

WSSH_DR: Weakly Supervised Sensitive Heatmap Framework to Classify and Localize Diabetic Retinopathy Lesions

Mohammed Almukhtar

University of Baghdad

Ameer Morad

University of Baghdad

Mustafa Albadri

University of Baghdad

MD Islam (✉ mdsamiul@ualberta.ca)

University of Alberta

Research Article

Keywords: Diabetic Retinopathy, DR Detection, Medical Image Classification, WSSH, DFCNN

Posted Date: April 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-433490/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

WSSH_DR: Weakly Supervised Sensitive Heatmap framework to classify and localize Diabetic Retinopathy lesions

Mohammed Al-Mukhtar¹, Ameer Hussein Morad², Mustafa Albadri³, MD Samiul Islam⁴

¹IEEE Member, University of Baghdad, Computer Center, Iraq mohammed.abdul@cc.uobaghdad.edu.iq

²University of Baghdad, Al Khwarizmi College of Engineering, Iraq ameer@kecbu.uobaghdad.edu.iq

³University of Baghdad, Computer Center, Iraq, m.albadri@uob.edu.iq

⁴University of Alberta, Computing Science, Ph.D. Student, Canada, mdsamiul@ualberta.ca

Article Info

Article history:

Received

Revised

Accepted

Keywords:

Diabetic Retinopathy

DR Detection

Medical Image Classification

WSSH

DFCNN

ABSTRACT

Vision loss happens due to diabetic retinopathy (DR) in severe stages. Thus, an automatic detection method applied to diagnose DR in an earlier phase may help medical doctors to make better decisions. DR is considered one of the main risks, leading to blindness. Computer-Aided Diagnosis (CAD) systems play an essential role in detecting features in fundus images. Fundus images may include blood vessel area, exudates, micro-aneurysm, hemorrhages, and neovascularization. In this paper, our model combines automatic detection for the diabetic retinopathy classification with localization methods depending on weakly-supervised learning. The model has four stages; in stage one, various preprocessing techniques are applied for smoothing the data set. In stage two, the network had gotten deeply to the optic disk segment for eliminating any exudate's false prediction because the exudates had the same color pixel as the optic disk. Stage three, the network is fed through training data to classify each class label. Finally, the layers of the convolution neural network are re-edited, and the layers are used to localize the impact of DR on the eye's patient. The framework tackled the matching technique between two essential concepts where the classification problem depends on the supervised learning method. In comparison, the localization problem was obtained by the weakly supervised method. An additional layer known as weakly supervised sensitive heat map (WSSH) was added to detect the ROI of the lesion at a test accuracy of 98.65%.

Corresponding Author:

MD Samiul Islam,

Department of Computing Science,

University of Alberta,

Email: mdsamiul@ualberta.ca

1. INTRODUCTION

World Health Organization WHO considered Diabetic Retinopathy on the priority list of most fatal and common complication diseases. Regular dilated eye examinations were carried out for all patients with diabetes as per WHO recommendation. That is to identify any eye pathology at early stages to enable early intervention and reducing the risk of progression to sight loss [1].

The main classification phases were reported to detect the severity of DR. The DR is a microvascular disease developing either as non-proliferative DR (NPDR) or as proliferative DR (PDR). The main variation between them is based on the presence of neovascularization NPDR. The retinal microvasculature is compromised, and its control on permeability becomes susceptible to microaneurysm [2]. NPDR is subdivided into classes: 0- No apparent retinopathy 1- Mild: small outpunching in the tiny blood vessels' walls appear in the retina. 2- Moderate: The disease's progression where the blood vessels nourish the retina swell, distort, and lose their ability to transport blood. 3- Severe: occur within numerous hemorrhages and microaneurysms in 4 quadrants of the retina, the cotton wool spots appear in 2 or more quadrants and Intraretinal microvascular abnormalities in at least 1 of retina's quadrant.

Soft exudates (SE) or Cotton Wool Spots (CWS) appear due to the arteriole occlusion. This debris accumulation has a different shape that appears as woolly white lesions in the Retinal Nerve Fiber Layer (RNFL). On the other hand, hard exudates (HE) are vivid yellow or white-colored entities on the retina. These entities appeared waxy and had sharp edges against the background from blood vessels. Hard exudates are evolved due to blood leakage from veins, and exudates had circular shapes around vessels.

PDR is an advanced stage of DR. It happens when flimsy and fragile blood vessels grow peculiarly from the retina into the vitreous. Where it represents a high risk of leakage that is considered the leading cause of blindness problem. PDR can be characterized by neovascularization on the retina and posterior surface of the vitreous and can lead to retinal detachment [3,4]. These new blood vessel developments are abnormal that lead to leaking the blood inside the retina. The Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) [5] defined PDR, with the risk of sight loss dramatically increasing with progression through these groups.

In this paper, the pre-trained Convolutional Neural Network (CNN) developed to a Multilayer convolutional Neural Network Model (MCNN). CNN and MCNN are a first-hand cooperative detection and localization layer by classifying the input image and gathering mapping features from image calculation by applying a weakly supervised sensitive heat map (WSSH). Our project has many contributions; first & foremost, the structure accomplishes a DR raw dataset matching procedure to guide the matching movement heat mapping near the exudate's regions. The consolidated network is a step-set localization procedure by computing the lesion's point in the validation phase with matching flow-data training as a shape before instructing the weakly supervised (WS) for APTOS localization. On the other hand, Image classification's accuracy for the whole network was increased by separating its computation in typical images and DR images.

The rest of this paper is organized as follows. In Section-2, the former researches work related to the CNN, and DR predictions & localization are discussed. Section-3 seizes the suggested methodology, which consists of a deep learning framework, an Efficient Net model: and weakly supervised localization techniques. Section-4 illustrates the dataset description, the experimental settings, and the result in analysis. In summary, a conclusion is given.

2. LITERATURE REVIEW

Vision-2020 is a global initiative that indicates 'the Right to Sight' that established a partnership between the World Health Organization (WHO) and the International Agency for the Prevention of Blindness (IAPB), launched in 1999. The two aims of eliminating avoidable blindness and decrease visual impairment as a global public health problem. However, Automatic predicting approaches are supposed to be developed to improve the pathological examination's results' efficiency and reliability. Previously, numerous models were applied to obtain the best classification with segmentation, even with localize techniques such as the DWCE algorithm. It proposed for segmentation of digitized mammograms and implemented along with edge detection and morphological feature classification [6]; nevertheless, it is still a complicated issue and tedious task while consuming time, size of lesions, and the similarity between levels of the pixels that appear in hard exudates (HE) and optic disk (OD).

Basically, the Convolutional neural network is an efficient branch of the neural network representing the core of deep learning. CNN has been expanded its utility from detection to localization and segmentation [7]. Moreover, it makes significant breakthroughs in computer vision. Improving the representation ability is a crucial problem in designing the model structure for handling complex tasks such as attribute analysis. Typical image processing methods such as contrast enhancement [8], histogram analysis [9], edge detection [10], and matched filtering [11] have been applied to identify different types of DR lesions.

2.1. CNN CLASSIFICATION

Several algorithms were proposed to solve the classification problem, such as Decision Tree, Support Vector Machine, and Random Forest to detect bright and dark DR lesions. The region-level classification was the best technique compared with pixel-level, considering the shape and size of the lesion. Pires et al. [12] investigated a data-driven method to extract powerful abstract representations and applied it to investigate the ability of transfer learning (TL) in the context of DR screening at a particular area on the image. The presented technique adopts to denote the patient immediately from the retinal examination pixels, without preliminary feature extraction for the image or lesion edges detection from an initial basic configuration.

A deep learning-based solution was created to highlight various methods explored to realize a robust framework. Lama et al. [13] a novel red lesion detection method based on a new set of shape features named Dynamic Shape Features. Although it is an automatic telemedicine scheme, it decreased the AUC to 0.899 on Messidor and used a Random Forest (RF) classifier to discriminate between lesions and non-lesions spots. Shanthi et al. [14] modified the Alexnet architecture for classification problems on the Messidor dataset after preprocessing step, including the green channel to provide enhanced optic nerves and other vital features of the image by initialing into RGB channels. The goal of Alexnet is to compute a better computational aptitude to tackle the difficulties than other networks. Such as Le-Net, Conv-Net. The presented technique utilizes principal component analysis (PCA) to separate the optic disc OD from the fundus dataset.

2.2. Automatic detection of DR

The main advantage of automatic detection is assisting ophthalmologists to detect, classify, localize, or segment the damage region that DR causes and ensure the diagnosis at any phase of the class. Thus it is considered more efficient and saving time of detection.

Many studies have been proposed for DR detecting. These studies' central perspective is generally mentioned as three significant methods: Unsupervised Methods, Supervised Methods, and weakly supervised learning.

Many DR detection approaches for retinal images have been introduced in recent years. The most efficient models are related to deep learning techniques especially, the supervised method [15]. Supervised approaches make use of the training phase that is manually managed and segmented by ophthalmologists. Various supervised classifiers have been used for this purpose as Artificial Neural Networks (ANN), K-Nearest Neighbor (K-NN) classifier. The inclusion of contextual semantic environments enhanced the performance of the classifier. The Gaussian mixture model [16] produces segmentation by classifying each image pixel as vessel or no vessel depending on the pixel feature vector.

Unsupervised methods [17] apply the Support Vector Machines (SVM) model with the basic line detector to the green channel of the retinal image. It is created to evaluate the average grey pixel-level along the lines of fixed length passing through the target pixel at different orientations. Maninis et al. [18] proposed a CNN model called Deep Retinal Image Understanding DRIU algorithm that specializes in a base network for segmenting blood vessels and optic discs detection as an image-to-image regression task.

Several blood vessel segmentation methods have been achieved a satisfying result. Moreover, the segmentation approaches implemented to retinal image with enhancement methods which support a better accuracy. It applied a threshold method to obtain the edge binary map of DR. [19] Multilayered threshold method was proposed to segment the retinal Dataset's blood vessels to detect neovascularization. The new blood vessels improved by two-dimensional wavelets. Abnormal blood vessels are detected using a sliding window procedure. Also, segmentation of blood vessels from the retina images may improve the detection process of any small spots of lesions, leading to vision disorders [20][21].

3. METHODOLOGY

As shown in **Fig.1**, the proposed model to detect diabetic retinopathy consists of four main stages; the image dataset is divided into two subsets: training the CNN and the other for testing and evaluating the detection algorithm.

3.1. Pre-processing

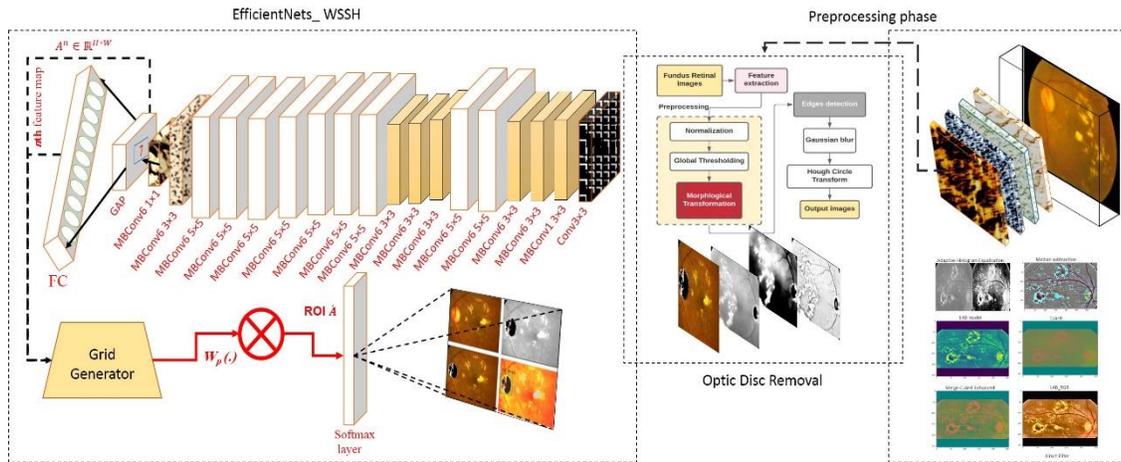


Figure 1 The schematic diagram of the proposed WSSH_DR model with Multi-scale features strategy, Modified preprocessing, ODR algorithm, and fully connected layer in EfficientNetB5 stage.

Algorithm-1: Preprocessing Algorithm

Input	Collect the distribution of labels depends on each class; the features X_i of the disease collected from the patient, $i=1,2, N$; The stored data of the disease in the database, class of the disease Y_i , $i = 0$ or 1 ; $M \times N$ size of the image, n each data contains the disease feature vector X_i , corresponding Filter size, H is histogram with column h_1, \dots, n .
Initialize	Gamma Correlation step by Eq. (1); and Adaptive histogram equalization by Eq. (2)
Output	Preprocessing steps increasing feature extracting for image Y

1: procedure

2: median subtraction: (First step)

3: $X \leftarrow$ Features of probability density function

4: for $i \leftarrow 1$ to M , $j \leftarrow 1$ to N do

5: Remove $X_{i-n/2-1, j+n/2}$ from $h_{j+n/2}$

6: Add $X_{i+n/2, j+n/2}$ from $h_{j+n/2}$

```

7:       $H \leftarrow H+hj+n/2-hj-n/2-1$ 
8:       $Y_{i,j} \leftarrow \text{median}(H)$ 
9:      end for
10: Compute histogram CDF: (Second step)
11:   for  $i:1$  to  $M*N+1$  do
12:      $g = [g_{max} - g_{min}] p(f) + g_{min}$ 
13:      $LAB \leftarrow (Y, \text{convert } BGR \text{ to } LAB)$ ;
14:      $Clahc \leftarrow \text{create } CLAHE \text{ (tile Grid Size)}$ 
15:   end for
16: Krisch filter (Third step)
17: Kernel  $[1: N] = N, S, E, W, NW, SW$  edge operator
    for  $i=1$  to  $n$  do
     $i = i+1$ 
    end for
18: Edge[i] = image maximum value  $\otimes$  Kernel[i]
    maximum value,  $M = \max(\text{kernel}[i])$ 
19: Compute the maximum intensity pixel and set other 0
20: Convolution output
    for  $k: =1$  to  $6$ ,  $i: =1$  to  $M$ ,  $j: =1$  to  $N$  do
     $[m,n] = \text{size}(\text{Edge}[i])$ 
    Create Y-image with  $M \times N$  size  $E = \text{Edge}[Y]$ 
    end for
21: Return:  $(E[i,j] = M) Y(i,j) = Y[i,j] + E[i,j]$ 
22: normalize function
    If  $Y >= 255$ , then  $Y(i,j) = 255$ 
    end if
23: Return Edge[Y]
24: end procedure

```

Image preprocessing is an essential step to eliminate erroneous illumination contrast and noisy pixels with cumulative contrast between the background and the retinal blood vessels. The selection of preprocessing techniques affects the results accuracy on feature extraction and DR detection stages. Initially, the histogram of an image is used to produce a better image. The Gamma correlation Eq.1 is applied to mapping the relationship between a color value and its brightness. Then the adaptive histogram Eq.2 is used to improve the image's contrast. The preprocessing algorithm steps are described in the algorithm (1).

$$I = 255 * \left(\frac{I}{155} \right)^{\frac{1}{g}} \quad (1)$$

Where I is the input value, g is Luminance intensity, 255 represent maximum gray level, but the adaptive histogram computed as:

$$P_x(i) = P(x = i) = \frac{n_i}{n}, \quad 0 \leq i \leq L \quad (2)$$

Where x is detached grayscale image, and n_i be the numeral of occurrences of gray level i , L being the total number of gray levels in the image, n being the total number of pixels of the image, and $P_x(i)$ being the image's histogram for pixel value i , that normalized to $[0,1]$ as a black and white.

First step: Median filter is a creative technique applied to reduce or remove the impulsive noise and preserves edges of objects effectively, which does not require convolution. As described in algorithm-1, a median filter was applied to remove outliers from the image.

Second step:

- Converting image to LAB Color model
- Splitting the LAB image to different channels
- Applying contrast limited adaptive histogram equalization (CLAHE) to L-channel
- Merge the CLAHE enhanced L-channel with the a and b channel
- Converting images from the LAB Color model to the RGB model.

The Adaptive Histogram Equalization technology (AHE) is the technique of mapping every pixel in the image to an intensity value derived from the surrounded pixels by a transformation function. As computed by Eq.2, the main advantage of Contrast Limited AHE (CLAHE) in comparison with AHE is that it prevents over-amplification of noise in the image homogeneous parts. Basically, the CLAHE procedure [7] improves the contrast among tiny regions of the lesion.

However, Contrast enhancement produces a better image than the original by changing the pixel intensities. The point processing adopted the pixel's values in the original input image to produce the values of the corresponding Eq.3.

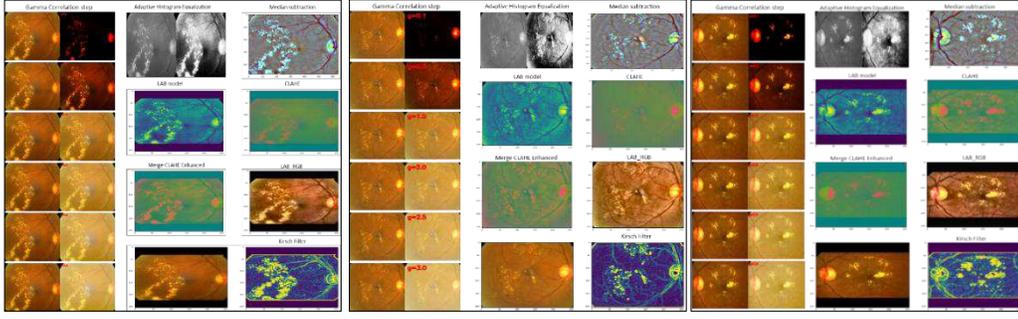


Figure 2 Preprocessing algorithm results

$$Y(i, j) = T[X(i, j)] \quad (3)$$

Where $X(i, j)$ is the original raw image, $Y(i, j)$ is the enhanced image, and T defines pixels in the enhanced image expressed in the equation. The transformation between the two images Some of the point processing techniques includes: contrast stretching, global thresholding, histogram equalization; particular mask processing methods contain averaging filters, sharpening filters, local thresholding.

Third step: Kirsch filter is a non-linear edge detector that finds the maximum edge strength in a few predetermined directions. The main advantage of the kirsch filter is the efficiency of detecting maximum edges. The result analysis for this filter illustrates the last of preprocessing, and it used a convolution table as a 3×3 table of pixels. Fig.2 is illustrating all the results of preprocessing algorithm-1.

3.2. Optic disk removal

The optic disk removal algorithm's main advantage is producing the images without regions of interest (ROI) of optic disk (OD) pixels, thus increasing exudates detection performance as long as the exudates have a similar color of OD pixels, which appears with yellow level. The second advantage is solving the problem of false-positive detection. Following Eq. 4 will provide a mathematical background to normalize the grayscale image.

$$G(i, j) = L(i, j)L(i, j + 1) \quad (4)$$

$L(i, j)$ is the grayscale of point (i, j) , while images convert them to luminance components. As well as each pixel in grayscale has a value representing the shade of gray. The large images are resized into smaller images to avoid a long time for training. The gradient and angle for the image are computed by Eq.5.

$$G = \sqrt{G_x^2 + G_y^2}, \quad \theta = \tan^{-1}\left(\frac{G_y}{G_x}\right) \quad (5)$$

The pixels intensity for the first selected circle is computed, and then the optic disk is cutting using specific algorithms [21, 22]. The proposed Optic disk removal (ODR) algorithm-2 is enhanced, as shown in Fig. 1, by improving the contrast of the image components by (CLAHE) in the Global thresholding technique. Gaussian blur added after edges detection for removing any false pixel. While horizontal direction G_x and vertical direction G_y for mask processing. The pixel value is extracted from the original pixel value of each image. It is a more expensive process than simple point processing, though it is robust. The grid mask for an input image will produce an output image of the same size as the input array. Even the features extracted from all images provide the same size to provide data consistency – the process used to determine OD's size when r represented the radius of OD by Eq.6.

$$r^2 = (x - a)^2 + (y - b)^2 \quad (6)$$

$$x = a + r \cdot \cos(\theta), y = b + r \cdot \sin(\theta) \quad (7)$$

Where a, b are the center coordinates. The gradient $g(x, y)$ is a threshold version of $f(x, y)$ at some global threshold T . The global thresholding applied for the effective region outputs in a packed bit $(0, 1)$ format described in Eq. (8). The Global thresholding is creative as the degree of intensity separation between the two peaks represents the image's background and object pixels.

$$g(x, y) = \begin{cases} 1 & f(x, y) \geq T \\ 0 & \text{Otherwise} \end{cases} \quad (8)$$

Two morphological operators are applied (Opening and Closing) obtained from the essential tasks of erosion and dilation for binary images. Opening operation Eq.9, use to remove some of the small objects from the front pixels around the edges of ROI in the retinal image. Next operating, Closing operator Eq.10 removes small holes in the foreground by changing small islands of background

into the foreground; also, it reduces noise in binary images. To reduce the size of foreground objects, we can erode pixels given several iterations.

$$\text{Opening operation} : A \circ B = (A \ominus B) \oplus B \quad (9)$$

$$\text{Closing operation} : A \bullet B = (A \oplus B) \ominus B \quad (10)$$

Where \ominus and \oplus denote erosion and dilation, respectively. When, A is binary images, and B is the structuring element. For grayscale morphology, the structuring functions (Denoting an image by $f(x)$ and the structuring function by $b(x)$, the grayscale dilation f by b is given by Eq.11.

$$(f \oplus b)(x) = \text{SUP}_{y \in E} [f(y) + b(x - y)] \quad (11)$$

The gamma compression function must first be removed via gamma expansion (linearization) to transform the image to a linear RGB color space. Finally, The Erosion operator removes pixels from edges to make an object smaller, while the dilation operator adds pixels around its edges to make an object larger. The threshold method appears at the neighbors of a pixel and changes its state if the number of differing neighbor pixels exceeds a threshold. Fig.3 demonstrates the processing result of removing the optic disk and describes all images step by step.

4. Supervised learning CNN Works

Supervised learning model returns a function that predicts the correct result of $f(x)$, measured by generalized performance on F_{test} . It is generated by the learning algorithm using training data F_{train} . The classification model is the first part of the study. The ability to separate the objects from ground truth and translate it as classes by using their attributes was considered the most crucial target of many studies. The performance of any algorithm classifier should be evaluated to demonstrate the error rate of the network. Generally, other criteria used to assess the model's performance include sensitivity for true positive rate and specificity for true negative rate.

4.1. CNN for classification and localization (Algorithm 3)

Classification and localization procedure includes the following steps:

A-Initializing: The standard type of CNN is input sequential data $x = [x_1, \dots, x_T]$. T is the length ϵ of the sequence and $x_i \in \mathbb{R}^d$ at each step. The convolution step described by the dot product between a filter vector $u \in \mathbb{R}^{md}$ and a concatenation vector representation $x_{i:i+m-1}$, which defines the operation as in Eq. 12.

$$c_i = \varphi(u \cdot x_{i:i+m-1} + b) \text{ where } x_{i:i+m-1} = x_i \parallel x_{i+1} \parallel \dots \parallel x_{i+m-1} \quad (12)$$

Where (\cdot) represents the dot product, b and φ denote bias term and non-linear activation function, respectively. Where (\parallel) is the concatenation operation for two vectors.

Algorithm 2: Optic Disk removal algorithm

Input	Image with optic disk
Output	Image without optic disk
1: function extract mean features	
2: <i>equalize</i> ← equalize the histogram of gray scale image green ROI area	
3: The gradient G (i, j) is calculated by Eq. (5)	
4: <i>mask</i> ← Thresholding operation on equalize by Compute the center(x,y)(7) and radius of the eye by Eq.(6)	
5: Global Thresholding: Choose threshold T that separates object from background by Eq. (8)	
6: <i>OD Image</i> ← OD on red ROI with mask	
7: <i>Edge</i> ← Closing Morphological Transformation (OD Image) by Eq.(10)	
8: <i>OD Erased</i> ← Apply Opening Morphological Transformation for Edge by Eq.(9,10)	
9: Gaussian blur by Eq. (12)	
10: Hough Circle Transform by Eq. (13)	
11: end function	

B-Update Hyper-parameters

Its numbers of values derived via training the network have a significant effect on the model's predictive performance. It includes the number of layers, the kernel size as well as parameters that influence the gradient descent. The goal is to focus here that the parameters are governed and guide the learning method. The updating equation for the weights involving the hyper-parameters described in Eq.13 and 14:

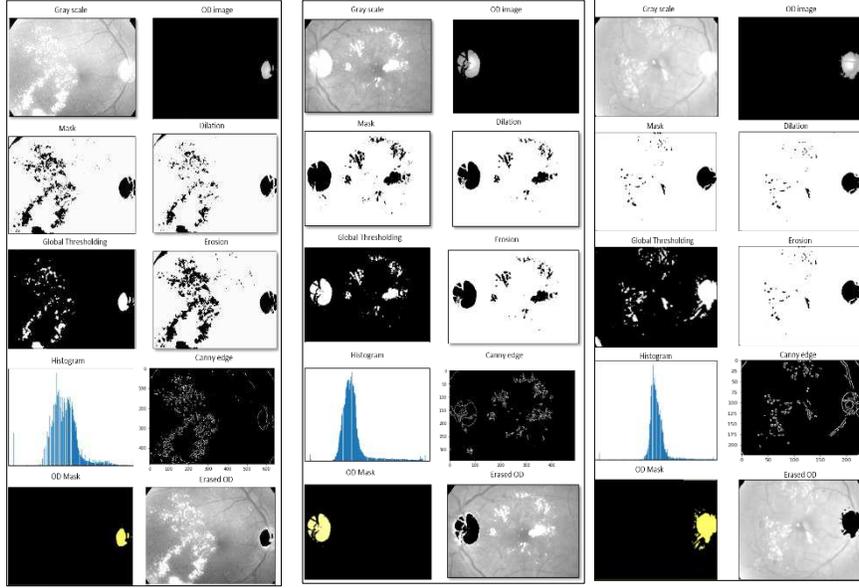


Figure 3 Optic Disk Removal detection

$$\omega \leftarrow \omega' = \omega - v \quad (13)$$

$$v \leftarrow v' = \gamma v + \eta \frac{\partial C}{\partial \omega} + \frac{\lambda}{2n} \sum_i \omega_i^2 \quad (14)$$

Where η is learning rate applied to determine the impact of each weight directly update. Momentum $\gamma \in (0,1)$ described the inertia of the gradient update that means how much weight value is assigned to the last gradient. Weight decay $\lambda \in (0,1)$ is a weight regularization. The most crucial step is that the learning rate should decrease over time depending on decreasing few epochs by a pre-defined factor or exponential decay.

C- EfficientNet model:

We revisit the EfficientNet [23] model, which defined the ConvNet layer N as a list of composed layers as Eq. 15. The goal is to develop the baseline in the model by leveraging a multi-objective neural design search that enhances the accuracy.

$$N = \odot_{i=1}^s F_i^{l_i} (X_{(H_i, W_i, C_i)}) \quad (15)$$

However, scaling up the compound layers of our model increases the accuracy of detection. Where many deep layers will lead the model to be slower and without any enhancing of detected. For that reason, we custom the layers of the model by fine-tuning and freezing others. On the contrary, the model is prevented from vanishing gradient problem that the gradient will be vanishingly small by preventing the weight from changing its value. Thus, it will completely stop the model from further training and produce poor results.

D- Normalization

- BN is an adaptive re-parameterization scheme and is used for solving the struggle of training deep models.
- Deep Neural Networks DNN will have the ability to optimize easily.
- The re-parameterization of any network can be effectively simple the problem of coordinating updates across all layers.

Mathematically, BN works to enhance overall the model by optimizing and update the parameters. Let us defined a mini-batch B of size n , when the normalization is applied to every activation independently. The specific activation x . Thus, the model has n values of this activation in the mini-batch, represented by $B = x_1 \dots x_n$. Normalize the batch of values $B = x_1 \dots x_n$ using the mean and variance of the batch compute as:

$$\begin{aligned} \mu_B &= \left(\frac{1}{n}\right) \sum_{i=1}^n x_i \\ \sigma_B^2 &= \left(\frac{1}{n}\right) \sum_{i=1}^n (x_i - \mu_B)^2 \end{aligned} \quad (16)$$

Where, μ_B is the mini-batch mean, and σ_B^2 is the mini-batch variance.

Algorithm 3: CNN for classification and localization procedure

Input	Y the preprocessed image feed to network, l convolutional layer, feature maps $Z^l(x, y, i)$, Input, sequential data $x = [x_1, \dots, x_T]$ T is the length ϵ of the sequence, Feature map $c_j = [c_{j1}, c_{j2}, \dots, c_{j_{l-m+1}}]$ and $x_i \in \mathbb{R}^d$ at each step:
Initialize	<ol style="list-style-type: none"> 1: set the number of object C, 2: set the number of the top sensitive region S, 3: Pre-trained a CNN model for classification, 4: Extract candidate regions by the selective search for all training data, 5: Convolution process, the dot product between a filter vector $u \in \mathbb{R}^{md}$ and concatenation vector representation $x_{i:i+m-1}$ defines the convolution operation by Eq. (12) 6: update hyperparameter, by Eq. (13)
Training:	<p>for $c = 1, \dots, C$ do</p> <p style="padding-left: 2em;">for $s = 1, \dots, S$ do</p> <ol style="list-style-type: none"> 7: compute N by Eq. (15), 8: Relu function $P^l(x, y, i) = \max(0, Z^l(x, y, i))$ 9: Normalized Image function by Eq. (16), 10: Calculate the Average pooling layer by Eq. (17.18), 11: learn Classification layer by Eq. (19), <p style="padding-left: 2em;">endfor</p> <ol style="list-style-type: none"> 12: aggregation formula to compute the overall score for object C_n in the image Eq. (20) 13: WSSH compute by Eq. (22), to getting the sensitive region S, 14: training the final object detection for class c; <p>endfor</p>
Testing:	<ol style="list-style-type: none"> 16: extract the sensitive candidate regions by WSSH for all test images. 17: Apply all the learnable object detectors for all the candidate regions of DR. 18: $S_{c_{WSSH}}$ represents images with final sensitive heat localize. 19: end function,

E- Global Average Pooling (GAP)

The proposed model's main problem is many parameters; consequently, the model requires a large amount of training data to avoid overfitting. Many different techniques succeeded in reducing the overfitting problem. One of them is the pooling process. The pooling process calculates the number of neurons in all convolutional layers and subsequently compares to the number of neurons stated.

It works to reduce the number of parameters drastically. Consequently, it reduces the training effort. The Average pooling layer has more efficiency to increase overall the accuracy of the model. It computed by Eq.17

$$Q^l(x, y, i) = P^l(x, y, i) \left(\gamma + \alpha \sum_{j=M^l} (P^l(x, y, i))^2 \right)^{-B} \quad (17)$$

Where $Q^l(x, y, i)$ is compute the response of the normalized activity from the ReLU with output $P^l_{(x,y,i)}$. This is achieved by multiplying the output with an inverse summation of squares plus an offset γ for all ReLU outputs within a layer l . The average pooling operator computes the mean response of each function channel obtained from the normalized output. It can be computed as Eq.18

$$R^l(x, y, i) = \frac{\sum_{x,y \in M(\bar{x}, \bar{y}, i)} Q^l(x, y, i)}{|M(\bar{x}, \bar{y}, i)|} \quad (18)$$

Summation weighted features were added in the latest convolutional layer of the model and adopted GAP layer performance to sum the average value of feature mapping. The average pooling operator computes the mean response of each feature channel obtained from the normalized output.

The GAP's drawback has an extra disposed to underestimate object sizes' values because it considers all the activations. GAP loss stimulates the system for recognizing objects compared to global max pooling, which encourages recognizing one discriminative slice. The map calculates an average maximized value for whole parts discrimination to the interpreter of an object.

F- Classification model

Y is fundus images with each image $A \in \mathbb{R}^{H \times W}$ and the classification model $f(\cdot)$ that assigns to each image a class label $y \in c$ where $c \subset \{A_0, A_1, A_2, A_3, A_4\}$, $Y_s = F(A; \omega_f)$ where Y_s is prediction and ω_f are the classification model parameters. The neural network learns the patterns from various images that should be included with detection. The most important part of model evaluation is the process of feature extracting. The network scans pixels of an image to recognize the object or the actual ground through this process. The main problem that affects adversely on the feature extraction is the noise included within images.

Thus, some preprocessing methods are added as mention in algorithm.1 on the Dataset to gather the highest values of features for learning our model. The probability of the classification step is computed using Eq.19, where w_k is the corresponding set of GAP layer output.

$$I_{c2}(x, y) = \sum_k \left(\frac{w_{k1}^c}{4} + w_{k2}^c \right) f_k(x, y) \text{ and } P_c = \frac{e^{S_c}}{\sum_c e^{S_c}} \quad (19)$$

S_c represents the summation of input and output of unit c at the final output layer of the fully connected network. $I_{c2}(x, y)$ is importance of unit (x, y) in special area. The class-sensitive heat saliency map is proposed to represent the values of $I_{c2}(x, y)$, $I_{c3}(x, y)$, $I_{c4}(x, y)$, and $I_{c5}(x, y)$.

4.2. Weakly Supervised DR Localization via WSSH

Generally, there are two approaches of determining localization, representing by bounding boxes or without bounding boxes labels. Weakly supervised localization (WSL) is still considered a challenging task that deals with image labels, splitting them into classes and creating the bounding boxes for all images. WSL concept used the automatically learning computing bounding boxes. The localization model is specified as $g(\cdot)$ that return s the position p of the ROI. the localization model parameters $f \omega P = g(A; \omega_g)$, $\hat{A} \in \mathbb{R}^{\hat{H} \times \hat{W}}$ Represent the ROI images obtained by $\hat{A} = W_p(A)$, $W(\cdot)$ is a warping operator. The main goal for weakly supervised localization layer is avoiding bounding box annotations, while the transformation matrix $W_p(\cdot)$ was applied to arrange and extract the primary samples for the location of the (ROI) grid. After the classification process, Class Sensitive Heat Map (Weakly Supervised Sensitive HeatMap) is derivative based on the activations of the feature maps of the latest convolutional layer. The activation of the unit at spatial location (x, y) in the n th feature map is denoted as $f_n(x, y)$. The class-sensitive heat map H_c for class c is defined by Eq.20.

$$H_c(x, y) = \sum_n w_n^c f_n(x, y) \quad (20)$$

$F(x, y) = \sum_n w_n^c f_n(x, y)$ when $A_{n1} = \frac{1}{N_1} \sum_{x,y=1}^n f_n(x, y)$, where N_1 represents the number of units in region r_1 . The pooling technique pools the n th feature map into units, A_{n2}, A_{n3}, A_{n4} , and A_{n5} . Then, all units will be fed into the softmax layer. The input of the softmax layer for category c is denoted as S_c and defines in Eq.21.

$$S_c = \sum_n w_{n1}^c A_{n1} + \sum_n w_{n2}^c A_{n2} \quad (21)$$

Where, c is the number of class, S is the score of probability, n feature map at $A^n \in \mathbb{R}^{H \times W}$.

$$S_{c(WSSH)} = \frac{1}{N_1} \sum_x \sum_y \sum_n W_n^c A_{x,y}^n \quad (22)$$

W_n^c is class feature weights, $A_{x,y}^n$ feature sensitive heat map, neural architecture search used to design a new baseline network, and scale it up to obtain high prediction performance, called by EfficientNets, perfectly solves the classification problem and faster than other models such as ResNet, DensNet, inception. Table (1) illustrates the comparison results of different models. Finally, revisiting the last layer of the model responsible for localizing the region of DR, fig.4 illustrates the Weakly Supervised Sensitive Heatmap for lesion localization without OD.

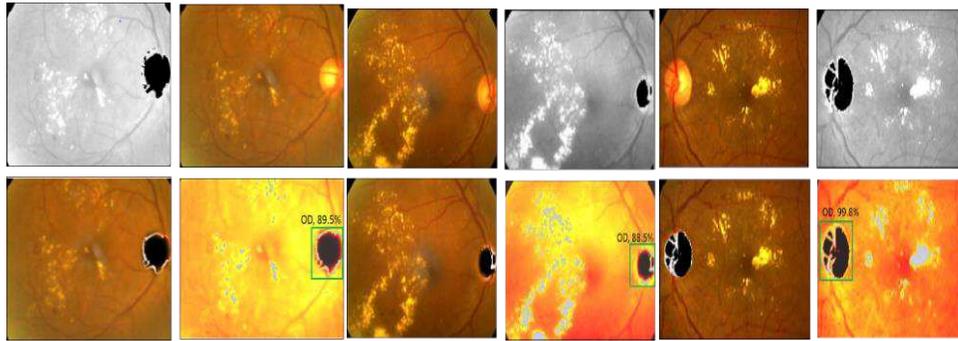


Figure 4: Weakly Supervised Sensitive Heat-map localization (WSSH)

5. Results and Evaluation

The parameter initialization methods set to train the whole network by used the momentum optimizer, while *MESSIDOR* [31] *Dataset* was described and published in 2008. However, it contains four different classes with variable intrinsic and extrinsic features. the network trained a total of 60 epochs with an initial weight decay of 0.0001 for learning rate and the momentum of 0.9. The learning rate sets at 0.1 at epoch 25, then 0.01 at epoch 37, then 0.001 at epoch 45, and 0.0001 to epoch 60, while the batch normalization is applied after the convolution layer and the non-linearity layer. Fig. 5 shows the accuracy of detection for our CNN-WSSH model for classification and localization of the lesions of DR. Table (2) compares the performance metrics for the test phase with different DR classes.

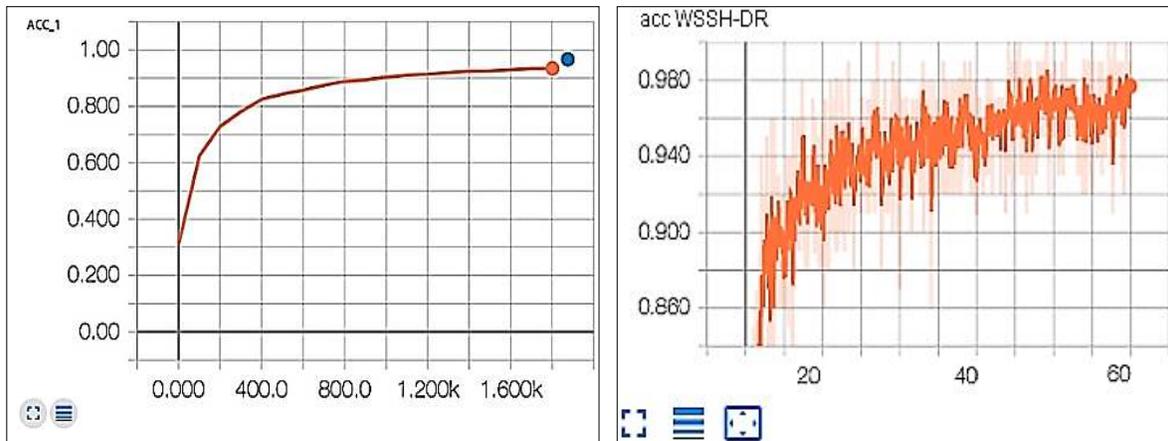


Figure 5 Accuracy of detection for WSSH-DR

Table (1): Level performance comparison with different techniques for Lesion detection.

Ref	Method	AUC
Kumar S B et al. [24]	ADDR classified DR principal	94.44%
		87.5%
Bhatia et al. [25]	SVM	90.76%
	PNN	87.69%
García et al. [26]	SVM	82
Haar et al. [27]	–	72
Gondal et al. [28]	CNN-CAM	0.954
Shanthi et al. [14]	Modified Alexnet	95.6%
Dash et al. [29]	Star networked pixel tracking	95.83%
	2-D Gabor wavelet to	94.69%
	SVM	93.2%

Verma et al. [20]	Random Forests technique (normal cases) (moderate and severe) [4]	90% 87.5%
Niemeijer et al. [30]	-	95
CNN_EfficientNetB5 (ours)	CNN-WSSH	98.4%
Densenet201 (ours)	CNN- Classification	94.5%
ResNet50 (ours)	CNN- Classification	89.7%

Table (2) Performance metrics for test phase with different DR classes.

Input class	Specificity	Sensitivity	Precision	Accuracy
DR A0 (healthy retina)	99.3	95.3	95.3	99.6%
DR A1 MILD NPDR	98.3	88.0	93.0	99.2%
DR A 2 MODERATE NPDR	98.6	90.1	94.0	98.6%
DR A3 SEVERE NPDR	96.6	96.0	86.0	97.6%

6. Conclusion

Early diagnosis of subclinical DR will offer appropriate recognition and control for patients at a greater DR progression danger. Currently, strict regulation of blood glucose levels and other changes of risk factors are the only prevention schemes at early DR stages. The ideal technique is feature extraction by using some preprocessing algorithm for increasing feature learning that feeds the network. In contrast, an optic disk removal algorithm is added to cutting the optic disk circular because it makes a false-positive result with Exudates regions. As well as, OD has appeared in the same pixel color of Exudates. A localizing part has been detecting spots of exudates clearly and avoiding false learning error at the training part. All preprocessing steps with preparing conspicuous data and removing optic disc have significantly impacted the accuracy and avoided many errors in the learning phase. It can be concluded that any model detects appropriately depend on the data used in the learning stage. Therefore, if the network feeds with the dark, tenebrous, dim Dataset, it will decrease the model's efficiency.

Competing interests

The author(s) declare no competing interests.

Data availability

The datasets analyzed during the current study are *MESSIDOR* [31]

REFERENCES

- [1] World Health Organization. Prevention of blindness from diabetes mellitus: report of a WHO consultation in Geneva, Switzerland, 9-11 November 2005. World Health Organization; 2006.
- [2] Lois N, McCarter RV, O'Neill C, Medina RJ, Stitt AW. Endothelial progenitor cells in diabetic retinopathy. *Frontiers in endocrinology*. 2014 Apr 9;5:44.
- [3] Fong DS, Aiello L, Gardner TW, King GL, Blankenship G, Cavallerano JD, Ferris FL, Klein R. Retinopathy in diabetes. *Diabetes care*. 2004 Jan 1;27(suppl 1):s84-7.
- [4] Gardner TW, Abcouwer SF, Barber AJ, Jackson GR. An integrated approach to diabetic retinopathy research. *Archives of ophthalmology*. 2011 Feb 14;129(2):230-5.
- [5] Klein R, Knudtson MD, Lee KE, Gangnon R, Klein BE. The Wisconsin Epidemiologic Study of Diabetic Retinopathy XXII: the twenty-five-year progression of retinopathy in persons with type 1 diabetes. *Ophthalmology*. 2008 Nov 1;115(11):1859-68.
- [6] Petrick N, Chan HP, Sahiner B, Wei D. An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection. *IEEE Transactions on Medical Imaging*. 1996 Feb;15(1):59-67.
- [7] Ghesu FC, Krubasik E, Georgescu B, Singh V, Zheng Y, Hornegger J, Comaniciu D. Marginal space deep learning: efficient architecture for volumetric image parsing. *IEEE transactions on medical imaging*. 2016 Mar 7;35(5):1217-28.
- [8] Zeljković V, Bojic M, Zhao S, Tameze C, Valev V. Exudates and optic disk detection in retinal images of diabetic patients. *Concurrency and Computation: Practice and Experience*. 2015 Jan;27(1):172-92.
- [9] Kumar PS, Kumar RR, Sathar A, Sahasranamam V. Automatic detection of exudates in retinal images using histogram analysis. In *2013 IEEE Recent Advances in Intelligent Computational Systems (RAICS) 2013 Dec 19 (pp. 277-281)*. IEEE.
- [10] Youssef D, Solouma NH. Accurate detection of blood vessels improves the detection of exudates in color fundus images. *Computer methods and programs in biomedicine*. 2012 Dec 1;108(3):1052-61.
- [11] H. A. Nugroho, D. A. Dharmawan, I. Hidayah, and L. Listyalina. Automated microaneurysms (mas) detection in digital colour fundus images using matched filter. In *Proceeding - 2015 International Conference on Computer, Control, Informatics and Its Applications: Emerging Trends in the Era of Internet of Things, IC3INA 2015*, pages 104–108, 2016.
- [12] Pires, R., Avila, S., Wainer, J., Valle, E., Abramoff, M.D. and Rocha, A., 2019. A data-driven approach to referable diabetic retinopathy detection. *Artificial intelligence in medicine*, 96, pp.93-106.
- [13] Seoud L, Hurtut T, Chelbi J, Cheriet F, Langlois JP. Red lesion detection using dynamic shape features for diabetic retinopathy screening. *IEEE transactions on medical imaging*. 2015 Dec 17;35(4):1116-26.

- [14] Shanthi T, Sabeenian RS. Modified Alexnet architecture for classification of diabetic retinopathy images. *Computers & Electrical Engineering*. 2019 Jun 1;76:56-64.
- [15] Sinthanayothin C, Boyce JF, Cook HL, Williamson TH. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *British journal of ophthalmology*. 1999 Aug 1;83(8):902-10.
- [16] Soares JV, Leandro JJ, Cesar RM, Jelinek HF, Cree MJ. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Transactions on medical Imaging*. 2006 Aug 21;25(9):1214-22.
- [17] Ricci E, Perfetti R. Retinal blood vessel segmentation using line operators and support vector classification. *IEEE transactions on medical imaging*. 2007 Oct 1;26(10):1357-65.
- [18] Maninis KK, Pont-Tuset J, Arbeláez P, Van Gool L. Deep retinal image understanding. In *International conference on medical image computing and computer-assisted intervention* 2016 Oct 17 (pp. 140-148). Springer, Cham.
- [19] Singh N, Kaur L. A survey on blood vessel segmentation methods in retinal images. In *2015 International Conference on Electronic Design, Computer Networks & Automated Verification (EDCAV) 2015 Jan 29* (pp. 23-28). IEEE.
- [20] M. U. Akram, I. Jamal, A. Tariq and J. Imtiaz, "Automated Segmentation of Blood Vessels for Detection of Proliferative Diabetic Retinopathy", *IEEE-EMBS International Conference on Biomedical and Health Informatics, China*, Jan. 2012.
- [21] Rizon M, Haniza Y, Puteh S, Yeon A, Shakaff M, Abdul Rahman S, Sugisaka M, Sazali Y, M Rozailan M, Karthigayan M. Object detection using circular Hough transform.
- [22] Almazroa A, Burman R, Raahemifar K, Lakshminarayanan V. Optic disc and optic cup segmentation methodologies for glaucoma image detection: a survey. *Journal of ophthalmology*. 2015 Sep;2015.
- [23] Tan M, Le QV. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*. 2019 May 28.
- [24] SujithKumar SB, Singh V. Automatic detection of diabetic retinopathy in non-dilated RGB retinal fundus images. *International Journal of Computer Applications*. 2012 Jan 1;47(19).
- [25] Bhatia K, Arora S, Tomar R. Diagnosis of diabetic retinopathy using machine learning classification algorithm. In *2016 2nd International Conference on Next Generation Computing Technologies (NGCT) 2016 Oct 14* (pp. 347-351). IEEE
- [26] García M, López MI, Álvarez D, Hornero R. Assessment of four neural network based classifiers to automatically detect red lesions in retinal images. *Medical engineering & physics*. 2010 Dec 1;32(10):1085-93.
- [27] TerHaar F. Automatic localization of the optic disc in digital colour images of the human retina. *Utrecht University*. 2005 Dec 16.
- [28] Gondal WM, Köhler JM, Grzeszick R, Fink GA, Hirsch M. Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images. In *2017 IEEE international conference on image processing (ICIP) 2017 Sep 17* (pp. 2069-2073). IEEE
- [29] Dash J, Bhoi N. A survey on blood vessel detection methodologies in retinal images. In *2015 International Conference on Computational Intelligence and Networks 2015 Jan 12* (pp. 166-171). IEEE.
- [30] Niemeijer M, van Ginneken B, Russell SR, Suttorp-Schulten MS, Abramoff MD. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Investigative ophthalmology & visual science*. 2007 May 1;48(5):2260-7.
- [31] Decenciere E, Zhang X, Cazuguel G, La, B, Cochener B, Trone C, Gain P, OrdóñezVarela J-R, Massin P, Erginay A. 2014. Feedback on a publicly distributed image database: the Messidor database. *Image Analysis and Stereology* 33(3):231–234 DOI 10.5566/ias.1155

BIOGRAPHIES OF AUTHORS

	<p>Mohammed Al-Mukhtar</p> <p>Mohammed Al-Mukhtar received the M.Sc. degree in Information and Communication Engineering from Huazhong University of Science and Technology HUST, Wuhan, China in 2019. Currently, he is working as a Lecturer at the University of Baghdad. His research interest includes Medical Image Analysis with deep learning techniques, Computer Vision, Object Recognition, and Object Segmentation.</p>
	<p>Ameer Hussein Morad</p> <p>Dr. Ameer H. Morad is Associate Professor in Computer engineering. He works at the Department of Information and Communication, Al Khwarizmi College of Eng., University of Baghdad. Presently he is involved in research work in Image Processing, Computer Vision, Visual Cryptography, and Network security.</p>



Mustafa Albadri

MSC in electrical engineering from the University Of Denver, United States, interested in power electronics, IoT networks, analysis image and wide band-gap semiconductors applications.



MD Samiul Islam

MD Samiul Islam is a Ph.D. scholar at the University of Alberta and currently is working in Multimedia Research Center in Canada. He works at the Department of Computing Science, University of Alberta, Canada, as a RA/TA. Presently he is involved in research work in Computer Vision, Machine Learning, Computer Networking, and Radar Image Processing. He completed his MSc in Information & Communication Engineering from Huazhong University of Science & Technology, China.

Figures

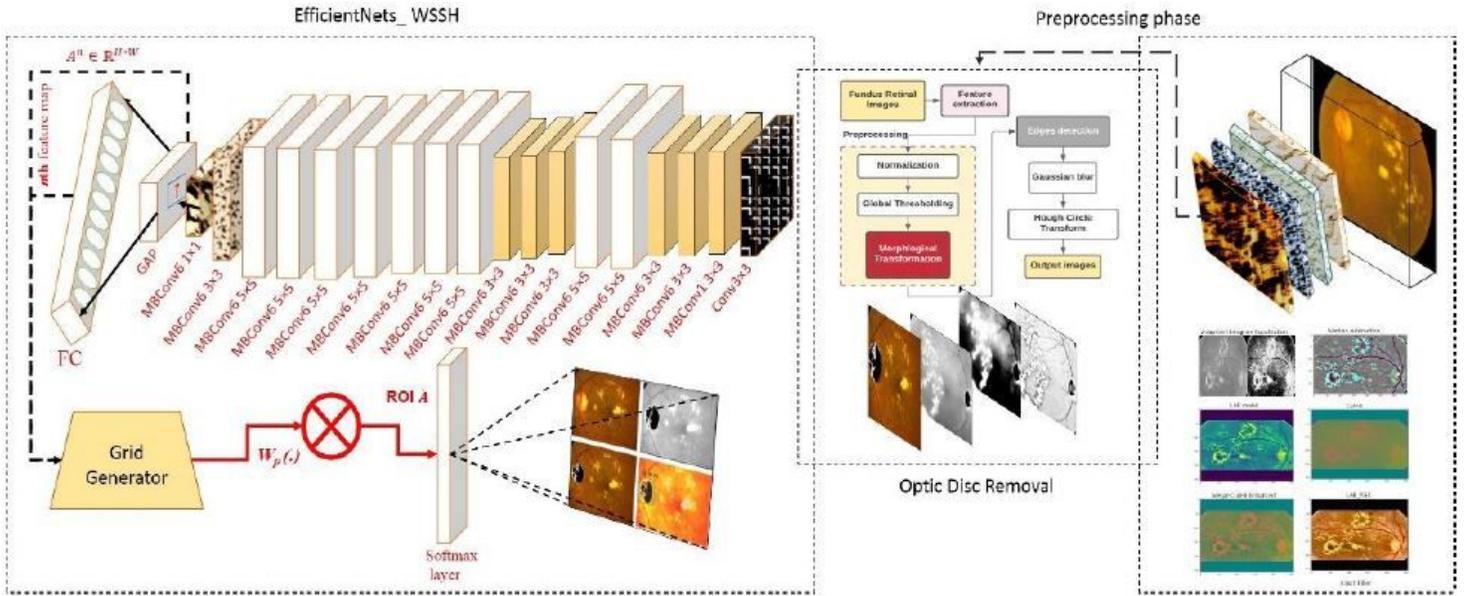


Figure 1

"Please see the Manuscript PDF file for the complete figure caption".

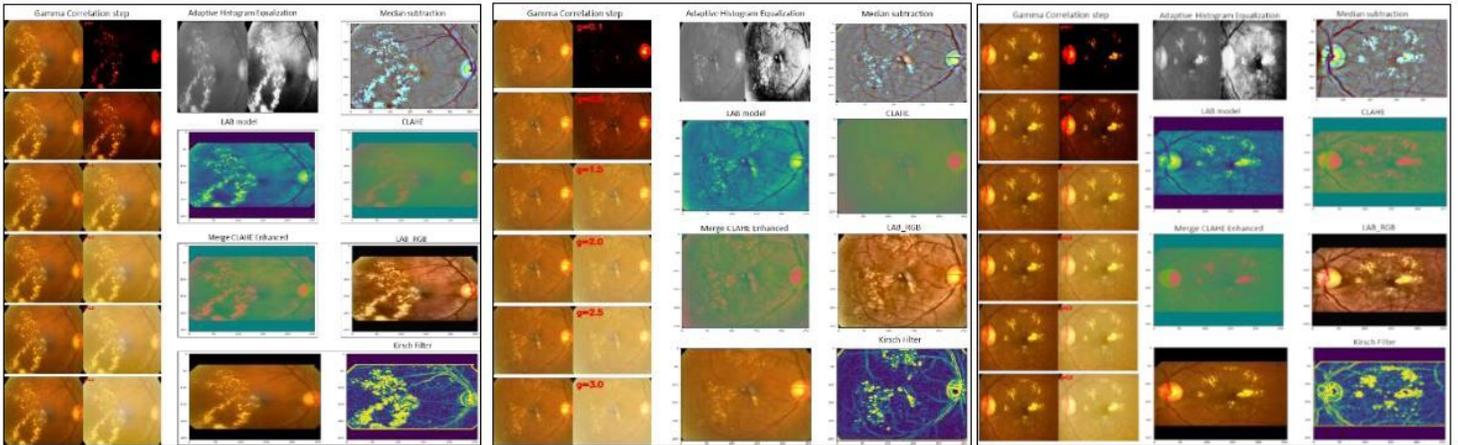


Figure 2

Preprocessing algorithm results

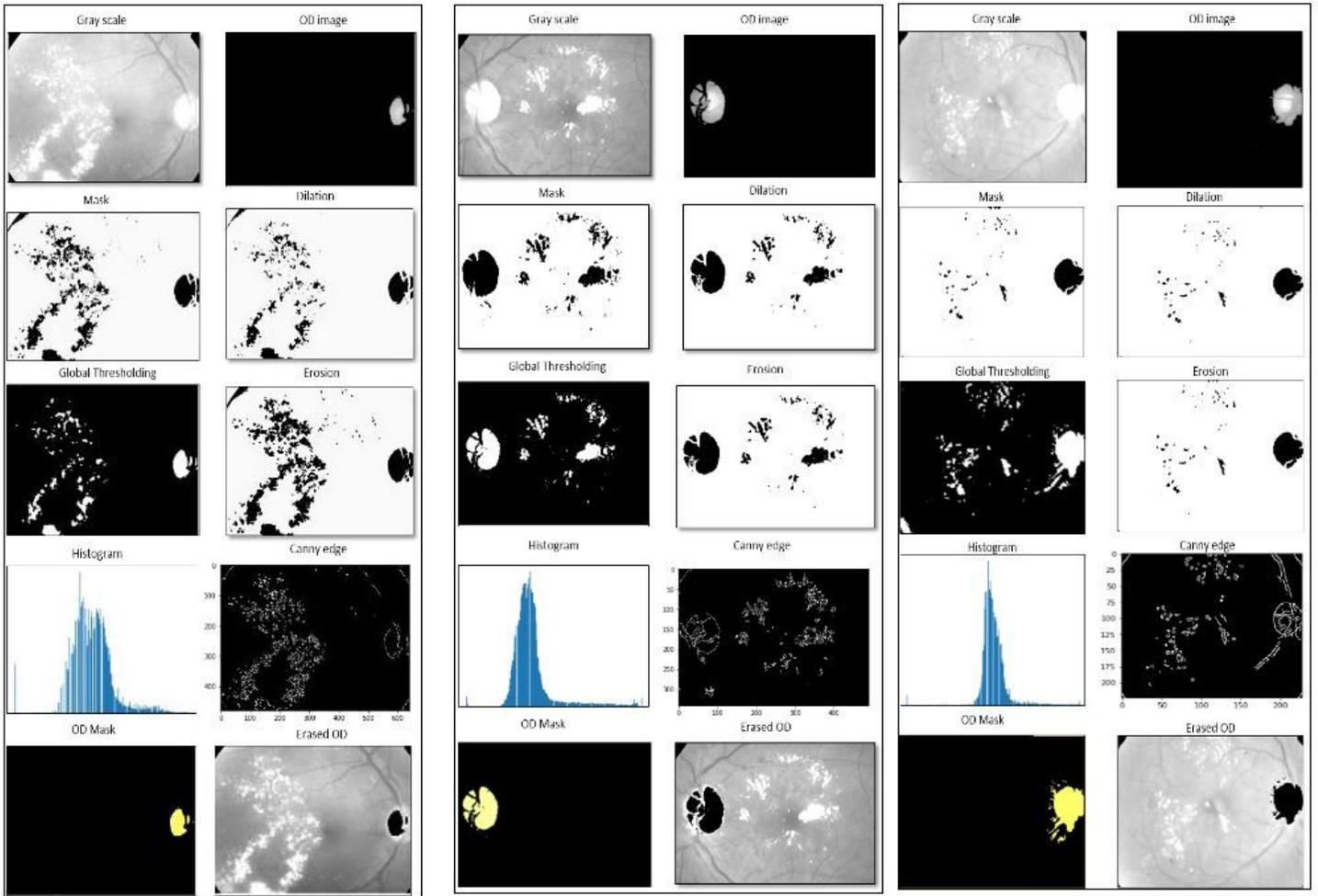


Figure 3

Optic Disk Removal detection

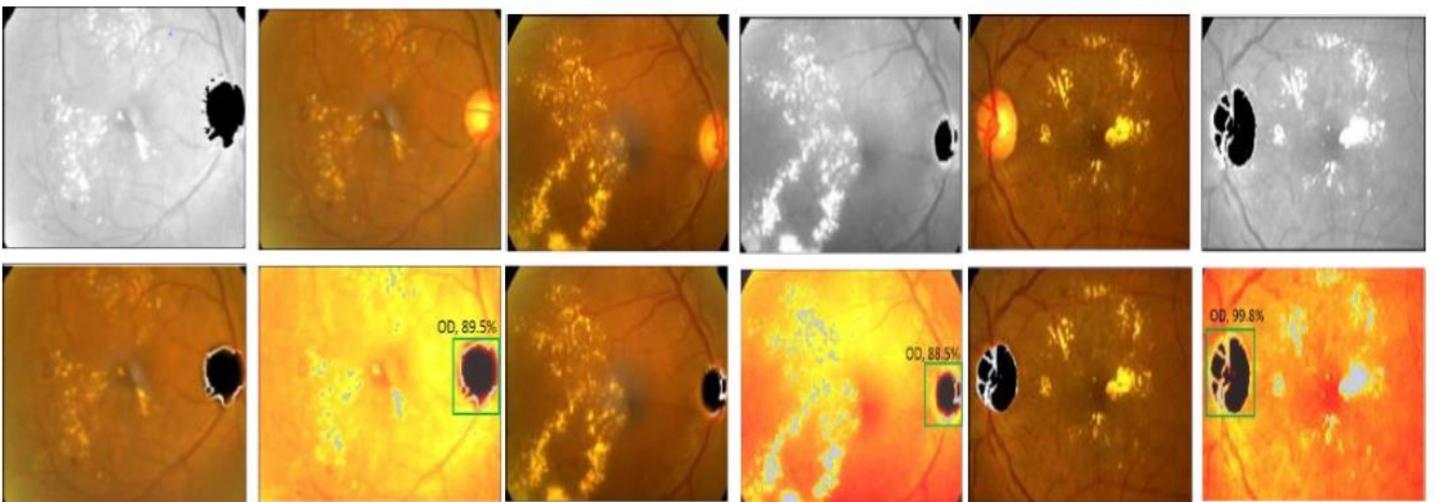


Figure 4

"Please see the Manuscript PDF file for the complete figure caption".

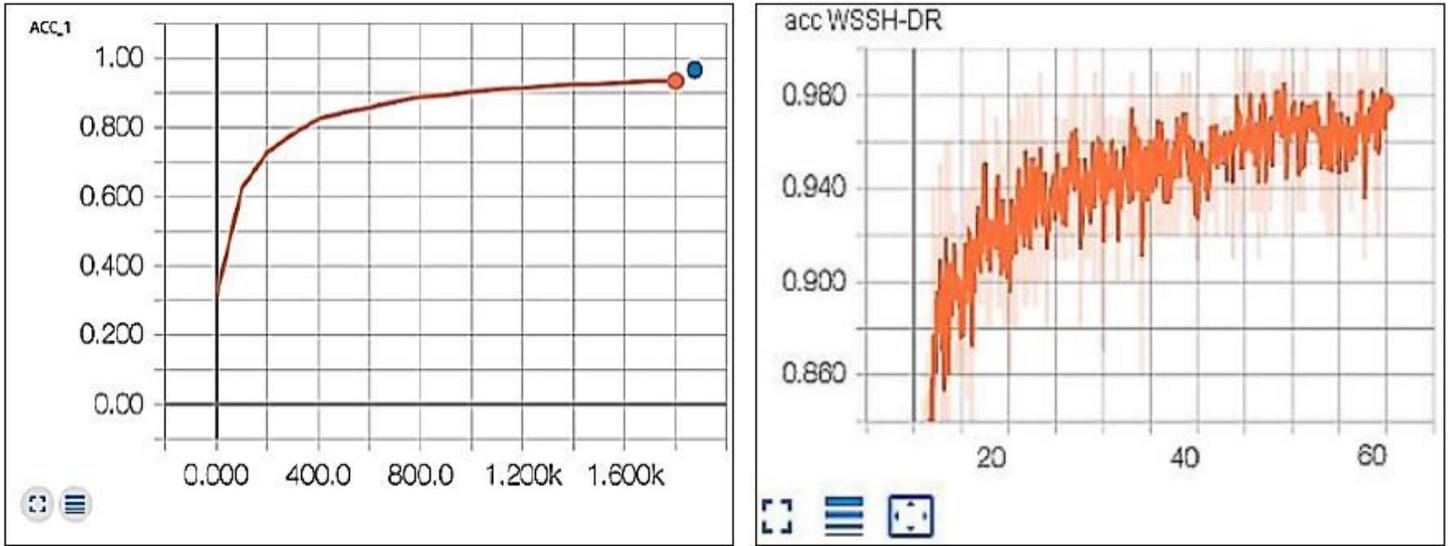


Figure 5

"Please see the Manuscript PDF file for the complete figure caption".