Assessment dermoscopy images of skin lesion using U-Net segmentation for clinicians teledermatology

A Ciptaningrum¹, I K E Purnama², R F Rachmadi³
¹Department of Electrical Engineering
²Department of Electrical Engineering and Department of Computer Engineering
³Department of Computer Engineering
Institut Teknologi Sepuluh Nopember Surabaya

E-mail: adiratna.18071@mhs.its.ac.id

Abstract. Standard solutions to the process of assessing skin lesions are based on a medical pathology examination using dermoscopy images. Fortunately, abovementioned diagnosis and treatment often show a defect-prone, yet in the hands of accomplished health care professionals. Due to the rapid development of skin cancer lesions, computational analysis is required. In order to minimize the probability of error, a segmentation task is required in which multiple medical analyzes can be conducted. These frameworks are generally required versatile and need assistance from advanced computing power. The proposed system is tested with datasets for dermoscopy images of clinical signs. U-Net segmentation method provides greater segmentation result IoU 94.37, Dice Coefficient 88.11, precision 90.87, recall (sensitivity) 91.82, accuracy 94.55, loss 16.8, and F1-score 91.34.

1. Introduction
Skin cancer is known as irregular proliferation of skin cells, most commonly occurs on sun-exposed skin, and is the most prevalent type of malignancy in humans. The most dangerous skin cancers are melanoma, squamous cell carcinoma, and basal cell carcinoma. Malignant melanoma is identified as melanoma, the most severe form of skin cancer [1]. Malignant melanoma arises in melanocytes, where the cells produce melanin and add color to the skin. This case has a rising trend in the world, particularly for women under the age of 40. However, this recently retired cancer will be treated if it is detected early. As we know, doctors inspect and capture their patients with visual examination assisted by hand-held dermoscopy. The possibility visual analysis would catch both the macroscopic and the microscopic. The macroscopic captures the optical close-up, and the microscope is also known as the dermoscopy [2].

Dermoscopy is a way to examine the skin using bright light and enhancement, and also using polarizing or immersion fluids to minimize surface reflections. Dermoscopy has only improved the rate of diagnosis by visual examination. The dermoscopy images image plays an important role in decision-making to help the diagnosis of a disease. This is attributed to the fact that the image offers different details found in it for study. Quantitative analysis of digital images can yield high throughput images, such as image segmentation tasks. This segmentation approach is the most preferred condition for automatic image processing, and most researchers define and analyze each lesion in the study to help clinicians measure how they respond to different treatments [3].

Recently, the image segmentation task is a crucial step needed to establish the scale and ratio that is used for clinically relevant purposes. As we know, the production of automated segmentation is difficult due to wide differences in texture, shape and colour. Various types of image processing
methods are provided, such as regional growth, thresholding, morphological operations, k-means clustering, and most of their accuracy are less common types. In fact, the noise and background of the image make it impossible to achieve thorough and accurate segmentation. In addition, artifacts may be produced during image processing and the contrast is weak [3]. Actually, the most preferred approach is to take a deep-learning section of the lesion. One of the learning architectures for the segmentation of images is the U-Net. In general, this approach uses a downsampling encoder and an upsampling decoder architecture by separating multiple parameters for each level. This architecture has the potential to run at a much greater random sample without requiring any extra effort. U-Net is also suitable for sample processing and automated large-scale recording. U-Net is more competitive than current methodologies in terms of architecture and pixel-based image segmentation generated by convolutional neural network layers. It is also efficient with small dataset images. The value of this network system is that it can not only effectively segment the targeted lesion target and efficiently analyze and objectively examine the dermoscopy image of the skin lesion, but also help to increase the accuracy of medical image assessment. This study consequently provides a systematic literature review of U-net-based medical image segmentation, focusing on the effective segmentation implementation of U-net for medical imaging systems.

2. Related Works

On that same basic principle of a research study, a deep learning algorithm is the most preferred method of image segmentation to solve this issue. For such segment lesion region, the features are systematically trained in deep learning because they achieve significant output of the metrics, segment the region, and require a huge number of annotated data. The Convolutional Neural Network, known as CNN, has been able and efficient in large amounts of data. Previous studies by C. Hernandez et.al. [4] Use Function Pyramid Network (FPN) and merge neural nets with VGG-style to predict the cell mask. F. Araújo et.al, meanwhile. [5] used CNN to estimate the irregular region of cells in the pap test.

In addition, Bi et al.[6] suggested segmentation of skin lesions using multi-stage, completely convoluted networks (FCNs). The multi-stage method involved localized early-stage coarse-appearance learning and also detailed boundary-characteristic learning in the next stage. In his work, a parallel integration strategy was put in order to make a merger of the outcome that they reported to have improved detection. However, various PH2 datasets outperformed their system by around 90.66% and still show modest change. T. Tran et al. [7] use of these Seg-Net in white blood cells (leukocytes) and red blood cells (erythrocytes) in peripheral blood tests shows a global accuracy of 89.45%. Between multi cell or cell nuclei segmentation, U-Net is the most recommended and always provides state-of-the-art segmentation outcomes.

In comparison, R. Hollandi et.al. [8] used two approaches in the form of R-CNN and U-Net to estimate cell nuclei, and the algorithm proposed to measure the level of accuracy. And then, in the analysis of Pan et.al[9], simplified reconstruction techniques based on CNN were used, the nuclei of which is roughly segmented from the background. In Yang et al.[10] ‘s work, Fully Convolutional Neural Network (F-CNN) was used for an iterative k-terminal trying to cut algorithm for the segmentation of glia cells. The F-CNN has used a fully connected layer with a convolution layer at the end, even though other layers extract the necessary features from input data, so this method is best suited. Akram et al.[11] demonstrated a concept for a cell bounding box and used F-CNN for cell segmentation to achieve good cell segmentation accuracy based on spatial details. Hou et al.,[12] applied CNN to the nuclei section using the Generative Adversarial Network (GAN) module to generate synthetic patches dependent on weight. However, segmentation collapsed when the nuclei is close to the dark zone. Hatipoglu et al.,[13] used the spatial and temporal connections of extracellular and cell pixels. It can segment cells by generating a number of pixels in the window framework to fit neighboring pixels. U-Net is the most favored approach since there has been a major change in the segmentation and classification of the medical image.

There are three reasons U-Net is used for this work. First, the ambiguity of its boundaries makes it very hard to differentiate the gland from surrounding tissue with intraprostatic tissue heterogeneity further contributing to under or oversegmentation. Second, examinations on different dermoscopic skin
lesion with use of different imaging protocols lead to wide variations in signal intensity. Third, the skin lesion has a wide variety of sizes, textures of tissue, either due to anatomical differences within patients or due to the presence of pathology.

3. Material and Methods

3.1. Dataset

The data of dermoscopic skin lesion is provided by PH2 dataset [14]. There are three unique diagnose of skin lesion of dermoscopic such as melanoma, nevus, and seborrheic keratosis. Melanoma is malignant skin tumor that derived from melanocytes (melanocytic). It can be seen in Figure 1.

![Figure 1. Melanoma.](image1)

Nevus is benign skin tumor that derived from melanocytes (melanocytic) can be seen in Figure 2. And then, seborrheic keratosis is benign skin tumor, derived from keratinocytes (non-melanocytic) can be seen in Figure 3.

![Figure 2. Nevus.](image2)

![Figure 3. Seborrheic keratosis.](image3)
Lesion information contains the original image, combined with a gold standard (definitive) diagnosis, which have seven groups such as melanocyte nevi, melanoma, benign keratosis, basal cell carcinoma, actinic keratoses, vascular lesions, dermatofibroma, and so-called ground truth. This dataset includes 200 images as training data of which 160 images are naevus (atypical naevus and normal naevus), as well as 41 images of melanoma, and each lesion is followed by a binary mask identified by an expert. These images resized as $256 \times 256$ to maintain training. U-Nets were chosen due to their proven ability to segment microscopic images.

3.2. The Structural Of U-Net Models
U-Net is also known as the original semantic segmentation network, who have generated incredible results in the segmentation of medical images. This architecture can be expansively thought of as an encoder network followed by a decoder network. By integrating layers, the encoder bit by bit eliminates the spatial dimension. And then bit by bit the decoder recovers the information of the object and the spatial dimension. Besides that, there are also fast connections, where the decoder encoder allows the decoder to better recover the information of the object. As a highly invariant, the U-Net provides a stable and precise segmentation of its form, where the features of the images are combined to obtain high-resolution data. In addition, the pooling operator is replaced by an upsampling operator to increase the output resolution. Adding more layers will help remove additional data input features and improve significantly accuracy. The architecture of the U-net architecture can be seen in Figure 1.

![U-Net Architecture](image1)

**Figure 4.** U-Net Architecture.

The breakdown of U-Net architecture can be seen in Figure 4. Each mechanism consists of two convolutional layer, and the number of channels varies from 1 to 64.

![Input Image](image2)

**Figure 5.** Input Image.
As a convolutional process, the depth of the image will increase. The red arrow pointing downward is the max pooling process which is half the downsizing of the image. The downsizing is the size reduced from $572 \times 572 \rightarrow 568 \times 568$ is due to padding, but the implementation still using the same padding.

![Figure 6](image)

**Figure 6.** The process is repeated 3 more times.

In Figure 6 can be seen that U-Net reaches at the bottom. Still two convolutional layers are built, but with no max pooling. The image at this bottom has been resized to $28 \times 28 \times 1024$. The next step is expansive path, can be seen in Figure 7.

![Figure 7](image)

**Figure 7.** The process is repeated 3 more times.

In the expansive path, the image is upsized to its original size, and need transpose convolution. This transpose convolution is an upsampling technic that expands the size of images. Basically, the transpose convolution does some padding on the original image followed by a convolution operation. After the transposed convolution, the image is upsized from $28 \times 28 \times 1024$ to $56 \times 56 \times 512$. In this stage, the image is merged with the appropriate image from the contracting line and makes an image in size $56 \times 56 \times 1024$ to combine the information from the previous layers in order to get a more precise prediction.

![Figure 8](image)

**Figure 8.** The process is repeated 3 more times.

The uppermost of the architecture is the last step to reshape the image to make prediction requirements. The last layer is a convolution layer with one filter of size $1 \times 1$. And then, the rest left one is the same for neural network training. During training are used deep models based on the popular U-Net architecture. Adam optimization was chosen to update the weights with binary cross entropy as
loss function. In this method, loss function is used combination of binary cross-entropy and binary Dice loss to train the model. The Dice loss is to make segmented lesion, so it will equal to ground truth.

3.3. Loss Function
In semantic segmentation image is actually a pixel-level classification. Dice loss is to make segmenting the target of skin lesion to be equal to each other with ground truth as much as possible. The Eq. (1) of Dice coefficient is:

\[
\text{Dice} = 2 \times \frac{y_a \cap y_b}{|y_a| + |y_b|}, \text{where} \, \text{dice} \in (0,1)
\]  

(1)

In order to maximize the penalty, the lower the estimation result of the Dice coefficient is done by logarithmic (log) operations and negative values are taken as a loss function, e.g. Log-Dice loss. Then in (2) is shown where \(y_a\) is the ground truth value of skin lesion image and \(y_b\) is the predicted value.

\[
\text{LogDL} = -\log(2 \times (y_a \cap y_b))/(|y_a| + |y_b|)
\]  

(2)

In addition, Focal Loss may inhibit the contribution of easily segmented pixel regions to model loss, thereby enhancing regions that are not easily segmented. The theorem can therefore be shown in (2) as below.

\[
\text{Focal Loss} = y_a \times \log(y_b) \times (1 - y_b)^\lambda
\]  

(3)

Consider the two equations, Log-Dice loss and Focal loss are combined as a loss function for model training by parameter \(\lambda\). The \(\lambda\) specifies the proportion of Log-Dice loss and Focal loss in the overall loss function. The optimization of \(\lambda\) can also be chosen as corresponding to the values of F1-score and Intersection over Union (IoU). It can be seen in (3).

\[
\text{Loss} = \lambda \times \text{LogDL} + (1 - \lambda) \times \text{Focal Loss}
\]  

(4)

4. Result and Discussion
The proposed framework of U-Net to efficiently extract relevant features in the training process. This architecture clearly outperforms traditional graphic-based segmentation method. In this on experimental is validated and testing using PH2 dataset, which have 201 images with its dimension \(256 \times 256\). However, the new transformed images that will be added after image augmentation process.

Figure 9. Original Image.
The performance of this segmentation is evaluated using performance metrics. The metrics based on accuracy, precision, recall, F1-score, and IoU which has a representation of each. In this work, the accuracy means the ratio of the accurately segmented the lesion and background pixels to all the pixels in the image. Between the total of segmented skin lesion pixels, the precision represent of segmented skin lesion in label images, and it is called positive predictive value. Later, recall defines the percentage of the total numbers of lesion images on the label segmented by the proposed technique or called sensitivity. The F1-score is used to calculate the average accuracy and recall. And the last one, IoU, is the correlation between the predicted values and the ground values of truth. In the Eq. (5) (7) (8) (9), TP, FP, FN and TN are true positive, false positive, false negative and true negative.

\[
\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{5}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{6}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{7}
\]

\[
F1 - \text{Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}
\]

\[
\text{IoU} = \frac{|\gamma_0 \cap \gamma_1|}{|\gamma_0| + |\gamma_1| - |\gamma_0 \cap \gamma_1|} \tag{9}
\]

In Eq. (4) First, the $\lambda$ parameter of loss function is selected. And then, the best optimization method and loss function are chosen. The second one, skin lesion segmentation in datasets are processed using the proposed semantic segmentation as the model. The highest output and training network is chosen to boost the algorithm. The next step, then, is to evaluate the efficiency of the enhanced model with the same dataset. In this experiment, the comparison of Log-Dice loss and Focal loss with Adam is modified as an optimization. A massive quantity of data is required to create an efficient information classifier. An enhancements in image is expected to improve the performance of this network. In this research, an enhancement is described, one for random rotation and one for horizontal flipping.
This model performs well if just looking at the resultant images. In this model various of the epoch were tested in 125, 250, 750, and 1500.
Table 1. Experimental Performance On Skin Lesion Of U-Net 125 Epoch.

| Table Column 50 Epoch | Train Set | Test Set | Validation Set |
|-----------------------|-----------|----------|----------------|
| IoU                   | 97.47     | 93.75    | 94.73          |
| Dice Coefficient      | 86.58     | 80.87    | 82.19          |
| Precision             | 96.85     | 89.58    | 92.00          |
| Recall (Sensitivity)  | 97.52     | 91.31    | 92.14          |
| Accuracy              | 98.22     | 94.03    | 94.93          |
| Loss                  | 10.25     | 18.49    | 16.55          |
| F1-Score              | 97.18     | 90.44    | 92.03          |

Table 2. Experimental Performance On Skin Lesion Of U-Net 250 Epoch.

| Table Column 250 Epoch | Train Set | Test Set | Validation Set |
|------------------------|-----------|----------|----------------|
| IoU                    | 98.28     | 94.27    | 95.12          |
| Dice Coefficient       | 90.20     | 84.08    | 85.29          |
| Precision              | 97.35     | 89.51    | 91.91          |
| Recall (Sensitivity)   | 98.39     | 92.40    | 92.74          |
| Accuracy               | 98.65     | 94.22    | 95.06          |
| Loss                   | 7.41      | 16.88    | 15.03          |
| F1-Score               | 97.87     | 90.93    | 92.32          |

Table 3. Experimental Performance On Skin Lesion Of U-Net 750 Epoch.

| Table Column 750 Epoch | Train Set | Test Set | Validation Set |
|------------------------|-----------|----------|----------------|
| IoU                    | 98.95     | 94.28    | 95.48          |
| Dice Coefficient       | 93.78     | 86.86    | 88.43          |
| Precision              | 98.40     | 91.12    | 92.94          |
| Recall (Sensitivity)   | 98.77     | 91.04    | 92.49          |
| Accuracy               | 99.11     | 94.48    | 95.34          |
| Loss                   | 4.65      | 16.73    | 14.00          |
| F1-Score               | 98.59     | 91.08    | 92.71          |

Table 4. Experimental Performance On Skin Lesion Of U-Net 1500 Epoch.

| Table Column 1500 Epoch | Train Set | Test Set | Validation Set |
|-------------------------|-----------|----------|----------------|
| IoU                     | 99.15     | 94.37    | 95.50          |
| Dice Coefficient        | 95.00     | 88.11    | 89.52          |
| Precision               | 98.80     | 90.87    | 93.11          |
| Recall (Sensitivity)    | 98.82     | 91.82    | 92.60          |
| Accuracy                | 99.25     | 94.55    | 95.43          |
| Loss                    | 3.74      | 16.80    | 13.90          |
| F1-Score                | 98.81     | 91.34    | 92.85          |
As we can see in the Table 1, 2, 3 and 4, the bigger of epoch then more bigger Dice Coefficient result. So it represents the validated image segmentation performance has similar object size. In Figure 12, we compared the epoch of of train set, test set, and validation set on different epoch.

![Training Set Comparison](image1.png)

**Figure 13.** Comparison of metrics on train set.

In Figure 13. On train set we can see that the higher epoch then smaller the loss value whereas the Dice Coefficient and IoU more higher. However, it is the opposite of the test set where the higher the epoch value, the bigger the loss value. It can be seen in Figure 14.

![Test Set](image2.png)

**Figure 14.** Comparison of metrics on test set.

![Validation Set](image3.png)

**Figure 15.** Comparison of metrics validation set.
After conducting model training in the training set, then make predictions in the unseen test set and visualize the predicted output, so it can be seen in Figure 17.

![Image](image1.png)

**Figure 16.** Visualize predicted lesion.

Currently, the predicted outputs are blurry enough because of the predicted pixel values are in the range 0 to 1. To make clear edge predictions, this work enhances the image by rounding the pixel value to $1 > 0.5$.

![Image](image2.png)

**Figure 17 (a).** Comparing the prediction after enhancement. **Figure 17 (b).** Comparing the prediction after enhancement.

After visualing comparison of the prediction after enhancement, this work applying the mask. The output can be seen in Figure 18.

![Image](image3.png)

**Figure 18 (a).** Output after applying mask: Predicted mask. **Figure 18 (b).** Output after applying mask: Segmented output.

5. **Conclusion**

In this paper, we have proposed U-Net models as segmentation of dermoscopic skin lesion image. This experiment shows the segmentation results related to the proposed approach. In this experiment metrics metrics plays pivotal role during evaluating the performance such as IoU, Dice Coefficient, precission, recall (sensitivity), accuracy, and F1-score. The proposed architecture improves the performance of dermoscopic skin lesion segmentation in each area, such as visualization, training, testing, and validation. This study only focuses for segmentation task using U-Net architecture on skin lesion dataset. Further improvement can be made by fine-tuning the hyper-parameters of U-Net. The future hope, U-Net algorithm can be tested on the other publicly available segmentation data medical and biology domains.
6. References

[1] Mermelstein R and Riesenberg L 1992 Changing knowledge and attitudes about skin cancer risk factors in adolescents. *Health Psychology* **11** 6 p 371

[2] Harvey N T and Wood B A 2018 A practical approach to the diagnosis of melanocytic lesions. *Arch Pathol Lab Med* **143** pp 789 - 810 doi: 10.5858/arpa.2017-0547-RA

[3] Errichetti E and Stinco G 2016 Dermoscopy in general dermatology: A practical overview. *Dermatol Ther* **6** pp 471 - 507 doi: 10.1007/s13555-016-0141-6

[4] Hernández C X, Sultan M M and Pande V S 2018 *Using Deep Learning for Segmentation and Counting within Microscopy Data*. arXiv:1802.10548 [cs, q-bio]

[5] Araújo F H D, Silva R R V, Ushizima D M, Rezende M T, Carneiro C M, Campos Bianchi A G and Medeiros F N S 2019 Deep learning for cell image segmentation and ranking. *Comput Med Imaging Graph* **72** pp 13 - 21

[6] Yu L, Chen H, Dou Q, Qin J and Heng P A 2017 Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans. Med. Imag* **36** 4 pp 994 – 1004, Apr 2017

[7] Tran T, Kwon O, Kwon K, Lee S and Kang K 2018 Blood cell images segmentation using deep learning semantic segmentation. *IEEE International Conference on Electronics and Communication Engineering (ICECE)* pp 13 - 16

[8] Hollandi Ret al 2019 *A Deep Learning Framework for Nucleus Segmentation using Image Style Transfer*. bioRxiv. 2019: 580605

[9] Pan X et al 2017 Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks. *Neurocomputing* **229** pp 88 - 99

[10] Yang L, Zhang Y, Guldner I H., Zhang S and Chen D Z 2016 3d Segmentation of glial cells using fully convolutional networks and k-terminal cut. *Med Image Comput Comput Assist Interv of Lect Notes Comput Sci.* **9901** pp 658 - 666 (Springer)

[11] Akram S U, Kannala J, Eklund L and Heikkil¨a J 2016 Cell segmentation proposal network for microscopy image analysis. *DLMIA of Lect Notes Comput Sci.* **10008** pp 21 - 29

[12] Hou L et al 2019 Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images. *Pattern Recognition* **86** pp 188 - 200

[13] Hatipoglu N and Bilgin G 2017 Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships. *Medical & Biological Engineering & Computing* **55**

[14] Mendonça T P M, Ferreira J S, Marques A R, Marcal and Rozeira J 2013 PH2 - A dermoscopic image database for research and benchmarking. *Proc. 35th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)* (July 2013) pp 5437 – 5440