, the dice score achieved in segmentation of skin lesions is 91.6% and outperforms the most advanced algorithms [12].

A new approach consisting of two steps based on very deep CNNs was proposed by Lequan to resolve automatic melanoma detection difficulties in dermoscopic images: segmentation and classification. They obtained 0.949 accuracy, 0.897 dice, 0.829 jaccard scores [13].

Li and Shen proposed two different deep learning methods for tasks such as lesion segmentation, feature extraction, and lesion classification. These methods were tested on the ISIC 2017 dataset. For different tasks, accuracy values of 0.753, 0.848, and 0.912 were obtained [14].

Yuan, Chao, and Lo used deep convolution-deconvolution neural networks (CDNN) for segmentation of skin lesions on dermoscopic images. And they obtained an average of 0.784 Jaccard indices in the dataset [15].

Tang et. al. suggested a U-Net based skin lesion segmentation method. In the application results, they obtained 93.03% and 86.93% Dice coefficient, 89.40% 79.26% Jaccard index values for two different datasets [16].

Unlu and Cinar, proposed a study on the classification of skin images. They performed binary and multiple classifications using the AlexNet model. They achieved an average of 81% accuracy, best 86% on average [17].

In the second part of the paper, the developed model is introduced. The dataset used for the developed method and preprocessing stages are mentioned. The architecture used in the segmentation and classification of skin images is explained with a deep learning-based system and interpreted together with the application results. The outcomes obtained and the studies planned for the future are listed in the last section.

Material and Method

In this paper, deep learning methods that give successful results are used. First with colored skin images taken from dermoscopy devices using U-Net, which is one of the conventional neural network models, lesion segmentation is performed. Then with VGGNet, classification is completed. In the Application Results section, the results are interpreted in detail with tables and graphics. In this section, dataset, preprocessing, the structure of the developed model, and CNN layers used in the developed model are examined. An overview of the proposed method is as in Figure 1.

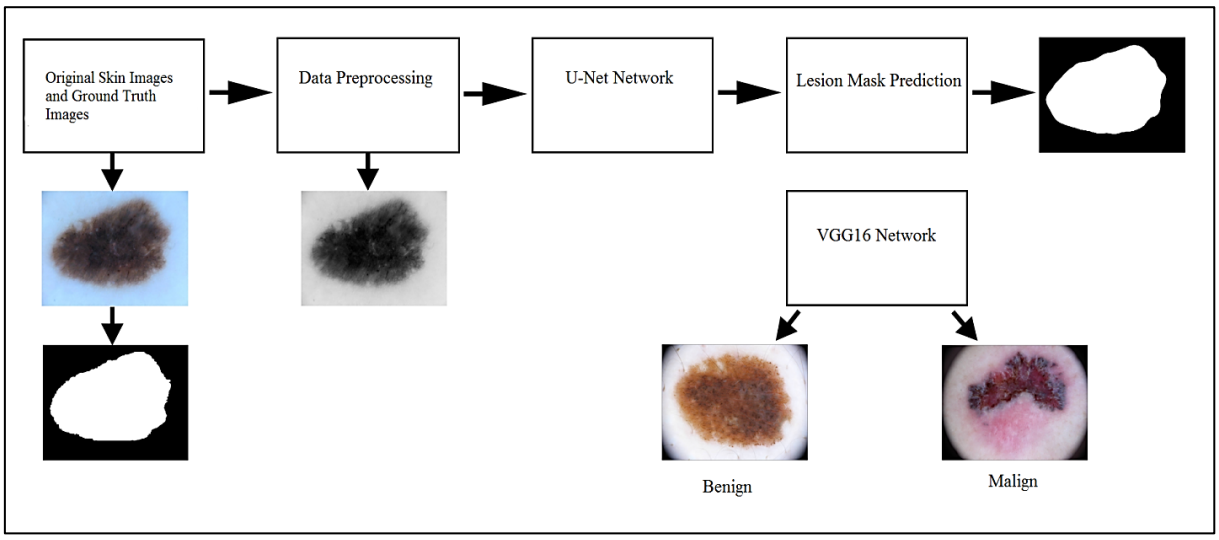


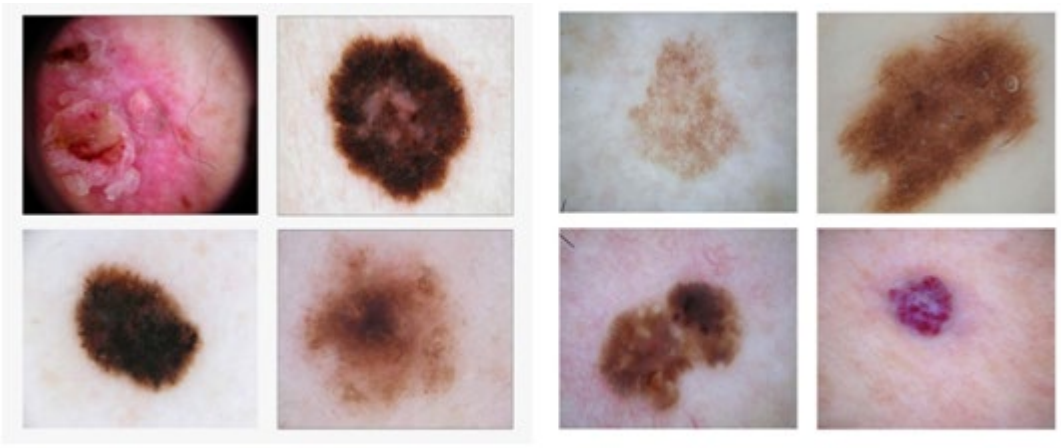
Figure 1. General view of the proposed method

Dataset

The ISIC (International Skin Imaging Collaboration) publicly accessible archive was used as a dataset in this paper [18]. 1317 skin images taken from the ISIC archive were used for segmentation and classification steps. Of these, 920 are grouped for training, 268 for testing, and 129 for validation. Images are labelled as benign or malignant and have a binary image mask in the image to label the lesion. The distribution of the images in the dataset is as in Table 1. Some examples of the original images are as in Figure 2.

Table 1. The distribution of dataset

Dataset Category	Malign	Benign
Train	153	767
Test	70	198
Validation	25	104



a. *Melanoma* **b.** *Non-melanoma*

Figure 2. a. Diseased images, b. Healthy images [18]

Preprocessing

Some preprocessing techniques are applied to the images before the segmentation step. These; Mean subtraction: A mean subtraction is applied to the image properties to center the RGB values cloud to approximately zero from the input data across each dimension of the image. The average subtraction calculation of an RGB image is as given in equations 1,2 and 3.

$$Rnorm(x,y) = R(x,y) - Rmean(x,y) \quad (1)$$

$$Gnorm(x,y) = G(x,y) - Gmean(x,y) \quad (2)$$

$$Bnorm(x,y) = B(x,y) - Bmean(x,y) \quad (3)$$

Image normalization: It is obtained by dividing each RGB size of the input images by standard deviation, normalizing the values normalized from the original 0 and 255-pixel values to 1 and 0. The linear normalization of a digital image is performed

$$Output_{channel} = 255 * \frac{(Input_{channel} - min)}{(\max - \min)} \quad (4)$$

Image cropping and resizing: To be approved by the architecture, the input images are preprocessed, but the image is dimensioned as required and resized to the original image as 64x80 pixels for U-Net and 224x224 pixels for VGG-16. Image examples before and after the preprocessing are as in Figure 3.

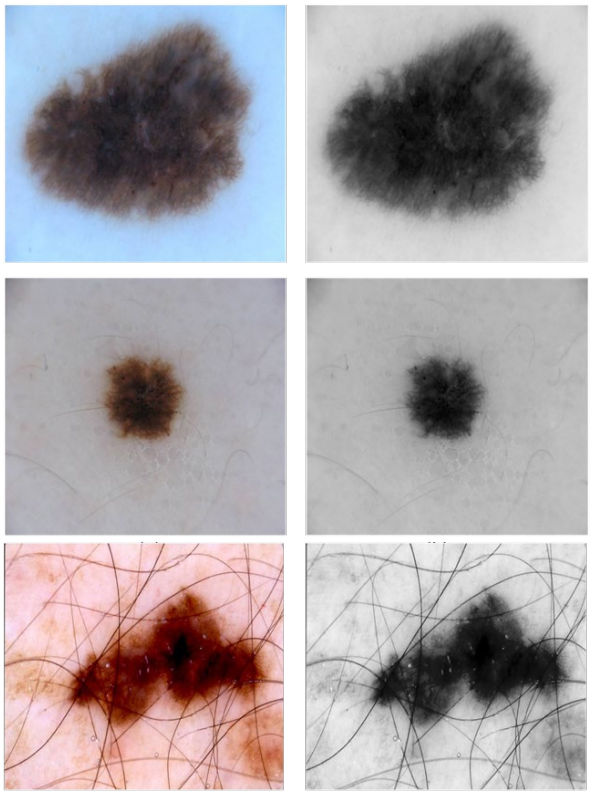


Figure 3. Original skin images b. Preprocessed skin images

Developed Model

U-Net Architecture

In the segmentation step of the developed system, the data are trained using one of the U-Net deep learning models. The structure of the developed model is given in Figure 4. The architecture of the U-Net is based on a Completely Convolutional Network (FCN) and has been updated to enhance medical imaging segmentation. There are two major differences from FCN. The U-net is symmetric and a concatenation operator instead of the complete operator applies the skip connections between the sub-sampling and the sampling path. Because of its symmetric nature, there are many feature maps that allow the transmission of information on the sampling path of the network. It consists of a downsampling

path (left side) and an expanding sampling path (right side). The downsampling path is a typical convolutional network. It consists of 4 blocks and each block contains 2 3x3 convolution layers + activation functions (with batch normalization) and one 2x2 maximum pooling layer. The number of feature maps is doubled in each pooling process. The purpose of this downsampling path is to get the content of the input image for the segmentation process. This contextual data is then transferred via skip links to the sampling path. The expansion / upsampling path consists of 4 blocks. These blocks include the stride 2 deconvolution layer, the feature map from the lower sampling path, the 3x3 convolution layer + activation function (by batch normalization) steps. U-Net obtains the general information required to predict a good segmentation map by combining the information in the last sampling path with the information in the sampling path [19].

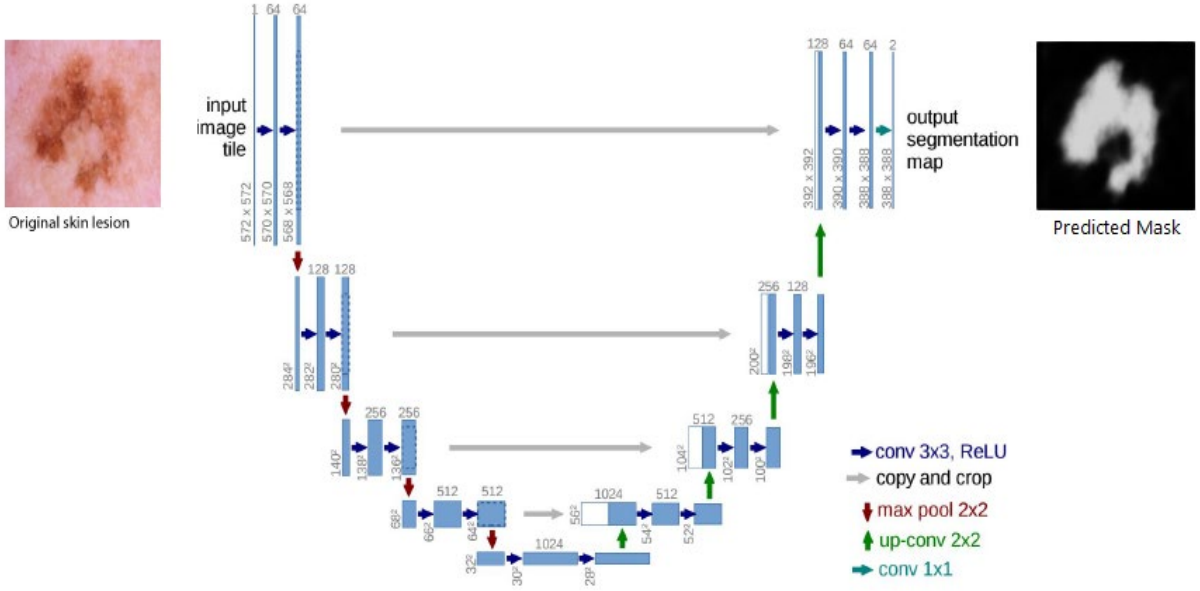


Figure 4. U-Net Architecture

VGGNet Architecture

The developed method uses VGGNet convolutional neural network architecture and transfer learning in the classification step. The proposed method model starts with weights from the VGG-16 trained in a larger dataset (like ImageNet), a process known as transference learning. The advantage of using transfer learning is that the model is pre-trained. Thus, the features required for classification are already learned. The benefit of training only to the top layers of the network is that it realizes the removal of more specific features. Therefore, the first four layers in the original VGG-16 architecture are initialized with weights used in the ImageNet dataset. With the weights registered during the extraction of features, the fifth and final convolution layers block is started. VGGNet is a widely used architecture for convolutional neural networks [20]. The structure of this model is as in Figure 5 and Table 2. It has two different structures, VGG-16 and VGG-19. In the VGG16 model, the input of the first convolution layer is a fixed size 224 x 224 RGB image. The image is passed through a convoluted layer stack with 3 x 3 filters. There are five maximum pooling layers after the convolution layers. Max-pooling is performed on a 2×2 pixel window, with stride 2. Three Fully Connected (FC) layers, each with 4096 channels, come last. The last layer is the softmax layer. Flattened linearization (ReLU) is used as the activation function in all layers.

$$Softmax: \sigma(x)_i = \frac{e^{x_i}}{\sum_{j=1}^{k} e^{x_j}}, i = 1,2,\dots,k \quad (5)$$

$$ReLU: f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (6)$$

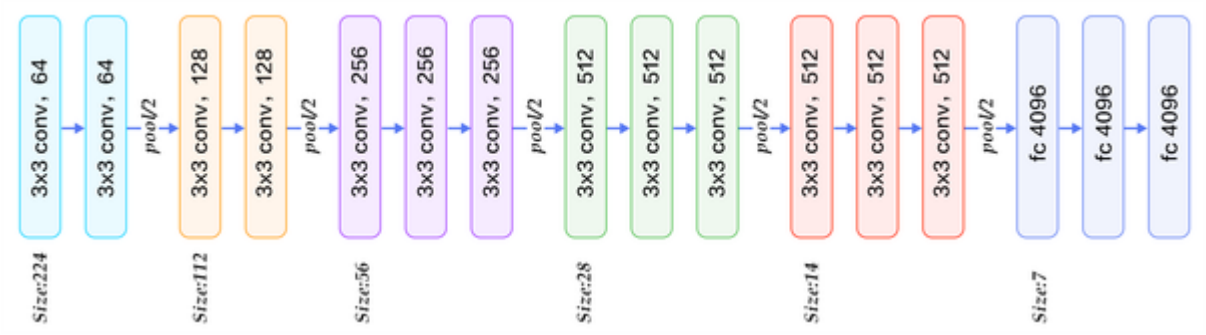


Figure 5. VGG16 architecture

Table 1. VGG16 layers

	Layer	Feature Map	Size	Kernel size	Stride	Activation
Input	Image	1	224 x 224 x 3	-	-	-
1	2 X Convolution	64	224 x 224 x64	3 x3	1	relu
	Max Pooling	64	112 x 112 x 64	3 x3	2	relu
3	2 x Convolution	128	112 x 112 x128	3 x3	1	relu
	Max Pooling	128	56 x 56 x 128	3 x3	2	relu
5	2 X Convolution	256	56 x 56 x 256	3 x3	1	relu
	Max Pooling	256	28 x 28 x 256	3 x3	2	relu
7	3 X Convolution	512	28 x 28 x 512	3 x3	1	relu
	Max Pooling	512	14 x 14 x 512	3 x3	2	relu
10	3 X Convolution	512	14 x 14 x 512	3 x3	1	relu
	Max Pooling	512	7 x 7 x 512	3 x3	2	relu
13	FC	-	25088	-	-	relu
14	FC	-	4096	-	-	relu
15	FC	-	4096	-	-	relu
Output	FC	-	1000	-	-	Softmax

Application and Results

In this paper, it is aimed to segmentation and classify skin lesion images. The dataset for the segmentation process is organized into training, training masks and test folders. Binary mask images obtained from specialist dermatologists are given as ground truth files. The preprocessed skin images are trained with 32 batch sizes for 200 epochs in the U-Net network. Original skin images and the estimated masks obtained as a result of the model are given in Figure 6. The segmentation results are obtained by a calculation between the binary mask and the ground truth mask automatically generated for each test mask. Some of the metrics used in the evaluation stage of the segmentation and their definitions are as follows:

Accuracy calculates the number of accurate estimates and divides by the total number of samples, as given in equation (7).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{7}$$

The Dice coefficient is calculated as in equation (8) by comparing the predicted segmentation (X) image corresponding to the ground truth (Y) image, in pixels.

$$Dice\ Coefficient = \frac{2 * |X \cap Y|}{|X| + |Y|} \tag{8}$$

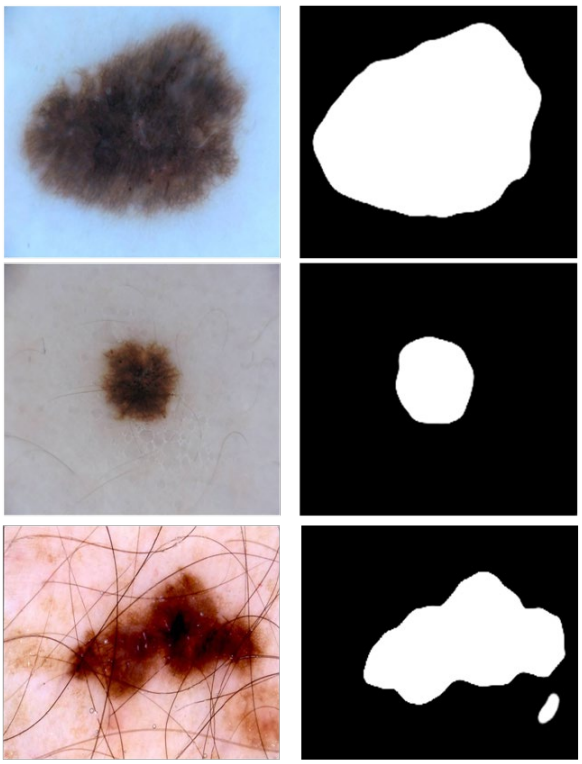
The Jaccard similarity coefficient compares predictions (A) with ground truth data (B) to see which examples are common and which are different, as given in equation (9).

$$Jaccard\ index\ (A,B) = \frac{|A \cap B|}{|A \cup B|} \quad (9)$$

The average and best evaluation metrics obtained as a result of segmentation of the skin images are as in Table 3.

Table 2. Segmentation results

	Accuracy	Dice Coefficient	Jaccard index
Average	0.76	0.70	0.80
Best	0.93	0.79	0.93



(a) (b)

Figure 6. a. Original Skin Images b. Predicted Masks

Classification is made by making fine-tuning to the pre-trained VGG16 network in the two-class dataset. As a result of the classification, the accuracy rate of the test data is 85% and the loss rate is 0.47. Accuracy and loss rates graphs are as in Figure 7. In addition to accuracy and loss rates, evaluation metrics such as precision and sensitivity are obtained.

These are a simple but useful way to measure the quality of estimates. If we briefly define these metrics:

Precision is the criterion indicating success in a situation that is positively predicted. The equation is calculated by the formula in (10).

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (10)$$

Sensitivity shows how successfully positive states are predicted. Equality is calculated by the formula given in (11).

$$Sensitivity = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (11)$$

The metrics obtained as a result of classification are as in Table 4.

Table 3. Classification results

Accuracy	Loss	Precision	Sensitivity
0.855435	0.478412	0.8423	0.8243

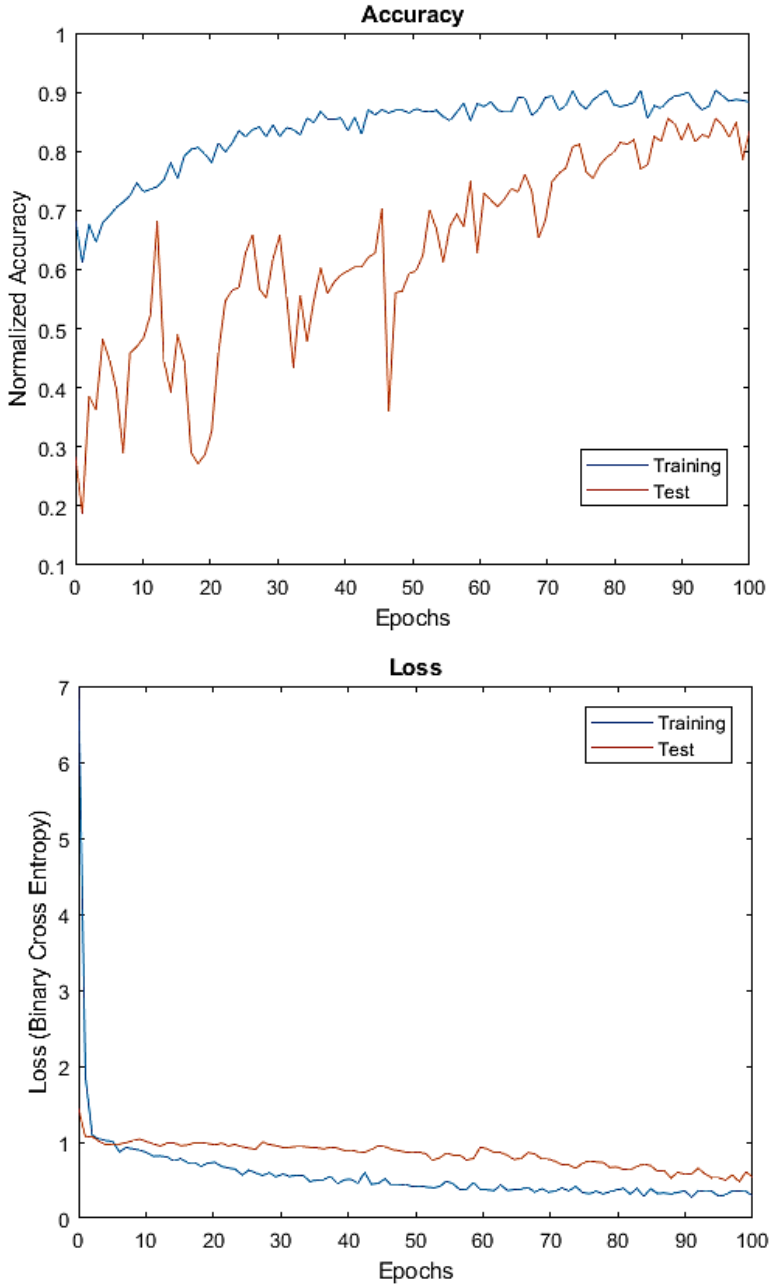


Figure 7. Accuracy and loss rates

With the verification dataset, the confusion matrix is obtained to evaluate the success of the model. For machine learning classification problems with two or more classes, the complexity matrix is a performance measure. It shows the performance of each classification model in the validation set. It is a table with 4 different combinations for predicted and real values. Estimates and actual values of the target quality are compared [21]. These 4 combinations are TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative).

TP: It is a positive prediction of a situation that normally exists positively.

TN: The existing situation is negative and the forecast is negative. In other words, the wrong situation was correctly predicted incorrectly.

FP: The existing situation is negative but it is estimated as positive.

FN: The existing situation is positive but it is estimated as negative.

Using the confusion matrix, we can estimate how many of the images we provided as verification belong to the malignant class and how many of them belong to the benign class. The result of the confusion matrix is as in Figure 8. It correctly classified 21 out of 25 images given as malignant. 95 of 104 images given as benign are correctly classified. This shows that our model has good classification performance.

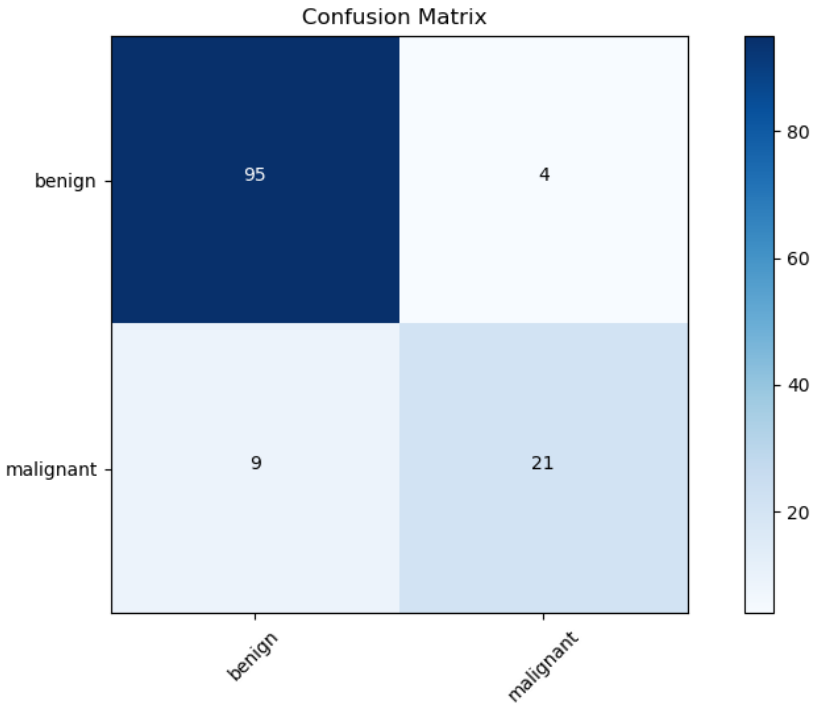


Figure 8. Confusion matrix

The goal of grouping the dataset as training and testing in deep learning is to prevent overfitting and evaluate the model's output on the dataset that it hasn't seen before The Cross-validation technique is used to minimize the errors caused by the distribution of the dataset during the training and testing phase. When the dataset is randomly divided into 'k' classes, cross-validation or 'k-fold cross-validation' is. One of these groups is used as a test set and the rest as a training set [22].

A new test dataset was obtained by selecting a different number of images first from the training data, which was previously set as 70% of the dataset, and then a different number of images. A new test dataset was obtained from the training data, which was set as 70% of the dataset, by first selecting five images and then a different number of images. While the accuracy value of this new dataset is 95%, the loss rate is 0.17. The result of the confusion matrix is as in Figure 9. Some of the metrics obtained are as in Table 5. Accuracy and loss rate graphs are given in Figure 10. Some studies used in the segmentation and classification of melanoma are examined in Table 6.

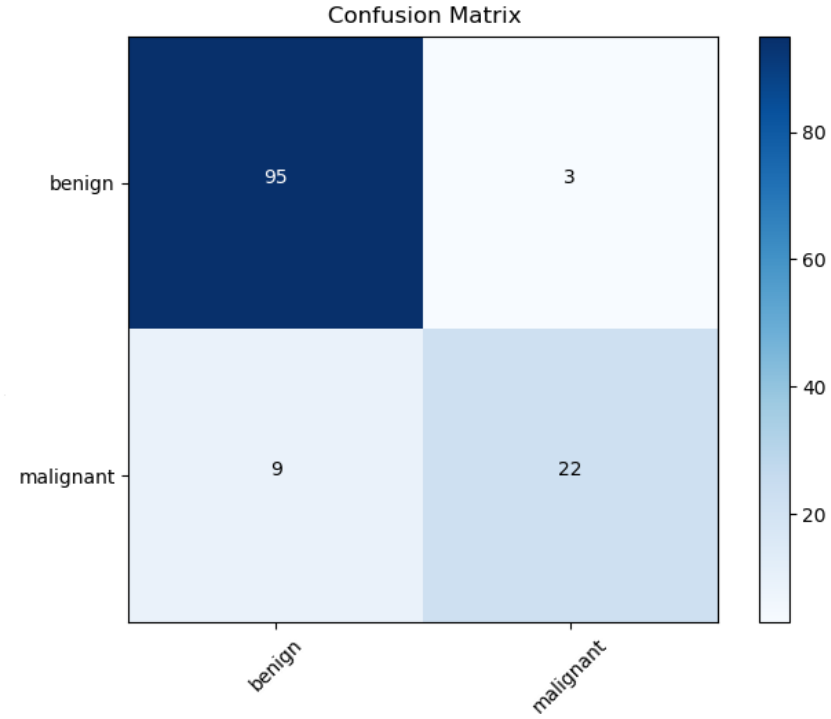


Figure 9. Confusion matrix

Table 4. Second Group Data Set Classification Results

Accuracy	Loss	Precision	Sensitivity
0.958696	0.173723	0.88	0.7096

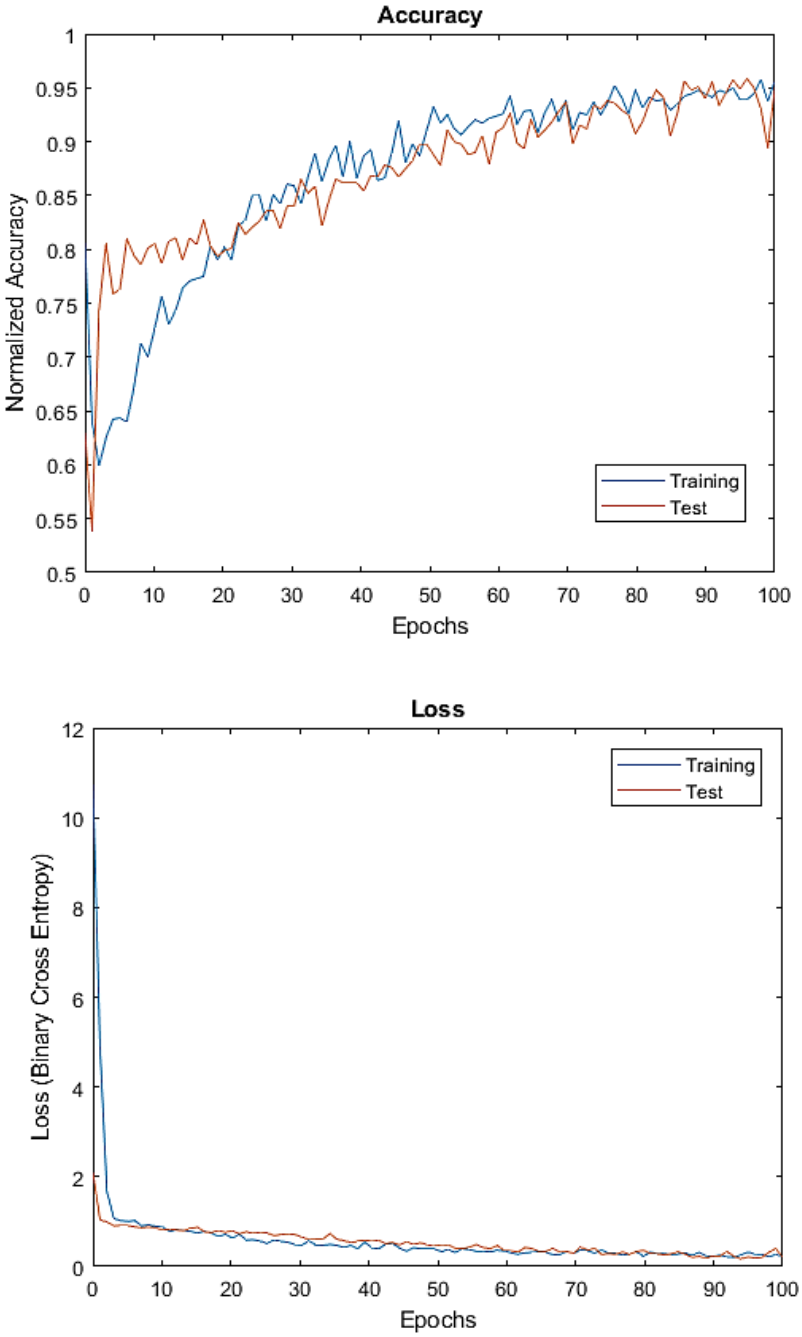


Figure 10. Accuracy and loss rates

As can be seen from the graphs, education and test success continued to increase very closely, and on the one hand, the loss rate decreased in parallel with the success rate. This shows that although the number of data is low, the model learned better than the previous one.

Table 5. Studies on skin lesion segmentation and classification

References	Method	Acc.	Dice	Jaccard
Ganster et al.	kNN, SVM	87%	-	-
Celebi et al.	SVM Border Detection	92%	-	-
Kawahara et al.	CNN	85.8%	-	-
Codella et al.	Deep Learning Sparse coding SVM	73.9%	93.1% -	-
Esteva et al.	CDNN	55.4%	72,1% -	-
Chen	U-Net Vgg-UNet	-	-	0.82 0.43
Nasr-Esfahani et al.	DFCN	-	91.6	-
Yu et al.	FCRN	94.9%	0.897	0.829
Li and Shen	FCRN CNN	84.8%	75.3% - 91.2%	-
Yuan et al.	CDNN	-	-	0.784
Tang et al.	U-Net	- -	0.93 0.86	0.89 0.79
Unlu and Cinar	AlexNet	85% ,81%	-	-
Proposed Method	U-Net VGGNet	93%	0.79 95%,85%	0.93

Conclusion

Automated analysis of dermoscopic images can help dermatologists in clinical decision making and even help patients evaluate skin lesions outside the hospital. To perform an automated analysis of dermoscopic images, the separation of skin lesions from the normal region is usually the first step. In this paper, a method with a deep network architecture has been proposed to determine the lesion site from skin images. U-Net architecture is used for lesion determination. As a result of segmentation applied to images that have been pre-processed, a good jacquard similarity coefficient was obtained. Even better results have been obtained that are quite similar to the ground truth images. As can be seen in Figure 6, even better results have been obtained that are quite similar to ground truth images.

As a result of the classification performed in the pre-trained VGG16 network, the accuracy rate of the test data is 85.5% and the loss rate is 0.47. By applying the cross-validation technique, images that the model has not seen before were given and reclassification was performed. 95.86% accuracy and approximately 0.17 loss rate were achieved. Figure 7 and Figure 10 show the accuracy and loss function curves.

Findings obtained as a result of this study show that deep learning structures can be an important tool for the diagnosis of melanoma. It is think that the performance of the system will increase by further expanding the dataset used, increasing the class labels and trying different classification, and segmentation methods.

References

[1] Cancer facts and figures 2016 , http://www.cancer.org/acs/groups/content/@research/documents /document/acspc-047079.pdf

[2] Lancet T. "Melanoma research gathers momentum", The Lancet, vol. 385, no. 9985, pp. 2323; 2015.

[3] Xu H, Hwang TH. “Automatic Skin Lesion Segmentation Using Deep Fully Convolutional Networks”;2018

[4] Bengio Y. “Learning deep architectures for AI. Foundations and trends, in Machine Learning”, Dept. IRO, Universite de Montreal, 2(1): 1-127; 2009.

[5] Jayasumana S. Semantic Image Segmentation with Deep Learning.

[6] Ganster H, Pinz P,Rohrer R, Wildling E, Binder M and Kittler H. Automated melanoma recognition, IEEE transactions on medical imaging, 20, no. 3, pp. 233-239;2001.
[7] Celebi ME, Kingravi HA, Uddin B, Iyatomi H, Aslandogan YA, Stoecker WV and Moss RHA methodological approach to the classification of dermoscopy images, Computerized Medical Imaging and Graphics, 31, no. 6, pp. 362-373; 2007.
[8] Kawahara J, BenTaieb A and Hamarneh G. Deep features to classify skin lesions. *IEEE* 13th International Symposium on Biomedical Imaging, Prague, Czech Republic; 13-16 April 2016.
[9] Codella, N, Cai, J, Abedini M, Garnavi R, Halpern A, and Smith JR. Deep learning sparse coding and svm for melanoma recognition in dermoscopy images, International Workshop on Machine Learning in Medical Imaging, pp. 118-126;2015.
[10] Esteva A, Kuprel B, Roberto A, Novoa, Justin, K., Susan, M., Swetter, Helen M., Blau, and Sebastian, T. Dermatologist-level classification of skin cancer with deep neural Networks. Research letter, 542, pp.115-118; 2017.
[11] Dandi, C, Xu M, Chao C, Jingyuan C, Zhuoran X and Fei W. U-Net Ensemble for Skin Lesion Analysis towards Melanoma Detection; 2018.
[12] Nasr-Esfahani, E, Rafiei S, Jafari MH, Karimi N, Wrobel J.S, Soroushmehr SM and Najarian K. Dense Fully Convolutional Network for Skin Lesion Segmentation. pp.1-9;2017.
[13] Lequan, Y, Hao C, Qi D, Jing Q and Pheng-Ann H. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks, IEEE Transactions on Medical Imaging, 36(4); 2017.
[14] Li Y, Shen L. "Skin Lesion Analysis towards Melanoma Detection Using Deep Learning Network", Sensors; Feb 2018.
[15] Yuan Y, Chao M and Lo YC. "Automatic skin lesion segmentation with fully convolutional-deconvolutional Networks" IEEE Journal of Biomedical and Health Informatics; 2018.
[16] Tang P, Liang Q, Yan X, Xiang S,Sun W, Zhang D, Coppola G. "Efficient skin lesion segmentation using separable-Unet with stochastic weight averaging" ELSEIVER Computer Methods and Programs in Biomedicine, 178 , 289–301;2019.
[17] Unlu EI, Cinar A. Classification of skin ımages with respect to melanoma and non-melanoma using the deep neural network. IOSR Journal of Engineering (IOSRJEN) 8(12): 35-40;2018.
[18] International Skin Imaging Collaboration,https://www.isic-archive.com.
[19] Ronneberger O, Fischer P and Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, Lncs, 9351: 234—241; 2015.
[20] Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image recognition, Conference Paper at ICLR. In arXiv preprint arXiv:1409.1556; 2014.
[21] Understanding Confusion Matrix,https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62.
[22] Hold-out vs. Cross-validation in Machine Learning https://medium.com/@eijaz/holdout-vs-cross validation-in-machine-learning-7637112d3f8f .