Deep Learning in Medical Applications: Lesion Segmentation in Skin Cancer Images Using Modified and Improved Encoder-Decoder Architecture

Ranpreet Kaur, Hamid GholamHosseini, and Roopak Sinha

ECMS, Auckland University of Technology, Auckland 1010, New Zealand
ranpreet.kaur@aut.ac.nz

Abstract. The rise of deep learning techniques, such as a convolutional neural network (CNN) in solving medical image problems, offered fascinating results that motivated researchers to design automatic diagnostic systems. Image segmentation is one of the crucial and challenging steps in the design of a computer-aided diagnosis system owing to the presence of low contrast between skin lesion and background, noise artifacts, color variations, and irregular lesion boundaries. In this paper, we propose a modified and improved encoder-decoder architecture with a smaller network depth and a smaller number of kernels to enhance the segmentation process. The network performs segmentation for skin cancer images to obtain information about the infected area. The proposed model utilizes the power of the VGG19 network’s weight layers for calculating rich features. The deconvolutional layers were designed to regain spatial information of the image. In addition to this, optimized training parameters were adopted to further improve the network’s performance. The designed network was evaluated for two publicly available benchmarked datasets ISIC, and PH2 consists of dermoscopic skin cancer images. The experimental observations proved that the proposed network achieved the higher average values of segmentation accuracy 95.67%, IoU 96.70%, and BF-score of 89.20% on ISIC 2017 and accuracy 98.50%, IoU 93.25%, and BF-score 84.08% on PH2 datasets as compared to other state-of-the-art algorithms on the same datasets.

Keywords: Image segmentation · Medical image analysis · Skin cancer · Deep learning

1 Introduction

Medical image analysis is the process of analyzing image information generated in clinical settings. The primary target behind image analysis is to obtain relevant information for diagnosis. The recent developments in computer vision techniques have made medical image analysis one of the exciting research areas. Deep learning is one of the popular choices of researchers to use in solving medical problems, such as lung cancer [1], breast tissue detection [2], glaucoma eye diagnosis [3], and blood pressure monitoring [4].
research focuses on the extraction of lesion areas from skin cancer images that would further analyze the type of cancer, such as benign and malignant. Skin cancer is the most commonly found problem around the world that occurs due to overexposure to ultraviolet sun rays. The incidence of skin cancer has been overgrowing, and it is the 19th most common cancer worldwide [5], thus become an issue of great concern in the healthcare community.

Skin cancer is the excessive growth of abnormal cells in the outermost skin layer known as the epidermis. The cancerous cells overgrow into the body and damage surrounding tissues to form a malignant tumor. The main type of skin cancer is squamous cell carcinoma, basal cell carcinoma, and melanoma [8], as shown in Fig. 1. Melanoma is rare but the deadliest form of other skin cancer types because the cancerous cells multiply unexpectedly. The risk factors that contribute towards the formation of melanoma include fair complexion, overexposure to sun rays, sunburn, genetic history, and weak immune system [6, 7].

![Skin cancer types](image)

**Fig. 1.** Skin cancer types: (a) Basal cell carcinoma (b) Squamous cell carcinoma (c) Melanoma.

Researchers are on track to develop an efficient and cost-effective software tool to detect cancer at an early stage, which can significantly reduce the mortality rate. Deep learning approaches are widely applied in image analysis applications. For the accurate diagnosis of medical problems, it is pertinent to understand image patterns that are possible through the image segmentation process. Segmentation is one of the crucial steps of image analysis to differentiate object descriptions from their background [9]. Manual skin image segmentation is a tedious and time-consuming task in a clinical environment to identify each patient’s skin patterns. Thus, there is a tremendous need for developing an automatic segmentation approach which is the primary aim of the current research. To design an automated classification system, a precise and efficient lesion segmentation approach is one of the key requirements.

We introduced three contributions in this paper, firstly four blocks of down-sampling layers consisted of convolutional, batch normalization, leaky ReLU, and pooling layers were designed as encoder block to extract spatial information of the image and to store max-pooling indices. Corresponding to each down-sampling layer, the decoder section was designed to perform up-sampling. The indices computed by down-sampling layers
were used by up-sampling layers in the decoder section to regain the spatial resolution. Secondly, a leaky ReLU layer was used instead of the ReLU layer to add more balance and faster learning. At last, the proposed framework was effectively evaluated for a large database to make it more general for other applications.

2 Related Study

Many attempts have been made to extract lesions using traditional image processing techniques, machine learning algorithms, and the latest deep learning approaches. There is no standard method of segmenting images because the area of interest and object description varies from application to application. Hence selection of segmentation technique depends upon the type of images given in the input. Segmentation of lesions is a challenging task due to the large variations in terms of color, size, shape, location, and texture. The approaches that have been applied in the field of lesion segmentation are categorized into six groups, edge-based methods [10, 11], thresholding-based methods [12–16], clustering methods [17, 18], active contour methods [19–21], and supervised algorithms [22–24]. These old image processing methods have become indispensible in solving very complex and challenging image segmentation problems. In 2013, an initial attempt was made to employ a deep learning approach for the segmentation of X-ray images [25]. They applied patch-wise classification on raw images to identify bone tissues and selected only rib regions to reduce training time. A similar approach was applied to fundus images for the segmentation of blood vessels proposed by Melinscak et al. in 2015 [26]. They used a deep convolutional neural network that made use of max-pooling layers instead of subsampling or downsampling layers and achieved an average accuracy of 94%.

In 2015, Long et al. proposed the first end-to-end pixel-wise segmentation technique, a fully convolutional neural network (FCN) [27]. They adapted AlexNet, VGGNet, and GoogleNet classification models for semantic segmentation by replacing convolutional layers with fully connected layers to generate segmentation maps. In 2015, Hong et al. proposed a semi-supervised learning approach, deconvolution networks (DeconvNet), as an extension of FCN networks [28]. This architecture decoupled classification and segmentation tasks by adding bridging layers to construct class-oriented feature maps that reduce training and inferencing time. The main problem with FCN was pooling layers, that increased field of view and discarded the actual information. In 2017, an idea of encoder-decoder-based architecture was presented by Badrinarayanan et al. [29] that has a network layout similar to VGG16. In this network, the encoder section reduces spatial dimensionality with pooling layers and the decoder gradually regains the object information and spatial details. U-Net is another famous network inspired by encoder-decoder architecture which is successfully applied for the segmentation of medical images [30]. More details of these networks are given in review papers [31, 32]. Some research work studied the impact of changing network architecture using a transfer learning approach in terms of layer structure [33–35], optimizing training parameters [36–39], aiming to reach an optimal solution that reduces computation time, power and increase network accuracy.
Motivated by the recent developments in deep learning algorithms, we aimed to propose a modified encoder-decoder framework to perform segmentation. Instead of applying any pre-processing or post-processing approach to improve segmentation accuracy, we focused on designing an end-to-end learning network that can process raw images contaminated with different noise levels and can accurately extract a region of interest. In this paper, we explored the VGG19 network [40], which was originally designed for the ImageNet classification task. The network applied convolutional and pooling operation over an entire image to extract features. A cross-entropy loss function was used to predict the pixel-wise labels. The lesion occupies a small portion of the whole image that tends loss function to be biased towards background information. Thus, a weighted cross-entropy loss function was used to improve segmentation results. The designed framework was evaluated on two publicly available datasets, ISIC, and PH2.

3 Materials and Methods

3.1 Encoder-Decoder Framework

The proposed model was inspired by VGG19 [40] deep neural network, whose layers were modified and improved to design deep encoder-decoder architecture for lesion segmentation. The CNN architecture designed for skin cancer segmentation is shown in Fig. 2. The idea of an encoder-decoder network was originally designed for doing semantic segmentation of road scene applications. The proposed network was designed by changing network depth, layer configuration, and optimizing hyperparameters. The first half of the architecture called the encoder section having four blocks that further consists of multiple convolutional, batch normalization, leaky ReLU, and max-pooling layers. In contrast, the lower half was the decoder section having deconvolutional layers, batch normalization, leaky ReLU layer, and max-unpooling layers corresponding to each layer in the encoder section, followed by a cross-entropy loss function-based segmentation layer. In this network, the pooling indices computed by the max-pooling layers of the encoder section were used to perform up-sampling in the decoder section. The fully connected layers in the previous networks were removed, which made the encoder section smaller and reduced the number of learnable parameters.

In the encoder section, a convolutional operation was performed with a set of kernels to generate feature maps. These were then batch normalized and the negative values were set to any fixed value by applying element-wise leaky rectified linear units (PReLU). Following this, a max-pooling operation was implemented with a $2 \times 2$ window size, a stride of 2, and padding 0. The max pooling was used to reduce the size of feature maps while preserving significant characteristics. It is often placed between two convolutional layers and is responsible for decreasing the number of parameters and calculations in the network. Pooling operation can be applied using the equation as:

$$FM[x, y] = f_p \times I_{x,y}$$

$FM[x, y]$ shows the output feature map, $f_p$ is the pooling function, $I_{x,y}$ is the input feature map from the previous layer. Max pooling calculates a maximum value from the portion overlapped by the kernel. The repetitive use of max-pooling and striding
at consecutive layers reduces the spatial resolution of the output feature map. Some architectures such as fully convolutional networks used a ‘transposed convolutional layer’ but it results in more time and memory. In the original VGG19 network, there was the successive use of several convolutional layers to achieve high translation invariance for more robust classification. However, there was a loss of spatial resolution which is not suitable for segmentation. Therefore, it is significant to preserve boundary information after sub-sampling layers. To achieve this, max-pooling indices were stored, i.e. the location of maximum features was memorized and used in the decoder section to perform upsampling that regains spatial resolution. The high dimensional feature maps were fed to a multi-class softmax pixel-wise classifier that generates K-channel probabilities for each pixel, K is the number of classes. We add here, Deconvnet, SegNet and U-Net shares similar kind of architecture, but have some differences in terms of layer organization, hyperparameters, and training. Deconvnet is hard to train end-to-end due to the presence of fully connected layers, SegNet is very time consuming and U-Net does not reuse pooling indices.

### 3.2 Network Architecture Details

The network consists of 60 layers designed from scratch broadly divided into encoder and decoder sections, as shown in Fig. 2. The layers of the encoder block follow the VGG19 layout with some additional layers, whereas the corresponding decoder block is our contribution. The first section has 29 layers that extract object details by downsampling the feature maps and the second section has respective deconvolutional layers to upsample the image information. The network consists of a total of 60 layers, 134 layer connections, and generates an output feature map of different sizes after each layer, and the network details are demonstrated in Table 1.

![Fig. 2. The proposed CNN architecture for skin cancer segmentation](image)

**Encoder Section.** This is the first half of the section responsible for down-sampling of feature maps and extracts features of input images using the following layers.
**Convolutional Layer:** This layer in the encoder section generates a feature map by sliding a kernel over an entire image and determines the product between the kernel and underlying region of an image using the formula:

\[
FM[x, y] = \sum_{i=1}^{l} \sum_{j=1}^{l} [I_{x-i, y-j}] \times K[i, j]
\]  

(2)

FM[x, y] represents the output feature map, [x, y] shows the pixel positions in the spatial domain, l is the kernel size, I is the input image, and K is the kernel or template. The other parameters used in this layer are kernel size, number of kernels, stride, and padding. The kernel size chosen for the current network is 3 × 3, as this size is suitable for images and covers all pixels positions of an image. Moreover, the larger kernel size leads to more time and memory consumption. Each encoder performs a convolutional operation with multiple kernels to produce a feature map. The higher number of kernels helps extract very fine details of the image, but it also takes a longer execution time. We chose the number of kernels as 64 kernels; however, in original VGG19 networks, the number of kernels at each convolutional layer varies from 64 to 512. Then, the stride value is taken as ‘1’ that defines the step size of the kernel when sliding over an image. The last parameter is padding selected as ‘1’ which adds extra ones at the end of the image border to prevent cropping.

**Batch Normalization Layer:** The outputs of the convolutional layer, i.e. output feature map are fed to batch normalization. This layer does not reduce the size of the feature map. It allows each layer to learn on a more stable distribution of inputs, i.e. rescales inputs between 0 and 1 and accelerate the training of the network.

**PReLU Layer:** It applied the threshold function to each element in the input and multiplied all negative values by a fixed scalar. This layer passes the output element as the input to the next layer directly if it is positive; otherwise, its output to value multiplied by a fixed scalar using the following activation function:

\[
PReLU(x) = \begin{cases} 
    \text{scalar} \times x, & x < 0 \\
    x, & x \geq 0 
\end{cases}
\]  

(3)

**Max-pooling Layer:** This layer reduces the spatial size of the network by decreasing the number of parameters and computation, also called down-sampling. Max-pooling layer in the current network used window size ‘2 × 2’, padding [0 0 0 0] and stride value [2 2]. The stride value in the max-pooling layer denotes the number of shifts over the input matrix and padding inserts extra zeros so that the filter exactly fits the input image.

**Decoder Section.** The decoder in the second half of the network performs the reverse operation of pooling and reconstructs the original image resolution. The deconvolutional layers used the same number of kernels with the same size, stride, and padding value to obtain an activation map through up-sampling layers that swap forward and backward pass of a convolution. As a result, a dense feature map is produced that represents high image resolution. Then batch normalization and PReLU layers are placed in the same pattern as in the encoder section. After that, the max-pooling indices applied that were calculated during the first section to obtain the same image resolution. In this way, the network learns from global information and fine details to perform segmentation.
Table 1. Architecture and parameter details of the proposed network.

| Encoder-section | Decoder-section |
|------------------|-----------------|
| Layer            | Filter | Output size | Layer            | Filter | Output size |
| Conv1_1          | 3 × 3   | 300 × 300 × 64 | UnPool4         | 2 × 2   | 37 × 37 × 64 |
| Conv1_2          | 3 × 3   | 300 × 300 × 64 | De_Conv4_2      | 3 × 3   | 37 × 37 × 64 |
| Pool1            | 2 × 2   | 150 × 150 × 64 | De_Conv4_1      | 3 × 3   | 37 × 37 × 64 |
| Conv2_1          | 3 × 3   | 150 × 150 × 64 | UnPool3         | 2 × 2   | 75 × 75 × 64 |
| Conv2_2          | 3 × 3   | 150 × 150 × 64 | De_Conv3_2      | 3 × 3   | 75 × 75 × 64 |
| Pool2            | 2 × 2   | 75 × 75 × 64  | De_Conv3_1      | 3 × 3   | 75 × 75 × 64 |
| Conv3_1          | 3 × 3   | 75 × 75 × 64  | UnPool2         | 2 × 2   | 150 × 150 × 64 |
| Conv3_2          | 3 × 3   | 75 × 75 × 64  | De_Conv2_2      | 3 × 3   | 150 × 150 × 64 |
| Pool3            | 2 × 2   | 37 × 37 × 64  | De_Conv2_1      | 3 × 3   | 150 × 150 × 64 |
| Conv4_1          | 3 × 3   | 37 × 37 × 64  | UnPool1         | 2 × 2   | 300 × 300 × 64 |
| Conv4_2          | 3 × 3   | 37 × 37 × 64  | De_Conv1_2      | 3 × 3   | 300 × 300 × 64 |
| Pool4            | 2 × 2   | 18 × 18 × 64  | De_Conv1_1      | 3 × 3   | 300 × 300 × 2 |

Pixel Classification Layer and Weighted Cross-Entropy Loss Function. The output of the decoder blocks given to the pixel classification layer that generates each pixel’s posterior probability using weighted cross-entropy loss [41]. This loss function measures the error between prediction scores \( P \) and targets \( T \). The weighted cross-entropy loss between \( P \) and \( T \) can be calculated as:

\[
\text{Loss} = \frac{1}{N} \sum_{i=1}^{K} \sum_{n=1}^{N} w_i T_{ni} \log(P_{ni})
\]  

(4)

Here, \( N \) is the number of observations, \( K \) is the number of classes, and \( w \) is a vector of weights determined by the network for each class. The entropy loss is mostly used to evaluate the performance of medical image segmentation in comparison to the ground truth images.

Network Training Algorithm. Training a deep neural network for 5694 images with different lesion shapes was a challenging task. As explained in the previous section, the network consists of 60 layers where 8 convolutional and 4 pooling layers perform encoding function and corresponding 8 deconvolutional and 4 unpooling layers to regain the original information. In the end, the classification layer classifies each pixel into 2 classes, i.e. lesion or background, and generated a posterior probability map for each pixel. Few parameters were fine-tuned for the proposed network to execute effective training.

Adaptive Moment Estimation (ADAM) Optimizer: This optimization algorithm emerged from gradient descent and momentum to update the network training parameters, weights, and bias aiming to minimize loss value. It computes the adaptive learning rate for each
parameter of the network, unlike stochastic gradient descent that calculates a single learning rate for all weights updates. The original gradient descent for updating the network in the direction of the negative value of loss function given as:

$$\theta_{i+1} = \theta_i - \alpha \nabla L(\theta_i)$$  \hspace{1cm} (5)

Here, $i$ denotes the iterations, $\alpha > 0$ is the learning rate taken as 0.01, $\theta$ is the parameter vector and $L(\theta)$ is the loss function. In the standard stochastic gradient method, the gradient of the loss function, $\nabla L(\theta)$ is evaluated for the entire training dataset. In contrast, the ADAM optimizer adds a term called ‘momentum’ to do an element-wise average of parameter gradients and their squared values. The above equation can be rewritten as:

$$m_i = \beta_1 m_{i-1} + (1 - \beta_1) \nabla L(\theta_i)$$  
$$v_i = \beta_2 v_{i-1} + (1 - \beta_2)[\nabla L(\theta_i)]^2$$  \hspace{1cm} (6)

The optimizer used moving average to update the network as:

$$\theta_{i+1} = \theta_i - \frac{\alpha m_i}{\sqrt{v_i} + \varepsilon}$$  \hspace{1cm} (7)

$\beta_1$ and $\beta_2$ are gradient decay factor set as ‘0.95’ and squared gradient decay factor set as ‘0.99’ respectively and the learning rate $\alpha$ was specified as 0.03.

### 4 Simulations and Results Discussion

#### 4.1 Dataset Preparation

Data collected from the ISIC 2018 archive repository [42], was nearly 5694 images of BCC, Melanoma, Squamous, and Nevus skin cancer types. Approximately 3100 images from ISIC 2017 challenge were used for testing purposes. Another dataset used for testing was PH2 [43], which consists of almost 200 images, including ground truth images. Ground truth images are necessary for evaluating segmentation results to predict whether the approach correctly classified all the pixels into the foreground and background. The original images had dimensions $1022 \times 767 \times 3$ which is too large to process and can slow down the processing of the system. Thus images were downsized to $300 \times 300 \times 3$. Different image size was taken into consideration, such as $224 \times 224$, $227 \times 227$, and $256 \times 297$, but the $300 \times 300$ had given optimal performance. To prepare datasets for training and validation, the whole dataset is split into the ratio of 80%, and 20% respectively (Fig. 3).
4.2 Quality Metrics

The performance evaluation metrics used in our experiments are accuracy, intersection over union (IoU), and BF score. The values of these parameters are expected to be high for efficient segmentation and these are calculated on the testing dataset. The parameters TP, TN, FP, FN denotes the true positives, true negatives, false positives, and false negatives.

**Accuracy** is the number of corrected pixels identified over the total number of pixels.

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

**Intersection Over Union (IoU)** is the ratio of correctly classified pixels to the total number of pixels that are assigned a class by the ground truth and the predictor.

\[
\text{IoU} = \frac{TP}{TP + FP + FN}
\]

**BF Score** computes the Boundary F1 contour matching score between the predicted segmentation in prediction and the true segmentation in the ground truth.

\[
\text{BF Score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{Precision} + \text{recall}}
\]

4.3 Comparison with Other State-of-the-Art Methods

To evaluate the effectiveness of the network, it is compared with other state-of-the-art methods from the last four years (2016–2019) and it is observed that the designed network outperformed as compare to others. The qualitative segmentation results using the proposed network are displayed for the input images (see Fig. 4). It showed that the lesion region was extracted accurately, however, some irrelevant background object information exists for a few highly noisy images.
The proposed network was trained for 5694 RGB images and tested for 3300 images different from the training set to measure the performance parameters. The comparative analysis with other similar research studies based on these metrics on the ISIC dataset is illustrated in Table 2. It clearly shows the higher performance of the proposed approach as compared to others in terms of accuracy and IoU also called a Jaccard index which should be expected higher for better segmentation results. The literature studies chosen for the comparison are based on the same dataset, i.e., ISIC 2017 and PH2. Our network achieved a high value of this parameter with a difference of 0.43% as compared to other mentioned studies. Moreover, it gained almost higher accuracy and BF score except given in [39] but it outperformed other studies with the highest Jaccard score of 96.7% and accuracy of 95.6%.

### Table 2. Performance-based on the ISIC 2017 dataset and compared to other similar studies.

| Method                     | Accuracy% | BF% | IoU% |
|----------------------------|-----------|-----|------|
| Proposed Approach          | 0.956     | 0.892 | 0.967 |
| FCRN [39]                  | 0.947     | 0.927 | 0.962 |
| VGG16 [39]                 | 0.903     | 0.794 | 0.707 |
| GoogleNet [39]             | 0.911     | 0.848 | 0.776 |
| Modified SegNet [44]       | –         | 0.853 | 0.771 |
| Supervised Saliency Detection [45] | 0.930 | 0.839 | 0.749 |
| Jeremy Kawahara [34]       | 0.930     | 0.837 | 0.752 |
| FCN with Jaccard loss function [46] | 0.934 | 0.849 | 0.765 |
| DCI-PSL [47]               | 0.940     | 0.856 | 0.777 |
Deep Learning in Medical Applications 49

Table 3 illustrates parameters on the PH^2 dataset, and it shows a higher performance of the proposed network over other approaches from the years 2016–2019 in terms of accuracy and IoU. Our model is gaining higher accuracy and overlapping index (IoU) as compared to other state-of-the-art methods which show the network is efficient in segmenting lesions from dermoscopic images on the PH^2 dataset.

Table 3. Performance analysis based on the PH^2 dataset and compared to other similar studies.

| Method                        | Accuracy% | BF% | IoU% |
|-------------------------------|-----------|-----|------|
| The proposed approach         | 0.985     | 0.840 | **0.932** |
| Supervised Saliency Detection [45] | 0.979   | **0.952** | 0.923 |
| DCL-PSI [47]                  | 0.953     | 0.921 | 0.859 |
| MFCN [48]                     | 0.940     | 0.906 | 0.839 |

The impact of choosing the different depths of encoder-decoder architecture is illustrated in Table 4. It shows that architectural depth below 4 given a lower performance in terms of segmentation accuracy, whereas, depth higher than 4 given little increase but leads to higher execution time. Thus, ‘4’ was selected as an optimal network depth for the proposed model.

Table 4. Performance of the network by the varying depth of the architecture and evaluated on the ISIC 2017 dataset.

| Encoder-Decoder Depth | Accuracy % | Execution Time in hours (approximately) |
|-----------------------|------------|----------------------------------------|
| 2                     | 63.56      | 3                                      |
| 3                     | 79.23      | 5.5                                    |
| 4                     | **95.67**  | **6**                                  |
| 5                     | **96.08**  | 11                                     |

5 Discussion and Conclusion

In this paper, we presented a modified deep learning framework to meet the challenges of automatic lesion segmentation on dermoscopic images. The new network leverages the increased segmentation accuracy despite highly noisy data and varying lesion shapes. The design of a new network is inspired by the VGG19 classification network and encoder-decoder models. The layers of the encoder section are organized taking the VGG19 framework as a base with some additional layers such as leaky ReLU in each block. To boost the network’s performance, a leaky ReLU layer, fine-tuned training parameters such as kernel size, number of filters, learning rate, momentum, and batch
size, and the weighted cross-entropy loss function was used. The network was trained for ISIC 2018 dataset and performance was evaluated for two datasets, ISIC 2017 and PH2. It was concluded that the proposed network yields better performance as compared to the other state-of-the-art methods. It is illustrated in Tables 2 and 3 that our proposed model outperformed previous studies by gaining higher values of accuracy, and IoU index. The proposed network can be efficiently applied in clinical settings to understand lesion patterns for the early diagnosis of the type of skin cancer. In some instances, there is a lack of properly segmented region due to the presence of intense hairlines, asymmetrical shapes, and color variations. Therefore, there is room for further improvement in segmentation results by applying any pre-processing techniques and improving network configuration. Moreover, in the future, our research will focus on the design of a more optimal segmentation network having low execution time and high accuracy.

References

1. Skourt, B.A., El Hassani, A., Majda, A.: Lung CT image segmentation using deep neural networks. Procedia Comput. Sci. 127, 109–113 (2018)
2. Zhang, Y., et al.: Automatic breast and fibroglandular tissue segmentation in breast MRI using deep learning by a fully-convolutional residual neural network U-net. Acad. Radiol. 26(11), 1526–1535 (2019)
3. Yu, S., Xiao, D., Frost, S., Kanagasingam, Y.: Robust optic disc and cup segmentation with deep learning for glaucoma detection. Comput. Med. Imaging Graph. 74, 61–71 (2019)
4. Wu, D., Xu, L., Zhang, R., Zhang, H., Ren, L., Zhang, Y.-T.: Continuous cuff-less blood pressure estimation based on combined information using deep learning approach. J. Med. Imaging Health Inform. 8(6), 1290–1299 (2018)
5. Skin Cancer-Index (2018). https://www.isic-archive.com. Accessed 2019
6. What is Melanoma Skin Cancer? https://www.cancer.org/cancer/melanoma-skin-cancer/about/what-is-melanoma.html. Accessed 2019
7. Bogo, F., Peruch, F., Fortina, A.B., Peserico, E.: Where’s the lesion?: Variability in human and automated segmentation of dermoscopy images of melanocytic skin lesions. In: Dermoscopy Image Analysis, pp. 82–110. CRC Press (2015)
8. Types of Skin Cancer. https://www.everydayhealth.com/skin-cancer/types. Accessed 2020
9. Sridevi, M., Mala, C.: A survey on monochrome image segmentation methods. Procedia Technol. 6, 548–555 (2012)
10. Barcelos, C.A.Z., Pires, V.: An automatic based nonlinear diffusion equations scheme for skin lesion segmentation. Appl. Math. Comput. 215(1), 251–261 (2009)
11. de Souza Ganzeli, H., Bottesini, J.G., de Oliveira Paz, L., Ribeiro, M.F.S.: Skan: Skin scanner-system for skin cancer detection using adaptive techniques. IEEE Latin Am. Trans. 9(2), 206–212 (2011)
12. Norton, K.A., et al.: Three-phase general border detection method for dermoscopy images using non-uniform illumination correction. Skin Res. Technol. 18(3), 290–300 (2012)
13. Qaisar Abbas, M., Celebi, E., Fondón, I.: Computer-aided pattern classification system for dermoscopy images. Skin Res. Technol. 18(3), 278–289 (2012)
14. Garnavi, R., Aldeen, M., Celebi, M.E., Varigos, G., Finch, S.: Border detection in dermoscopy images using hybrid thresholding on optimized color channels. Comput. Med. Imaging Graph. 35(2), 105–115 (2011)
15. Ma, Z., Tavares, J.M.R.: A novel approach to segment skin lesions in dermoscopic images based on a deformable model. IEEE J. Biomed. Health Inform. 20(2), 615–623 (2015)
16. Cavalcanti, P.G., Scharcanski, J.: A coarse-to-fine approach for segmenting melanocytic skin lesions in standard camera images. Comput. Methods Programs Biomed. 112(3), 684–693 (2013)
17. Shan, P.: Image segmentation method based on K-mean algorithm. EURASIP J. Image Video Process. 2018(1), 1–9 (2018). https://doi.org/10.1186/s13640-018-0322-6
18. Choudhry, M.S., Kapoor, R.: Performance analysis of fuzzy C-means clustering methods for MRI image segmentation. Procedia Comput. Sci. 89, 749–758 (2016)
19. Kasmi, R., Mokrani, K.: Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule. IET Image Proc. 10(6), 448–455 (2016)
20. Vasconcelos, F., Medeiros, A., Peixoto, S., Filho, P.: Automatic skin lesions segmentation based on a new morphological approach via geodesic active contour. Cognit. Syst. Res. 55, 44–59 (2019)
21. Bayraktar, M., Kockara, S., Halic, T., Mete, M., Wong, H.K., Iqbal, K.: Local edge-enhanced active contour for accurate skin lesion border detection. BMC Bioinform. 20(2), 91 (2019). https://doi.org/10.1186/s12859-019-2625-8
22. Manjón, J.V., et al.: MRI white matter lesion segmentation using an ensemble of neural networks and overcomplete patch-based voting. Comput. Med. Imaging Graph. 69, 43–51 (2018)
23. Tan, T.Y., Zhang, L., Lim, C.P., Fielding, B., Yu, Y., Anderson, E.: Evolving ensemble models for image segmentation using enhanced particle swarm optimization. IEEE Access 7, 34004–34019 (2019)
24. Alshayeji, M.H., Al-Rousan, M.A., Ellrath, H., Abed, S.: An efficient multiple sclerosis segmentation and detection system using neural networks. Comput. Electr. Eng. 71, 191–205 (2018)
25. Cernazanu-Glavan, C., Holban, S.: Segmentation of bone structure in X-ray images using convolutional neural network. Adv. Electr. Comput. Eng 13(1), 87–94 (2013)
26. Melničak, M., Prentašić, P., Lončarić, S.: Retinal vessel segmentation using deep neural networks. In: 10th International Conference on Computer Vision Theory and Applications (VISAPP 2015) (2015)
27. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
28. Hong, S., Noh, H., Han, B.: Decoupled deep neural network for semi-supervised semantic segmentation. In: Advances in Neural Information Processing Systems, pp. 1495–1503 (2015)
29. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39(12), 2481–2495 (2017)
30. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
31. Okur, E., Turkan, M.: A survey on automated melanoma detection. Eng. Appl. Artif. Intell. 73, 50–67 (2018)
32. Hesamian, M.H., Jia, W., He, X., Kennedy, P.: Deep learning techniques for medical image segmentation: achievements and challenges. J. Digit. Imaging 32(4), 582–596 (2019). https://doi.org/10.1007/s10278-019-00227-x
33. Yang, X., Zeng, Z., Yeo, S.Y., Tan, C., Tey, H.L., Su, Y.: A novel multi-task deep learning model for skin lesion segmentation and classification. arXiv preprint arXiv:1703.01025 (2017)
34. Kawahara, J., BenTaieb, A., Hamarneh, G.: Deep features to classify skin lesions. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1397–1400. IEEE (2016)
35. Nida, N., Irtaza, A., Javed, A., Yousaf, M.H., Mahmood, M.T.: Melanoma lesion detection and segmentation using deep region based convolutional neural network and fuzzy C-means clustering. Int. J. Med. Inform. 124, 37–48 (2019)

36. Nasr-Esfahani, E., et al.: Dense fully convolutional network for skin lesion segmentation arXiv preprint arXiv:1712.10207 (2017)

37. Vesal, S., Ravikumar, N., Maier, A.: SkinNet: a deep learning framework for skin lesion segmentation arXiv preprint arXiv:1806.09522 (2018)

38. Zhang, X.: Melanoma segmentation based on deep learning. Comput. Assist. Surg. 22(s1), 267–277 (2017)

39. Yu, L., Chen, H., Dou, Q., Qin, J., Heng, P.-A.: Automated melanoma recognition in dermoscopy images via very deep residual networks. IEEE Trans. Med. Imaging 36(4), 994–1004 (2016)

40. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition arXiv preprint arXiv:1409.1556 (2014)

41. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)

42. The International Skin Imaging Collaboration. www.isic-archive.com. Accessed 2019

43. Dermofit Image Library. https://licensing.edinburgh-innovations.ed.ac.uk/. Accessed 2019

44. Ninh, Q.C., Tran, T.-T., Tran, T.T., Tran, T.A.X., Pham, V.-T.: Skin lesion segmentation based on modification of SegNet neural networks. In: 2019 6th NAFOSTED Conference on Information and Computer Science (NICS), pp. 575–578. IEEE (2019)

45. Jahanifar, M., Tajeddin, N.Z., Asl, B.M., Gooya, A.: Supervised saliency map driven segmentation of lesions in dermoscopic images. IEEE J. Biomed. Health Inform. 23(2), 509–518 (2018)

46. Yuan, Y., Chao, M., Lo, Y.-C.: Automatic skin lesion segmentation using deep fully convolutional networks with Jaccard distance. IEEE Trans. Med. Imaging 36(9), 1876–1886 (2017)

47. Bi, L., Kim, J., Ahn, E., Kumar, A., Feng, D., Fulham, M.: Step-wise integration of deep class-specific learning for dermoscopic image segmentation. Pattern Recogn. 85, 78–89 (2019)

48. Bi, L., Kim, J., Ahn, E., Kumar, A., Fulham, M., Feng, D.: Dermoscopic image segmentation via multi-stage fully convolutional networks. IEEE Trans. Biomed. Eng. 64, 2065–2074 (2017)