A Deep Learning Approach for Retinal Image Feature Extraction

Mohammed Enamul Hoque[1]*, Kuryati Kipli[1], Tengku Mohd Afendi Zulcaffle[1], Abdulrazak Yahya Saleh Al-Hababi[2], Dayang Azra Awang Mat[1], Rohana Sapawi[1] and Annie Anak Joseph[1]

[1]*Department of Electrical and Electronics Engineering, Faculty of Engineering, Universiti Malaysia Sarawak, 94300 UNIMAS, Kota Samarahan, Sarawak, Malaysia*
[2]*Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, 94300 UNIMAS, Kota Samarahan, Sarawak, Malaysia*

ABSTRACT

Retinal image analysis is crucially important to detect the different kinds of life-threatening cardiovascular and ophthalmic diseases as human retinal microvasculature exhibits remarkable abnormalities responding to these disorders. The high dimensionality and random accumulation of retinal images enlarge the data size, that creating complexity in managing and understating the retinal image data. Deep Learning (DL) has been introduced to deal with this big data challenge by developing intelligent tools. Convolutional Neural Network (CNN), a DL approach, has been designed to extract hierarchical image features with more abstraction. To assist the ophthalmologist in eye screening and ophthalmic disease diagnosis, CNN is being explored to create automatic systems for microvascular pattern analysis, feature extraction, and quantification of retinal images. Extraction of the true vessel of retinal microvasculature is significant for further analysis, such as vessel diameter and bifurcation angle quantification. This study proposes a retinal image feature, true vessel segments extraction approach exploiting the Faster RCNN. The fundamental Image Processing principles have been employed for pre-processing the retinal image data. A combined database assembling image data from different publicly available databases have been used to train, test, and evaluate this proposed method. This proposed method has obtained 92.81% sensitivity and 63.34 positive predictive value in extracting true vessel segments from the top first tier of colour retinal images. It is expected to

Mohammed Enamul Hoque, Kuryati Kipli, Tengku Mohd Afendi Zulcaffle, Abdulrazak Yahya Saleh Al-Hababi,
Dayang Azra Awang Mat, Rohana Sapawi and Annie Anak Joseph

integrate this method into ophthalmic diagnostic tools with further evaluation and validation by analysing the performance.

Keywords: Cardiovascular disease, convolutional neural network, deep learning, feature extraction, retinal imaging

INTRODUCTION

Medical imaging offers the way of visual inspections of diseases state. Retinal image analysis is recognised as a significant part of the medical imaging discipline as some of the severe cardiovascular diseases such as Diabetic Retinopathy (DR), Hypertensive Retinopathy (HR), and Ischemic Stroke (IS) can be detected by analysing the degradation of retinal microvasculature in a non-invasive manner (Abbasi-sureshjani et al., 2016; James, 2000; Witt et al., 2006). Researches revealed that some distinct funduscopic disorders such as arteriovenous (AV) nicking, exudates, Cotton Wool Spots (CWS), vessel widening, microaneurysm, changes in bifurcation angles, and focal arteriolar narrowing in the retina are found as closely associative to the above-mentioned cardiovascular diseases though the different vascular risk factors and blood pressure are in control (De Silva et al., 2011; Henderson et al., 2011). Ong et al. (2013) demonstrated that the risk of stroke and hypertensive retinopathy are optimally assistive. However, Baker et al. (2008) and Wang et al. (2011) investigated that even this can be the cause of stroke fatality in people who are not suffering from the stroke risk factors. Therefore, deviations in the retinal artery and vein diameter are strongly recommended as the great cause of stroke (Kipli et al., 2018).

Image Processing (IP) techniques are being exploited to extract the human retina's qualitative and quantitative image features. However, most of the developed IP-based techniques are not fully automated and time-consuming, which are still considered the limitations of retinal image analysis. In biomedical imaging such as AI-based radiology, IP is excitingly contributing as the radiological diagnosis depends on different imaging modalities. As ophthalmology has a salient similarity with radiology, AI is being explored to develop new methods to assist ophthalmic practitioners in predicting fatal cardiovascular disease and other notable threatening events for vision loss. However, random acquisition of retinal image data and its high dimensionality create a heap of retinal data. This massive data accumulation is throwing the data management challenge to ophthalmologists. DL has been introduced to develop intelligent tools by integrating various task-driven AI algorithms to manage these tremendous-sized data in a more acceptable, safe, and efficient way.

The Artificial Neural Network (ANN) technique is employed to construct Deep Neural Network (DNN) exploiting multiple layers that analyse the image feature hierarchy from higher to lower level vice-versa (Goodfellow et al., 2016; Buduma & Locascio, 2017). The multi-layered DL structures process nonlinear data to analyse and classify many data patterns, extract and classify both supervised and unsupervised data features. The small

units of ANN, artificial neuron, produce real-valued activations that have been utilised to construct DNN forming parameterized functions such as Rectified Linear Units (ReLUs), sigmoid, Tangent Hyperbolic (tanh), and softmax (Abadi et al., 2016; Ghesu et al., 2016; Schmidt-Erfurth et al., 2018). The integrated artificial neurons in each layer of DNN are needed to train for defraying information with high-level representation and more abstraction from the network's first to the last layers. Convolutional Neural Network (CNN) is a form of feed-forward DNN consisting of convolution, pooling, and fully connected layers developed primarily to deal with image data. Convolutional Neural Network employs backpropagation to learn complex image features hierarchies and patterns automatically and adaptively.

The AI applications based on CNN are being developed extensively for retinal image analysis, especially for disease state gradation, microvasculature segmentation, and feature extraction. There are two types of recently developed DL algorithms for ophthalmic abnormality detection such as image-based and lesion-based. The lesion-based algorithms are trained with previously known features such as exudates, haemorrhages, and microaneurysms. Image-based algorithms are also known as black-box algorithms, and this kind of system is trained with manually graded retinal images that generate an output indicating the disease state (Fenner et al., 2018). To classify the Age-related Macular Degeneration (AMD), Grassmann et al. (2018) developed a DL algorithm exploiting six different CNN models such as Visual Geometry Group (VGG), Inception-V3, AlexNet, ResNet V-2, ResNet, and GoogLeNet that obtained 94.30% accuracy and 84.20% sensitivity on Cooperative Health Research on the Region of Augsburg (KORA) data set. Niemeijer et al. (2007) and Abràmoff et al. (2016) developed a CNN algorithm to detect DR signs, exudates, haemorrhages, and neovascularisation, combining VGG and AlexNet that recorded 90.7% accuracy, 82.7% sensitivity, and 96.8% sensitivity, 87% specificity, 0.98 AUC respectively. Gargeya & Leng (2017) and Pratt et al. (2016) proposed a customised CNN model, Gulshan et al. (2016) used Inception-V3 for DR detection, and Ting et al. (2017) developed a CNN model to detect referable and vision-threatening DR while Takahashi et al. (2017) modified GoogLeNet for DR grading. The proposed model of Ting et al. (2017) showed 90.5%, 100%, 96.4%, 93.2% sensitivity and 91.6%, 91.1%, 87.2%, 88.7% specificity for referable DR, vision-threatening DR, glaucoma and AMD detection respectively. The performance of Gargeya & Leng (2017) was recorded as 94% sensitive and 98% specific to DR detection. The work of Ting et al. (2017) seems inconsistent for the implementation in ophthalmic tools as their data set was not graded by the experts. Their work did not consider the great DR signs such as microaneurysms, haemorrhages to analyse and identification of macular oedema was poor, which are the strong limitations of the model of Ting et al. (2017).

In retinal image analysis, segmentation is an important step as the segmented image is further utilised for qualitative and quantitative feature detection. Melinsca et al. (2015)

Mohammed Enamul Hoque, Kuryati Kipli, Tengku Mohd Afendi Zulcaffle, Abdulrazak Yahya Saleh Al-Hababi,
Dayang Azra Awang Mat, Rohana Sapawi and Annie Anak Joseph

and Zhu et al. (2017) developed deep max-pooling CNN and Extreme Learning Machine (ELM) based retinal image segmentation algorithms that obtained 94.66% and 96.00% accuracy, respectively. The work of Zhu et al. (2017) has been evaluated on Retinal Images for Screening (RIS) and found as time effective. For semantic segmentation of retinal image Dense U-net, a DL approach has been introduced. Wang et al. (2019) proposed a Dense U-net model employing image patch-based technique for segmentation that obtained 0.9511 and 0.9538 accuracy, 0.7986 and 0.7914 sensitivity, and 0.9736 and 0.9722 specificity for DRIVE and STARE databases, respectively. Wang et al. (2019) model used a sequential reconstruction strategy to reconstruct the segmented patches at the output end. An optimized deep CNNN approach had been introduced by Badawi & Fraz (2019) for AV classification and obtained the best accuracy, 98%, for the AVRDV dataset. Exploiting Recurrent Residual CNN (RRCNN) and Recurrent CNN (RCNN) Alom et al. (2018) developed U-Net-based semantic segmentation algorithms, while Oliveira et al. (2018) developed a CNN model for data augmentation and prediction. Both algorithms of Badawi & Fraz (2019) and Oliveira et al. (2018) obtained the best result for the STARE database. However, the sensitivity of Oliveira et al. (2018) has slightly deviated while performing cross-training on STARE and DRIVE datasets. In the vessel segmentation model of Wang et al. (2015), CNN and Random Forest (RF) have been ensembled where CNN was dedicated to detecting hierarchical features, and RF contributed as a classifier. Wang et al. (2015) evaluated the DRIVE and STARE database and obtained 0.9767 and 0.9813 accuracy, 0.8173 and 0.8104 sensitivity, and 0.9733 and 0.9791 specificities for both databases, respectively. Maji et al. (2015) combined Denoising Auto-Encoder (DAE) and RF to develop a hybrid DL model for vessel detection and showed 93.27% accuracy. Guo et al. (2019), Mo and Zhang (2017) and Yan et al. (2019) proposed supervised CNN models, and Lahiri et al. (2016) proposed an unsupervised Deep Neural Ensemble Network. The work of Guo et al. (2019) and Lahiri et al. (2016) obtained 95.60% and 95.33% accuracy on the DRIVE database, respectively, while the work of Yan et al. (2019) showed the best outcome, 0.9638 accuracies, 0.7735 sensitivity, and 0.9857 specificities for STARE database. Though the work of Mo & Zhang (2017) showed robustness in segmentation and faster processing speed, the algorithm showed slightly lower accuracy for a cross-training on CHASE DB1 and DRIVE database, respectively. It is assumed that the wider arteriolar orientation, poor vessel contrast, and non-uniform background illumination of the CHASE DB1 image can cause a deviation in the accuracy.

Retinal Image feature identification is crucially important for visual analysis of the impairment in the microvascular structure of the human retina. In order to decide on cardiovascular disease or other ophthalmic abnormalities from retinal image analysis, diagnostic methods must have consistency in detecting the interesting features such as CWS, haemorrhages, exudates, vessel widening, and microaneurysm. Most of the retinal

image feature detection algorithms based on DL have been developed to detect retinal lesions. To detect haemorrhages, Van Grinsven et al. (2016) trained their 5-layered CNN model with both selective and non-selective retinal image samples. Van Grinsven et al. (2016) evaluated their Selective Sampling CNN (SeSCNN) and Non-Selective Sampling CNN (NSeSCNN) model's performance on Kaggle and Messidor databases for different false positive values. It was recorded that the SeSCNN performed better than NSeSCNN.

The developed Neural Network (NN) architecture combining Logistic Regression (LR) with Radial Basis Function (RBF) of García et al. (2009a) obtained 88.1% accuracy, 70.4% positive predictive value, and 100% sensitivity in retinal hard exudate detection. The image-based exudate detection of Osareh et al. (2009) obtained 96% sensitivity, and 94.6% positive predictive value, and this performance is better than their lesion-based detector. Tan et al. (2017) evaluated their 10-layered CNN model for exudates, microaneurysms, and haemorrhages detection on CLEOPETRA that showed 87.58% sensitivity for exudate detection and 62.37% sensitivity for haemorrhages and microaneurysm detection. The Fuzzy C-means and ANN-based DL model of Osareh et al. (2003) was 93% sensitive to exudate detection and 95% sensitive to detect lesions in retinal images. To detect both image-based and lesion-based hard exudates, Van Grinsven et al. (2016) investigated Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), RBF, and NN classifier.

For retinal image feature detection cross-entropy function, regularisation for MLP and RBF had been used in the investigation of Van Grinsven et al. (2016). Among these models, MLP obtained better results, 97.01% accuracy, 100% sensitivity, and 92.59% positive predictive value, compared to SVM, RBF, and NN classifier. The training of MLP was more complex than RBF. It could be happened due to the slow convergence as the effects of different weights and the presence of nearly flat regions in the error function of MLP. As the legacy of Van Grinsven et al. (2016), García et al. (2010) investigated Majority Voting (MV) schema (MV), SVM, RBF, and MLP for microaneurysm and haemorrhages detection. According to the investigation of García et al. (2010), MV and RBF can be recognised as successful feature detectors. However, training these two detectors is expensive, and between these two, RBF can be considered the best feature detector.

It is important to comprehend the suitable AI methods before implementation to obtain the expected results as implementing these methods is complex, and we need to train the methods recurrently. Supervised, unsupervised, and reinforcement learning are the available techniques for training the AI method. In terms of data processing proficiency, SVM is considered the most suitable and popular supervised learning algorithm compared to other existing algorithms such as ANN, K-nearest Neighbour (KNN), Naïve Bayes classifier, Decision Tree, Fuzzy Logic, and Random Forest. On the other hand, clustering algorithms and association rules are suitable for noisy and low-quality data processing that are the most used unsupervised learning algorithm to develop DL models for medical data processing.

Mohammed Enamul Hoque, Kuryati Kipli, Tengku Mohd Afendi Zulcaffle, Abdulrazak Yahya Saleh Al-Hababi, Dayang Azra Awang Mat, Rohana Sapawi and Annie Anak Joseph

The selection of DL methods to develop AI machines for retinal image analysis must be empirical regarding the retinal image data characteristics, number of parameters for training the algorithms, and length of the training period. For medical image analysis tools, ANN is suitable to implement in DL compared to the logistic regression because of the non-linear operation ability of ANN in high dimensional image data processing, dealing with noisy data, and securing higher prediction accuracy. Recurrent Neural Network (RNN) and CNN have been more efficient in cardiovascular disease state prediction, and abnormal feature detection by analysing retinal and brain MR images (Krittanawong et al., 2017).

The development of appropriate DL-based retinal image segmentation and feature extraction methods is crucially important to understand the complex hierarchical microvasculature of the human retina that can maximise the retinal abnormalities detection result. The highly varied retinal image dataset due to the poor acquisition method can affect the performance of DL models. To avoid this complication, appropriate annotation of characteristics features is significant while preparing the training data. According to the literature study, no DL-based retinal image feature detection method has been reported that can segment retinal microvasculature and extract qualitative and quantitative image features simultaneously. The existing retinal image feature detection algorithms had been developed to detect retinal lesions such as hemorrhages, microaneurysms, and exudates as a single feature. Various CNN models such as recurrent CNN, deep CNN utilise different non-linear functions that obtained comparatively better feature extraction results than the logistic regression approach.

The existing retinal image feature detection methods of   García et al. (2009a), García et al. (2010), García et al. (2009b), and Osareh et al. (2009) obtained the best detection performance for image-based criterion compared to the lesion-based criterion. It can be happen due to the lesion's low pixel intensity as the lesion is annotated alone from the whole image. Moreover, pixel-wise ground truth estimation in the lesion-based criterion is cumbersome that can degrade the training accuracy and consequently obtain poor feature detection performance. Though the CNN-based feature detection model showed better performance than the statistical analysis, there are some limitations of this newly developed technology in terms of overfitting training data and lengthy processing due to the utilisation of more parameters. Reducing hidden layers from the CNN architecture and increasing training data can potentially solve the data overfitting problem.

All the recently introduced DL models for feature detection have been investigated to select the suitable CNN architecture for this study. This paper proposes a DL approach for retinal image feature extraction employing the Faster-RCNN method. This proposed method has been designed to extract true vessel segments as retinal image features from the multiple locations of the first top tier of colour retinal images. To train, test, and evaluate the performance of this proposed method, colour retinal images from different publicly

available databases have been extracted. The development of this DL approach is aimed to integrate into our previous work the IP algorithm for retinal vessel diameter quantification of different interesting locations of retinal images for creating a fully automated vessel diameter quantification method (Hoque et al., 2019; Hoque et al., 2018; Kipli et al., 2020). The applied methodology, including training and testing for the development of this proposed DL method, is explained in detail in the following section. The obtained results are briefly described, and the evaluation and critical analysis of the performance are also added consecutively. Figure 1 illustrates the graphical representation of the Faster RCNN model.
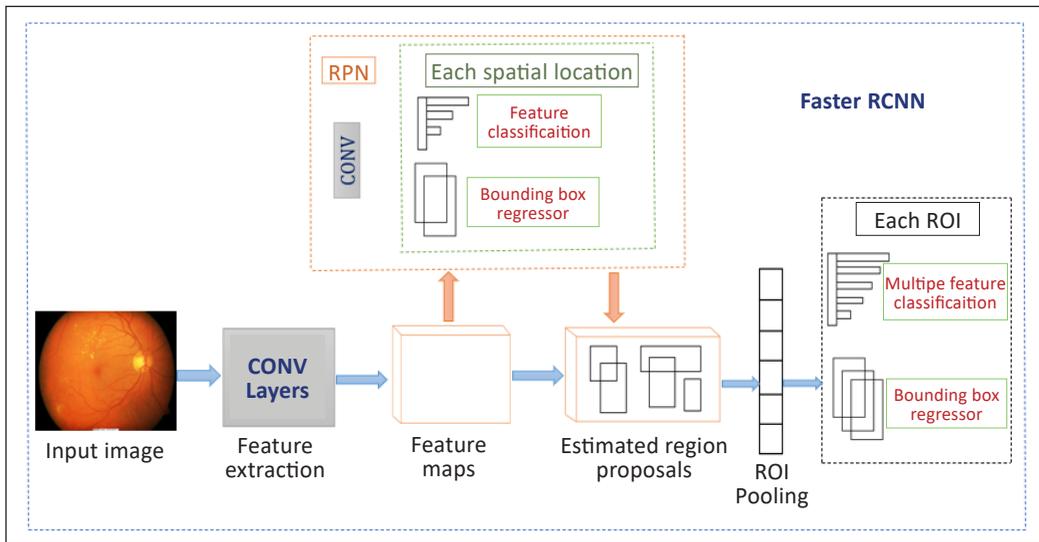

Figure 1. Graphical representation of Faster RCNN model

MATERIALS AND METHODS

The existing IP techniques, and the recently developed DL method, Faster-RCNN, were combined to develop this proposed method. A collection of normal, healthy, and abnormal, pathological, images from the different databases such as High-Resolution Fundus Image Database (HRFID), Digital Retinal Images for Vessel Extraction (DRIVE), Structured Analysis of the Retina (STARE), and MESSIDOR database were used in this research to train, test and validate the proposed DL model. Four hundred fifty images had been used for training and testing this algorithm. These 450 images were divided to form training and testing datasets where the training dataset contained 270 images, and the testing dataset contained 180 images. Both the training and testing dataset consisted of normal and pathological retinal images. Figure 2 describes the involved steps in this proposed method.

At the preprocessing stage, Contrast Limited Adaptive Histogram Equalization (CLAHE) was employed to enhance the quality of the input image. After that, the enhanced image was resized following the DL algorithm's requirement, [224 224 3], and estimated
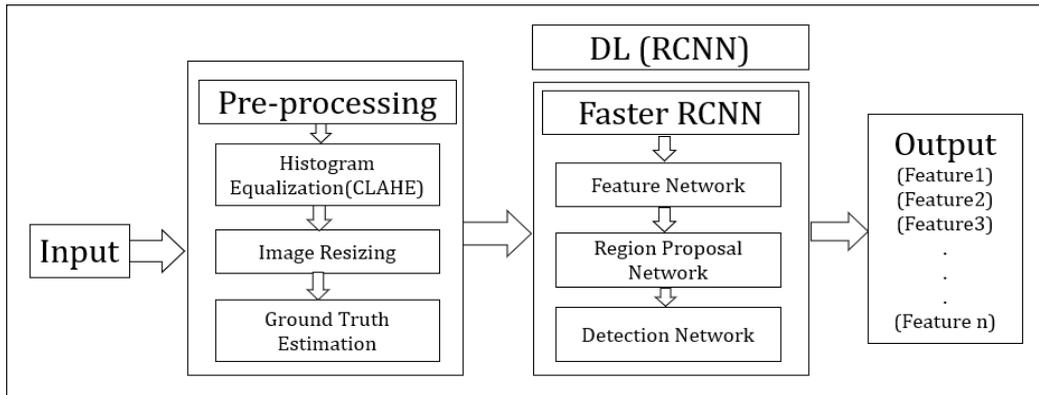
Figure 2. Proposed method

the GroundTruth from that image to extract the features, true vessel, using image labeller apps of MATLAB.

The main vessels of the top first tier were considered Region of Interest (ROI) of all the images. The multiple vessel segments of that ROI were annotated as the ground truth by using bounding boxes. These images with annotated ground truth were augmented to use as training data for securing better performance. A previously developed Faster-RCNN had been trained in this project to detect the expected features of retinal images. Figure 3 demonstrates the graphical representation of the proposed method.
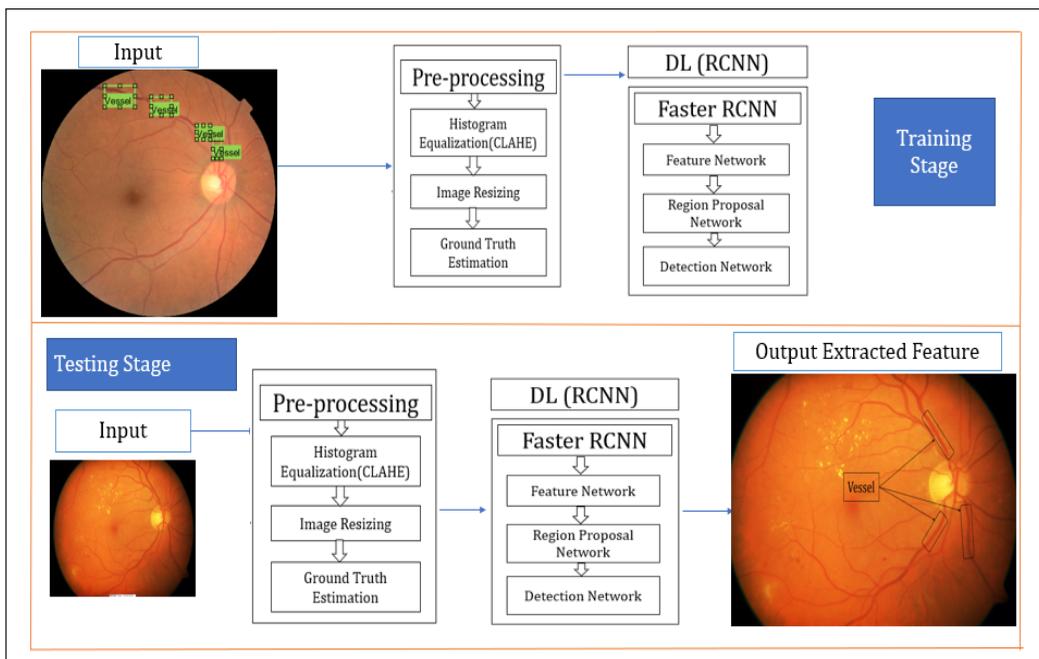

Figure 3. Graphical representation of the proposed method

DL Network (Region-based Convolutional Neural Network, RCNN)

Faster RCNN is the updated version of RCNN developed to detect multiple objects from a single image. To construct DL architecture for this research, Faster RCNN had been used. Faster RCNN uses a separate network, Region Proposal Network (RPN), to detect the region proposals from the feature maps provided by the convolutional layer. The RoI pooling layer reshapes the detected region proposals. It passes these to the fully connected layer, softmax, and linear regression layer that classifies the features and predicts the offset values for the bounding boxes. The Faster RCNN model is the composition of three different modules that are fully convolutional network, feature network, to generate feature maps from the input image, RPN to generate bounding boxes that contain different features or objects extracted from feature maps, and Detection Network that takes input from both RPN and feature network to detect the expected features. This entire system for feature detection is a single unified network.

Feature Network (Convolutional Layer)

The tasks were performed in three different stages to develop the proposed model in this research. First, a Fast R-CNN had been created, and further an RPN was added, and finally, the RPN and detection network was trained. To perform the convolutional operation, a Fast R-CNN, ResNet-50 (Residual Network-50), was trained to produce the feature maps further fed into RPN to generate the region proposals. ResNet-50 is a model of CNN that consists of 50 layers of different fundamental operations of CNN methodology such as convolution, pooling, activation, and fully connected layers. The input size of the images that is suitable for the network is *224×224×3.* The kernel sizes for the initial convolution and max-pooling of the ResNet were considered as *7×7* and *3×3,* respectively, with stride 2 for both convolution and max pooling. The network architecture starts after initial steps that consist of three different residual blocks, and each of the blocks contains three convolution layers performed with 64, 64, and 256 Kernel, respectively. In order to design the deeper RestNet, the bottleneck architecture was used instead of using all 3×3 convolution layers as a standard residual block. In the bottleneck architecture of the residual block, three convolution layers, *1×1, 3×3,* and *1×1* convolution, is stacked one over another for each residual function. 1×1 is used to reduce the input dimension before performing the 3×3 convolution, and then another *1×1* convolution layer is used to preserve the original shape. The size of the stride, stride 2, reduces the height and width of the input to half and doubles the channel width for the following stages. There are 4, 6, and 3 residual blocks in stages 2,3, and 4, respectively. The network has an average pooling layer followed by a fully connected layer as the final layer.

Region-based Proposal Network (RPN)

RPN contains 3 convolution layers that take the feature map as input are generated from the feature network and produces region proposals with bounding boxes containing the potential features. In order to generate the region proposal, a sliding window of *3×3* size kernel was used for each location of the feature map and *9 (K=9)* anchor boxes with three different scales of 128, 256, and 512, and 3 aspect ratios of 1:1, 1:2, and 2:1 were used for each location. In addition, a box-class layer, cls layer, results in 2 K scores that the anchor boxes contain an object or not, and a box regression layer, reg layer, results in 4K for the coordinates of K boxes. Figure 4 illustrates the operational block diagram of RPN.
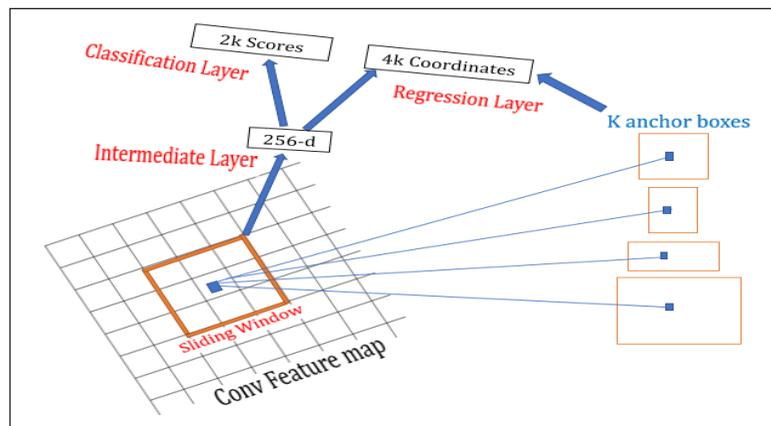

Figure 4. Operational block diagram of RPN

Detection Network

Fast RCNN had been adopted to construct the detection network where two sibling layers get input from the feature network and RPN. The output proposals from the pooling layer were fed to the classification layer, softmax classification, and linear regression layer, bounding box regression layer of the detection network as a batch. The softmax classification layer classifies the RoI pooling layer output, RoI bounding box, by computing the probability distribution, *p= (p_0.........p_k)* over the *K+1* class throughout the fully connected layer. The bounding box regression layer is responsible for predicting the bounding boxes, by computing the regression offsets for each object class. The regression layer generates 4 bounding box offsets that can be explained as $t_i^k = (t_x^k, t_y^k, t_w^k, t_h^k)$ where *i=x,y,w,h,* and *(x,y)* denotes the coordinates of the top left corner of the bounding box and *w* and *h* denote the width and height of the bounding box respectively.

Training DL Network

To train the proposed DL method stochastic gradient descent training algorithm had been used with an initial learning rate of 1×10^{-3} Twenty epoch and 400 iterations for each

epoch were performed to accomplish the whole training process. The mini-batch size was set to 3. As the Faster RCNN had been used in this research, the standard cost function was used to calculate the training loss. The whole training procedures for the entire network are briefly described in the following sections.

Training RPN

The number of anchors was reduced to train the RPN. Initially, the anchors were assigned by a binary class label threshold. Further, the values of Intersection over Union (IoU) of primary and predicted bounding boxes of anchors were considered to reduce the number of anchors. The IoU measures the overlap between the primary and the predicted boundary boxes. If the value of IoU is greater than 0.4, then the anchor is assigned as a positive label while the anchor is assigned as a negative label if the value of IoU is lower than 0.3. The rest anchors that do not satisfy this condition were not considered for the RPN training process. Thus, the multi-task training loss for RPN combines the losses in classification and regression operation calculated by Equation 1.

$$L(\{P_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p^* L_{reg}(t_i, t_i^*) \qquad (1)$$

Here λ is the balancing parameter to balance the weights of L_{cls} and L_{eg} roughly. L_{cls} is the classification loss which is log loss over two different classes of anchor, object, or not an object. The term i represents the index number of mini-batch, where p_i denotes the output label from the classification layer for i^{th} anchor and p_i^* denotes the ground-truth label. The label of both p_i and p_i^* is binary, 1 or 0 where 1 indicates a boundary box is an object, and 0 indicates the boundary box is not an object. L_{reg} is the regression loss considered for calculation if the anchor is an object, where t_i^* is the regression target, ground-truth coordinate for the regression layer, and t_i outputs of the learned regression layer.

After the sampling of anchors, the region proposals with that sampled anchors were fed to the Region of Interest (RoI) pooling. The RoI pooling layer employs max pooling to extract the fixed-sized feature maps with the size of (N, 7, 7, 512) for each proposal. Here N, is the number of Region proposals from RPN. Figure 5 illustrates the operation of the RoI pooling layer.

Training Detection Network

The threshold values of IoU were set to 0.1 to 0.5 to label the ROIs as background To train the detection network. The RoI was labelled as foreground when the IoU is above 0.5. Bounding box targets, ground truth boxes were also generated following the same approach of RPN. The multitask loss L was calculated for each labelled RoI to estimate
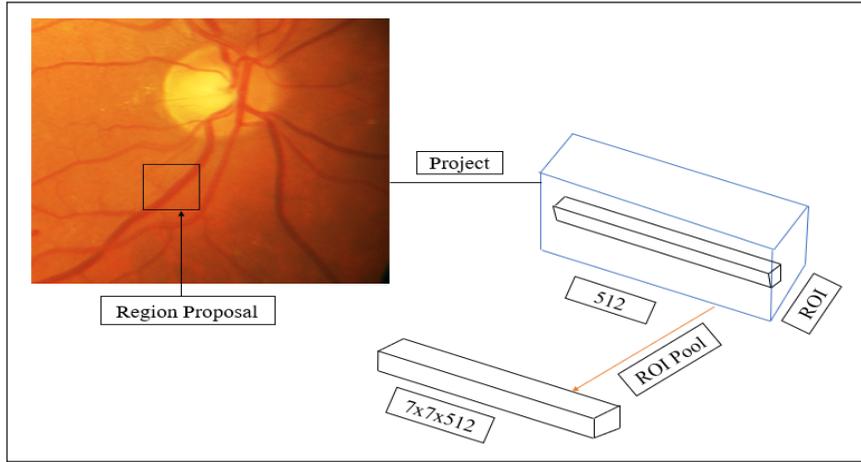
Figure 5. Operation of RoI pooling layer

the loss during training the detection network. The loss in the classification layer had been calculated by a cross-entropy or log loss, while the smooth L_1 loss had calculated regression loss for bounding box regression. The multitask loss L was calculated by Equation 2,

$$L(p, u, t^u, v) = L_{cls}(p, u) + \lambda[u \geq 1]L_{loc}(t^u, v) \tag{2}$$

Here $L_{cls}(p,u)$ is log loss and $L_{cls} = -logp_u$ where u is the true feature class in the bounding box. $L_{loc}(t^u,v)$ is the loss of the regression layer, which is calculated by smooth L_1 loss. The balancing parameter is denoted by λ, u represents the class, and v represents the bounding box regression targets for class u. The term *[u≥1]* determines the true bounding box regression targets. When $u=0$, the regression loss is considered 0 because there were no ground-truth boxes for the background. The bounding box regression loss, $L_{loc}(t^u,v)$, was computed by Equation 3,

$$L_{loc}(t^u, v) = \sum_{i \in \{x,y,w,h\}} smoothL_1(t_i^u - v_i) \tag{3}$$

Where smooth L_1 was determined by Equation 4,

$$SmoothL_1(x) = \begin{cases} 0.5x^2 & if|x|<1 \\ |x|-0.5 & otherwise, \end{cases} \tag{4}$$

Smooth loss L_1 was used because it is less sensitive to outliers, and $\lambda=1$ was set to balance the two losses. Finally, the ground truth regression target v_i was normalised for the mean as zero and unit variance.

The 4 Step alternating training method (Ren et al., 2017) was further employed to the RPN and Detection Network simultaneously to share the weights of convolution layers of the two networks between themselves. In the first step of this training method, the RPN

was trained as described in the section, **Training RPN**. Then, an ImageNet pre-trained model was used to initialise the RPN and fine-tuned end-to-end for the region proposal task in step 1. Next, the proposals generated in step 1 were used by Fast RCNN to train the detection network in step 2, which the ImageNet pre-trained model also initialised. Although, as in these training steps, the two networks do not share convolution layers, the detection network was used to initialise to train RPN in the third step. In this step, the shared convolution layers were fixed and fine-tuned the unique layers to RPN. In the final step, step 4, the unique convolution layers of Fast RCNN were also fine-tuned following the fixed shared convolutional layers. In this way, a unified network, Faster RCNN, was formed where both networks share the same convolution layers.

RESULT

The training performance was measured for each iteration of each epoch of the training stage. The training loss and the Root Mean Squared Error (RMSE) after the final epoch were recorded as 0.3446 and 0.17. It was observed that the training performance of the proposed network had increased over time and reached 99.4% training accuracy after 7 training epochs. After the completion of training, the proposed method had been tested on two different datasets, healthy and unhealthy datasets. The healthy dataset had been created extracting images from DRIVE, and the images for the unhealthy dataset were taken from Kaggle. These two unknown datasets consisted of 700 images. The proposed method was able to detect the true vessel from both healthy and unhealthy images. Figure 6 demonstrates the images, healthy and unhealthy, with the true detected vessel. The testing performance of the proposed method for these two unknown datasets was also recorded. Table 1 represents the testing performance of the proposed method for the unknown dataset.
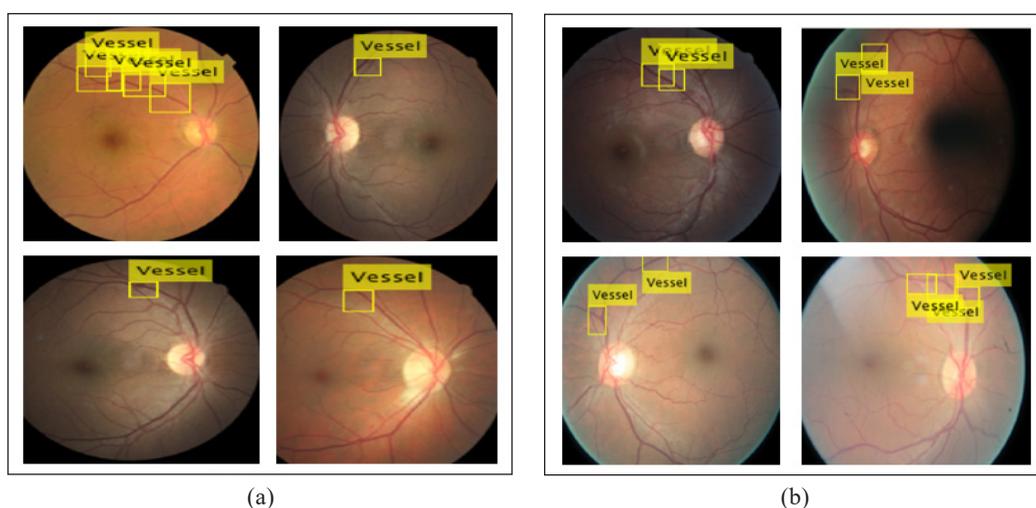

(a) (b)

Figure 6. Test images: (a) healthy and (b) unhealthy with the true detected vessel

In Table 1, columns 1,2 and 3 represent the datasets, Se, and PPV of the proposed method. The network successfully detected the true vessel from the interesting location of those images. Figure 7 shows the retinal images with the true detected vessel from multiple locations.

Table 1
Testing performance of the proposed method for the unknown dataset

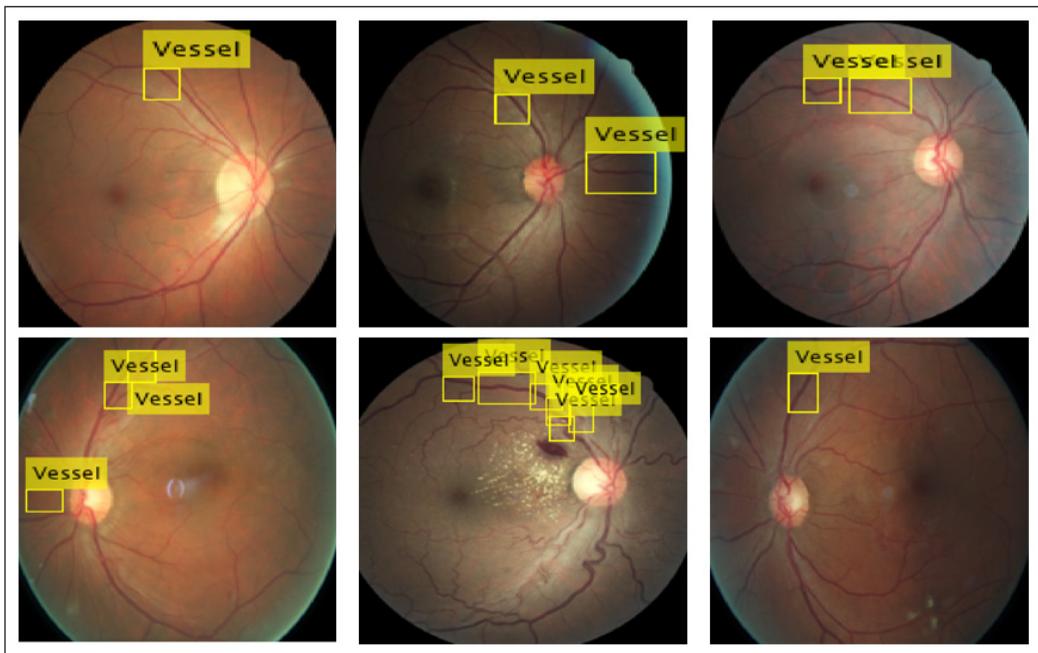| Dataset	Se (Mean value, %)	PPV (Mean value, %)
Healthy Dataset	90.53	61.22
Unhealthy Dataset	88.16	60.67

Figure 7. Retinal Images with the true detected vessel from multiple locations

Sensitivity (Se) and Positive Predictive Value(PPV) had been calculated to evaluate the performance of the proposed DL method. In addition, the testing dataset that was pre-processed for use in this research had been used for testing and calculating the performance measurement matrices. Table 2 illustrate the performance comparison between the recently developed methods and the proposed method.

In Table 2, columns 1,2,3 and 4 representing the author's information, applied method, features extracted, and result, respectively. According to column 3 of Table 2, it is seen that most of the methods were developed to detect different features such as haemorrhages, exudates, microaneurysms, and the proposed method focused on true vessel detection.

Table 2
Comparison of proposed method performance and existing methods

Author	Method Applied	Extracted Feature	Result (Mean Value %)	
			Se	PPV
(Tan et al., 2017)	CNN (10 Layered)	Exudates	87.58	
		Haemorrhages Microaneurysm	62.57	
(García et al., 2010)	Logistic Regression (MLP, RBF, SVM and combining these three NNs using an MV schema)	Haemorrhages, Microaneurysm		
		Image-based	100	56
		Lesion-based	86.01	51.99
		Retinal Exudates (Image-Based)	96.00	94.60
(García, Sánchez, Poza, et al., 2009)	Logistic Regression RBFNN	Retinal Hard Exudates (Lesion-Based)	92.10	86.40
		Retinal Hard Exudates (Image-Based)	100	70.4
(García, Sánchez, López, et al., 2009)	Logistic Regression MLP, RBFNN, SVM	Retinal Hard Exudates (Lesion-Based)		
		MLP	88.14	80.72
		RBF	88.49	77.41
		SVM	87.61	83.51
		Retinal Hard Exudates (Image-Based)		
		MLP	100	92.59
		RBF	100	81.48
		SVM	100	77.78
(Van Grinsven et al., 2016)	SeSCNN NSeSCNN	Haemorrhages, SeSCNN (FP*1)	78.60	
		Haemorrhages, SeSCNN(FP0.1)	51.10	
		Haemorrhages NSeSCNN(FP1)	75.30	
		Haemorrhages, NSeSCNN (FP0.1)	31.60	
Proposed Method	**Faster RCNN**	**True vessel**	**92.81**	**63.34**

Though all methods are dedicated to detecting different features, their performance is tabulated here as the retinal image feature detector.

DISCUSSION AND CONCLUSION

Faster RCNN is primarily introduced for multiple object detection from a single image. In this proposed method, this Faster RCNN has been used as a feature detector dedicated to extracting true vessels from interesting locations of retinal images. As this study is the pre-step of developing an automated retinal vessel diameter quantification algorithm, the ROIs for true vessels from different locations are considered during ground truth estimation following the recommendation of local clinical experts and the manually marked ROIs

of (Al-Diri et al., 2008). It is seen from the performance analysis table, Table 2 that the proposed method obtained 92.81% sensitivity and 62.34% PPV. Furthermore, the proposed method showed better performance as a feature extractor compared to the lesion-based feature extractor of García et al. (2010) and García et al. (2009b), and the exudates detector of Van Grinsven et al. (2016).

Though the proposed method's performance is slightly lower than some of the mentioned feature extractors in Table 2, the result is still comparable and considered for further development. In Figure 5, the unhealthy images contain the diabetic lesion, so the microvascular structure cannot be extracted more precisely. The performance of the proposed method with the unknown datasets was also satisfactory. The proposed method obtained 90.53% and 88.16% Se, and 61.22% and 60.67% PPV for both healthy and unhealthy datasets (Table 1), respectively. It is seen that the performance of the proposed method for anonymous healthy and unhealthy data is slightly lower than the data that were pre-processed initially for training and testing the proposed method. For the healthy dataset, this deviation was happened due to the algorithm's execution with the anonymous data that were not pre-processed initially. The cause of the deviation in the performance of a proposed method for an unhealthy dataset is due to retinopathies and lesions.

Attaining maximum results is challenging due to the highly complex and hierarchical structure of retinal images. The proposed method had been trained with a comparatively small dataset that consists of 270 images containing healthy and pathological signs. As it is suggested to train the DL method with a large dataset, data augmentation had been performed to secure the best training performance. The obtained testing performance of the proposed method showed consistency as the testing dataset contained both normal retinal images and images with abnormality. Data pre-processing is considered one of the most crucial parts of DL algorithm development. Therefore, it is highly expected to use primary data and perform a better operation for data pre-processing such as enhance data quality by histogram equalisation and efficient ground truth estimation for training, testing, and validating the proposed method.

To increase the robustness of the proposed method, utilising a large dataset is highly recommended, and training options should be empirical. It is recommended to configure the training option by setting with more epochs, high Verbose-Frequency to secure the maximum result. Due to the inefficient retinal image acquisition, annotating the background and foreground appears to be a major challenge, and the parameter values of positive and negative overlapping need to be focused on for more accurate feature detection. In order to secure the best detection result, different ranges for positive and negative overlapping have been explored in this study. It was investigated that the range [0 0.3] for negative overlapping and [0.4 1] for positive overlapping is the best suit for training the proposed algorithm. As the retinal vessel detection is considered a small object and the difference

between the background and true vessel pixel values is less, the positive overlap range needs to be larger to distinguish the expected objects during training the system. Though the obtained result is satisfactory, this study has some limitations, such as lack of primary retinal image data and highly configured hardware with efficient graphics computing units that are crucially important for higher detection accuracy and speed up the training process. To further development this proposed method, it is recommended to use a large dataset of real images containing healthy retinal images and retinal images with the abnormality. It is expected to integrate this proposed method in clinical tools with further development, evaluation, and validation.

ACKNOWLEDGMENT

REFERENCE

Abadi, M., McMahan, H. B., Chu, A., Mironov, I., Zhang, L., Goodfellow, I., & Talwar, K. (2016). Deep learning with differential privacy. *Proceedings of the ACM Conference on Computer and Communications Security*, *24-28-Octo*(Ccs), 308-318. https://doi.org/10.1145/2976749.2978318

Abbasi-sureshjani, M. F. S., Romeny, H., & Sarti, A. (2016). Analysis of vessel connectivities in retinal images by cortically inspired spectral clustering. *Journal of Mathematical Imaging and Vision*, *56*(1), 158-172. https://doi.org/10.1007/s10851-016-0640-1

Abràmoff, M. D., Lou, Y., Erginay, A., Clarida, W., Amelon, R., Folk, J. C., & Niemeijer, M. (2016). Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Investigative Ophthalmology and Visual Science*, *57*(13), 5200-5206. https://doi.org/10.1167/iovs.16-19964

Al-Diri, B., Hunter, A., Steel, D., Habib, M., Hudaib, T., & Berry, S. (2008). A reference data set for retinal vessel profiles. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 2262-2265). IEEE Publishing. https://doi.org/10.1109/IEMBS.2008.4649647

Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M., & Asari, V. K. (2018). *Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation*. ArXiv Publishing.

Badawi, S. A., & Fraz, M. M. (2019). Multiloss function based deep convolutional neural network for segmentation of retinal vasculature into arterioles and venules. *BioMed Research International*, *2019*, Article 4747230. https://doi.org/10.1155/2019/4747230

Baker, M. L., Hand, P. J., Wang, J. J., & Wong, T. Y. (2008). Retinal signs and stroke: Revisiting the link between the eye and brain. *Stroke*, *39*(4), 1371-1379. https://doi.org/10.1161/STROKEAHA.107.496091

Buduma, N., & Locascio, N. (2017). *Fundamentals of deep learning: Designing next-generation machine intelligence algorithm*. O'Reilly Media Inc.

De Silva, D. A., Manzano, J. J. F., Liu, E. Y., Woon, F. P., Wong, W. X., Chang, H. M., Chen, C., Lindley, R. I., Wang, J. J., Mitchell, P., Wong, T. Y., & Wong, M. C. (2011). Retinal microvascular changes and subsequent vascular events after ischemic stroke. *Neurology*, *77*(9), 896-903. https://doi.org/10.1212/WNL.0b013e31822c623b

Fenner, B. J., Wong, R. L. M., Lam, W. C., Tan, G. S. W., & Cheung, G. C. M. (2018). Advances in retinal imaging and applications in diabetic retinopathy screening: A review. *Ophthalmology and Therapy*, *7*(2), 333-346. https://doi.org/10.1007/s40123-018-0153-7

García, M., López, M. I., Álvarez, D., & Hornero, R. (2010). Assessment of four neural network based classifiers to automatically detect red lesions in retinal images. *Medical Engineering and Physics*, *32*(10), 1085-1093. https://doi.org/10.1016/j.medengphy.2010.07.014

García, M., Sánchez, C. I., Poza, J., López, M. I., & Hornero, R. (2009a). Detection of hard exudates in retinal images using a radial basis function classifier. *Annals of Biomedical Engineering*, *37*(7), 1448-1463. https://doi.org/10.1007/s10439-009-9707-0

García, M., Sánchez, C. I., López, M. I., Abásolo, D., & Hornero, R. (2009b). Neural network based detection of hard exudates in retinal images. *Computer Methods and Programs in Biomedicine*, *93*(1), 9-19. https://doi.org/10.1016/j.cmpb.2008.07.006

Gargeya, R., & Leng, T. (2017). Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*, *124*(7), 962-969. https://doi.org/10.1016/j.ophtha.2017.02.008

Ghesu, F. C., Krubasik, E., Georgescu, B., Singh, V., Zheng, Y., Hornegger, J., & Comaniciu, D. (2016). Marginal space deep learning: Efficient architecture for volumetric image parsing. *IEEE Transactions on Medical Imaging*, *35*(5), 1217-1228. https://doi.org/10.1109/TMI.2016.2538802

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Grassmann, F., Mengelkamp, J., Brandl, C., Harsch, S., Zimmermann, M. E., Linkohr, B., Peters, A., Heid, I. M., Palm, C., & Weber, B. H. F. (2018). A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. *Ophthalmology*, *125*(9), 1410-1420. https://doi.org/10.1016/j.ophtha.2018.02.037

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J. L., & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA - Journal of the American Medical Association*, *316*(22), 2402-2410. https://doi.org/10.1001/jama.2016.17216

Guo, S., Wang, K., Kang, H., Zhang, Y., Gao, Y., & Li, T. (2019). BTS-DSN: Deeply supervised neural network with short connections for retinal vessel segmentation. *International Journal of Medical Informatics*, *126*, 105-113. https://doi.org/10.1016/j.ijmedinf.2019.03.015

Henderson, A. D., Bruce, B. B., Newman, N. J., & Biousse, V. (2011). Hypertension-related eye abnormalities and the risk of stroke. *Reviews in Neurological Diseases*, *8*(404), 1-9. https://doi.org/10.3909/rind0274

Hoque, M. E., Kipli, K., Zulcaffle, T. M. A., Mat, D. A. A., Joseph, A., Zamhari, N., Sapawi, R., & Arafat, M. Y. (2019). Segmentation of retinal microvasculature based on iterative self-organizing data analysis technique (ISODATA). In *2019 International UNIMAS STEM 12th Engineering Conference (EnCon)* (pp. 59-64). IEEE Publishing. https://doi.org/10.1109/EnCon.2019.8861259

Hoque, M. E., Kipli, K., Zulcaffle, T. M. A., Sapawi, R., Joseph, A., Abidin, W. A. W. Z., & Sahari, S. K. (2018). Feature extraction method of retinal vessel diameter. In *2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES)* (pp. 279-283). IEEE Publishing. https://doi.org/10.1109/IECBES.2018.8626660

James, M. (2000). Cost effectiveness analysis of screening for sight-threatening diabetic eye disease. *BMJ*, *320*(7250), 1627-1631. https://doi.org/10.1136/bmj.320.7250.1627

Kipli, K., Hoque, M. E., Lim, L. T., Mahmood, M. H., Sahari, S. K., Sapawi, R., Rajaee, N., & Joseph, A. (2018). A review on the extraction of quantitative retinal microvascular image feature. *Computational and Mathematical Methods in Medicine*, *2018*, Article 4019538. https://doi.org/10.1155/2018/4019538

Kipli, K., Hoque, M. E., Lim, L. T., Zulcaffle, T. M. A., Sahari, S. K., & Mahmood, M. H. (2020). Retinal image blood vessel extraction and quantification with Euclidean distance transform approach. *IET Image Processing, 14*(15), 3718-3724.. https://doi.org/10.1049/iet-ipr.2020.0336

Krittanawong, C., Zhang, H. J., Wang, Z., Aydar, M., & Kitai, T. (2017). Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology*, *69*(21), 2657-2664. https://doi.org/10.1016/j.jacc.2017.03.571

Lahiri, A., Roy, A. G., Sheet, D., & Biswas, P. K. (2016). Deep neural ensemble for retinal vessel segmentation in fundus images towards achieving label-free angiography. In *2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 1340-1343). IEEE Publishing. https://doi.org/10.1109/EMBC.2016.7590955

Maji, D., Santara, A., Ghosh, S., Sheet, D., & Mitra, P. (2015). Deep neural network and random forest hybrid architecture for learning to detect retinal vessels in fundus images. In *2015 37th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 3029-3032). IEEE Publishing. https://doi.org/10.1109/EMBC.2015.7319030

Melinsca, M., Prentasic, P., & Loncaric, S. (2015). Retinal vessel segmentation using deep neural networks. In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications (VISAPP-2015)* (pp. 577-582). Science and Technology Publications. https://doi.org/10.5220/0005313005770582

Mo, J., & Zhang, L. (2017). Multi-level deep supervised networks for retinal vessel segmentation. *International Journal of Computer Assisted Radiology and Surgery*, *12*(12), 2181-2193. https://doi.org/10.1007/s11548-017-1619-0

Niemeijer, M., Van Ginneken, B., Russell, S. R., Suttorp-Schulten, M. S. A., & Abràmoff, M. D. (2007). Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Investigative Ophthalmology and Visual Science*, *48*(5), 2260-2267. https://doi.org/10.1167/iovs.06-0996

Oliveira, A., Pereira, S., & Silva, C. A. (2018). Retinal vessel segmentation based on fully convolutional neural networks. *Expert Systems with Applications*, *112*, 229-242. https://doi.org/10.1016/j.eswa.2018.06.034

Ong, Y. T., Wong, T. Y., Klein, R., Klein, B. E., Mitchell, P., Sharrett, A. R., Couper, D. J., & Ikram, M. K. (2013). Hypertensive retinopathy and risk of stroke. *Hypertension, 62*(4), 706-711. https://doi.org/10.1161/HYPERTENSIONAHA.113.01414

Osareh, A., Mirmehdi, M., Thomas, B., & Markham, R. (2003). Automated identification of diabetic retinal exudates in digital colour images. *British Journal of Ophthalmology, 87*(10), 1220-1223. http://dx.doi.org/10.1136/bjo.87.10.1220

Osareh, A., Shadgar, B., & Markham, R. (2009). A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images. *IEEE Transactions on Information Technology in Biomedicine*, *13*(4), 535-545. https://doi.org/10.1109/TITB.2008.2007493

Pratt, H., Coenen, F., Broadbent, D. M., Harding, S. P., & Zheng, Y. (2016). Convolutional neural networks for diabetic retinopathy. *Procedia Computer Science*, *90*(July), 200-205. https://doi.org/10.1016/j.procs.2016.07.014

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

Schmidt-Erfurth, U., Sadeghipour, A., Gerendas, B. S., Waldstein, S. M., & Bogunović, H. (2018). Artificial intelligence in retina. *Progress in Retinal and Eye Research*, *67*(May), 1-29. https://doi.org/10.1016/j.preteyeres.2018.07.004

Takahashi, H., Tampo, H., Arai, Y., Inoue, Y., & Kawashima, H. (2017). Applying artificial intelligence to disease staging: Deep learning for improved staging of diabetic retinopathy. *PLoS ONE*, *12*(6), 1-11. https://doi.org/10.1371/journal.pone.0179790

Tan, J. H., Fujita, H., Sivaprasad, S., Bhandary, S. V., Rao, A. K., Chua, K. C., & Acharya, U. R. (2017). Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network. *Information Sciences*, *420*(August), 66-76. https://doi.org/10.1016/j.ins.2017.08.050

Ting, D. S. W., Cheung, C. Y. L., Lim, G., Tan, G. S. W., Quang, N. D., Gan, A., Hamzah, H., Garcia-Franco, R., Yeo, I. Y. S., Lee, S. Y., Wong, E. Y. M., Sabanayagam, C., Baskaran, M., Ibrahim, F., Tan, N. C., Finkelstein, E. A., Lamoureux, E. L., Wong, I. Y., Bressler, N. M., … & Wong, T. Y. (2017). Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA - Journal of the American Medical Association*, *318*(22), 2211-2223. https://doi.org/10.1001/jama.2017.18152

Van Grinsven, M. J. J. P., Van Ginneken, B., Hoyng, C. B., Theelen, T., & Sánchez, C. I. (2016). Fast convolutional neural network training using selective data sampling: Application to hemorrhage detection in color fundus images. *IEEE Transactions on Medical Imaging*, *35*(5), 1273-1284. https://doi.org/10.1109/TMI.2016.2526689

Wang, C., Zhao, Z., Ren, Q., Xu, Y., & Yu, Y. (2019). Dense U-net based on patch-based learning for retinal vessel segmentation. *Entropy*, *21*(2), 1-15. https://doi.org/10.3390/e21020168

Wang, J. J., Baker, M. L., Hand, P. J., Hankey, G. J., Lindley, R. I., Rochtchina, E., Wong, T. Y., Liew, G., & Mitchell, P. (2011). Transient ischemic attack and acute ischemic stroke: Associations with retinal microvascular signs. *Stroke*, *42*(2), 404-408. https://doi.org/10.1161/STROKEAHA.110.598599

Wang, S., Yin, Y., Cao, G., Wei, B., Zheng, Y., & Yang, G. (2015). Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing*, *149*(PB), 708-717. https://doi.org/10.1016/j.neucom.2014.07.059

Witt, N., Wong, T. Y., Hughes, A. D., Chaturvedi, N., Klein, B. E., Evans, R., McNamara, M., McG Thom, S. A., & Klein, R. (2006). Abnormalities of retinal microvascular structure and risk of mortality from ischemic heart disease and stroke. *Hypertension*, *47*(5), 975-981. https://doi.org/10.1161/01.HYP.0000216717.72048.6c

Yan, Z., Yang, X., & Cheng, K. T. (2019). A three-stage deep learning model for accurate retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics*, *23*(4), 1427-1436. https://doi.org/10.1109/JBHI.2018.2872813

Zhu, C., Zou, B., Zhao, R., Cui, J., Duan, X., Chen, Z., & Liang, Y. (2017). Retinal vessel segmentation in colour fundus images using Extreme Learning Machine. *Computerized Medical Imaging and Graphics*, *55*(2017), 68-77. https://doi.org/10.1016/j.compmedimag.2016.05.004