

# Automated Skin Lesion Segmentation using VGG-UNet

Anwar Jimi

*MECATronique Team, CPS2E Laboratory  
National Superior School of Mines  
Rabat, Morocco  
anwar.jimi@enim.ac.ma*

Hind Abouche

*MECATronique Team, CPS2E Laboratory  
National Superior School of Mines  
Rabat, Morocco  
hind.abouche@enim.ac.ma*

Nabila Zrira

*ADOS Team, LISTD Laboratory  
National Superior School of Mines  
Rabat, Morocco  
zrira@enim.ac.ma*

Ibtissam Benmiloud

*MECATronique Team, CPS2E Laboratory  
National Superior School of Mines  
Rabat, Morocco  
benmiloud@enim.ac.ma*

**Abstract**—Skin cancer is a serious worldwide health worry with high mortality rates and high grimness. For this reason, to successfully diagnose skin lesions, a computer-aided automatic diagnostic system is required. One of the most crucial methods to do that is the segmentation of skin lesions. In this paper, we present a new model that integrates two architectures, the U-Net and the VGG19. Furthermore, to improve the results of segmentation, we also employ image preprocessing, including the Dull-Razor algorithm for hair removal and Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve the image contrast. Moreover, we evaluated our model on three datasets: ISIC 2016, ISIC 2017, and ISIC 2018. Our suggested model achieved satisfactory results compared to the state-of-the-art.

**Index Terms**—Skin lesion segmentation, U-Net, VGG19, Dull-Razor, CLAHE

## I. INTRODUCTION

Skin cancer, known as melanoma, is highly metastatic and has a high rate of mortality. Despite the fact that it only causes 1% of cases of skin cancer, melanoma causes the majority of skin cancer mortality. In the United States, 106,110 new cases of melanoma were projected to be detected by 2021, resulting in 7,180 deaths [1]. When caught early, melanoma has an estimated five-year survival rate of over 99%, and when it is discovered at an advanced stage, it has a survival rate of just 14% [2]. Early detection is therefore crucial for treatment and metastasis prevention, which enhances prognosis.

One common non-invasive imaging method is dermoscopy. It enhances the skin image's brightness and size to make skin spots more visible [3]. The visual effect of deep skin can be improved by reducing the surface reflex, which gives more information about the skin problem. Despite the fact that dermoscopy has been shown to be more accurate [4], correct dermoscopy image interpretation is typically time-consuming, difficult, and subject to observer bias. As a result, computer analysis techniques [5] have been developed to assist dermatologists in increasing the efficiency and the accuracy of dermoscopy image diagnosis.

IEEE/ACM ASONAM 2022, November 10-13, 2022  
978-1-6654-5661-6/22/\$31.00 © 2022 IEEE

Automatic segmentation of skin lesions is an important step in Computer-Aided Diagnosis (CAD) [6]. However, because skin lesions vary significantly in shape, size, and color, this task remains difficult. Furthermore, the borders of certain lesions are uneven and hazy. Thus, today, computer vision and image processing approaches are being used to improve dermoscopy in order to develop tools that are capable of correctly diagnosing lesions, with the goal of improving access to reliable data to assist doctors. This enhancement can be implemented in a number of ways, including the detection of lesions, their borders, and colors, as well as the segmentation of different types of lesions.

Deep learning based on Convolutional Neural Networks (CNNs) has gained prominence recently in machine learning and computer vision, especially in semantic image segmentation [7]. In this paper, we propose a new automatic approach for skin lesion segmentation using the VGG19 pre-trained model combined with U-Net. In addition, we used image inpainting for hair removal and the CLAHE method to enhance the contrast of the image, with the aim of improving the segmentation result.

The major contributions of our paper are described as follows:

- 1) We propose a VGG-UNet deep neural network that combines the VGG19 and U-Net architecture for segmentation of skin lesions;
- 2) We remove the hair from images using Dull-Razor algorithm;
- 3) In order to enhance the image contrast, we use CLAHE technique;
- 4) Three distinct datasets were utilized to evaluate our model.

The following is a summary of this paper. An overview of skin cancer is presented in Section 2. The state-of-the-art of skin lesion segmentation is briefly introduced in Section 3. Section 4 describes the used datasets and our proposed

approach. Section 5 illustrates implementation details and experimental results. Finally, we conclude this paper in Section 6.

## II. SKIN CANCER

The most dangerous kind of skin cancer is melanoma [8]. It spreads easily to any organ and expands swiftly. Skin cells called melanocytes are the source of melanoma. These cells create the dark pigment known as melanin, which gives skin its color. Though it only accounts for around 1% of all skin malignancies, melanoma is the most common death from skin cancer. Early melanomas are often recoverable, so it's crucial to be able to identify them. Melanoma can present as raised bumps, scaly patches, open sores, or moles. Table I illustrates the indicators which are offered by the "ABCDE" memory aid from the American Academy of Dermatology [9] to determine if a lesion on the skin can be melanoma:

- Asymmetry** : The two halves are not identical;
- Border** : There are rough edges;
- Color** : With varying tones of brown, black; gray, red, or white, the color is mottled and irregular;
- Diameter** : The spot is larger than the eraser's tip (6.0 mm);
- Evolving** : The spot is either brand-new or is altering in size, shape, or color.

TABLE I: Comparison between melanoma and normal lesion

Indicator	Melanoma	Normal
Asymmetry (A)	Asymmetrical	Symmetrical
Border (B)	Uneven	Even
Color (C)	Multiple colors	One color
Diameter (D)	Larger than $\frac{1}{4}$ inch	Smaller than $\frac{1}{4}$ inch
Evolving (E)	Changing in size, color, shape	Ordinary mole

Moreover, Fig.1 shows the comparison between melanoma and non melanoma skin lesion based on the ABCDE rule.

## III. RELATED WORK

The relevant research on the topic of skin lesion segmentation is described and presented in this section. It focuses on recent methods that segment lesions using deep learning techniques.

A large part of the new works are principally founded on deep convolutional neural networks [10] that cause a noticeable boost in the effectiveness of the results of semantic segmentation. In 2015, Ronneberger *et al.* [11] presented the U-Net network for segmenting medical images. The symmetric encoders and decoders of the U-Net neural network have demonstrated excellent performance in the area of medical imaging. The architecture comprises a contracting path to capture context, and a symmetrically expanding path that permits exact localization. Yuan *et al.* [12] developed a deep Fully Convolutional Networks (FCN) approach for segmenting skin lesions. They modified the well-known FCN approach by using the Jaccard distance as a loss function in order to eliminate the requirement for reweighting samples. On the ISBI 2016 and PH2 datasets, their technique had overall accuracies of 95.5% and 93.8%. Akyel *et al.* [13] presented

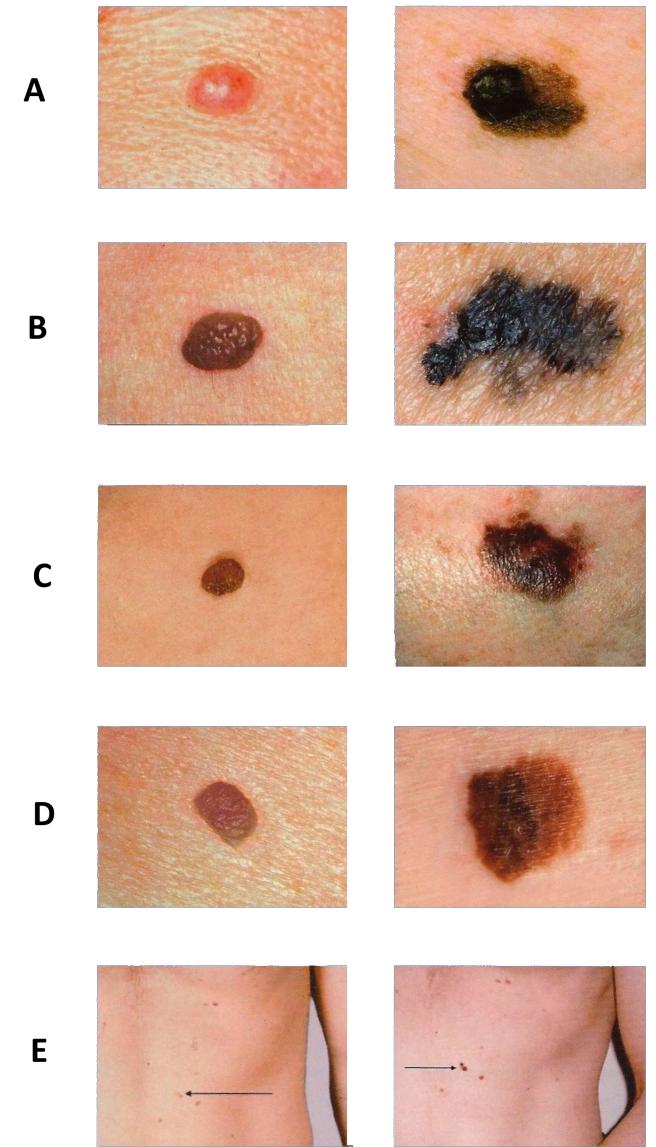


Fig. 1: Right: normal lesion. Left: melanoma lesion

a novel FCN8-based method for hair removal and image segmentation in skin cancer. ResNetC was added to FCN8 to increase success. ResNetC is another model given ResNet.

Inspired by the U-Net architecture, Tanga *et al.* [14] proposed a skin lesion segmentation method according to the separable-U-Net combined with Stochastic Weight Averaging (SWA). This model in particular makes use of the separable convolutional block and U-Net architectures, which are incredibly capable of capturing context feature channel correlation and higher semantic feature information to improve pixel-level discrimination. Also, they used SWA to avoid the overfitting problem. The suggested methodology's segmentation performance is evaluated on the ISBI 2016, ISBI 2017, and PH2 datasets that reached an accuracy of 97.16%, 94.31%, and 96.69%, respectively. Liu *et al.* [15] provided an enhanced

DCNN model for segmenting skin lesions based on U-Net. The overfitting issue is avoided by using the batch normalization layer, and the dilated convolution layer is employed to increase the perceptive field all through training without affecting resolution. Sanjar *et al.* [16] proposed an FCN and its architecture using a modified U-Net. They used a block of convolutional layers followed by a parametric rectified linear-unit non-linearity as part of a bilinear interpolation technique for upsampling. A dropout is used after each convolution block to prevent overfitting. The authors were able to segment skin lesions with a 94% accuracy and an 88.33% dice coefficient. Zafar *et al.* [17] proposed an automated approach for lesion boundary segmentation. This approach, known as Res-UNet, integrates the U-Net and ResNet architectures. In addition, they used image inpainting to remove the hair, which considerably improved the segmentation results. The ISIC 2017 test set and the PH2 datasets were used to evaluate the model. The results achieved a Jaccard index of 0.772 on the ISIC 2017 test set and 0.854 on the PH2 dataset. Iranpoor *et al.* [18] introduced a CNN with an upgraded U-Net architecture. In the suggested approach, the authors used a pre-trained architecture in the encoding phase and modified some of the pooling layers. After these modifications, the architecture's efficiency is greatly increased. The suggested method achieved 92% accuracy on the given dataset. Nguyen *et al.* [19] suggested a productive deep learning-based method for segmenting skin lesions. In particular, the paper proposed an enhanced U-Net for tasks involving skin lesion segmentation. In order to do this, they utilized EfficientNetB4 in the encoder part of the original U-Net. Additionally, residual blocks from the ResNet architecture are used to build the decoder portion of the proposed network. Barin and Güraksin [20] proposed a novel hybrid deep learning architecture based on a FCN. The encoder part of this architecture uses deep learning architectures ResNet18 and AlexNet, and the decoder part uses three deconvolution layers. Ramadan and Aly [21] provided a new DGCU-Net (Dual Gradient-Color U-Net) model with dual encoder channels that incorporate the color and gradient data of the skin lesion image. They also introduced five atrous convolutional layers with various dilation rates in the bottleneck stage to replace the 2D convolutional layer. The proposed model interconnects the encoder and decoder paths via a spatial and channel-wise attention module. Three datasets, ISIC 2017, ISIC 2018, and PH2, are used to test the proposed model.

#### IV. MATERIALS AND METHODS

In this section, we provide more details about the used datasets as well as our proposed approach.

##### A. Dermoscopic images

We used dermoscopic skin images from a publically available datasets acquired at the International Skin Imaging Collaboration (ISIC). These datasets include dermoscopy images with ground truth masks that have been annotated by leading

dermatologists. Fig. 2 provides examples of images from these datasets.

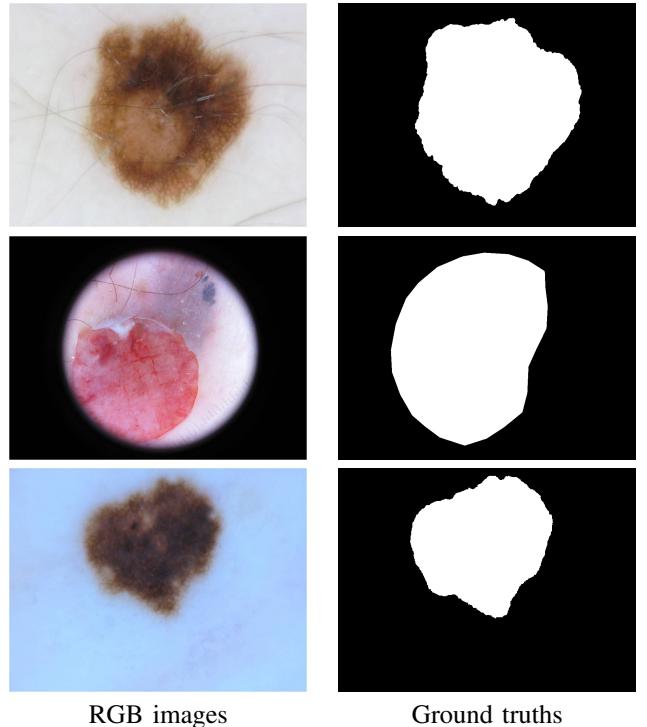


Fig. 2: Examples of skin lesion images from ISIC datasets

*a) ISIC 2016:* We exclusively make use of the ISIC 2016 lesion segmentation task dataset [22]. It comprises a testing set of 379 images and a training set of 900 dermoscopic images. There is a corresponding segmentation ground truth for each image in both the training and the testing sets.

*b) ISIC 2017:* The ISIC 2017 [2] contains 2,000 training images, 600 test images, and 150 images for validation. Images in the challenge dataset range in size from  $540 \times 722$  to  $4,499 \times 6,748$  pixels. A skin lesion segmentation ground truth for each image has been manually annotated by the segmentation task expert.

*c) ISIC 2018:* The ISIC 2018 dataset [23], [24] includes dermoscopic images that range in size from  $767 \times 576$  to  $6682 \times 4401$  and are 8-bit RGB. The dataset includes 2,594 training images.

##### B. Proposed Approach

For the fully automatic segmentation of skin lesions, we created a segmentation model based on two architectures: VGG19 and U-Net. To enhance the efficiency of the architecture, we also used various image preprocessing techniques. As shown in Fig. 3, the approach is divided into three major steps, including preprocessing, segmentation, model evaluation, and comparison with the state-of-the-art.

*1) Image Preprocessing:* Preprocessing is used to improve the quality of the images. Preprocessing enhances segmentation performance, according to earlier approaches [25]. Consequently, we utilized various preprocessing techniques such as

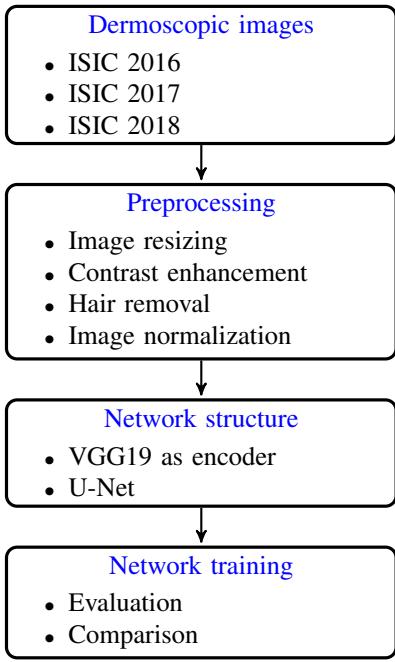


Fig. 3: Flowchart of our proposed approach

image resizing and normalization, hair removal, and contrast enhancement before being provided to our model as input.

*a) Image resizing:* is an important part of the computer vision preprocessing process. Deep learning models generally train more quickly on tiny images. The neural network must learn from four times as many pixels in a larger input image, which lengthens the architecture's training period. Dermoscopic images come in different sizes. Therefore to account for these individual variations, images and their corresponding ground truths were resized to  $256 \times 256$  resolution.

*b) Contrast enhancement:* CLAHE is a useful technique for enhancing contrast that successfully raises the contrast of the images. The CLAHE method was created using the principles of Adaptive Histogram Equalization (AHE) [26]. AHE is a great contrast-enhancing technique for medical imaging. The adaptive approach varies from conventional histogram equalization in that it improves local contrast. The image is divided into different blocks, and each section's histogram equalization is computed. As a result, AHE generates numerous histograms, each of which represents a different element of the image. It improves the local contrast and edge definition in all clearly defined areas of the image. In relatively homogeneous areas of an image, AHE overamplifies the noise. To prevent this problem, we used CLAHE as depicted in Fig. 4.

*c) Hair removal:* Dermoscopy images encounter many skin lesion artifact types, such as hair. It might present some problems while segmenting skin lesions. To resolve these problems, a preprocessing method called Dull-Razor [27] is used to find and remove hairs. This entails applying several morphological procedures to the image to produce a mask that includes the hairs. The following steps present the Dull-Razor algorithm:

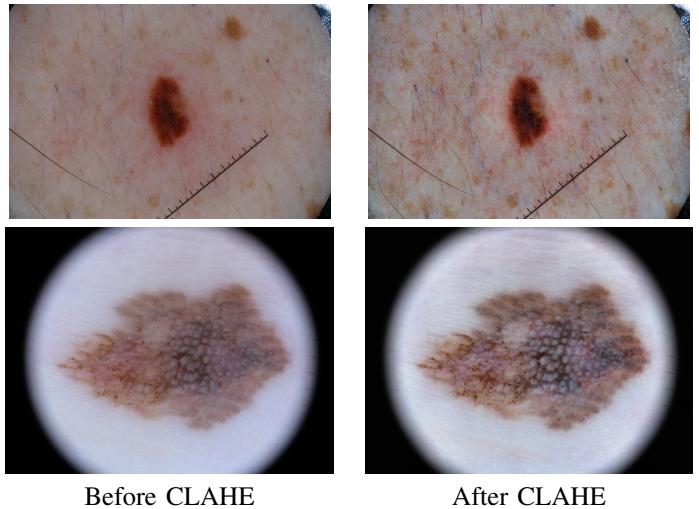


Fig. 4: CLAHE algorithm

- 1) Convert the original image to grayscale;
- 2) Apply advanced morphological operations (i.e., Black-hat);
- 3) Create a mask that includes all the hairs in the image, and use binary thresholding;
- 4) Replace the pixels shared by the original image and the mask with those from the latter.

A structural element with a cross shape of  $17 \times 17$  is described. Black-Hat filtering is accomplished by taking the original image and subtracting the closing of the image.

The image produced after performing the closing technique on a grayscale image is removed from the original image to produce hair like structures. Then, we applied a threshold value of "10" to the image produced by the Black-Hat filter results in the creation of a binary mask of the hair elements.

The hair removal algorithm restores the original image by swapping out the hair structures with the surrounding pixels. This method is frequently used to recover outdated or noisy images. The mask created after thresholding was applied to the image to be inpainted, and the desired result was attained by inpainting the hairy regions that had been retrieved using the neighboring pixels as shown in the Fig.5.

*d) Image normalization:* The goal of normalization is to scale the input data so that it is centered on zero in all dimensions by altering the range of pixel values between 0 and 1. In order to normalize an image, its mean value is subtracted from the image, which is then divided by the image's standard deviation.

*2) Network Structure:* The VGG19 network's architecture [28], as its name suggests, consists of 19 CNN layers and three fully connected layers. Basically, this architecture consists of three different types of layers:

- Convolution layers to extract the feature from the image using a variety of filter types and numbers;
- Max-pooling layers to reduce image size and extract the feature from the feature map created from these filters

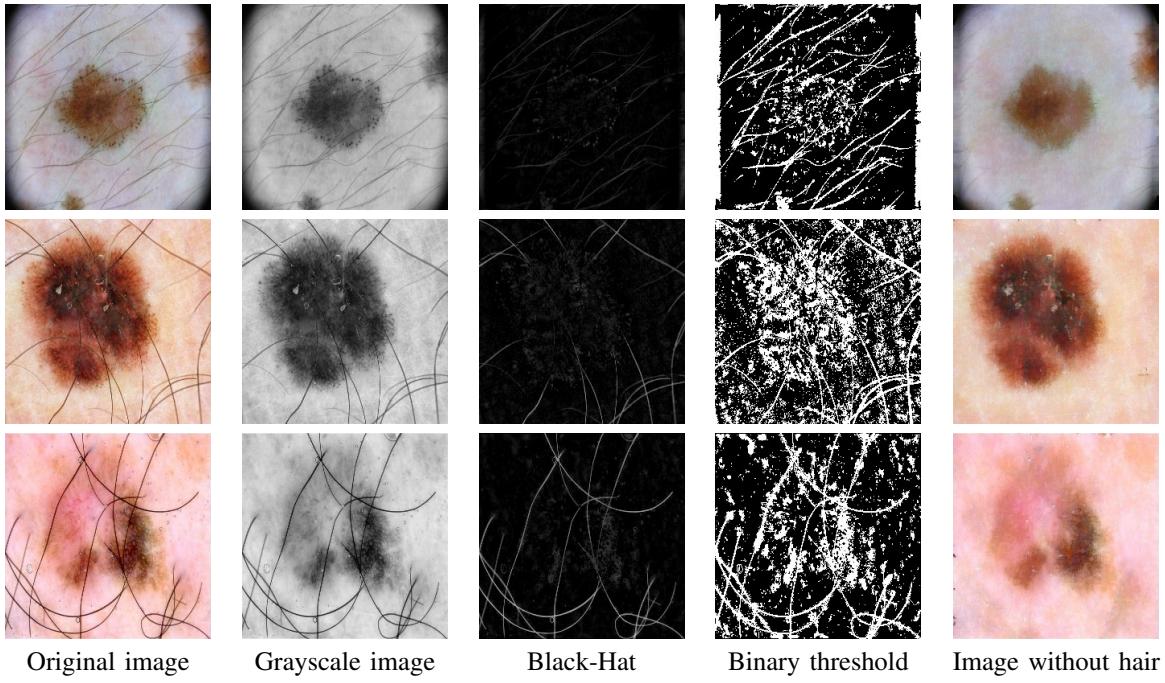


Fig. 5: Hair removal algorithm

contained in the convolution layer;

- Flatten layers to generate a 1D tensor from the batches of feature maps;
- Fully-connected layers are the last layers in the convolutional neural network. The VGG19 consists of three fully connected layers. The first ones contain dense units, while the last one contains the softmax layer.

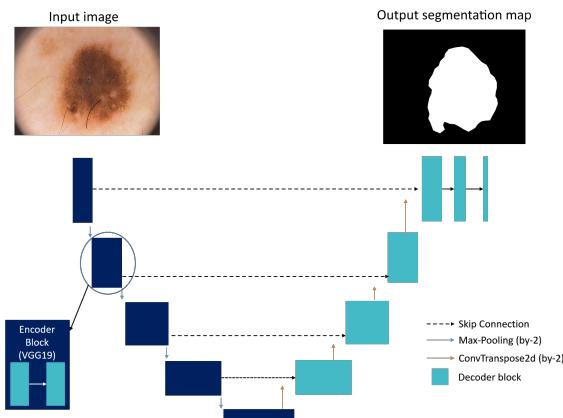


Fig. 6: Our network architecture VGG-UNet

The U-Net architecture is one of the most frequently employed techniques for semantic segmentation. Encoding and decoding are the two main parts of this network. In this paper, the structure of the network includes two parts. In the first section, context is captured by employing four blocks of downsampling based on the VGG19 network. The second

section is based on the U-Net, which is composed of four up-sampling blocks and uses a symmetric expanding path to provide exact localization. In this path, we add some Dropout and BatchNormalization layers to prevent the overfitting problem. The summary of our VGG-UNet architecture is illustrated in Fig. 6 and Table II.

*3) Network Training:* We train our model around 100 epochs with early stopping technique to prevent the overfitting. The learning rate is decreased if, after 10 epochs, the model loss does not reduce. After nearly 40 epochs, our model came to an end. Transfer learning was used to train the model using our dataset, utilizing pre-trained weights acquired through training on the ImageNet dataset. The hyperparameters used to train our model are presented in the Table III.

## V. EXPERIMENTS AND RESULTS

### A. Implementation Details

In Google Colab, we developed our network using TensorFlow and Python on GPUs T4 and P100. All training and testing phases were performed in the same environment, utilizing the TensorFlow 2.5.0 deep learning framework and Python 3.5 as the programming language.

### B. Segmentation Metrics

In order to evaluate semantic segmentation techniques in the literature the following measures have been employed:

- Accuracy (AC) is a review of how well the lesion image was segmented overall.

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

TABLE II: Architecture of the VGG-UNet

Layer	Output shape	Parameter
img (InputLayer)	[(None, 256, 256, 3)]	0
block1-conv1 (Conv2D)	(None, 256, 256, 64)	1792
block1-conv2 (Conv2D)	(None, 256, 256, 64)	36928
block1-pool (MaxPooling2D)	(None, 128, 128, 64)	0
block2-conv1 (Conv2D)	(None, 128, 128, 128)	73856
block2-conv2 (Conv2D)	(None, 128, 128, 128)	147584
block2pool (MaxPooling2D)	(None, 64, 64, 128)	0
block3-conv1 (Conv2D)	(None, 64, 64, 256)	295168
block3-conv3 (Conv2D)	(None, 64, 64, 256)	590080
block3-conv4 (Conv2D)	(None, 64, 64, 256)	590080
block3-pool (MaxPooling2D)	(None, 32, 32, 256)	0
block4-conv1 (Conv2D)	(None, 32, 32, 512)	1180160
block4-conv2 (Conv2D)	(None, 32, 32, 512)	2359808
block4-conv3 (Conv2D)	(None, 32, 32, 512)	2359808
block4-conv4 (Conv2D)	(None, 32, 32, 512)	2359808
block4-pool (MaxPooling2D)	(None, 16, 16, 512)	0
block5-conv1 (Conv2D)	(None, 16, 16, 512)	2359808
block5-conv2 (Conv2D)	(None, 16, 16, 512)	2359808
block5-conv3 (Conv2D)	(None, 16, 16, 512)	2359808
block5-conv4 (Conv2D)	(None, 16, 16, 512)	2359808
conv2d-transpose (Conv2DTranspose)	(None, 32, 32, 128)	589952
concatenate (Concatenate)	(None, 32, 32, 640)	0
dropout (Dropout)	(None, 32, 32, 640)	0
conv2d-1 (Conv2D)	(None, 32, 32, 128)	737408
batch-normalization-1 (BatchNormalization)	(None, 32, 32, 128)	512
activation-1 (Activation)	(None, 32, 32, 128)	0
conv2d-transpose-1 (Conv2DTranspose)	(None, 64, 64, 64)	73792
concatenate-1 (Concatenate)	(None, 64, 64, 320)	0
dropout-1 (Dropout)	(None, 64, 64, 320)	0
conv2d-3 (Conv2D)	(None, 64, 64, 64)	184384
batch-normalization-3 (BatchNormalization)	(None, 64, 64, 64)	256
activation-3 (Activation)	(None, 64, 64, 64)	0
conv2d-transpose-2 (Conv2DTranspose)	(None, 128, 128, 32)	18464
concatenate-2 (Concatenate)	(None, 128, 128, 160)	0
dropout-2 (Dropout)	(None, 128, 128, 160)	0
conv2d-5 (Conv2D)	(None, 128, 128, 32)	46112
batch-normalization-5 (BatchNormalization)	(None, 128, 128, 32)	128
activation-5 (Activation)	(None, 128, 128, 32)	0
conv2d-transpose-3 (Conv2DTranspose)	(None, 256, 256, 16)	4624
concatenate-3 (Concatenate)	(None, 256, 256, 80)	0
dropout-3 (Dropout)	(None, 256, 256, 80)	0
conv2d-7 (Conv2D)	(None, 256, 256, 16)	11536
batch-normalization-7 (BatchNormalization)	(None, 256, 256, 16)	64
activation-7 (Activation)	(None, 256, 256, 16)	0
conv2d-8 (Conv2D)	(None, 256, 256, 1)	17

- Jaccard index (JS) is a union over intersection of segmented lesions and ground truth masks [29].

$$JS = \frac{TP}{TP + FN + FP} \quad (2)$$

- Dice Coefficient (DC) is the similarity between the predicted results and the annotated ground truths [30].

$$DC = \frac{2 \times TP}{2 \times (TP + FN + FP)} \quad (3)$$

- Sensitivity (SE) shows the percentage of correctly iden-

TABLE III: Hyperparameters maintained during training

Parameter	Value
Input size	$256 \times 256 \times 3$
Batch size	32
Learning rate	$10^{-4}$
Optimizer	Adam
Epochs	100
Loss function	Binary Crossentropy

tified skin lesion pixels.

$$SE = \frac{TP}{TP + FN} \quad (4)$$

- Specificity (SP) represents the percentage of pixels segmented as non skin lesions.

$$SP = \frac{TN}{TN + FP} \quad (5)$$

### C. Comparative Experiments

a) *Comparison on the ISIC 2016 dataset:* On the ISIC 2016 dataset, we trained and evaluated the suggested network. We compared our results against those of the current deep learning approaches as the Table IV shows. Our provided network produced beneficial results.

TABLE IV: Model performance on the ISIC 2016

Methods	AC(%)	JS(%)	DC(%)	SE(%)	SP(%)
U-Net [11]	94.3	81.2	89.5	90.7	96.2
Recurrent U-Net [31]	93.7	79.3	88.4	89.6	96.5
Attention U-Net [32]	94.4	81.1	89.0	90.8	96.3
VGG-UNet	<b>96.15</b>	76.00	88.90	<b>92.69</b>	<b>97.54</b>

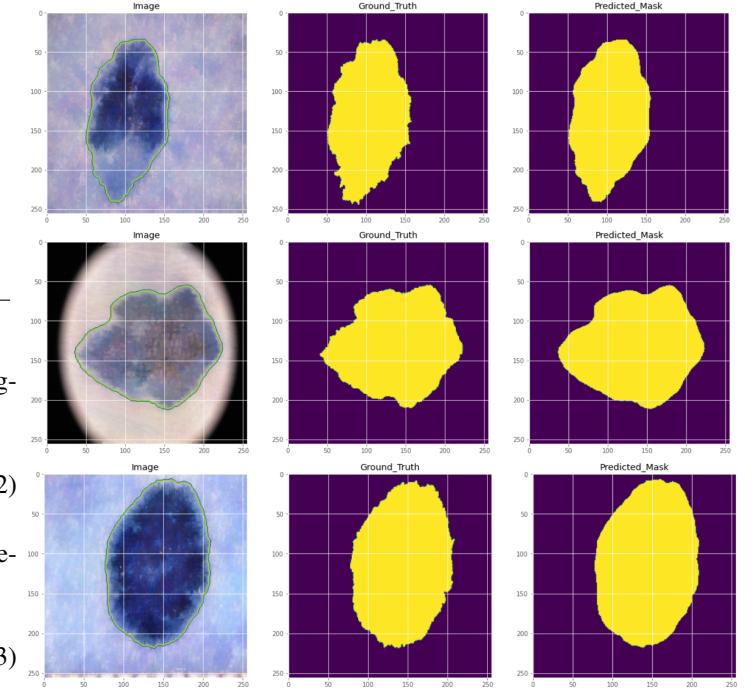


Fig. 7: Segmentation results of our model on ISIC 2016 dataset

Also, Fig. 7 illustrates the results of visualizing the suggested network. The outcomes also demonstrated the highest performance of our suggested network VGG-UNet.

*b) Comparison on the ISIC 2017 dataset:* On the ISIC 2017 dataset, we further trained and tested the suggested network. A comparison of the segmentation performance of the proposed network and existing techniques is shown in Table V.

TABLE V: Model performance on the ISIC 2017

Methods	AC(%)	JS(%)	DC(%)	SE(%)	SP(%)
U-Net [11]	92.2	74.4	82.1	79.8	97.5
Recurrent U-Net [31]	90.5	64.3	78.2	81.6	95.3
Attention U-Net [32]	91.3	69.2	88.9	76.5	97.6
Separable-UNet [14]	94.31	79.26	86.93	89.53	96.32
<b>VGG-UNet</b>	<b>96.88</b>	70.44	<b>87.22</b>	<b>90.14</b>	<b>98.55</b>

Fig. 8 shows the visualization of the segmentation results on the ISIC 2017 dataset. The outcomes also demonstrated the highest performance of our suggested network, VGG-UNet.

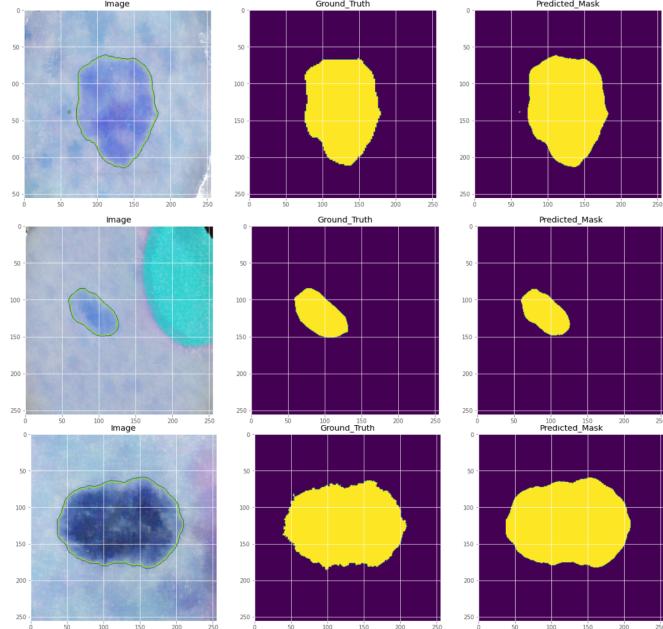


Fig. 8: Segmentation results of our model on ISIC 2017 dataset

*c) Comparison on the ISIC 2018 dataset:* We further evaluated the model on the ISIC 2018 dataset and compared our segmentation results with the current state-of-the-art methodologies to determine how robust our suggested model was. The results are shown in Table VI below. Our approach produced excellent outcomes.

TABLE VI: Model performance on the ISIC 2018

Methods	AC(%)	JS(%)	DC(%)	SE(%)	SP(%)
U-Net [11]	89.0	54.9	64.7	70.8	96.4
Recurrent U-Net [31]	88.0	58.1	67.9	79.2	92.8
Attention U-Net [32]	89.7	56.6	66.5	71.7	96.7
FCN-ResAlexNet [20]	94.65	84.17	91.40	95.85	87.86
<b>VGG-UNet</b>	<b>97.00</b>	75.30	89.17	91.97	<b>98.35</b>

Fig. 9 provides a visual representation of our suggested network's segmentation of skin lesions. Experimental renderings can also be used to see how well the approach works.

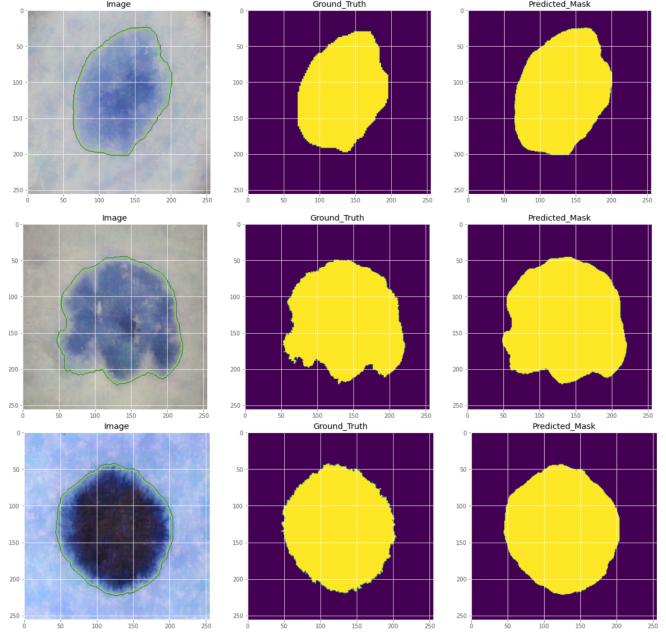


Fig. 9: Segmentation results of our model on ISIC 2018 dataset

## VI. DISCUSSION

The main goal of this work is to develop an automatic segmentation of skin lesions that represents a crucial stage for CAD systems. In particular, as deep learning-based approaches require a large amount of data for effective segmentation, our model was trained with three datasets. When compared to other deep learning-based approaches, the results demonstrated that the suggested method achieved promising results. The suggested VGG-UNet combines the characteristics of U-Net and VGG, which improves the performance of the basic version of the U-Net. Moreover, we applied the CLAHE technique and Dull-Razor algorithm to improve the quality of dermoscopic images.

The main advantage of the proposed approach is the combination of the strength of the deep learning algorithm and the preprocessing approach. We also include Dropout and BatchNormalization to improve the efficiency of our model by preventing overfitting.

## VII. CONCLUSION AND PERSPECTIVES

In this paper, we proposed a combination of the U-Net and VGG19 architectures. The approach used a powerful hair removal algorithm and contrast-enhancing with CLAHE as preprocessing steps. Three publicly available benchmark datasets were used in the experiments. The obtained results revealed that the model had higher gains in segmentation and produced better accurate segmentation results than both the initial U-Net network and subsequent modified U-Net networks.

For future research, we will use Vision Transformers (ViT) for lesion segmentation. Also, we will create a software application to help the dermatologist to segment the skin lesion for further diagnosis.

## REFERENCES

- [1] K. D. Miller, A. P. Ortiz, P. S. Pinheiro, P. Bandi, A. Minihan, H. E. Fuchs, D. Martinez Tyson, G. Tortolero-Luna, S. A. Fedewa, A. M. Jemal *et al.*, “Cancer statistics for the us hispanic/latino population, 2021,” *CA: a cancer journal for clinicians*, vol. 71, no. 6, pp. 466–487, 2021.
- [2] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [3] M. Binder, M. Schwarz, A. Winkler, A. Steiner, A. Kaider, K. Wolff, and H. Pehamberger, “Epiluminescence microscopy: a useful tool for the diagnosis of pigmented skin lesions for formally trained dermatologists,” *Archives of dermatology*, vol. 131, no. 3, pp. 286–291, 1995.
- [4] M. Vestergaard, P. Macaskill, P. Holt, and S. Menzies, “Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting,” *British Journal of Dermatology*, vol. 159, no. 3, pp. 669–676, 2008.
- [5] K. Doi, “Computer-aided diagnosis in medical imaging: historical review, current status and future potential,” *Computerized medical imaging and graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.
- [6] K. Korotkov and R. Garcia, “Computerized analysis of pigmented skin lesions: a review,” *Artificial intelligence in medicine*, vol. 56, no. 2, pp. 69–90, 2012.
- [7] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [8] G. Capdehourat, A. Corez, A. Bazzano, R. Alonso, and P. Museé, “Toward a combined tool to assist dermatologists in melanoma detection from dermoscopic images of pigmented skin lesions,” *Pattern Recognition Letters*, vol. 32, no. 16, pp. 2187–2196, 2011.
- [9] F. Nachbar, W. Stoltz, T. Merkle, A. B. Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, “The abcd rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions,” *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551–559, 1994.
- [10] M. Attia, M. Hossny, S. Nahavandi, and A. Yazdabadi, “Skin melanoma segmentation using recurrent and convolutional neural networks,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 292–296.
- [11] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [12] Y. Yuan, M. Chao, and Y.-C. Lo, “Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance,” *IEEE transactions on medical imaging*, vol. 36, no. 9, pp. 1876–1886, 2017.
- [13] C. Akyel and N. ARICI, “Hair removal and lesion segmentation with fcn8-resnetc and image processing in images of skin cancer,” *Bilişim Teknolojileri Dergisi*, vol. 15, no. 2, pp. 231–238, 2022.
- [14] P. Tang, Q. Liang, X. Yan, S. Xiang, W. Sun, D. Zhang, and G. Coppola, “Efficient skin lesion segmentation using separable-unet with stochastic weight averaging,” *Computer methods and programs in biomedicine*, vol. 178, pp. 289–301, 2019.
- [15] L. Liu, L. Mou, X. X. Zhu, and M. Mandal, “Skin lesion segmentation based on improved u-net,” in *2019 IEEE Canadian conference of electrical and computer engineering (CCECE)*. IEEE, 2019, pp. 1–4.
- [16] K. Sanjar, O. Bekhzod, J. Kim, J. Kim, A. Paul, and J. Kim, “Improved u-net: fully convolutional network model for skin-lesion segmentation,” *Applied Sciences*, vol. 10, no. 10, p. 3658, 2020.
- [17] K. Zafar, S. O. Gilani, A. Waris, A. Ahmed, M. Jamil, M. N. Khan, and A. Sohail Kashif, “Skin lesion segmentation from dermoscopic images using convolutional neural network,” *Sensors*, vol. 20, no. 6, p. 1601, 2020.
- [18] R. Iranpoor, A. S. Mahboob, S. Shahbandegan, and N. Baniasadi, “Skin lesion segmentation using convolutional neural networks with improved u-net architecture,” in *2020 6th Iranian Conference on Signal Processing and Intelligent Systems (ICSPI)*. IEEE, 2020, pp. 1–5.
- [19] D. K. Nguyen, T.-T. Tran, C. P. Nguyen, and V.-T. Pham, “Skin lesion segmentation based on integrating efficientnet and residual block into u-net neural network,” in *2020 5th International Conference on Green Technology and Sustainable Development (GTSD)*. IEEE, 2020, pp. 366–371.
- [20] S. Barin and G. E. Güraksin, “An automatic skin lesion segmentation system with hybrid fcn-resalexnet,” *Engineering Science and Technology, an International Journal*, p. 101174, 2022.
- [21] R. Ramadan and S. Aly, “Dgcu–net: A new dual gradient-color deep convolutional neural network for efficient skin lesion segmentation,” *Biomedical Signal Processing and Control*, vol. 77, p. 103829, 2022.
- [22] D. Gutman, N. C. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, “Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (isbi) 2016, hosted by the international skin imaging collaboration (isic),” *arXiv preprint arXiv:1605.01397*, 2016.
- [23] N. Codella, V. Rotemberg, P. Tschanl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti *et al.*, “Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic),” *arXiv preprint arXiv:1902.03368*, 2019.
- [24] P. Tschanl, C. Rosendah, and H. Kittler, “Data descriptor: the ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions,” 2018.
- [25] K. De Raad, K. A. van Garderen, M. Smits, S. R. van der Voort, F. Incekara, E. Oei, J. Hirvasniemi, S. Klein, and M. P. Starmans, “The effect of preprocessing on convolutional neural networks for medical image segmentation,” in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 655–658.
- [26] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, “Adaptive histogram equalization and its variations,” *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987.
- [27] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, “Dullrazor®: A software approach to hair removal from images,” *Computers in biology and medicine*, vol. 27, no. 6, pp. 533–543, 1997.
- [28] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [29] D. M. Powers, “Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation,” *arXiv preprint arXiv:2010.16061*, 2020.
- [30] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain tumor segmentation using convolutional neural networks in mri images,” *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.
- [31] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, “Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation,” *arXiv preprint arXiv:1802.06955*, 2018.
- [32] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, “Attention u-net: Learning where to look for the pancreas,” *arXiv preprint arXiv:1804.03999*, 2018.