Journal of Engineering Science and Technology
Vol. 14, No. 5 (2019) 2522 - 2539


# AUTOMATIC DETECTION OF BACKGROUND DIABETIC RETINOPATHY DISEASE USING HYBRID MULTILEVEL THRESHOLDING AND CONVOLUTIONAL NEURAL NETWORK

ERWIN[1,*], NAUFAL RACHMATULLAH[2], WULANDARI SAPUTRI[3]

[1,3]Department of Computer Engineering. Universitas Sriwijaya, Indonesia
[2]Department of Informatics. Universitas Sriwijaya, Indonesia
Jl.Raya Palembang-Prabumulih KM.32, 30662, Indralaya, Indonesia
*Corresponding Author: erwin@unsri.ac.id

## Abstract

Background Diabetic Retinopathy is a medical condition that causes retinal damage due to diabetes mellitus. This paper discusses a novel combination method, namely multilevel thresholding with a convolutional neural network. Multilevel thresholding resulting in disease characteristics by performing morphological vessel segmentation accurately. The segmentation process was carried out automatically by performing resizing on the grayscale image then histogram equalization. After that, the histogram equalization result was performed a Gaussian filter and Sobel edge detection to detect blood vessels. Lastly, the segmentation process was carried out using Otsu's multilevel thresholding, and then the classification phase using convolutional neural network was performed. The proposed method tested by using Stare datasets. In order to measure the performance of this technique, specifications, sensitivity, and accuracy, used. The results showed 100% specifications, 60% sensitivity and 92% accuracy using 2 level thresholding and at 3 levels of thresholding results 100% in Specifications, Sensitivity and Accuracy.

Keywords: Background diabetic retinopathy, Convolutional neural network, Multilevel thresholding, Otsu.

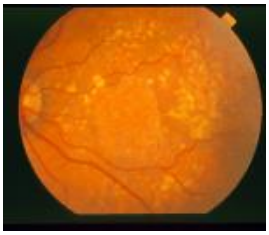
## 1. Introduction

Background Diabetic Retinopathy (BDR), known as diabetic non-proliferative (NPDR), is the initial stage of diabetic eye disease. Fluid leaking from a blood vessel can damage the retina so that the retina becomes swollen. This disease could attack humans in the age group of 20-64 years [1]. Besides age, someone who has high blood pressure and cholesterol can also worsen retinopathy. BDR is also a dangerous disease that can cause blindness in the human eye. According to the Global Prevalence of Diabetes, in 2000 it was estimated that there are 171 million people with diabetes worldwide and would continue to grow to 366 million by 2030. Meanwhile, in Indonesia, the World Health Organization estimates that diabetics will increase from 8.4 million in 2000 to 21.3 million in 2030, which make Indonesia become the fourth country with the highest number of diabetics in the world [2].

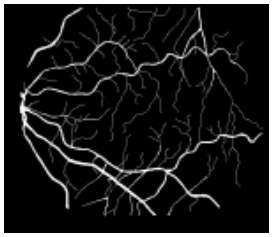
The structure of the retinal blood vessels is important to identify the characteristics of diabetic retinopathy. There are several methods that can segment the retinal blood vessels automatically. Hasan et al. [3] and Bao et al. [4] segment blood vessels by using a mathematical morphology approach. However, the processing time required during segmentation was less efficient. Vessel or vascular segmentation is an important process for obtaining the characteristics of retinal disease. The initial clinical sign of BDR in micro-aneurysms is the appearance of small red dots on the retinal layer, which caused by pericyte in the capillary wall. Bleeding that occurs has bright red, enlarged retinal veins and hardening of the exudate. In order to get the retinal vessel, a Multilevel Thresholding technique is used.

Kim et al. [5] used the convolutional neural network (CNN) method combined with the concept of the biological structure of Retina. However, the problem that arises in applying this method is that the proposed model is not sufficient to explain the mechanism of human perception. In future work, it is necessary to resolve higher levels of perception and cognition. Next, Chen et al. [6] study performed acute ischemic lesion segmentation on diffusion-weighted imaging using convolutional neural networks. The proposed approach still has several limitations. First, the semantic segmentation of objects in images on various scales remains a fundamental, unsolved challenge. Second, training and testing are not end-to-end, which reduces system efficiency and only considered the false positives value. Tan et al. [7] performed blood vessel segmentation, fovea and retinal vascularization using CNN. Unfortunately, this technique has two disadvantages when segmentation using CNN it is first to do segmentation by classifying the membership of each pixel, while most other problems only identify membership of an image or image segment. Second, membership of a pixel can be ambiguous, especially along the boundary. Figure 1(c) shows the difficult image of blood vessel segmentation using the Piece-wise Threshold method and Fig. 1(d) shows the failure of verve optical detection results with the fuzzy method.

We propose a novel method namely hybrid multilevel thresholding and CNN to perform automatic classification of BDR disease on the retina. Hybrid multilevel thresholding provides a detailed recognition of retinal blood vessels. This technique used to separate disease features in detail with backgrounds, which are not part of the characteristics of the disease. The segmentation process applied with different threshold values iteratively. The contribution of this paper is that this method can automatically segment and classify BDR disease with high specifications, sensitivity, and accuracy compared to the previous method.

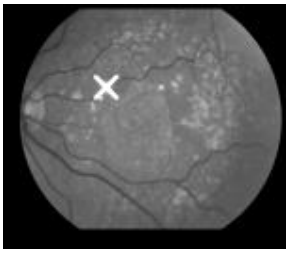


**(a) Original image.**



**(b) Result of ground truth segmentation by expert.**

**(c) Segmentation results using Piece-wise Threshold.**



**(d) Failure to detect results from locating the optic nerve using the fuzzy convergence.**

**Fig. 1. Examples of datasets.**

## 2. Related Works

Previous researches have been conducted in order to classify BDR disease by using various methods. Zhang and Chutatape [8] conducted BDR classification by applying two steps Improved Fuzzy C-Means to segment candidates for bright lesion areas. The hierarchical Support Vector Machine (SVM) method is applied to classify bright non-lesion areas, exudates and cotton spots. This paper test two different approach, bottom-up bright lesion and top-down dark lesion. The bottom-up approach showed 97% and 96% for sensitivity and specificity, while the other gives 88% for sensitivity and 84% for specificity. In order to perform bleeding detection, a top-down strategy is carried out. Zhang et al. [9] proposed a blood vessel detection using the haemorrhage lesion method to improve haemorrhage lesions detection in images of diabetic retinopathy. The results obtained reach 70% of accuracy. Sreng et al. [10] also performed BDR detection using boundary measurement and maximum entropy thresholding. The aim is to extract Optic Disk (OD) and exudates for early detection of diabetic retinopathy. The results obtained were 93% in OD and 89% in the detection of exudates. Verma et al. [11] used the Random Forests technique to detect blood vessels, identifying bleeding and classifying various stages of diabetic retinopathy into normal, moderate and non-proliferative diabetic retinopathy (NPDR). Accuracy for a normal class is 90% while accuracy for NPDR moderate and heavy classes is 87.5%. Haleem et al. [12] used the Scanning Laser Ophthalmoscopes (SLO) method, which can be used for early detection of retinal disease. The overall accuracy results are 92%. Yazid et al. [13] used edge sharpening to sharpen edges and to simplify the segmentation process for cotton spots and exudates through the reduction of blood vessels. Table 1 shows the summary for each work.

**Table 1. Multilevel thresholding segmentation with different learning levels and iterations.**

| Method | Result |
|---|---|
| **Fuzzy C-Means + SVM** | Bottom up approach:<br>Sensitivity = 97% and Specificity = 96%<br>Top-bottom approach:<br>Sensitivity = 88% and specificity = 84% |
| **The hemorrhage lesion method** | 70% of accuracy |
| **Boundary measurement and maximum entropy thresholding** | 93% for OD detection and 89% in the detection of exudates |
| **Random Forests** | 90% of accuracy for NPDR and 87.5% for heavy classes |
| **Scanning Laser Ophthalmoscopes (SLO) method** | 92% of accuracy |

## 3. Research Method

### 3.1. Research framework

In general, the first step of the proposed method is the preprocessing stage. It involved image resizing, grey scaling, and contrast enhancement using a Histogram Equalization (HE). The purpose of contrast enhancement is to increase retinal blood vessels contrast so it becomes easier to recognize. A Gaussian filter was applied to give a blur effect into the blood vessels images. After that, a Sobel edge detection technique was performed to see the edges of the blood vessels in the retina. The detected edge will go through a multilevel thresholding process to separate the foreground and background. In order to find the best threshold for blood vessel segmentation, we applied different threshold values. The segmentation result was used as an input for the classification step. The classification process used Convolutional Neural Network architecture to classify the BDR disease.

A Convolutional Neural Network (CNN) is one of deep learning architecture that is very well to implement in computer vision applications because of their ability to produce the hierarchical representations of an image using local operations [14]. CNN has a special structure for feature extraction. CNN input is a 2-D feature, and the classification results are expressed as probabilities. CNN is widely used in image processing, pattern recognition and classification [15]. This method provides a level of invariance for shifting, scale and rotation as local receptive fields allow access to neuron units or processing to basic features such as oriented edges or angles. There are 3 types of layers in the convolutional network, namely: Convolution, Pooling, and Fully connected.

The framework for this study can generally be seen in Fig. 2, which will include preprocessing, segmentation and detection.

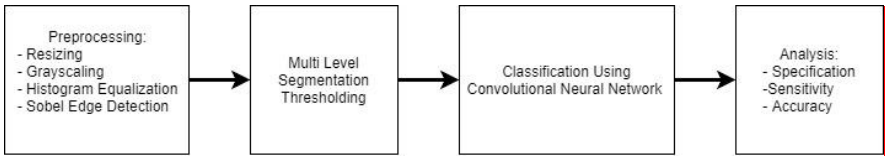



**Fig. 2. Framework of the research.**

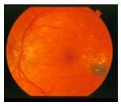
### 3.2. Dataset

In this study, Hoover et al. [16] provided the STARE dataset and Hoover and Goldbaum [17] as the founder of the Stare project from the University of California, which provides clinical data by Shiley Eye Center. The image size in this database is 700×605 with 8-bit colour channels. The dataset consists of 137 images from 5 classes. The five classes used in this research are Background Diabetic Retinopathy, Central Retinal Vein Occlusion, Coats, Normal and Proliferative Diabetic Retinopathy.

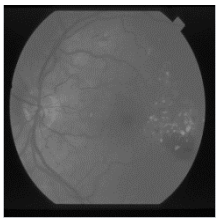
### 3.3. Preprocessing

In the retinal image used there are colour and lighting degradation. Colour and lighting differences can interfere with the process of segmentation and classification. Besides that, the size of the image used is also too large so that the computing process will be long. In order to overcome this problem, it is necessary to do the pre-processing stages. The pre-process stages carried out in this study were: Resize, grayscale, HE, Gaussian blurring, and detection of Sobel edges. The resizing image was used to make the image become smaller than the original image. This process needed to reduce the computational process. We use interpolation approach to resize the image. The histogram equalization method is used to improve image quality by improving the distribution of histograms in the input image. The HE method used in this study is to normalize the histogram so that the sum of histogram bins is 255. The next preprocessing step is to do a Gaussian blurring operation to remove noise from the image. The last stage in the preprocessing process is single edge detection that aims to detect edges in the image so that it facilitates the segmentation process. The results of each pre-processing step are shown in Fig. 3.

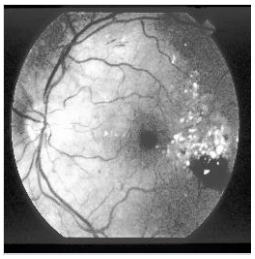
**(a) Input image 768×584.**



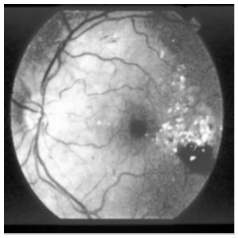
**(b) Resize 255×255.**



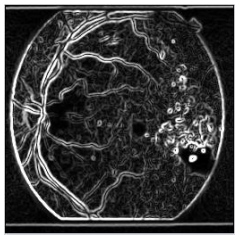
**(c) Grayscale.**



**(d) Histogram equalization.**



**(e) Gaussian filter.**



**(f) Sobel detection.**

**Fig. 3. Preprocessing.**

### 3.4. Segmentation multilevel thresholding

Segmentation is the process of separating original images into constituents or regional objects. Segmentation aims to separate each component from the image [18]. The quality of the segmentation process will be determined from the high level of accuracy obtained. The better the object that is recognized, the higher the accuracy obtained [19]. One of the most important things in the segmentation process is thresholding. Image thresholding is one of image segmentation technique that divides segments or partitions one image into several objects to detect various areas, which have the same colour or grey level intensity [20]. In multilevel thresholding, the main task is to choose the appropriate threshold value [21].

Thresholding value is divided into two types, local and global. The global threshold that is often used is Otsu thresholding. This thresholding technique was introduced in 1979 by Otsu [22]. In Otsu algorithm, the threshold value was obtained by maximizing the variance between two classes (foreground and background classes). The technique is used to determine the value of grayscale intensity, namely: $L$ {1, 2, ..., $L$). At that level, the probability of the occurrence of an image is given by Eq. (1) [23]:

$$p_a = \frac{Sa}{P_{tot}},\ p_{a \geq 0},\ \sum_{a=1}^{L} p_a = 1, \tag{1}$$

where: $i$ = gray level intensity ($L$) ($0 \leq i \leq L$-1), $a$ = component, the value specified from $i$ or Red Green Blue (RGB), $p_{tot}$ = the total pixel of the image, $S_a$ = histogram (the number of pixels corresponds to the intensity in $a$), $p_a$ = probability of distribution.

If a picture is bimodal, the histogram of the image is divided into two classes $C_a$ and $C_b$ and at certain levels, $C_a$ classes contains level 0 to S and $C_b$ class contains other levels with $S$ + 1 to $L$.

Probability ($W_A$ and $W_B$), distribution for two classes $C_a$ and $C_b$ explained with Eq. (2):

$$C_a = \frac{p_1}{W_A(S)}, \ldots \frac{p_S}{W_A(S)}, \text{ and}$$

$$C_b = \frac{p_{S+1}}{W_B(S)}, \ldots \frac{p_L}{W_B(S)}, \tag{2}$$

where $W_A(S) = \sum_a^S p_a$ and $W_B(S) = \sum_{a=s+1}^{L} p_a$

Value of $C_a$ and $C_b$ is class from $\vartheta_a$ and $\vartheta_b$ obtained from Eq. (3):

$$\vartheta_a = \sum_{a=1}^{S} \frac{ap_a}{W_A(S)}, \text{ and}$$

$$\vartheta_b = \sum_{a=s+1}^{S} \frac{ap_a}{W_B(S)}, \tag{3}$$

Suppose the average intensity value of all images is $\vartheta_s$ resulting from the calculation as follows Eq. (4) [23]:

$$W_A\vartheta_a + W_B\vartheta_b = \vartheta_{s,} \text{ and}$$

$$W_A + W_B = 1 \tag{4}$$

The total variants of the two levels are as follows Eq. (5):

$$\sigma_c^2 = \sigma_a + \sigma_b, \tag{5}$$

where $\sigma_a = W_A(\vartheta_a - \vartheta_s)$ and $\sigma_a = W_B(\vartheta_b - \vartheta_s)^2$

Vascular segmentation is carried out using the multilevel thresholding method [24]. In the Multilevel Thresholding technique, different threshold values applied. We use 5 thresholds to produce a binary level image. In the first stage, the $T_{MAX}$ threshold value is calculated from the retinal image histogram so that storing pixels in segmented images is greater than $T_{MAX}$. Then the thinning operator is used for the $I_{THIN}$ process, the image from $I_{SEG}$ that has been segmented will produce a one-pixel blood vessel. The edge image from $I_{EDGE}$ will highlight pixels to all identified retinal veins. Equation (6) used for each pixel p on $I_{THIN}$ [24]:

$$Edge(p) = \frac{1}{2}\sum_{i=1}^{s}|I_{THIN(P_{imod\,8})} - I_{THIN(P_{i-1})}| \quad (6)$$

where $P_0$ to $P_7$ will represent pixels, which will be referred to in the hourly long pixel sequence, which will define eight neighbours of $P$ and $I_{THIN}(P)$ is the pixel value. $I_{THIN}$ $(P)$ = 1 for pixel retinal blood vessels and zero for all other conditions. Edge $(P)$ = 1 and Edge $(P)$ = 2 for the edge of the vein. Then it will be reduced by a threshold value of 1 to calculate the segmented value for the next iteration.

Steps in the Otsu thresholding process can be seen in Fig 4.

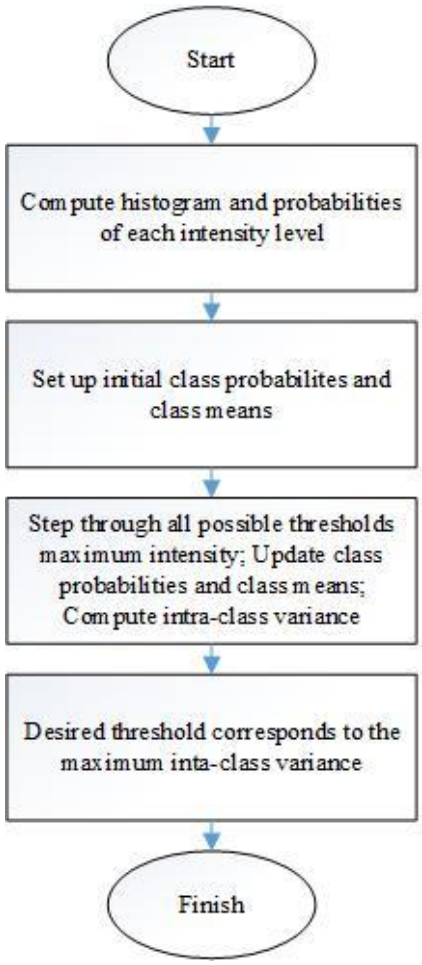



**Fig. 4. Process diagram multilevel thresholding Otsu.**

## 4. Convolutional Neural Network

The deep learning architecture model used in this study is the Convolutional Neural Network (CNN). CNN is a development of the Multilayer Neural Network model designed to be able to process two-dimensional data. CNN is a type of neural network that is biologically inspired by the human visual system that combines feature extraction and classification stages. CNN is a good alternative to image classification. The main advantage of this classification architecture is the implicit ability for automatic feature extraction.

Convolutional Neural Network takes 2-D images as an input (*I*), which will be convolved with a kernel (*K*). The kernel used to extract features from the images. The relationship between the input image (*I*) and kernel (*K*) as shown in Eq. (7):

$$(I * K)_{xy} = \sum_{i=1}^{h} \sum_{j=1}^{w} K_{ij} \cdot I_{x+i-1,y+j-1} \tag{7}$$

where K was represented as follows:

$$K = \begin{pmatrix} K_{-h_1,-h_2} & \cdots & K_{-h_1,h_2} \\ \vdots & K_{0,0} & \vdots \\ K_{h_1-h_2} & \cdots & K_{h_1,h_2} \end{pmatrix}$$

The result of convolution process was called feature maps, which shown in Fig. 5.

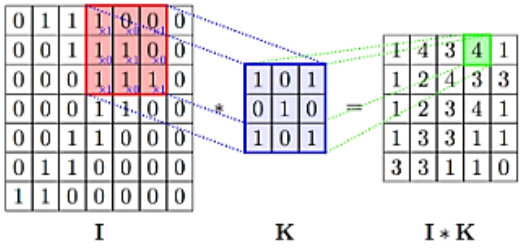



**Fig. 5. Illustration of convolution process.**

In the convolution process, there are several parameters that can be used such as, number of filters (f), stride (s), and padding (p). The number of filters correlates with the number of feature maps produced. The greater the number of filters used, the more features that can be extracted from the image. Stride is the number of pixel transfers when the kernel/filter searches for input pixels. When stride = 1, the filter moves 1 pixel per shift. Convolution operations using the kernel in the input image results in a convolution matrix, which has a smaller size than the input image. This process results in information loss from the input image. The use of padding can overcome this by adding a value of 0 around the input image. In order to calculate output size quickly Eq. (8) can be applied:

$$Output = \frac{(W-K+2P)}{S} + 1 \tag{8}$$

where W: Input size, ***K***: Filter size, ***P***: Padding, ***S***: Stride.

Like other neural networks, a convolution neural network also uses an activation function to make output non-linear. In the case of a Convolutional Neural Network, the convolution output will be passed through the activation function. The activation function used is rectified linear unit.

After the convolution layer, a process called pooling is applied. The pooling function was applied to reduce the number of parameters in feature maps so that the computational cost reduced. This process is useful for shortening the training process and controlling overfitting. Pooling operates was done by placing a window in a non-overlapping position on each feature map and storing one value per window. There are some pooling operations such as max pooling, min pooling and average pooling. The most common type of pooling is max pooling, which takes the maximum value in each window. This window size needs to be determined first. This reduces the size of the feature map while at the same time maintaining important information.

The last layer in the convolutional neural network is the fully connected layer. The fully connected layer is a type of layer with all of the node in each layer connected. Fully Connected layer like a standard neural network layer, where the layer has a weight matrix W and bias b. In fully connected layer was given $m_1^{(l-1)}$ feature map with the size of $m_2^{(l-1)}\ x\ m_3^{(l-1)}$ as an input unit $i^{th}$ the layer can be calculated with Eq. (9):

$$m_1^{(l)} = f z_i^{(l)} \quad with\ z_i^{(j)} = \sum_{j=1}^{m_1^{(l-1)}} \sum_{r=1}^{m_2^{(l-1)}} \sum_{s=1}^{m_3^{(l-1)}} w_{i,j,k}^{(l)} (Y_j^{(l-1)})_{r,s} \tag{9}$$

where $\boldsymbol{w_{i,j,k}^{(l)}}$ indicates the weight that connects the unit into the position (r,s).

In this study, Lecun et al. [25] develop the architectural model used, which is LeNet architecture, one of the first convolutional neural networks to help drive the field of Deep Learning. Lecun et al. [25] proposed the CNN architecture as shown in Fig. 6. Layer Cl and C3 are convolution layers, S2 and S4 are subsampling layers and Cs and S6 are fully connected layers and the number of neurons in the output layer is the number of classes.

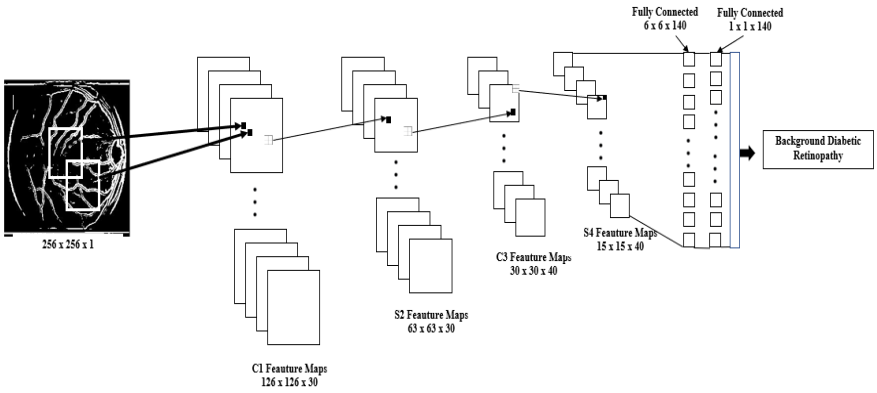



**Fig. 6. Architecture of CNN using LeNet-5.**

The LeNet-5 architecture consists of 5 convolution layers and 1 fully connected. At the C1 layer, the pixel entered is 256×256×1. Each feature map has 5×5 kernels to be included in 256×256 images by overlapping with one bias. In layer C1 it will produce a convolution layer with 126×126 pixels with 2 strides. In layer C1 there is 30 features map. One feature map contains 5×5 kernels. So that the trained parameters can be calculated (5×5+1 bias)×30 FM will produce 780 parameters

that can be trained. The total parameters in this layer are 126×126× (5×5+1 bias)×30 resulting in 12,383,280 connections.

In the subsampling layer (pooling layer) the feature map provided is also the same, namely 30-feature maps. In s1 the receptive field does not overlap. Due to there will be stride of 2 while for the pooling layer using 2×2 max pooling. So that it will produce 63×63 pixels. Each feature map in the subsampling layer has 1 bias and also 1 weight. Parameters that can be trained on S2 can be calculated as 30(1+1) = 60 parameters that can be trained and have 63×63×(2×2+1)×30 = 595,350 connections. This layer's input-output relationship is presented in the following Eq. (10).

$$\boldsymbol{v_j = w_j \sum_{i=1}^{4} \frac{y_i}{4} + b_j} \tag{10}$$

The next layer is the convolution layer (C3) with 40 kernels. Each neuron from FM will connect the subsampling S2 layer to several areas. Due to the incomplete connection scheme will keep the number of connections interconnected. More importantly, different feature maps are forced to extract different features because they will get different sets of inputs. So, every layer of S2 will go to the S3 layer with 10 connections.

The next layer is S4, which consists of 40 FM also with 5×5 kernel size. Layer s4 has 1 weight and 1 bias. So that overall S4 has 40(1+1) = 80 parameters that are trained. The given pooling is also the same as 2×2 with stride 2. Connections on S4 can be calculated with 5×5(2×2+1)×40, so that the connection has as many as 5,000 connections. This fully connected layer has 6×6×40 neurons with a total of 1,440 neurons. Parameters can be calculated with 1440×(40×25+1) so as to produce 1,497,600 parameters that can be trained with the full connection. Layer F6 is the last fully connected layer that has 84 neurons. Therefore, the number of parameters can be calculated 84×(1440+1) so that it produces 121,044 training and connection parameters.

Table 2 shows how the map features of the C3 layer are connected to the S2 layer feature map [26].

**Table 2. Feature maps connection between C3 and S2.**

| **Features maps in S2** | **Features maps in C3** | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** |
| **0** | x | - | - | - | x | x | x | - | - | x | x | x | x | - | x | x |
| **1** | x | x | - | - | - | x | x | x | - | - | x | x | x | x | - | x |
| **2** | x | x | x | - | - | - | x | x | x | - | - | x | - | x | x | x |
| **3** | - | x | x | x | - | - | x | x | x | x | - | - | x | - | | x |
| **4** | - | - | x | x | x | - | - | x | x | x | x | - | x | x | - | x |
| **5** | - | - | - | x | x | x | - | - | x | x | x | x | - | x | x | x |

The last layer is the SoftMax layer. This layer is used to identify the image followed by the fully connected layer. The training algorithm used for classification is Backpropagation. In every CNN convolution layer, a single filter will produce a feature map that is applied to different spatial positions. In the subsampling layer, there is weight and bias that can be trained using a local average coefficient, which causes independent parameters from the subsampling layer to be less than the convolution layer. In the CNN, training process actually requires less computing time.

However, when storing weights, it will require a relatively long processing time. Another thing to note is that CNN performs extraction of implicit features and invariant distortion to a certain extent, so that CNN adjusts to the process of identification, especially pattern recognition. In the problem of artificial neural networks, network architecture must be trained repeatedly [27].

## 5. Result and Discussion

The training process is carried out using 100, 1000 and 10,000 iterations with a resolution of 255×255 pixels. The stare dataset is divided into two sets: training sets and test sets and each data set contain thirty images, which will be divided into 25 training data and 5 testing data.

Table 3 shows the result of the multilevel thresholding segmentation process using 2, 3, 4 and 5 thresholds. In order to measure the segmentation performance, we applied the Mean Squared Error (MSE) calculation. MSE calculated segmentation error from the proposed method with manually segmentation image. Table 4 shows MSE from the multilevel thresholding.

**Table 3. Multilevel thresholding segmentation with different learning levels and iterations.**

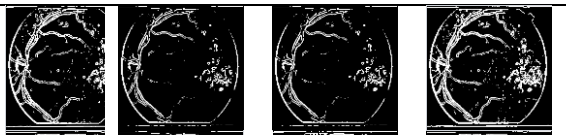
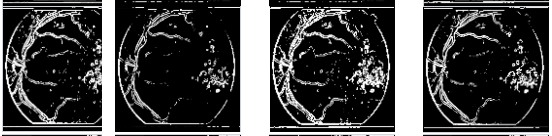
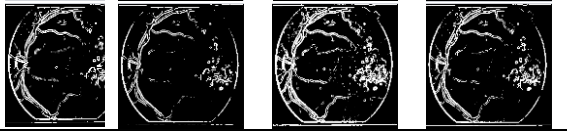
| Learning rate | Iteration | Level of Th | | | |
|---|---|---|---|---|---|
| | | Th 2 | Th 3 | Th 4 | Th 5 |
| **0.001** | 100 |  | | | |
| **0.01** | 1000 |  | | | |
| **0.1** | 10000 |  | | | |

**Table 4. The value of mean-squared error at each threshold.**

| Learning rate | Iteration | Mean squared error | | | |
|---|---|---|---|---|---|
| | | Level Th | | | |
| | | Th 2 | Th 3 | Th 4 | Th 5 |
| **0.001** | 100 | 98.2378 | 98.9995 | 99.3763 | 98.8698 |
| **0.01** | 1000 | 99.0980 | 99.0241 | 99.2372 | 98.5403 |
| **0.1** | 1000 | 98.8671 | 98.1812 | 98.4236 | 98.2440 |

The CNN model has been implemented using the learning level of 0.001, 0.01, and 0.1 with the number 100, 1000, and 10000. The minimum error used is 0.1 presented in Table 5.

**Table 5. CNN detection results with different learning rate and iterations.**

| Learning rate | Iteration | Detection results | | | |
|---|---|---|---|---|---|
| | | **Th 2** | **Th 3** | **Th 4** | **Th 5** |
| **0.001** | 100 | BDR | BDR | BDR | BDR |
| **0.01** | 1000 | BDR | BDR | BDR | BDR |
| **0.1** | 1000 | BDR | BDR | BDR | BDR |

We performed a testing scenario using 2, 3, 4 and 5 thresholds. The quantitative evaluation of the segmentation algorithm is done by comparing the detection results with the appropriate ground truth images. Comparisons produce statistical measures that can be summarized using the confusion matrix as in Table 6. True positives mean that images are correctly classified as a method and manual detection is used as ground truth. False positives are data that are properly separated. The true negative is the actual data is marked as BDR and false negatives are images that are predicted to be wrong.

**Table 6. Performance analysis using confusion matrix.**

| | | **Actual** | |
|---|---|---|---|
| | | **Positive** | **Negative** |
| **Predicted** | Positive | True Positive ($T_p$) | False Positive ($F_p$) |
| | Negative | True Negative ($T_N$) | False Negative ($F_N$) |

From Table 6, we calculate specifications, sensitivity, and accuracy. Specifications provide a percentage of non-BDR, while sensitivity will give the correct percentage of BDR detection, and accuracy will give a percentage of matching methods used. Equations (11) to (13) used is as follows.

$$\boldsymbol{Specification} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FP}} \tag{11}$$

$$\boldsymbol{Sensitivity} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \tag{12}$$

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+TN+FP+FN}} \tag{13}$$

where *TP* is true positive, *TN* is true negative, *FP* is false positive and *FN* is a false negative. Table 7 is a table of results from BDR automatic detection performance.

The results of BDR classification accuracy obtained from experiments using the stare dataset shown in Fig. 7.

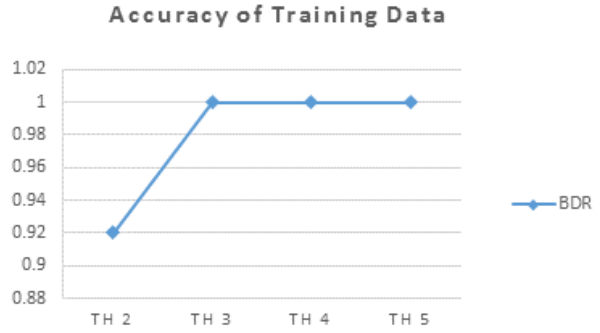



**Fig. 7. Classification performance using CNN.**

## 6. Analysis and Evaluation

In Table 4, is the result of applying multilevel thresholding and classification using CNN. In Th 2, the non-BDR specification is 100% because there is no error when making a prediction. on sensitivity when detecting there are two BDR imagery that cannot be detected so that it produces 60%. The accuracy produced in Th 2 is 92%. In Th 3, i.e., increasing binary resulted in specifications, Sensitivity and 100% Accuracy. At Th 4 and Th 5 produce 100% too. It can be analyzed that, the higher threshold is used, the accuracy will also increase.

Table 8 shows a comparison between the proposed method and other methods. Sharath Kumar et al. [28] proposed the Two-Field Mydriatic Fundus photography, in which, used bi-cubic interpolation and waterfall model classification method. The result showed that the proposed method still needed to improve the detection rate of BDR cases, which will increase the sensitivity of the system. Similarly, the specification of the system can be improved by devising an algorithm for quantifying the image quality and information that can be used for dynamic selection of a red lesion detection algorithm.

**Table 8. Comparison of proposed method.**

| Methods | Specification (%) | Sensitivity (%) | Accuracy (%) |
|---|---|---|---|
| **Sharath Kumar et al. [28] Two-Field Mydriatic Fundus Photography** | 80% | 50% | 60% |
| Kaur and Sinha **[29] morphological filter** | 86.39% | 91.28% | NA |
| **Kanth et al. [30] artificial neural network** | NA | NA | 94.11% |
| **Gartner et al. [31] back propagation mathematical** | 99% | 69% | 99% |
| Sopharak **et al. [32] morphology methods** | 71.35% | 98.95% | 98% |
| **Proposed Method using Th 2** | 100% | 60% | 92% |
| **Proposed Method using Th 3** | 100% | 100% | 100% |
| **Proposed Method using Th 4** | 100% | 100% | 100% |
| **Proposed Method using Th 5** | 100% | 100% | 100% |

Kaur and Sinha [29] applied the Morphological filter method. Blood vessel segmentation using Morphological Filters in colour retinal images. It was found that the appearance of vessels is highly sensitive in the grayscale image containing only the green wavelength. For the segmentation of vessels, therefore, only the green channel of the RGB colour image was used. Compared to the Matched Filter and Gabor Filter to detect line-like features, Morphological Filter delivered a better result. Values of all filter parameters were selected based on vessel properties. Increasing the number of filter banks did not result in significant results improvement but increased the time processing in convolution operation. The resulting enhanced vessels were then subject to vessel pixel classification thresholding. It was found that the number of miss classified pixels was lower compared to matched filter methods using the same database.

While Kanth et al. [30] applied the Artificial Neural Network method. These results strengthen the idea that Multilayer Perceptron can be used efficiently as a classifier for detecting eye-related diseases in fundus images. Even with such results and progress, the proposed network won't give desired results in case the exudates areas at a particular section in fundus exceed that of the optical disc. With these limitations and results, work should be carried on to derive several more features and develop more efficient systems.

Gardner et al. [31] applied backpropagation. The resulting enhanced vessels were then subject to vessel pixel classification thresholding. It was found that the number of miss classified pixels was lower compared to matched filter methods using the same database. Because of their similar pixel values to vessels, retinal haemorrhages were most difficult to recognize. Compared to the ophthalmologist, the neural network was able to detect patients with retinopathy with sensitivity and specificity values comparable to other screening methods such as optometrists, polaroids from non-mydriatic cameras and diabetic physicians for whom sensitivity values for retinopathy detection varied from 61% to 97% and specificities from 78% to 96%. The results with the neural network were at least comparable to other systems providing 88.4 percent sensitivity and 83.5 percent specificity for diabetic retinopathy detection. By optimizing the system's sensitivity to 99 percent, by adjusting the threshold of the artificial neural network, the specificity was reduced to 69 percent and none of the patients requiring referral was missed. Although network training took up to 5 days of computing in some circumstances, once trained the network has the ability to analyze an image in a few seconds; therefore, it would not be difficult to perform a rapid evaluation of the fund of large numbers of individuals.

The last one, mathematical morphology methods applied by. There are also some incorrect exudate detections that are caused by exudate-like artefacts, noise artefacts in the image acquisition process, exudates that are close to blood vessels, or because the exudates appear very weak. Strong and high contrast choroidal blood vessels that appear and lie in the background of the retina are detected incorrectly as exudates. Missing faint exudates may not have affected the sensitivity much as even human experts are uncertain about some ambiguous regions. However, the performance of the algorithm can be improved if this set of low-contrast exudates can be detected. We may need to add more specific features to the system in order to detect these kinds of regions in the future.

Future work will address the issue of improving sensitivity by improving the results of other tasks, such as detecting optic disks and blood vessels and also attempting to locate faint and small exudates. In future, there may be a problem in separating pathologies from small vessels in work to expand the detection system to recognize microaneurysms and haemorrhages. Before detecting pathologies, we can detect vessels and subtract them from the image. If small vessels are missed during this step and confused with microaneurysms or haemorrhages, it may be possible to combine more than one detection technique to make a final decision on whether the detected area is either a vessel or a microaneurysm or a haemorrhage. At the moment, our algorithm cannot detect differences between hard and soft exudates. However, hard and soft exudates can be distinguished by their colour and the sharpness of their border so that this can be detected by tuning the edge filter and selection of features. These features are intended to be used in future detection algorithms.

## 7. Conclusion

In this study, we propose a new method namely Hybrid Multilevel Thresholding and Convolutional Neural Network. The indicator in the study was to evaluate the performance of Hybrid Multilevel Thresholding and Convolutional Neural Network is to use a confusion matrix consisting of Specifications, Sensitivity, and Accuracy. The experimental results that have been implemented by applying 2, 3, 4 and 5 years produce excellent accuracy. At the time of application with Th 2, the accuracy obtained was 92%. Then we increase the number with an accuracy of 100%. Comparison of the proposed method with other methods produces the highest Specifications, Sensitivity, and Accuracy values from the previous method. The shape of the pattern is very influential during the automatic detection process. The results of Specifications, Sensitivity and Accuracy on Th 2 are 100%, 60% and 92%. At 3, 4, and 5 the proposed method results in 100% in Specifications, Sensitivity and Accuracy.

## Acknowledgement

This paper is partly supported by Rector of University of Sriwijaya through Hibah Penelitian Skema Unggulan Kompetitif tahun 2019. This article's publication is supported by United States Agency for International Development (USAID) through the Sustainable Higher Education Research Alliance (SHERA) Program for Universitas Indonesia's Scientific Modeling, Application, Research and Training for City-centered Innovation and Technology (SMART CITY) Project, Grant#AID-497-A-1600004, Sub Grant#IIE-00000078-UI-1.

## Nomenclatures

| | |
|---|---|
| $a$ | Component RGB |
| $C_a$ | Class contains level 0 to S |
| $C_b$ | Class contains other level with S+1 to L |
| $Edge(p)$ | Edge of Image |
| $I$ | Image |
| $I_{THIN}$ | Operator Thinning |
| $i$ | Grey level |
| $K$ | Kernel |
| $L$ | Level Intensity |
| $m$ | Feature map |
| $P$ | Padding |
| $P_{tot}$ | Total pixel of image |
| $p_a$ | Probability of distribution |
| $S$ | Stride |
| $S_a$ | Histogram of component RGB |
| $T_{MAX}$ | Threshold Maximum |
| $W$ | Input size |
| $W_a$ | Probability distribution for $C_a$ |
| $W_b$ | Probability distribution for $C_b$ |
| $w$ | Weight matrix |

***Greek Symbols***

| | |
|---|---|
| $\vartheta$ | Average intensity value |

| | |
|---|---|
| σ | Variants |

**Abbreviations**

| | |
|---|---|
| BDR | Background Diabetic Retinopathy |
| CNN | Convolutional Neural Network |
| FN | False Negative |
| FP | False Positive |
| HE | Histogram Equalization |
| MSE | Mean Squared Error |
| NPDR | Non-Proliferative Diabetic Retinopathy |
| OD | Optic Disk |
| RGB | Red Green Blue |
| SLO | Scanning Laser Ophthalmoscopes |
| SVM | Support Vector Machine |
| TH | Threshold |
| TN | True Negative |
| TP | True Positive |

## References

1. Ghosh, R.; Ghosh, K.; and Maitra, S. (2017). Automatic detection and classification of diabetic retinopathy stages using CNN. *Proceedings of the 4th International Conference. Signal Processing and Integrated Networks (SPIN)*. Noida, India, 500-554.
2. Aras, R.A.; Lestari, T.; Nugroho, H.A.; and Ardiyanto, I. (2016). Segmentation of retinal blood vessels for detection of diabetic retinopathy : A review. *Communications in Science and Technology*, 1, 33-41.
3. Hasan, G.; El-Bendary, N.; Hassanien, A.E.; Fahmy, A., M.; Shoeb, A.M.; and Snasel, V. (2015). Retinal blood vessel segmentation approach based on mathematical morphology. *Proceedings of the International Conference on Communication, Management and Information Technology (ICCMIT).* Prague, Czech Republic, 612-622.
4. Bao, X.-R.; Ge, X.; She, L.-H.; and Zhang, S. (2015). Segmentation of retinal blood vessel based on cake filter. *BioMed Research International*, Article ID 137024, 1-11.
5. Kim, J.; Sangjun, O.; Kim, Y.; and Lee, M. (2016). Convolutional neural network with biologically inspired retinal structure. *Proceedings of 7th Annual International Conference on Biologically Inspired Cognitive Architectures (BICA).* New York, United States of America, 145-154.
6. Chen, L.; Bentley, P.; and Rueckert, D. (2017). Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. *NeuroImage: Clinical*, 15, 633-643.
7. Tan, J.H.; Achrarya, U.R.; Bhandary, S.V.; Chua, K.C.; and Sivaprasad, S. (2017). Segmentation of optic disc, fovea and retinal vasculature using a single convolutional neural network. *Journal of Computational Science*, 20, 70-79.
8. Zhang, X.; and Chutatape, O. (2005). Top-down and bottom-up strategies in lesion detection of background diabetic retinopathy. *Proceeding of the IEEE*

*Computer Society Conference Computer Vision and Pattern Recognition (CVPR).* San Diego, California, United States of America, 422-428.

9. Zhang, D.; Li, X.; Shang, X.; Yi, Y.; and Wang, Y. (2011). Robust hemorrhage detection in diabetic retinopathy image. *Proceedings of the 1st Asian Conference on Pattern Recognition*. Beijing, China, 209-213.
10. Sreng, S.; Takada, J.-I.; Maneerat, N.; Isarakom, D.; Varakulsiripunth, R.; Pasaya, B.; and Panjaphongse, M.D.R. (2013). Feature extraction from retinal fundus image for early detection of diabetic retinopathy. *Proceedings of IEEE Region 10 Humanitarian Technology Conference*. Sendai, Japan, 63-66.
11. Verma, K.; Deep, P.; and Ramakrishnan, A.G. (2011). Detection and classification of diabetic retinopathy using retinal images. *Proceedings of the Annual IEEE India Conference.* Hyderabad, India, 1-6.
12. Haleem, M.S.; Han, L.; Hemert, J.V.; Li, B.; and Fleming, A. (2015). Retinal area detector from scanning laser ophthalmoscope (SLO) images for diagnosing retinal diseases. *IEEE Journal Biomedical and Health Informatics*, 19(4), 1472-1482.
13. Yazid, H.; Arof, H.; and Mokhtar, N. (2010). Edge sharpening for diabetic retinopathy detection. *Proceedings of the IEEE Conference on Cybernetics and Intelligent Systems (CIS)*. Singapore, 41-44.
14. Cengil, E.; Cinar, A.; and Guler, Z. (2017). A GPU-based convolutional neural network approach for image classification. *Proceedings of the International Artificial Intelligence and Data Processing Symposium*. Malatya, Turkey, 1-6.
15. Zhang, M.; Diao, M.; and Guo, L. (2017). Convolutional neural networks for automatic cognitive radio waveform recognition. *IEEE Access*, 5, 11074-11082.
16. Hoover, A.D.; Kouznetsova, V.; and Goldbaum, M. (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transaction on Medical Imaging*, 19(3), 203-210.
17. Hoover, A.; and Goldbaum, M. (2003). Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *IEEE Transaction on Medical Imaging*, 22(8), 951-958.
18. Erwin; S.; and Saputri, W. (2018). New hybrid multilevel thresholding and improved harmony search algorithm for image segmentation. *International Journal of Electrical and Computer Engineering*, 8(6), 4593-4602.
19. Dass, R.; Priyanka; and Devi, S. (2012). Image segmentation techniques. *International Journal of Electronics and Communication Technology*, 3(1), 66-70.
20. Mishra, D.; Bose, I.; De, U.C.: and Pradhan, B. (2014). A multilevel image thresholding using particle swarm optimization. *International Journal of Engineering and Technology (IJET)*, 6(2), 1204-1211.
21. Guo, Z.; Yue, X.; Liu, G.; Wang, S.; and Li, K. (2015). An improved harmony search algorithm for multilevel image segmentation. *Innovative Computing, Information and Control Express Letter*, 9(9), 2531-2536.
22. Vala, H.J.; and Baxi, A. (2013). A review on otsu image segmentation algorithm. *International Journal of Advanced Research in Computer Engineering and Technology (IJARCET)*, 2(2), 387-389.

23. Singh, R.; Agarwal, P.; Kashyap, M.; and Bhattacharya, M. (2016). Kapur's and otsu's based optimal multilevel image thresholding using social spider and firefly algorithm. *Proceedings of the International Conference on Communication and Signal Processing (ICCSP).* Melmaruvathur, India, 2220-2224.
24. Nayak, C.; and Kaur, L. (2015). Retinal blood vessel segmentation for diabetic retinopathy using multilayered thresholding. *International Journal of Science and Research (IJSR),* 4(6), 1520-1526.
25. Lecun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of IEEE*, 86(11), 2278-2324.
26. Ahranjany, S.S.; Razzazi, F.; and Ghassemian, M.H. (2010). A very high accuracy handwritten character recognition system for Farsi/Arabic digits using convolutional neural networks. *Proceedings of the IEEE 5th International Conference Bio-Inspired Computing: Theories and Applications (BIC-TA).* Changsha, China, 1585-1592.
27. Triwijoyo, B.K.; Heryadi, Y.; Ahmad, A.S.; Sabarguna, B.S.; Budiharto, W.; and Abdurachman, E. (2018). Retina disease classification based on colour fundus images using convolutional neural networks. *Proceedings of the International Conference on Innovative and Creative Information Technology*. Salatiga, Indonesia, 1-4.
28. Sharath Kumar, P.N.; Deepak, R.U.; Sathar, A.; Sahasranamam, V.; and Rajesh Kumar, R. (2016). Automated detection system for diabetic retinopathy using two field fundus photography. *Proceedings of 6th International Conference on Advances in Computing and Communications.* Cochin, India, 486-494.
29. Kaur, J.; and Sinha, H.P. (2012). Automated detection of diabetic retinopathy using fundus image analysis. *International Journal of Computer Science and Information Technologies*, 3(4), 4794-4799.
30. Kanth, S.; Jaiswal, A.; and Kakkar, M. (2013). Identification of different stages of diabetic retinopathy using artificial neural network (IC3). *Proceedings of the Sixth International Conference Contemporary Computing*. Noida, India, 479-484.
31. Gardner, G.G.; Keating, D.; Williamson, T.H.; and Elliott, A.T. (1996). Automatic detection of diabetic retinopathy using an artificial neural network: A screening tool. *British Journal Ophthalmology*, 80(11), 940-944.
32. Sopharak, A.; Uyyanonvara, B.; Barman, S.; and Williamson, T.H. (2008). Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. *Computerized Medical Imaging and Graphics*, 32(8), 720-727.