Received: December 20, 2021. Revised: February 15, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Exudate Segmentation for Diabetic Retinopathy Using Modified FCN-8 and Dice Loss

Dinial Utami Nurul Qomariah[1] Handayani Tjandrasa[1]* Chastine Fatichah[1]

[1]*Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
* Corresponding author's Email: handatj@its.ac.id

Abstract: The severity of diabetic retinopathy can lead to blindness if undiagnosed and untreated. The presence of hard exudates is one of the diabetic retinopathy symptoms. Therefore, automatic segmentation of hard exudates can provide an important diagnosis for diabetic retinopathy. Due to the relatively small dimensions of the exudates and the availability of the optic disc that has similar color, the exudate segmentation is a challenge in itself. In this study, we propose a modification of the fully convolutional network model (FCN-8) by combining FCN-8 and shortcuts to improve the performance of FCN-8. Each shortcut consists of a convolutional layer and batch normalization to reduce input degradation. Prior to processing the hard exudates using a modified FCN-8, the optic disc was removed from the retinal image by detecting the area using Faster R-CNN based on the Alexnet architecture. For training and testing, we applied the IDRiD dataset to evaluate the performance of our proposed architecture. Experiments show that our proposed architecture provides accuracy, sensitivity, specificity of 98.18 %, 81.7%, and 98.37 % respectively. Our proposed method gives higher sensitivity compared to Autoencoder, U-Net, FCN-32, FCN-16, and FCN-8.

Keywords: Exudates, Diabetic retinopathy, Semantic segmentation, FCN-8, Dice loss.

1. Introduction

According to WHO, diabetes patients have reached 2.8 % of the entire world population in year 2000. Diabetes patients will keep increasing to up to 4.4 % in year 2030 [1]. Diabetes can cause various illness to its patients such as heart attack, stroke, glaucoma [2] and diabetic retinopathy [3]. Diabetic retinopathy is an illness caused by chronic diabetes. Diabetic retinopathy is suffered by many productive societies around the world [4, 5]. Diabetic retinopathy is characterized by vascular abnormalities in the retina [6, 7]. The severity of diabetic retinopathy is divided into two, namely non-proliferative and proliferative diabetic retinopathy [8, 9].

As diabetic retinopathy gets worse, the fatty blood vessels can rupture, causing hard exudates [10]. Hard exudates are yellowish in color and vary in size, and shape. Hard exudates can occur in the macula area, namely macular edema. Since hard exudates are an important factor for detecting diabetic retinopathy, the ability to segment hard exudates is very important for early detection and effective treatment of diabetic retinopathy. However, manual exudate segmentation using fundus images can take a long time and the results can be biased. Therefore, an automatic and accurate exudate segmentation is needed to reduce time consumed.

Several methods have been proposed to detect hard exudates. Eadgahi et al. used morphology operations for pre-processing, erasing optic disc areas, and detecting exudates [11]. Qomariah et al. [10] proposed top hat morphology and automatic threshold to segment the exudate areas. Gupta et al. applied adaptive intensity thresholding selected based on first order statistics and local thresholding to detect exudates [12]. Liu et al. implemented 3 stages to perform exudate segmentation, namely removing anatomic structures, exudate location and exudate segmentation. At the anatomic stage, a matched filter was applied to remove blood vessels and saliency to remove the optic disc. At the location stage, the random forest method is applied to

determine exudate and non-exudate locations, then the last stage is the exudate segmentation stage [13].

Although many previous methods have been proposed to segment exudates, many of the proposed methods were developed based solely on features such as size, color, texture, without taking into account the low contrast of the image. Low image contrast and variable exudate sizes make it difficult to get accurate exudate segmentation results. In addition, image variation makes it difficult for the model to determine the exact features of the exudate. Therefore, an efficient segmentation method is needed so that it can select features automatically and accurately and can automatically detect the location, size, and shape of the exudate lesion.

In recent years, the state of the art convolutional neural network (CNN) [14, 15], is the method frequently used for medical imaging, due to its superior medical imaging segmentation capability [16, 17]. CNN uses low level and high level features to obtain information from the images automatically. Benzamin et al. [18] used 8 convolutional layers to do feature extraction and 3 fully connected layers to detect exudates. Xue et al. [19] suggested a network named deep membrane to detect exudates, microaneurysms and the optic disc. All suggested methods performed satisfactory lesion segmentation. However, the network proposed did not consider the features in the network. We concluded that continuing features from the previous layer can result in more accurate exudate segmentation. Segmentation results can help doctors to diagnose early. As many as 90 % of the patients can avoid blindness if diabetic retinopathy diagnosis is made early [20].

Image segmentation classifies each pixel in each image with ground truth. CNN can be used for image segmentation, especially for medical images using image patching as training to determine the class of each pixel [21]. Long et al. [22] proposed a Fully Convolutional Network (FCN), where FCN replaces a fully connected layer for classification with a convolutional layer. The FCN consists of FCN-32, FCN-16, and FCN-8. FCN-32 uses 32 pixel strides in the prediction layer, thus limiting the size of detail in the upsampled results. Using FCN-32 speeds up computing during training, but using FCN-32 for medical image segmentation has spatial resolution problems because the output is generated in one upsampling process [22]. The FCN-32 architecture is improved by combining the 2x upsampled prediction layer with the feature map from the fourth pooling layer to obtain a finer prediction layer and then 16x upsampling to get the FCN-16 architecture. Next, the finer prediction layer is upsampled 2x and combined with the feature map from the third pooling layer and followed by upsampling 8x to get the FCN-8 architecture. In addition to the development of the FCN itself, there is another architectural development to overcome the problem of spatial resolution in the prediction layer detail, namely the Autoencoder [23] and U network (U-Net) [24]. U-Net architecture is similar to Autoencoder, consisting of a contracting path and an expansive path. Contracting blocks perform feature extraction and reduce the size of an input image during training. Expansive blocks restore the image to its original size by using upsampling. The difference between Autoencoder and U-Net is that U-Net uses a skip connection to pass information from the contracting layer to the expansive layer, to help obtain spatial resolution information at the output layer [25].

In addition, there is an optic disc that was first removed because it has a similar intensity and color to the exudates [26]. Accurate optic disc detection is very important to obtain accurate exudate segmentation. Accurate optic disc detection can also help to diagnose eye diseases such as glaucoma and papilledema. Several methods for detecting optic disc have been carried out, because the optic disc is an important part of detecting diabetic retinopathy. Pathan et al. [27] proposed contour based method to detect the optic disc. Al-Bander et al. [28] proposed Multiscale sequential CNN to detect fovea and the optic disc. Karkuzhali et al. [29] proposed adaptive thresholding to detect the optic disc.

CNN method can be used for object detection [30], region based convolutional neural networks (R-CNN) is used to detect the optic disc because of the similarity of features with exudates, such as color and intensity values. The R-CNN detector creates an area on an object using the edge box algorithm [31]. The area of the object is cut from the image and resized. Next CNN performs a classification on the cut and resized area. [32, 33]. One-by-one image training requires long computation time, in order to solve this problem fast R-CNN [34] also uses edge boxes algorithm to obtain the proposal region. However, unlike R-CNN, fast R-CNN collects every CNN feature that matches the proposal region. Using edge box to detect the proposal region is still considered inefficient, so instead of using an external edge box there is a detection method that uses the region proposal network (RPN), namely faster R-CNN [35]. Faster R-CNN detector adds RPN to generate proposal regions directly on CNN so that it produces proposal regions faster and more in line with training data.

Based on the considerations, detection and removal of the optic disc will be carried out first. We

use a combination of Faster R-CNN and Alexnet. Based on our understanding, faster R-CNN produces a proposal region that is faster and in accordance with the optic disc object. Alexnet architecture is used to derive relevant features from the optic disc. To obtain area from exudate, we propose architectural modification of FCN-8 by adding modified identity mapping as skip connection or shortcut and dice loss function to overcome unbalance pixels between exudates and background. The advantage of using the FCN-8 architecture with shortcuts is to obtain effective features simultaneously. Shortcuts can continue features in the previous convolution process so as to reduce computation time during training and improve segmentation performance. The proposed shortcut consists of one convolutional layer and batch normalization. Our proposed architecture consisting of a combination of FCN-8 and shortcuts, with a dice loss minimization function, will be compared with state-of-the-art deep learning methods.

This paper is structured as follows. The introduction, objective, and related work are described in section 1. Section 2 provides material and a proposed methodology for optic disc removal and exudate segmentation. Section 3 describes the experiments and analyzes the segmentation results. Section 4 provides conclusion and future work.

2. Material and methodology

This study conducted exudate segmentation using the Indian diabetic retinopathy image dataset (IDRID) dataset. The detection and removal of the optic disc was carried out first because of the similarity of the intensity values between the optic disc and exudates. The stages in this research are detection and erasing of the optic disc using faster R-CNN and Alexnet, enhancement, cropping, patching, training using FCN-8 modification, and testing. The flow chart of this research is shown in Fig. 1.

Table 1. Alexnet architecture.

Name	Filter / Channel	Stride / Padding	Output
Input	- / 3	-	227 × 227 × 3
Conv	11 × 11 / 96	4 / 0	55 × 55 × 96
Conv	5 × 5 / 256	1 / 2	27 × 27 × 256
Conv	3 × 3 / 384	1 / 1	13 × 13 × 384
Conv	3 × 3 / 384	1 / 1	13 × 13 × 384
Conv	3 × 3 / 256	1 / 1	13 × 13 × 256
FC6	- / 4096	-	1 × 1 × 4096
FC7	- / 4096	-	1 × 1 × 4096
FC8	- / 2	-	1 × 1 × 2
Softmax	-	-	1 × 1 × 2

2.1 Dataset

IDRID dataset is obtained from eye clinic in Nanded, India. IDRID dataset consists of 3 sub parts namely disease grading, segmentation and localization [5] [36]. In the segmentation section, IDRID is divided into training and testing images, each with 54 images and 27 images with ground truth. Each ground truth has been validated by two retinal specialists, and created manually by graduate students using the ascis software. The image was taken using a Kowa VX-10a digital fundus camera with 50 FOV with the size of 4288 x 2848.

2.2 Optic disc removing

Detection and erasing of the optic disc is carried out due to the similarity of color and intensity values with exudates. Optic disc detection was carried out using faster R-CNN [37] and Alexnet [38] architecture.

2.2.1. Alexnet architecture

Proposed by Alex Krizhevsky et al [38]. Alexnet has 5 convolution layers for feature extraction, 3 pooling layers for downsampling and 3 fully connected layers. The Alexnet architecture on the last layer, which is fully connected, divides classes into two, namely optic disc and non-optic disc. Alexnet has 2 dropout layers for regularization, which can reduce overfitting. Alexnet won the 2012 imagenet large scale visual recognition (ILSVRC) challenge with an error rate of up to 15.3 % [39]. Detail about Alexnet is shown in Table 1.

2.2.2. Faster R-CNN

Ren et al. [37] suggested faster R-CNN with two stages, namely extraction and training region using a region proposal network (RPN) and classification based on the features obtained. Faster R-CNN network training uses RPN, which was originally obtained from fast R-CNN and has been updated.

RPN consists of two networks, the first using Alexnet and the second consisting of a convolutional layer for feature extraction, regression box convolution, and smooth box regression output layer. RPN training aims to minimize the total loss function. Total loss function as in Eq. (1).

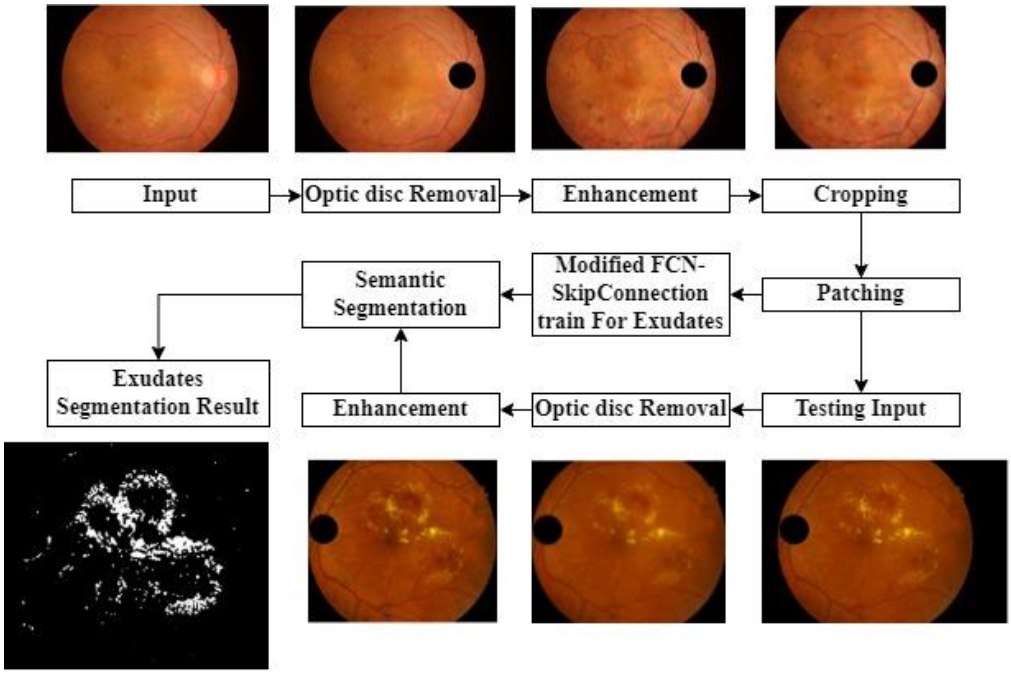


Figure. 1 Proposed method design

$$L = \frac{1}{n_c}\hat{L}_c + \lambda\frac{1}{n_{reg}}\hat{L}_{reg}, \quad (1)$$

where n_c is the size of mini-batch, n_{reg} is the number of anchor location, λ balancing parameter notation. $\hat{L}_c$ function is a loss in the classification stage, consisting of object class and non-object class, with the Eq. (2).

$$\hat{L}_c = \sum_k^M L_c\,(r_k, r_k^*), \quad (2)$$

where r_k is the prediction probability of k-th anchor in the mini batch, r_k^* the ground truth of the anchor. Loss regulation function $\hat{L}_{reg}$ is a function that calculates the value of the object's limiting box object, namely ground truth t_k with limiting box prediction t_k^*. The regularization function is written as in Eq. (3).

$$\hat{L}_{reg} = \sum_k^M r_k^*\,R(t_k - t_k^*) \quad (3)$$

2.3 Enhancement

Enhancement is used to increase the contrast of the image. Enhancement on training and testing images uses contrast limited adaptive histogram equalization (CLAHE) by applying it to each red, green, and blue (RGB) channels and then returning it to the RGB image.

2.4 Patching

Before entering the patching stage, the cropping stage is done to reduce the background area. Cropping used the size of 2848 x 3420, to get the overall retinal area.

Deep learning architecture is a learning method that requires a lot of training images. Therefore, the use of patching can maintain the quality of the image. This is essential because exudates lesions vary in size and spreads in the image. The use of patching can also increase the amount of training data, so that the architecture learning is more relevant. Patching used the size of 256 x 256, because that size is enough to detect exudates. The patching image is shown in Fig. 2.

2.5 Proposed architecture

2.5.1. FCN

In semantic segmentation, to get good results, it is very important to use low level details when maintaining high level semantic segmentation information. Training using deep learning is also very difficult, especially when the amount of training data is very limited and the computation time is quite long

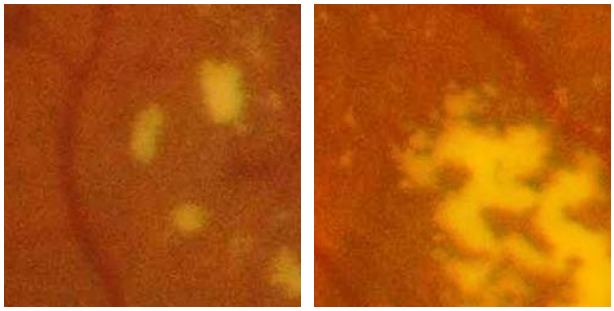


Figure. 1 Exudate patching

when the network architecture is too deep. One way is to apply augmentation data extensively as previously applied to the FCN architecture [22]. FCN restores the image to full size simultaneously during upsampling, this makes the FCN architecture effective and fast in the training process.

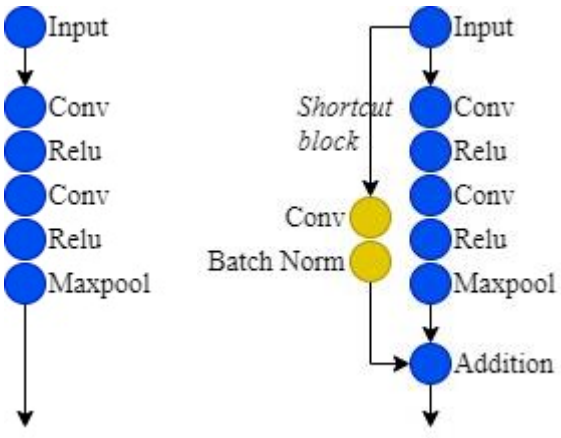
2.5.2. Shortcut/Identity mapping

Deeper network architecture will affect the performance of the network architecture. But getting deeper into a network architecture can lead to feature degradation. To overcome this problem, Kaiming et al., proposed an identity mapping or shortcut to overcome the problem of feature degradation, when features pass through the convolution layer. Shortcut layers can propagate directly from one unit to another. Shortcuts achieve fast error reduction and with lowest training during training [40]. Shortcut is also sufficient to address feature degradation issue [41]. Shortcuts are used in conjunction with unit residuals, with each unit being derived in Eqs. (4) and (5).

$$y_l = h(x_l) + \mathcal{F}(x_l W_l), \tag{4}$$

$$x_{l+1} = R(y_l). \tag{5}$$

Shortcut function and residual function are denoted as $h(x_l)$ and $\mathcal{F}(.)$. WIth x_l and x_{l+1} each as input and output in the layer l unit. Activation function is defined as $R(y_l)$. Plain FCN-8 consists of a stack of convolution and relu layers followed by maxpooling. The FCN-8 modification uses a shortcut consisting of a convolution layer stack and a normalization batch shown in Fig. 3.



(a) (b)

Figure. 2 Building blocks of modified FCN-8 and shortcut: (a) plain FCN-8 and (b) modified FCN-8 and shortcut

2.5.3. Modified FCN

FCN-8 has a simple architectural structure thus speeding up the computational process during training. Shortcuts combine low-level and high-level networks that function to forward information so that there is no feature degradation to achieve good performance. The proposal to combine FCN-8 and shortcuts can have two simultaneous advantages to improve architectural performance in semantic segmentation. Shortcuts are used in conjunction with FCN-8 units, with each unit being derived in Eqs. (6) and (7).

$$y_l = f_l(\{x_l, w_l\}) + h_l(x_l, w_l), \tag{6}$$

$$x_{l+1} = Mx(y_l), \tag{7}$$

The input and output of the encoder block are denoted by x_l and x_{l+1}, w_l and $f_l(.)$ are weights and functions in plain FCN-8. Shortcuts and max pooling functions are denoted by $h_l(.)$ and $Mx(y_l)$.

The modified architecture of FCN-8 and set shortcut uses 3 levels for exudates segmentation. The first level is the encoder level, which consists of plain FCN-8 and shortcuts for feature extraction. The second level is a bridge to connect the encoder and upsampling. The third level is an upsampling block that functions to restore the image to its original size. Upsampling has 3 blocks in which the first and second blocks use size 2 stride and the third block uses size 8 stride. Size 8 stride used in the third block serves to improve the level of smoothness and detail in the output. Convolution details in each block are shown in Table 2 and the proposed modified FCN-8 architecture shown in Fig. 4.

2.6 Dice loss

The unbalanced pixel comparison between exudate and background causes the use of cross entropy alone to be insufficient. Cross entropy is not able handle segmentation that has class imbalance, because the probability calculated by cross entropy is taken from the major class. In this study, the major class is the background and the minor class is the hard exudates. Instead of using the Cross entropy loss function that has been used in several previous proposed methods [42, 43], this study uses the dice loss function to deal with the imbalance between exudate and background. Dice loss is widely used in medical images [44] due to its ability to recognize minor object classes. The formula for the dice loss is explained in Eq. (8).

Table 2. Detail convolution layer in modified FCN-8

Name	Level	Convolutional Layer	Stride	Filter / Channel	Output
Input	-	-	-	-	$256 \times 256 \times 3$
Encoder	Plain FCN-8 first block	Conv1-1	1	3×3 / 64	$454 \times 454 \times 64$
		Conv1-2	1	3×3 / 64	$454 \times 454 \times 64$
		Conv2-1	1	3×3 / 128	227×227×128
		Conv2-2	1	3×3 / 128	227×227×128
		Conv3-1	1	3×3 / 256	113×113×256
		Conv3-2	1	3×3 / 256	113×113×256
		Conv3-3	1	3×3 / 256	113×113×256
	Shortcut first block	Conv1-1	1	3×3 / 64	454×454×64
		Conv1-2	2	3×3 / 128	227×227×128
		Conv1-3	2	3×3 / 256	113×113×256
	Plain FCN-8 second block	Conv4-1	1	3×3 / 512	56×56×512
		Conv4-2	1	3×3 / 512	56×56×512
		Conv4-3	1	3×3 / 512	56×56×512
		Convscore 1	1	1×1 / 2	56×56×2
	Shortcut second block	Conv2-1	1	3×3 / 512	56×56×512
	Plain FCN-8 third block	Conv5-1	1	3×3 / 512	28×28×512
		Conv5-2	1	3×3 / 512	28×28×512
		Conv5-3	1	3×3 / 512	28×28×512
		Convscore 2	1	1×1 / 2	28×28×2
	Shortcut third block	Conv3-1	1	3×3 / 512	28×28×512
Bridge	Bridge block	ConvFC6	1	7×7 / 4096	8×8×4096
		ConvFC7	1	1×1 / 4096	8×8×4096
		Convscore 3	1	1×1 / 2	8×8×2
Upsampling	Upsampling first block	De-Conv 1	2	4×4 / 2	18×18×2
	Upsampling first block	De-Conv 2	2	4×4 / 2	38×38×2
	Upsampling first block	De-Conv 3	8	16×16 / 2	312×312×2
Output	-	-	-	-	256×256×2

$$Loss = 1 - \frac{2TP}{2TP+FN+FP} = 1 - \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}, \quad (8)$$

where TP is the true positive value, FN is false negative, and FP shows the false positive value. N denotes the total number of pixels in the ground truth or predicted mask, p_i is the i-th pixel value from the binary segmentation mask, $p_i \in [0,1]$, and g_i is the i-th pixel value from the ground truth binary mask. Dice loss equation can be derived by the j-th pixel from the prediction. With the gradient resulted in Eq. (9).

$$\frac{\partial D}{\partial p_j} = 2\left[\frac{g_j(\sum_i^N p_i^2 + \sum_i^N g_i^2) - 2p_j(\sum_i^N p_i g_i)}{(\sum_i^N p_i^2 + \sum_i^N g_i^2)^2}\right]. \quad (9)$$

Using dice loss does not require weighting to balance the exudate and background pixels.

3 Experiments and analysis

To evaluate the performance of the proposed method, experiments using the IDRID dataset were conducted and compared with several methods.

The first step was to remove the area from the optic disc because of the similar intensity between the exudates and the background. Faster R-CNN and Alexnet architecture were used for detection of optic disc area. Faster R-CNN consists of pre-trained CNN and RPN.

The training images consist of 100 images taken from the disease grading data in the IDRID dataset. The results of the faster R-CNN model were then tested on training images and testing on sub-data segmentation. The combination of Faster R-CNN and Alexnet detected optic disc by generating bounding

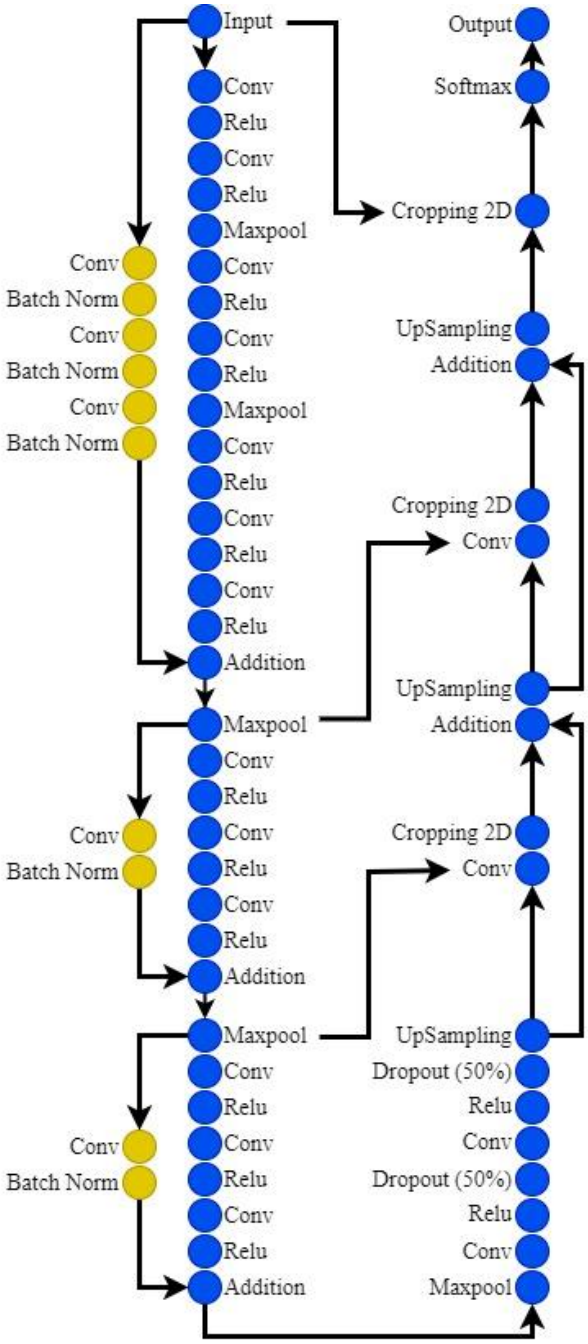
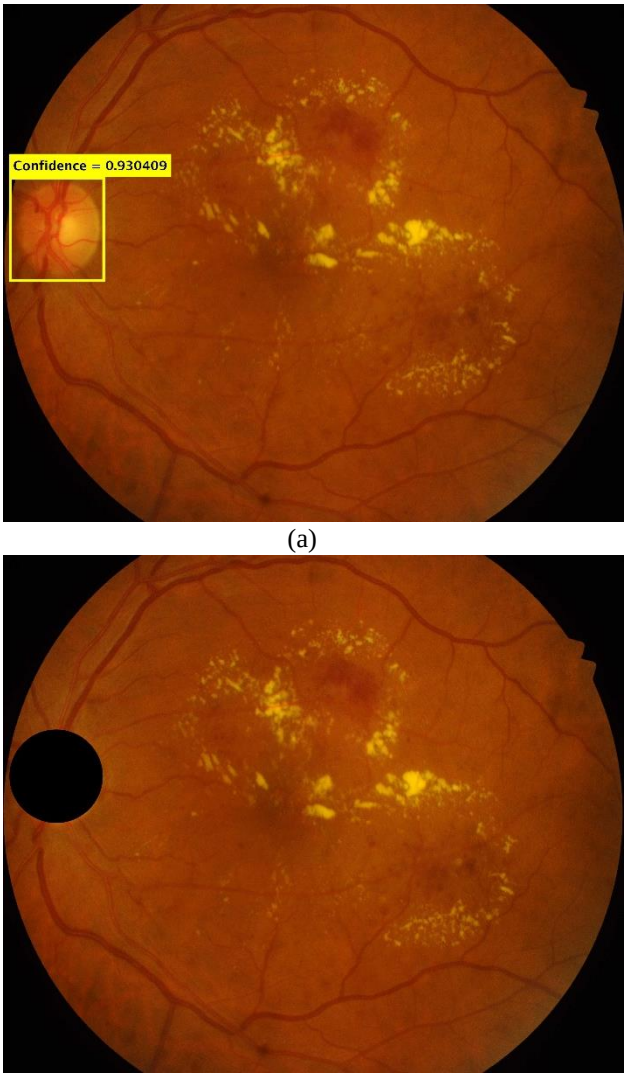


Figure. 3 Proposed modified FCN-8 and shortcut

box and score of confidence on the area of the optic disc. The optic disc was removed by using a circle area inside the box obtained from Faster RCNN. The result from detecting and removing optic disc is shown in Fig. 5.

The combination of faster RCNN and Alexnet can detect all optic discs in the data. The amount of training and testing data for the optic disc segmentation consists of 45 and 27 images, respectively, which will be used as training and testing data for exudate segmentation. CLAHE was used for image enhancement to increase contrast



(a)

(b)

Figure. 4 Visualization optic disc removing: (a) a result from Faster RCNN and (b) a result after removing OD

value for training and testing data. Next, the image was cropped only for the retinal area to reduce the background area and resized to 3584 × 3072. To increase the number of training images used during modified FCN-8 training, images were patched to a size of 256 × 256. The input training images were divided into 90% train set and 10 % validation set.

3.1 Parameter and setting

The modified FNC-8 architecture has several parameters. The architectural parameters use the stochastic gradient descent with momentum (SGDM) algorithm, with a learning rate of 0.05 and regularization 0.0001, mini batch size of 1, verbose frequency of 20, gradient threshold of 0.05 and maximum epoch of 50. The momentum size used is 0.9, the training data is divided into 90% train and 10 % as validation. All networks used for comparison with the proposed method used the same parameters.

For each testing image, after enhancing using CLAHE, cropping was done only on the retina area to reduce the background area.

3.2 Evaluation metrics

In this study, three measurements were used, namely sensitivity, specificity, and accuracy to measure the performance of the proposed architecture. Evaluation metrics are used to determine the success of detection of exudates lesion and background. Sensitivity (SE), specificity (SPE), and accuracy (AC) are measured based on a confusion matrix containing true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The calculations of the three metrics are given in Eqs. (10), (11), and (12).

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{10}$$

$$sensitivity = \frac{TP}{TP+FN} \tag{11}$$

$$specificity = \frac{TN}{TN+FP} \tag{12}$$

3.3 Comparison with baseline network

In this section, we evaluate the performance of the proposed architecture in comparison with the baseline network. Baseline architectures include autoencoder, U-Net, and FCN. The FCNs used are FCN types 32, 16 and 8. All methods are described below.

Autoencoder [23] introduces an encoder and decoder to segment binary images. There are 4 encoder blocks, 1 bridge block and 4 decoder blocks. The block encoder is for downsampling and feature extraction, whereas the bridge is to connect the block encoder and decoder. The decoder block is used to restore the image back to its original size. Autoencoder can also be applied to visual geometry group (VGG) by removing the fully connected layer for classification and replacing it with decoder blocks to restore the image to its original size and pixel classification layer is to classify object and background pixels.

U-Net [24] consists of encoder, bridge, and decoder blocks. The difference between autoencoder and U-Net is that there is a skip connection layer block that transmits information directly from the encoder block to the decoder, so that the decoder block is richer in information and helps to restore complete spatial information from the encoder block.

FCN [22] uses the convolution layer to perform downsampling and passes information to the upsampling layer. The upsampling layer on FCN type 32 immediately returns the image at the same size so that it can speed up the computational process during training, but it reduces the detail in the object detection process. Therefore, 32 pixel stride can limit the size and detail in the upsampling of the output. To overcome this problem, The FCN-32 architecture is enhanced by adding the 2x upsampled prediction layer with the output of fourth pooling layer to get a finer prediction layer and then 16x upsampling to obtain the FCN-16 architecture. Furthermore, the finer prediction layer is upsampled 2x and added with output of the third pooling layer, and followed by upsampling 8x to obtain the FCN-8 architecture.

Table 3 shows the comparison of the experiments using measurements of sensitivity, specificity and accuracy. The sensitivity result of the proposed method is 81.7 % higher than all the methods. The highest specificity value was obtained using the VGG autoencoder method of 99.99 % and the highest accuracy was obtained using the autoencoder method of 98.93 %, but our proposed method is still not too far away in terms of specificity and accuracy, reaching 98.87 % and 98.13 % respectively compared to the VGG autoencoder method. The sensitivity value, which reached 81.7 %, indicates that the proposed architecture can detect exudates lesion well.

Fig. 6 shows the results of exudate segmentation from Autoencoder, VGG autoencoder, U-Net, FCN-32, FCN-16, FCN-8, and the proposed method. It can be seen that the proposed method shows a better exudate detection rate with less noise compared to FCN-32, FCN-16, and FCN-8, as indicated by the blue box that has less area. The proposed method also detects the exudates area better than autoencoder, VGG autoencoder and U-Net, as indicated by the yellow box. The FCN-32, FCN-16, and FCN-8 methods were also able to detect the exudate area well, but the resulting segmentation results tend to be rounded, as indicated by the yellow box.

This study also presents a performance

Table 2. The evaluation of the baseline networks.

Method	**Performance (%)**		
	AC	**SE**	**SPE**
AutoEncoder	**98.93**	19.14	99.98
VGGAutoEncoder	98.90	19.88	**99.99**
U-Net	97.11	75.99	97.50
FCN-32	98.52	68.25	98.86
FCN-16	98.08	76.96	98.32
FCN-8	98.90	75.26	99.14
Proposed Method	98.18	**81.67**	98.37

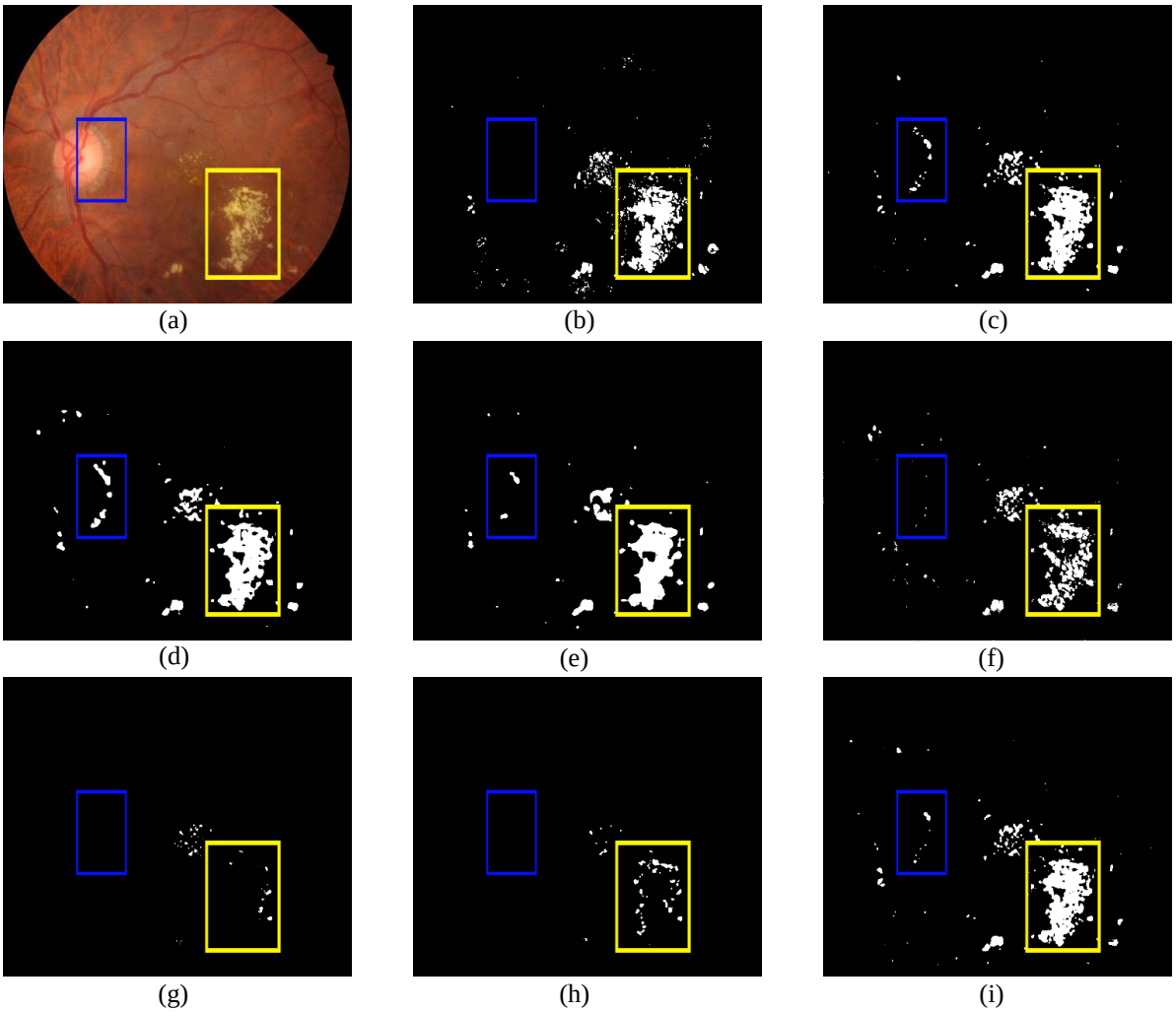


Figure. 5 Exudate segmentation results: (a) input image, (b) corresponding ground truth, (c) FCN-8, (d) FCN-16, (e) FCN-32, (f) U-Net, (g) AutoEncoder, (h) VGGAutoEncoder, and (i) proposed method

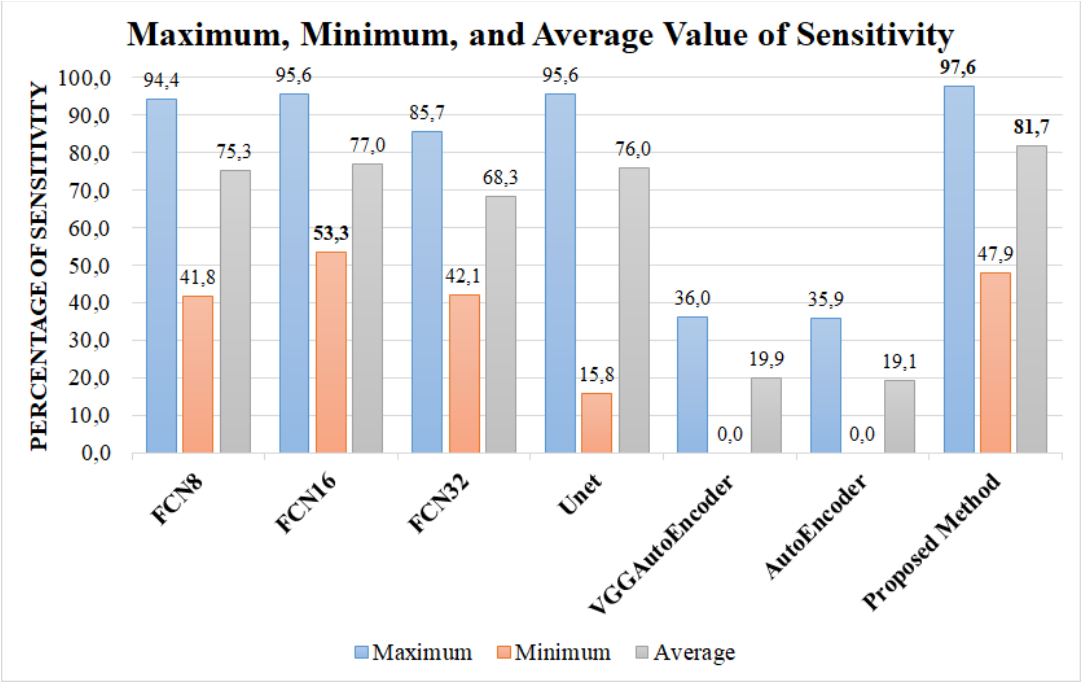


Figure. 6 Comparison of maximum, minimum, and average of sensitivity in different architecture on IDRID dataset

comparison chart based on descriptive statistics, such as the maximum, minimum, and average values of sensitivity values. It can be seen in Fig. 7 that the proposed method has a fairly high maximum value compared to other architectures. The minimum value of the proposed method is lower than the FCN-16 architecture, however the maximum and average value of the proposed method is still higher than all architectures. The proposed method is superior to all architectures. This means that the proposed method can maintain its performance to detect exudates, as can be seen from the high average sensitivity value of 81.7 %.

3.4 Comparison with existing method

The test results were compared with several state of the art methods to perform binary segmentation on exudates lesion. The state of the art methods are briefly described as follows.

Benzamin et al [18] used 8 convolutional layers to perform down sampling and feature extraction, 3 layers fully connected with the last layer to perform pixel hard exudate classification.

Xue et al [19] used a convolutional network called a dynamic membrane system with a hybrid structure. The hybrid structure consists of dynamic and communication channels between cells.

Table 4 shows the results of the comparison of the proposed method of segmentation of exudates lesion with two state of the art methods on the IDRID dataset. The results of the performance comparison are measured using 3 measurements, namely sensitivity, specificity, and accuracy. Compare to the Benzamin et al. [18] method and the Xue et al. [19] method, the sensitivity result of the proposed method reached the highest value of 81.7 %. While the highest value of specificity and accuracy was obtained by the method proposed by Xue et al. 99.6 % and 99.2 % respectively, our proposed method is not too far off for the measurement of specificity and accuracy compared Xue et al. method, which are 98.2 % and 98.4 %, respectively.

4 Conclusion

In this study, we proposed a modification of FCN-8 by adding a shortcut in the form of a modified identity mapping to continue features in semantic segmentation. We also proposed faster R-CNN using Alexnet network to detect optical disc area. The experimental results using the IDRID dataset show that our proposed method achieved a competitive performance compared to the state of the art methods, with the sensitivity value of 81.7 %.

For further research, we plan to develop a diabetic retinopathy classification system based on features extracted from the developed semantic segmentation model for blood vessels, microaneurysms, hemorrhages, and exudates, in order to obtain an integrated system of segmentation and classification of diabetic retinopathy lesions.

Table 3. The evaluation of the existing method.

Method	Performance (%)		
	AC	SE	SPE
Benzamin et al.[18]	96.6	41.4	98.3
Xue et al. [19]	**99.2**	77.9	**99.6**
Proposed method	98.2	**81.7**	98.4

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

The formulation of the proposed method, CNN architecture, experiments and preparation of the writing draft were contributed by Dinial Utami Nurul Qomariah. Conceptualization, problem formulation, research implementation processes, and article preparation were supervised by Handayani Tjandrasa. The formulation of the problem, the research implementation process and the article preparation process were supervised by Chastine Fatichah.

References

[1] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, “Global prevalence of diabetes: estimates for the year 2000 and projections for 2030”, *Diabetes care*, Vol. 27, No. 5, pp. 1047–1053, 2004.

[2] J. Tian, P. Suma, and T. C. Manjunath, “Use of Artificial Intelligence & Machine Learning with Deep Learning for Glaucoma Detection in Human Eyes & its Real Time Hardware Implementation”, *European Journal of Electrical Engineering and Computer Science*, Vol. 4, No. 2, 2020.

[3] N. Badariah, N. A. Mustafa, W. Zaki, W. M. D. W. Zaki, and A. Hussain, “A Review on the Diabetic Retinopathy Assessment based on Retinal Vascular Tortuosity”, In: *Proc. of International Colloquium on Signal Processing Its Applications (CSPA)*, pp. 127–130, 2015.

[4] A. N. Kollias and M. W. Ulbig, “Diabetic retinopathy: Early diagnosis and effective treatment”, *Deutsches Arzteblatt international*, Vol. 107, No. 5, pp. 75–83, 2010.

[5] P. Porwal, S. Pachade, R. Kamble, M. Kokare, G.

Deshmukh, V. Sahasrabuddhe, and F. Meriaudeau, "Indian Diabetic Retinopathy Image Dataset (IDRiD): A Database for Diabetic Retinopathy Screening Research", *Data*, Vol. 3, p. 25, 2018.

[6] J. J. Kanski, B. Bowling, K. K. Nischal, and A. Pearson, *Clinical ophthalmology : a systematic approach*, Elsevier/Saunders, Edinburgh; New York, 2011.

[7] D. U. N. Qomariah, H. Tjandrasa, and C. Fatichah, "Segmentation of Microaneurysms for Early Detection of Diabetic Retinopathy using MResUNet", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 3, pp. 359–373, 2021, doi: 10.22266/ijies2021.0630.30.

[8] D. U. N. Qomariah, H. Tjandrasa, and C. Fatichah, "Classification of Diabetic Retinopathy and Normal Retinal Images using CNN and SVM", In: *Proc. of International Conference on Information & Communication Technology and System*, pp. 1–6, 2019.

[9] H. Tjandrasa, R. E. Putra, A. Y. Wijaya, and I. Arieshanti, "Classification of non-proliferative diabetic retinopathy based on hard exudates using soft margin SVM", In: *Proc. of International Conference on Control System, Computing and Engineering*, pp. 376–380, 2013.

[10] D. U. N. Qomariah and H. Tjandrasa, "Exudate Detection in Retinal Fundus Images Using Combination of Mathematical Morphology and Renyi Entropy Thresholding", In: *Proc. of International Conference on Information & Communication Technology and System (ICTS)*, pp. 31–36, 2017.

[11] M. Eadgahi and H. Pourreza, "Localization of hard exudates in retinal fundus image by mathematical morphology operations", In: *Proc. of 2012 2nd International eConference on Computer and Knowledge Engineering, ICCKE 2012*, pp. 185–189, 2012.

[12] A. Gupta, A. Issac, N. Sengar, and M. K. Dutta, "An efficient automated method for exudates segmentation using image normalization and histogram analysis", In: *Proc. of 2016 Ninth International Conference on Contemporary Computing (IC3)*, pp. 1–5, 2016.

[13] Q. Liu, B. Zou, J. Chen, W. Ke, K. Yue, Z. Chen, and G. Zhao, "A location-to-segmentation strategy for automatic exudate segmentation in colour retinal fundus images", *Computerized Medical Imaging and Graphics*, Vol. 55, pp. 78–86, 2017.

[14] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges", *Journal of Digital Imaging*, Vol. 32, No. 4, pp. 582–596, 2019.

[15] M. Kim, J. Yun, Y. Cho, K. Shin, R. Jang, H. J. Bae, and N. Kim, "Deep Learning in Medical Imaging", *Neurospine*, Vol. 16, No. 4, pp. 657–668, Dec. 2019.

[16] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI", *Zeitschrift fur Medizinische Physik*, Vol. 29, No. 2, pp. 102–127, 2019.

[17] S. Vieira, W. H. L. Pinaya, and A. Mechelli, "Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications", *Neuroscience and Biobehavioral Reviews*, Vol. 74, pp. 58–75, 2017.

[18] A. Benzamin and C. Chakraborty, "Detection of Hard Exudates in Retinal Fundus Images Using Deep Learning", In: *Proc. of* 2018 *Joint 7th International Conf. on Informatics, Electronics and Vision and 2nd International Conf. on Imaging, Vision and Pattern Recognition*, pp. 465–469, 2018.

[19] J. Xue, S. Yan, J. Qu, F. Qi, C. Qiu, H. Zhang, M. Chen, T. Liu, D. Li, and X. Liu, "Deep membrane systems for multitask segmentation in diabetic retinopathy", *Knowledge-Based Systems*, Vol. 183, p. 104887, 2019.

[20] R. J. Tapp, J. E. Shaw, C. A. Harper, M. P. D. Courten, B. Balkau, D. J. M. Carty, H. R. Taylor, T. A. Welborn, and P. Z. Zimmet, "The prevalence of and factors associated with diabetic retinopathy in the Australian population", *Diabetes Care*, Vol. 26, No. 6, pp. 1731–1737, Jun. 2003.

[21] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. V. D. Laak, B. V. Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis", *Medical Image Analysis*, Vol. 42, pp. 60–88, 2017.

[22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation", In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 39, pp. 3431-3440, 2015.

[23] H. C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, "Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4D patient data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 8, pp. 1930–1943, 2013.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical

Image Segmentation", In: *Proc. of Medical Image Computing and Computer-Assisted Intervention - MICCAI*, pp. 234–241, 2015.

[25] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation", In: *Proc. of Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 179–187, 2016.

[26] H. Tjandrasa, A. Wijayanti, and N. Suciati, "Segmentation of the retinal optic nerve head using Hough transform and active contour models", *TELKOMNIKA (Telecommunication, Computing, Electronics and Control)*, Vol. 10, No. 3, pp. 531–536, 2012.

[27] S. Pathan, P. Kumar, R. Pai, and S. V. Bhandary, "Automated detection of optic disc contours in fundus images using decision tree classifier", *Biocybernetics and Biomedical Engineering*, Vol. 40, No. 1, pp. 52–64, 2020.

[28] B. Al-Bander, W. Al-Nuaimy, B. M. Williams, and Y. Zheng, "Multiscale sequential convolutional neural networks for simultaneous detection of fovea and optic disc", *Biomedical Signal Processing and Control*, Vol. 40, pp. 91–101, 2018.

[29] S. Karkuzhali and D. Manimegalai, "Robust intensity variation and inverse surface adaptive thresholding techniques for detection of optic disc and exudates in retinal fundus images", *Biocybernetics and Biomedical Engineering*, Vol. 39, No. 3, pp. 753–764, 2019.

[30] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", In: *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 580–587, 2014.

[31] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges", In: *Proc. of Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 391–405, 2014.

[32] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer New York Inc., 2008.

[33] N. Cristianini and J. S. Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods", *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Vol. 22, 2000.

[34] R. Girshick, "Fast R-CNN", In: *Proc. of IEEE International Conference on Computer Vision*, pp. 1440–1448, 2015.

[35] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", *Advances in Neural Information Processing Systems*, Vol. 28, 2015.

[36] P. Porwal, S. Pachade, M. Kokare, G. Deshmukh, J. Son, W. Bae, L. Liu, J. Wang, X. Liu, L. Gao, T. Wu, J. Xiao, F. Wang, B. Yin, Y. Wang, G. Danala, L. He, Y. Ho, Y. Chan, S. Jung, Z. Li, X. Sui, J. Wu, X. Li, T. Zhou, J. Toth, A. Baran, A. Kori, S. Saketh, M. Safwan, V. Alex, B. Harangi, B. Sheng, R. Fang, D. Sheet, A. Hajdu, Y. Zheng, A. Maria, S. Zhang, A. Campilho, B. Zheng, D. Shen, L. Giancardo, G. Quellec, and F. Mériaudeau, "IDRiD : Diabetic Retinopathy – Segmentation and Grading Challenge", *Medical Image Analysis*, Vol. 59, pp. 1–26, 2020.

[37] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 6, pp. 1137–1149, 2017.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", *Neural Information Processing Systems*, Vol. 60, No. 6. 2017.

[39] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg and L. F. Fei, "ImageNet Large Scale Visual Recognition Challenge", *International Journal of Computer Vision*, Vol. 115, No. 3, pp. 211–252, 2015.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks", In: *Proc. of Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9908 LNCS, pp. 630–645, 2016.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", In: *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[42] N. Sambyal, P. Saini, R. Syal, and V. Gupta, "Modified U-Net architecture for semantic segmentation of diabetic retinopathy images", *Biocybernetics and Biomedical Engineering*, Vol. 40, No. 3, pp. 1094–1109, 2020.

[43] J. Mo, L. Zhang, and Y. Feng, "Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks", *Neurocomputing*, Vol. 290, pp. 161–171, 2018.

[44] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net : Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation", In: *Proc. of International Conference on 3D Vision*, pp. 567–573, 2016.