# Melanoma multi class segmentation using different U-Net type architectures

Nojus Dimša[a] and Agnė Paulauskaitė-Taraševičienė [b]

[a] *Department of Applied Informatics, Kaunas University of Technology, Studentų 50, Kaunas, Lithuania* nojus.dimsa@ktu.edu

[b] *Department of Applied Informatics, Kaunas University of Technology, Studentų 50, Kaunas, Lithuania* agne.paulauskaite-taraseviciene@ktu.lt

**Abstract**

Automatic segmentation of skin lesions is the most important step towards the analysis of malignant melanoma, which is a specific kind of skin cancer. Deep learning is one of the most effective approaches to medical image processing applications. The encoder–decoder structures are good for segmentation tasks in particular the U-Net architecture, which is used as a basic architecture for the medical image segmentation networks. Recently, different variants of U-Net type architecture have been provided for improvement in terms of the segmentation results. Therefore, we focused on three U-Net type models, specifically U-Net, U-Net++ and MultiResUNet in order to evaluate their capability and performance on the multi class segmentation of melanoma.

**Keywords**

Melanoma, segmentation, deep learning, U-Net, image analysis

## 1. Introduction

There are two main types of skin cancer: melanoma and non-melanoma. Melanoma of the skin is the 19th most commonly occurring cancer in men and women [1]. There were nearly 300,000 new cases in 2018. American Cancer Society estimates 106.110 new cases of skin melanoma and 7180 deaths from it in 2021[2]. Melanoma misdiagnosis accounts for more malpractice claims than any cancer, excluding breast cancer. Misdiagnosis of melanoma in an early stage can reduce a patient's chances of survival. Although an international staging system for melanoma exists, diagnosing melanoma accurately and consistently is still a very challenging task. Variability prognosis and diagnosis for melanoma comes from subjective visual observations used for melanoma prognosis and diagnosis [3]. Only visual inspection has variable accuracy that leads the patient to undergo other tests and series of biopsies and complicates the treatment. Computer vision can help in improving accuracy and consistency of diagnosis. Many researchers have been working on the image processing techniques for skin cancer detection. Non-invasive diagnosis methods and algorithms can achieve excellent performance in segmenting skin lesions. Image features to perform skin lesion segmentation usually include shape [4], texture [5], edges [6], luminance [7], histogram thresholding [8], and so on. Many researches are being carried out on ABCD (Asymmetry, Border, Color, Diameter) rules for the melanoma skin cancer [9], [10], [11], [12]. Most often melanoma tends to show asymmetric forms with diverse colors and structures. However, melanoma can also show symmetric pattern or nonspecific pattern.

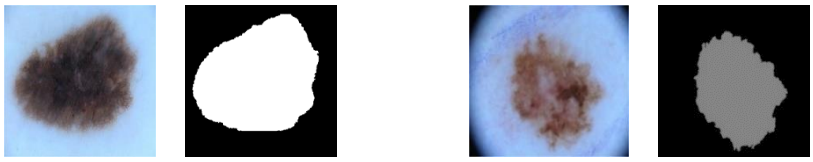
Convolutional neural networks (CNNs) as a special type of multi-layer neural networks have become one of the most widely used models of deep learning and have demonstrated a very high accuracy results in image segmentation tasks. Various CNN architectures are widely used in solving melanoma detection and segmentation tasks [13], [14], [15], [16], [11]. Some CNN are created specifically for segmentation of medical images. The purpose of medical image segmentation is to classify the pixels in an image, thereby recognizing abnormal areas such as tumors, cancer cells or lesions.

U-Net is one of the most used architectures for bio-image segmentation and produces promising results in the domain [17]. However, some researchers have noticed that classical U-Net model is still quite simple and convolution in the nodes can be improved. Therefore, some variants of U-Net have been provided such as R2U-Net [18], UNet 3+ [19], UNet++ [20], H-DenseUnet [21], TMD-Unet [22]

and others. This paper contributes to this field by performing the experimental investigation of multi class segmentation performance of three different U-Net type structures: classical U-Net, UNet++, and MultiResUNet [23].

## 2. Dataset

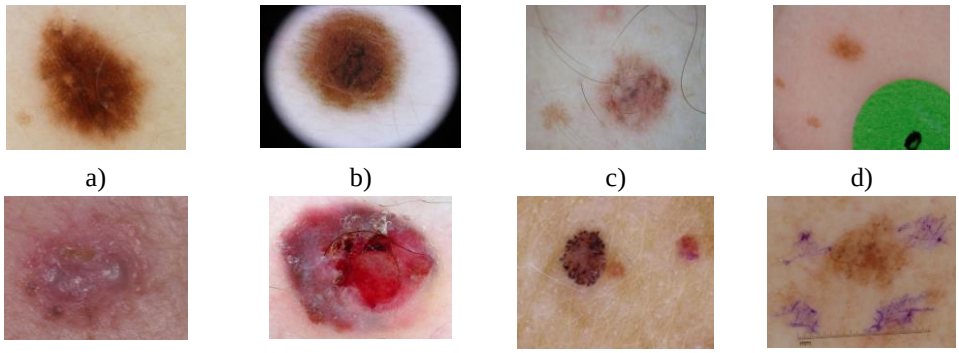
Dataset used for this research was obtained from ISIC (International Skin Imaging Collaboration) archive – an open-source public access archive of skin lesion images including melanoma. The goal of ISIC melanoma project is to help decrease melanoma-related deaths [24]. ISIC archive downloader written by Oren Tolmor and Gal Avineri was used to download the images and their masks (Figure 1). Masks are used to make recognition of object easier. However, only 1084 malignant images had mask provided, so for remaining images, manual creation of masks was performed. Images in the dataset are divided into benign and malignant. 9900 benign images and 4291 malignant images were augmented and used for the training. Data augmentation is used to reduce over-fitting and to increase size of the dataset. The following augmentations were randomly performed: image rotation from 0 to 360 degrees, width shift from 0 to 8% of total width, height shift from 0 to 8% of total height, a horizontal flip, a vertical flip and zoom inside of image between 0 and 20%. The same augmentation process was performed on training and validation sets.



a) b)

**Figure 1**: Benign skin lesion and it's mask a) and malignant skin lesion and it's mask b)

Black and white masks for malignant images were replaced with black and grey masks. This was done because multi class segmentation requires masks to be different color. Images were resized to 256 x 256 px. A single image can contain multiple skin spots (Figure 2c, d, g, h). However only the biggest one is taken for evaluation. Smaller marks are ignored (Figure 2c, g). In some images irrelevant spots are hidden (Figure 2d) or painted over (Figure 2h).



a) b) c) d)

e) f) g) h)

**Figure 2**: Image instances of benign (a-d) and malignant (e –h) melanocytic lesions

## 3. Methods

We utilize three U-Net type architectures, specifically U-Net, U-Net++, and MultiResUNet in order to compare their performance on the selected dataset.

## 3.1. U-Net architecture

U-Net was developed in 2015 for biomedical image segmentation [17]. Classical U-Net architecture consists of two paths: down-sampling and up-sampling. Down-sampling path consists of 4 blocks (Figure 3). Each block applies two 3x3 unpadded convolutions followed by a rectified linear unit ( ReLu) and a 2x2 max pooling operation with stride 2. However Leaky ReLu with α value of 0.3 was used instead. ReLu unit can end up in a state where it only outputs zeros and that will prevent parts of neural network from learning. At each block number of feature channels doubles. Up-sampling path also consists of 4 blocks. Each block in this path up-samples feature map, applies 2x2 convolution halving number of feature channels, concatenates it with feature map from down-sampling path, and applies two 3x3 unpadded convolutions followed by Leaky ReLu. In this implementation input layer receives 256x256 image with 256x256 mask. Final layer of network uses 1x1 convolution with softmax function to map feature vector to 2 classes. In total, architecture consists of 29 layers. 20 of them being convolutional followed by Leaky ReLu, 4 max-pooling layers in down-sampling path, 4 up-sampling layers in up-sampling path, and an output layer. For optimization Adam optimizer was used, with loss function being categorical cross-entropy.

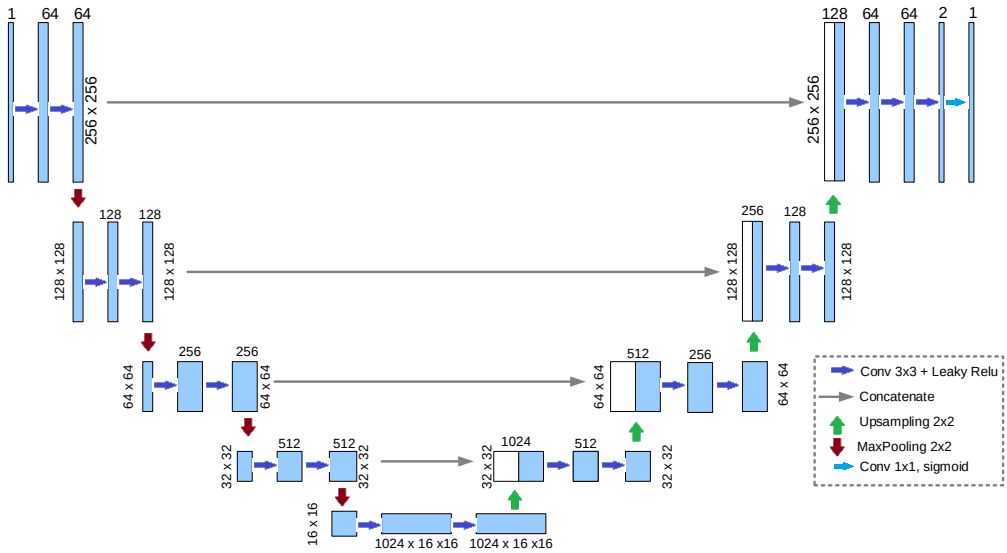



**Figure 3**: Structure of U-Net in case of 256x256 image and input of image and mask

## 3.2. U-Net++ architecture

U-Net++ architecture was created in 2020 in order to increase accuracy of medical image segmentation [20]. This architecture is based on nested and dense skip connections. The idea behind this architecture is that model can more effectively capture foreground object details when feature maps from down-sampling path are enriched before concatenation with the corresponding feature maps from up-sampling path.

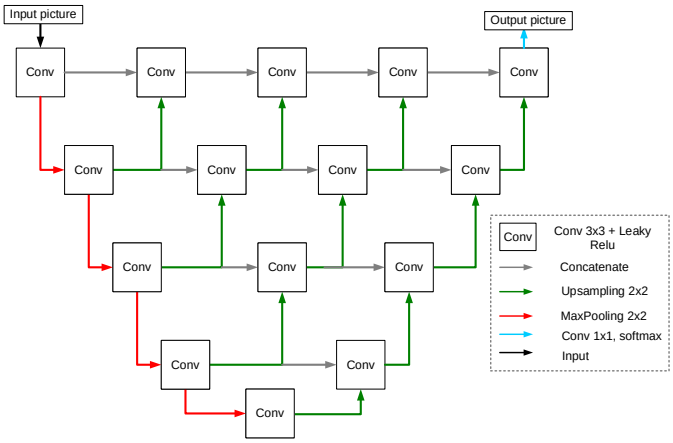



**Figure 4**: Structure of U-Net++ network.

The main difference between U-Net++ and U-Net is re-design skip connections they can be seen in Figure 4. Instead of just concatenating feature maps from down-sampling path and up-sampling path feature maps undergo a dense convolution block. Dense convolution block brings down-sampling path feature maps closer to feature maps of up-sampling path. This architecture allows the use of deep supervision, where the outputs from each up-sampling path's blocks would be average improving models accuracy but slowing it down. This feature of Unet++ was not used here therefore, it is not showed in Figure 4.

## 3.1. MultiResUNet

MultiResUnet aims to make U-Net suitable for multi-resolutional analysis by incorporating 3x3 and 7x7 convolutions in parallel to the 5x5 convolutions which are resembled by two 3x3 convolutions. Having 7x7 convolution would increase memory needed to train model so to go around that they will also be substituted by sequence of 3x3 convolutions [25]. Outputs of these convolutional blocks are concatenated. A residual connection and 1x1 convolution is added to keep additional spatial information see Figure 5.

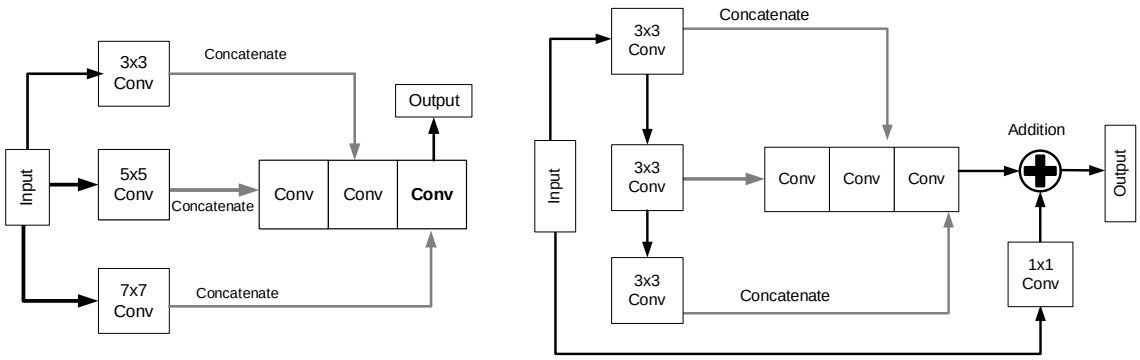



**Figure 5:** Memory intense and optimized MultiRes Blocks

In this U-Net architecture convolutional layers and a residual connection is added to shortcut connection so instead of just concatenating feature maps from down-sampling paths first they go through convolutional layers with residual connection and only then are concatenated with the up-sampling feature map Figure 6.

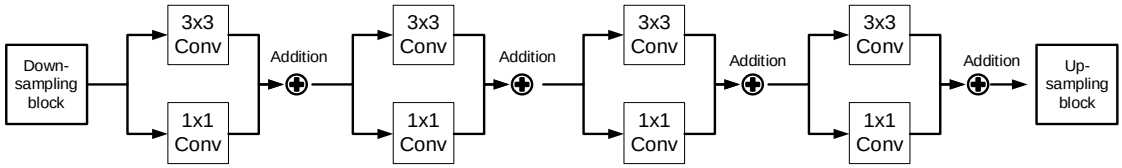
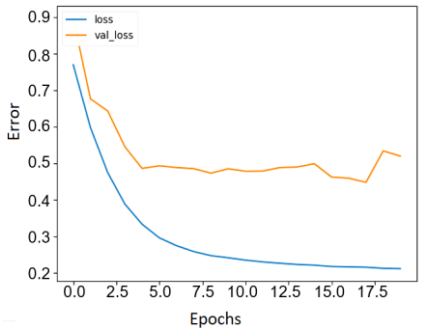



**Figure 6:** Res Path that replaces simple concatenation

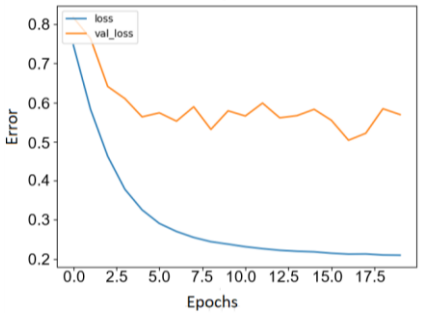
Each MultiRes model block has parameter $W$, which sets the number of filters of the convolutional layers in the block. $W = 1.67\ x\ U$ where $U$ is corresponding layer of U-net ($U$ can have values of 32, 64, 128, 256, 512). Number of filters become $\frac{w}{6}$, $\frac{w}{3}$, $\frac{w}{2}$ and are assigned to the three successive layers. Number of convolutional blocks used in the Res paths are gradually reduced from 4 to 1 as model goes deeper.

## 4. Experimental results

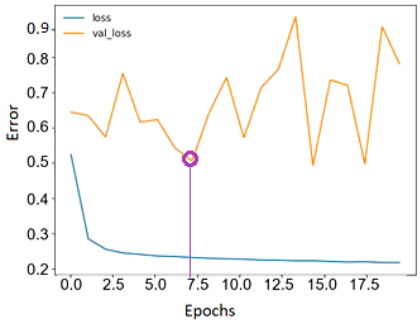
Each of three models was trained for 20 epochs each epoch taking 1773 iterations for training and 500 iterations for validation (Figure 7). Chosen batch size was 8 due to hardware limitations. Original U-Net managed to get lowest validation loss out of all however it also provides the lowest dice coefficient score. Unet++ managed to get highest accuracy score and lowest loss out of all models however it also got highest validation loss. Result depicted in the Figure 7c shows that MultiResUNet is overtrained after 7 epochs, and therefore the model was retrained for only 10 epochs. This retrained model got highest loss, validation accuracy and dice coefficient.



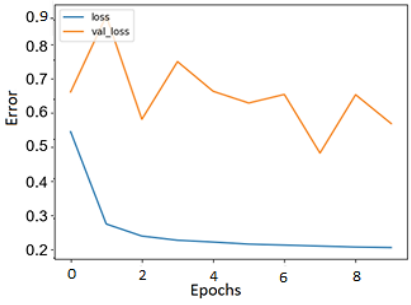

a) U-Net loss function dependencies




b) U-Net++ loss function dependencies




c) MultiResUNet loss function during 20 epochs




d) MultiResUNet loss function for 10 epochs

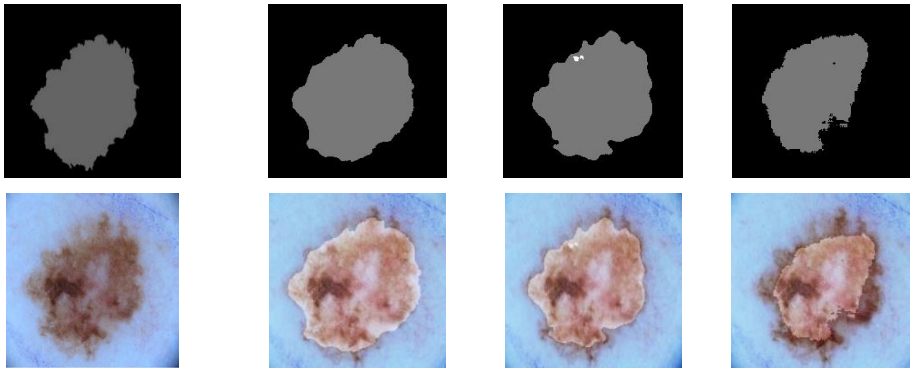
**Figure 7**: cross-entropy loss function for different architectures

In this study, we used five metrics to evaluate the model performance: accuracy, loss, valid accuracy, valid loss and dice coefficient. Dice coefficient, which quantifies the similarity between the model output and reference masks. Form the Table 1, we can observe that in the terms of accuracy and dice coefficient the most promising result (91.45% accuracy, 73.56% dice coefficient) was achieved by MultiResUNet, while the poorest performance was demonstrated by classic U-Net architecture. MultiResUNet performed 3.70% higher than U-Net and 2.61% than U-Net++ corresponding dice coefficient results.

**Table 1**
Comparison of segmentation accuracy by different U-Net type architectures

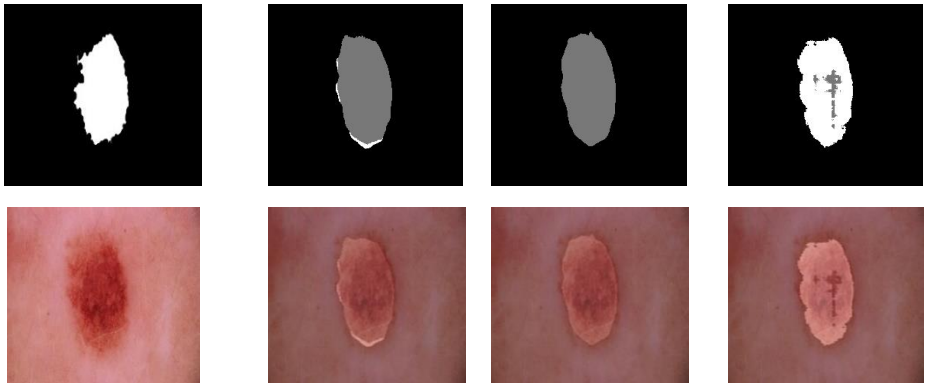
| Model | Epochs | Accuracy | Loss | Valid Accuracy | Valid Loss | Dice Coefficient |
|---|---|---|---|---|---|---|
| U-Net | 20 | 0.8465 | 0.2177 | 0.7450 | 0.5361 | 0.7096 |
| U-Net++ | 20 | 0.9223 | 0.1938 | 0.7986 | 0.5768 | 0.7176 |
| MultiResUnet | 20 | 0.9163 | 0.2181 | 0.7616 | 0.7997 | - |
| MultiResUnet | 10 | 0.9145 | 0.2258 | 0.8058 | 0.5385 | 0.7369 |

One instance of malignant and benign skin lesion from the dataset has been selected to demonstrate the performance of different U-Net type architectures (Figure 8 and Figure 9). Observing the segmentation results in the figure 8, we can see that the malignant skin lesion is detected using all three architectures. Observing the segmentation results in the figure 9, we can see that the benign skin lesion is detected only by MultiResUnet architecture. When comparing the boundaries of the mask image, we can observe that contours differ. In all cases models manage to locate and segment skin moles, however all models output wrong type of mask for both malignant and benign skin lesions reducing dice coefficient value.



a) original image and mask b) U-Net c) U-Net++ d) MultiResUnet

**Figure 8:** Example of malignant input image overlayed by output masks retrieved from three models



a) original image and mask b) U-Net c) U-Net++ d) MultiResUnet

**Figure 9:** Example of benign input image overlayed by output mask retrieved from three models

## 5. Conclusions and final remarks

In this paper, the experimental investigation of multi class melanoma segmentation of three U-Net type architectures: U-Net, U-Net++ and MultiResUNet is demonstrated. The selected methods were tested using skin lesion datasets, which contain malignant and benign skin lesions. The overall accuracy is more or less similar for all models used in the study: the worst performance 84.65% was demonstrated by the classical U-Net model, and the best accuracy 92.23% was achieved by U-Net++, but this result is higher only by 0,86% when compared to MultiResUNet. Experimental results have confirmed that U-Net type architectures are very successful in the single class medical image segmentation, and new modified architectures can slightly improve the performance results of classical U-Net. However, multi class segmentation of skin lesions remains a though task.

## References


[1] Skin cancer statistics, Melanoma of the skin is the 19th most common cancer worldwide, URL: http:www.wcrf.org/dietandcancer/cancer-trends/skin-cancer-statistics, 2021.

[2] O. Tessler, A. M. Lin, S. Patel, C. L. Dupin, Reconstruction Options for Trunk and Extremity Melanoma, Journal Melanoma, 2018, pp. 411-424.

[3] E. L. Davis, S. C. Shalin and A. J. Tackett. Current state of melanoma diagnosis and treatment Journal: Cancer biology and therapy. Vol. 20, No. (11), pp. 1366–1379, 2019. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6804807.

[4] D. Patino et al., Melanoma detection on dermoscopic images using superpixels segmentation and shape-based features, 15th International Symposium on Medical Information Processing and Analysis; Vol. 11330, 2020, URL: https://doi.org/10.1117/12.2545300

[5] F. Khan G. Nalinipriya. Texture Based Automated Segmentation of Skin Lesions using Echo State Neural Networks, Journal of Electrical Engineering and Technology, Vol 12(1), 2017, pp. 436-442.

[6] Local edge-enhanced active contour for accurate skin lesion border detection, BMC Bioinformatics, URL: biomedcentral.com

[7] Z. Liu, J. Zerubia, Skin image illumination modeling and chromophore identification for melanoma diagnosis, Physics in Medicine and Biology, Vol. 60, no. 9, 2015, pp. 3415-3431.

[8] R. Garnavi, Skin Lesion Segmentation Using Color Channel Optimization and Clustering-based Histogram Thresholding, Vol.3, no. 12, 2009.

[9] E. Harrington et al. "Diagnosing malignant melanoma in ambulatory care: A systematic review of clinical prediction rules", Vol. 7, no. 3, BMJ Open 2017, Vol. 7.

[10] Y. Togawa, Dermoscopy for the Diagnosis of Melanoma: An Overview, Dermoscopy, 2017, pp. 1-10.

[11] X. Zhang, Melanoma segmentation based on deep learning, Computer Assisted Surgery, DOI: 10.1080/24699322.2017.1389405, 2017, pp. 267-277.

[12] E. M. Senan, M. E. Jadhav, Analysis of Dermoscopy Images by Using ABCD Rule for Early Detection of Skin Cancer, ScienceDirect, 2021.

[13] A. Sagar, J. Dheeba. Convolutional Neural Networks for Classifying Melanoma Images, 2020, pp. 1-12, URL: https://www.biorxiv.org/content/10.1101/2020.05.22.110973v2.full.pdf.

[14] N. Nida, A. Irtaza, A. Javed, M. H. Yousaf, M. T. Mahmood. Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering, Int Joural Med Inform, 2019, pp. 37-48.

[15] S. Albahli et. al. Melanoma Lesion Detection and Segmentation Using YOLOv4-DarkNet and Active Contour, IEEE Access, 2020, pp.1-13.

[16] S. Banerjee, S. K. Singh, A. Chakraborty, A. Das and R. Bag. Melanoma Diagnosis Using Deep Learning and Fuzzy Logic, In Diagnostics, Vol. 10, 2020, pp.1-26.

[17] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, In Computer Vision and Pattern Recognition, 2015, pp. 1-8.

[18] Z. Lom, C. Yakopcic, T.M. Taha, V.K. Asari, Nuclei Segmentation with Recurrent Residual Convolutional Neural Networks based U-Net (R2U-Net). In Proceedings of the NAECON 2018—IEEE National Aerospace and Electronics Conference, USA, 2018, pp. 228–233.
[19] H. Huang, et.al., UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 1055–1059.
[20] Z. Zhou, Md M. R. Siddiquee, N. Tajbakhsh, and J. Liang. UNet++: A Nested U-Net Architecture for Medical Image Segmentation, 2018, pp. 1-8. URL: https://arxiv.org/pdf/1807.10165.pdf
[21] X. Li, et. al., UNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes. IEEE Trans. Med. Imaging 2018, 37, pp. 2663–2674.
[22] S. Tran, C.H Cheng, T.T. Nguyen, M. H. Le, and D. G. Liu, TMD-Unet: Triple-Unet with Multi-Scale Input Features and Dense Skip Connection for Medical Image Segmentation, Healthcare 2021, Vol. 9, No. 1, pp. 1-19.
[23] C.Y., Lee, S, Xie, P. Gallagher, Z. Zhang, Z. Tu, Deeply-supervised nets, Artificial Intelligence and Statistics, 2015, pp. 562–570.
[24] ISIC. The International Skin Imaging Collaboration, URL: ISIC Archive (isic-archive.com)
[25] N. Ibtehaz and M. S.Rahman, MultiResUnet : Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation, 2019, pp. 1-25. URL: https://arxiv.org/pdf/1902.04049.pdf