

An innovative approach for retinal blood vessel segmentation using mixture of supervised and unsupervised methods

Md. Abu Sayed¹ | Sajib Saha² | G. M. Atiqur Rahaman¹ | Tanmai K. Ghosh¹ |
Yogesan Kanagasingam²

¹ Computational Color and Spectral Image Analysis Lab, Computer Science and Engineering Discipline, Khulna University, Khulna, Bangladesh

² Australian e-Health Research Centre, Commonwealth Scientific and Industrial Research Organization (CSIRO), Perth, Western Australia, Australia

Correspondence

Md. Abu Sayed, Computational Color and Spectral Image Analysis Lab, Computer Science and Engineering Discipline, Khulna University, Khulna, Bangladesh.
Email: Sayed1520@cseku.ac.bd

Abstract

Segmentation of retinal blood vessels is a very important diagnostic procedure in ophthalmology. Segmenting blood vessels in the presence of pathological lesions is a major challenge. In this paper, an innovative approach to segment the retinal blood vessel in the presence of pathology is proposed. The method combines both supervised and unsupervised approaches in the retinal imaging context. Two innovative descriptors named local Haar pattern and modified speeded up robust features are also proposed. Experiments are conducted on three publicly available datasets named: DRIVE, STARE and CHASE DB1, and the proposed method has been compared against the state-of-the-art methods. The proposed method is found about 1% more accurate than the best performing supervised method and 2% more accurate than the state-of-the-art Nguyen et al.'s method.

1 | INTRODUCTION

Retinal blood vessel segmentation is a fundamental step in several retinal image analysis tasks. In automated detection of ophthalmic diseases, retinal blood vessels are typically segmented and eliminated in the pre-processing stage [1] so that it does not interfere with pathologies. Blood vessels are considered more reliable than other features when it comes to registering retinal images collected from different viewpoints or at different sites [2–6]. Retinal vessel segmentation and delineation of morphological attributes of retinal blood vessels, such as length, width, tortuosity and/or branching pattern and angles are utilized for diagnosis, screening, treatment and evaluation of various cardiovascular and ophthalmologic diseases such as diabetes, hypertension, arteriosclerosis and choroidal neovascularization [7]. The automatic generation of retinal maps has been used for optic disk identification and fovea localization. Retinal vascular tree is found to be unique for each individual and is used for biometric identification [7]. Figure 1 shows a typical colour fundus photo and its manual segmentation of blood vessels.

Manual segmentation of retinal blood vessels is a long and tedious task which also requires training and skill. It is commonly accepted by the medical community that automatic quantification of retinal vessels is the first step in the development

of a computer-assisted diagnostic system for ophthalmic disorders. A large number of algorithms and techniques have been published relating to the segmentation of retinal blood vessels over the last two decades. Still, there are challenges to address. Some of the important challenges are listed below [9]:

- (i) Segmenting retinal blood vessels in the presence of central vessel reflex.
- (ii) Segmenting blood vessels presenting in crossover and bifurcation regions.
- (iii) Segmenting the merging of close vessels.
- (iv) Segmenting the small and thin vessels.
- (v) Segmenting the blood vessels in the pathological region (dark lesion and bright lesion).

The recently proposed method by Nguyen et al. [9] address many of these challenges quite efficiently. However, it still lacks in accurately segmenting blood vessels in the presence of pathology. From that perspective, in this work, we aim to augment the method proposed by Nguyen et al. [9], so that blood vessels can be detected more accurately even with the presence of pathology. The proposed segmentation incorporates robust feature description and supervised learning steps with the Nguyen et al.'s method. Two innovative descriptors named

This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *IET Image Processing* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology

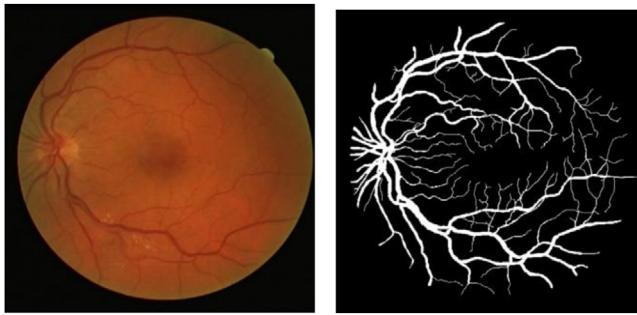


FIGURE 1 An example colour fundus image from the DRIVE dataset [8] (left) and it is manually segmented vessel by the expert grader (right)

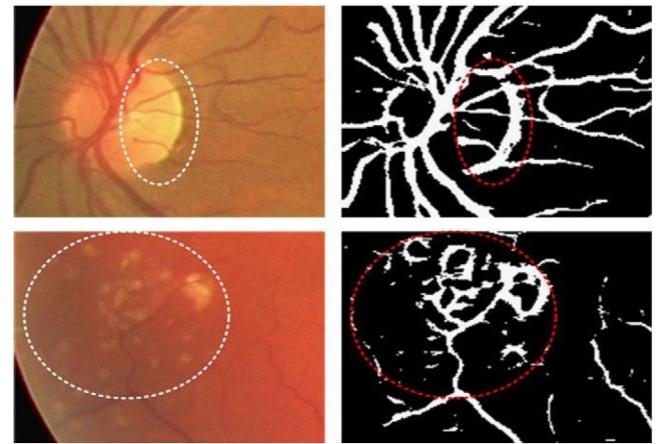


FIGURE 2 Pathology and false vessel detection. First column contains the portions of optic disc and pathological lesions in original image [8] and second column contains the segmented false vessels [9]

local Haar pattern (LHP) and modified speeded up robust features (mSURF) are also proposed.

2 | LITERATURE REVIEW

Over the past two decades, plenty of researches are conducted for segmenting retinal blood vessels. The state-of-the-art methods in the literature can be broadly divided into supervised and unsupervised methods. Supervised methods require a set of training images that are manually labelled by specialists [8]. The rule of vessel extraction is learned on the basis of gold standard by these methods. Unsupervised methods do not require any annotated label. Most of the unsupervised methods are solely based on basic image processing techniques such as mathematical morphology, matched filters, thresholding, vessel tracing, region growing, multiscale purposes etc.

Among unsupervised methods, Wang et al. [10] proposed a segmentation method based on a matched filter. In this method, vessels are enhanced using matched filtering with multiwavelet kernels. The multiscale hierarchical decomposition method is used to remove noise and locate the vessel. For detecting the edges of blood vessels borders, wavelet kernels are successfully applied in DRIVE and STARE datasets. The method can segment the vessels in the pathological lesion., however, thin or small vessels can be affected by fragmentation. Das et al. [11] proposed a thresholding-based technique for segmenting retinal vessels. The paper includes three steps for segmentation: (1) pre-processing; (2) segmentation; (2) post-processing. For pre-processing, contrast limited adaptive histogram equalization is used which enhances the quality of the retinal image. Mean-c clustering method is used for segmentation to extract retinal blood vessels. Mathematical morphology is used for post-processing which removes the isolated pixels. The method is prone to produce false segmentation in the presence of pathological lesions and in optic disc region. Ricci et al. [12] proposed a segmentation method based on basic line operators, which is one of the most significant unsupervised methods. Line detection is used to determine the discontinuity of the intensity value of an image. The green channel is extracted from the original retinal image. Because the green channel is less noisy and has better illumination. The red or blue channel is more prone to

noise. They have also poor illumination than the green channel. The basic line detector deals with the inverted green channel. In the inverted green channel, the vessels appear brighter than the background. A window of size $W \times W$ is taken for each pixel position. The method has limitations such as it has poor segmentation results in the presence of central vessel reflex, at bifurcation and crossover regions, the possibility of merging close vessels. For solving the limitations of Ricci et al.'s method, Nguyen et al. proposed multi-scale detection method in [9]. The method does not require images to be pre-processed prior to segmentation. The Method efficiently segments blood vessels:

- (i) In the presence of central vessel reflex.
- (ii) At bifurcation and crossover regions.
- (iii) In presence of merging of close vessels.

However, one major shortcoming of the method is that it cannot segment the blood vessels accurately, especially in the pathology regions as depicted in Figure 2.

Among supervised methods, Yang et al. [13] proposed a method based on mathematical morphology and a fuzzy clustering algorithm. Mathematical morphology is used for pre-processing which increases the smoothness and strength of the retinal image. Then, the fuzzy clustering algorithm is applied to the pre-processed image for the segmentation of the retinal images. To reduce the noise and weak edges purification is done. Wang et al. [14] proposed a blood vessel segmentation method based on the feature and ensemble learning. Convolutional neural network (CNN) is used for extracting trainable hierarchical feature and ensemble RFs work as a trainable classifier. Marin et al. [15] proposed a method based on NN. A 7D feature vector is extracted for each pixel of the pre-processed retinal images, which is then used to classify the pixel as vessel or not.

In [16], Somoro et al. proposed a deep learning model based on CNN with dice loss function. They implemented the CNN relying upon a modified version of U-Net called

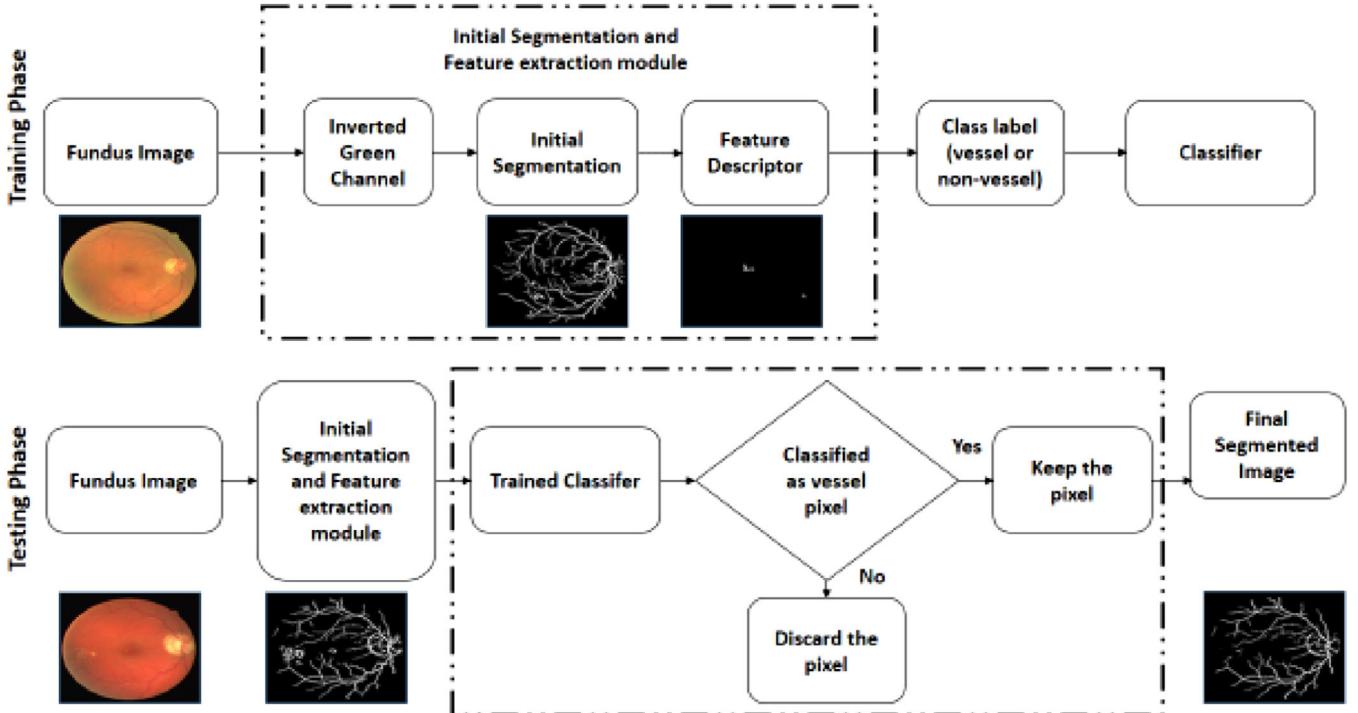


FIGURE 3 The proposed system architecture. Operations shown within the dotted box are performed pixel-wise

variational auto-encoder, where they replaced all pooling layers with progressive convolution and deeper layers. Guo et al. [17] presented a multi-scale deeply supervised convolutional neural network with short connection (BTS-DSN) model, where the bottom-top short connection transfers low-level semantic information to high level, and bottom-top short connection transfers mostly structural information to low level. Fusion of multi-level features enables their model to perform well in cross-training experiments. Jiang et al. [18] proposed a supervised method based on a pre-trained fully convolutional AlexNet through transfer learning to perform pixel-to-pixel and end-to-end semantic segmentation. After pre-processing the whole training and test set, each image was augmented by slicing it to 50×50 patches. Each patch represents whether the corresponding is vessel or non-vessel and utilized to fine tune the AlexNet in the training phase. In the testing phase, the resulted patches are merged into a full-sized segmented image. The method was experimented in four publicly available datasets extensively. Considering highly imbalanced ratio between thick and thin vessels, Yan et al. [19] proposed a three-stage deep learning model to segment retinal vessel accurately. They explicitly annotated thick vessels and thin vessels that were used in computing the cost function to be used in backpropagation step of each of three stages separately. It helped thick vessels to detect vessels with a simple model called ThickSegmenter and thin vessels with a FCN called ThinSegmenter. However, the ThinSegmenter results in generating thin vessels thicker than their usual size. The FusionSegmenter combines these two models to generate more accurate segmentation of the original image. Rammy et al. [20] proposed a conditional patch-

based generative adversarial network model to address the vessel segmentation issues in retinal images. In their two-stage model, they train a deep generative model to generate vessel maps which are later applied to the discriminator along with ground truth vessel masks. The discriminator discriminates the actual and generated vessel maps until it confuses both the input.

While supervised methods are generally found to perform better than unsupervised methods, this is not always the case, especially when the training data is limited [7]. Scarcity of datasets with ground truth is a fact in retinal imaging; that possibly explains why best performing supervised methods in vessel segmentation still performs poor in optic disc region and/or strong contrasted pathology regions [7]. Combination of basic image processing techniques and supervised approaches are often preferred [7].

3 | METHODOLOGY

The proposed method augments the multiscale line detector of Nguyen et al. [9] by incorporating a supervised step with it. Vessel segmentation is performed in 2 steps as shown in Figure 3. In the first step a preliminary segmentation of the vessels is performed using Nguyen et al.'s [9] method, which is an unsupervised approach. In the second step elimination of the falsely classified vessels are performed using supervised methods. Two innovative descriptors are proposed to describe the pixels of interest. Three different classifiers are trained to determine a pixel as true vessel or not.

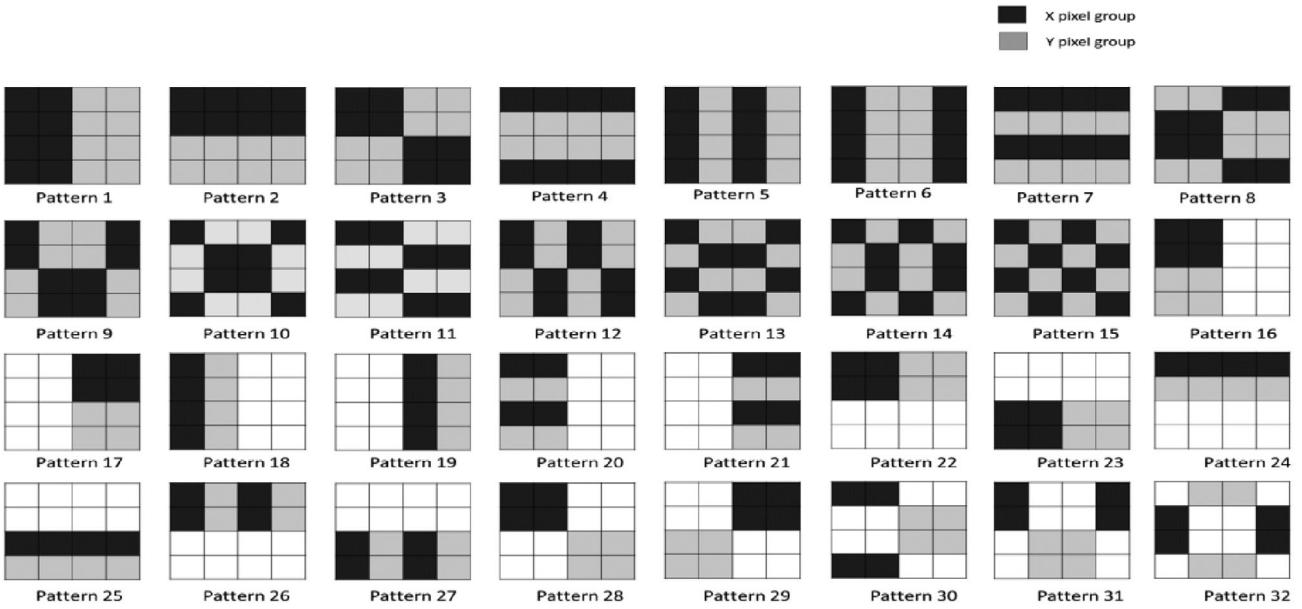


FIGURE 4 All of the 32-pixel patterns used to compute LHP descriptor

3.1 | Local Feature Descriptors

Descriptors named LHP, and mSURF are proposed to describe pixel of interests in the context of retinal imaging. Both LHP and mSURF rely upon Haar patterns while computing the descriptors, however, they vary on how the patterns are utilized and feature vectors are computed.

3.1.1 | LHP

LHP is inspired by the earlier works of Saha et al. [21] and Calonder et al. [22]. In [21, 22], it was shown that image patches can be effectively described based on a small number of pairwise intensity comparisons defined within the patch. Saha et al. used Haar patterns to define groups for intensity comparisons. Randomly selected pixel pairs are used by Caholonder et al. for intensity comparisons. Both methods produce a binary value (0 or 1) out of each comparison, which are finally concatenated to form the binary vector describing the patch. In this work instead of just comparing the intensity to produce 0 or 1, we calculate and store the actual intensity difference, which is to some extent similar to SURF [23]. In order to perform pixels grouping, we define a set of 32 patterns (represented in Figure 4) in this work. These patterns are reminiscent of Haar basis function [22].

A patch p of size 32×32 is considered around the pixel of interest, and a vector of size 128 bytes is calculated. Each byte of the vector is computed based on the intensity difference of two-pixel groups as defined below.

$$T(p, X, Y) = \bar{I}_X - \bar{I}_Y, \quad (1)$$

Here, \bar{I}_X and \bar{I}_Y represent the mean intensities of two different pixel groups X and Y belonging to the patch p .

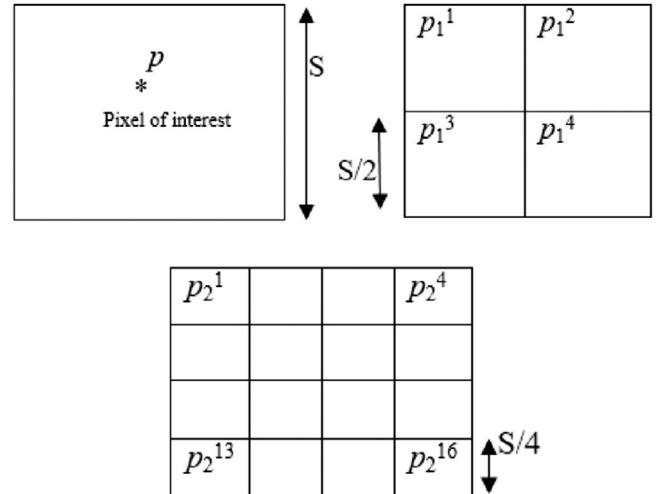


FIGURE 5 Decomposition of the main patch into sub-patches

The 128 bytes vector is generated in three stages and that defines the LHP descriptor. In the first step all the patterns in the set ξ are used to perform intensity comparisons.

In the third step each of the subpatches is further divided into four equal sized subpatches as shown in Figure 5. A subset ϑ of set η is used for comparisons in the third step. Depending on Haar patterns in the initial set ξ , two different versions of the LHP descriptor are devised.

LHP-32

Uses all of the 32 patterns shown in Figure 4 to define the initial set ξ . First 16 of these patterns are then used to define η in the second step. First four patterns are used to form ϑ in the third step.

LHP-16

Uses first 16 of the 32 patterns in Figure 4 to form the initial set ξ . All of these 16 patterns are also used in the second step. The third step uses the first three of these 16 patterns.

3.1.2 | mSURF

In order to extract useful information surrounding the pixel, mSURF relies on Haar wavelet responses likewise in [23]. However, indifferent to [23], here wavelet responses are computed at one scale, which is determined by the Euclidean distance between the optic disc and macula centres. At the same time, instead of using local gradient information for each keypoint or pixel of interest, a global orientation is used. The global orientation is computed based on the optic disc and macula centres.

Orientation Assignment

Prior to computing Haar wavelet responses, we identify a reproducible orientation of the image, which is then used to rotate the image. For that purpose, we first compute the centres of the optic disc and macula relying on the method proposed by Rust et al. in [24]. Let, (x_M, y_M) and (x_{OD}, y_{OD}) are the coordinates of the optic disk and macula center respective, then the reproducible orientation of the image θ is computed as,

$$\theta = \tan^{-1} \left(\frac{y_M - y_{OD}}{x_M - x_{OD}} \right).$$

In the second step the patch p is divided into four equal sized subpatches of 16×16 , a subset η of the pattern set ξ are used.

Image Resizing

Prior to computing wavelet responses we also resize the image. We compute the Euclidean distance,

$$E_i = \sqrt{(x_M - x_{OD})^2 + (y_M - y_{OD})^2}$$

between the optic disc and macula centres. Then the image resizing factor, s is determined as the ratio of E_i and E_{avg} , where E_{avg} is the average Euclidean distance between optic disc and macula and centres computed on 1000 selected images from EyePACS (<http://www.eyepacs.com/>).

Descriptor Components

A square region of size 36×36 around the pixel of interest is considered. This region is further split up into smaller 4×4 square sub-regions. For each sub-region, we compute Haar wavelet responses in the X and Y directions as illustrated in Figure 7. The wavelet responses and their absolute values are summed up over each subregion and a 4D vector $v = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$ is formed, where d_x, d_y are respectively the wavelet responses in the X and Y directions as depicted in Figure 6. The responses are computed at 3×3 regularly spaced intervals using a 4×4 window. The responses are then weighted with a Gaussian of $\sigma = 12$ centred at the

-1	-1	1	1
-1	-1	1	1
-1	-1	1	1
-1	-1	1	1

1	1	1	1
1	1	1	1
-1	-1	-1	-1
-1	-1	-1	-1

FIGURE 6 Haar wavelets

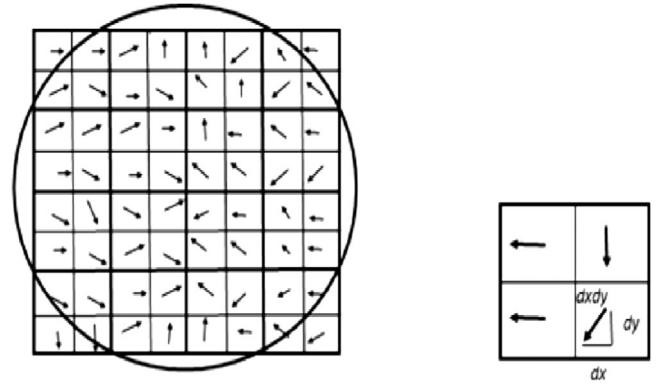


FIGURE 7 Feature description process of mSURF

pixel of interest. Vectors computed over all the sub-regions are then concatenated to form the descriptor of length 64 to represent the pixel. The descriptor is finally normalized to have unit length.

3.2 | Classifiers

The feature vectors of interest points in fundus images are trained and tested in three well-known classifiers.

3.2.1 | Random Forest Classifier

A random forest classifier [25] is trained to classify a pixel as a vessel or not. The training algorithm for random forests applies the general technique of bagging, to tree learners. Given a training set $X = x_1, x_2, \dots, x_n$ with responses $Y = y_1, y_2, \dots, y_n$, bagging repeatedly (K times) selects a random sample with replacement of the training set and fits trees to these samples:

For $k = 1$ to K :

- (i) Sample, with replacement, n training examples from X, Y ; call these X_k, Y_k .
- (ii) Train a classification or regression tree f_k on X_k, Y_k .

After training, predictions for unseen samples x' are made by taking the majority vote in the case of classification trees (Figure 8).

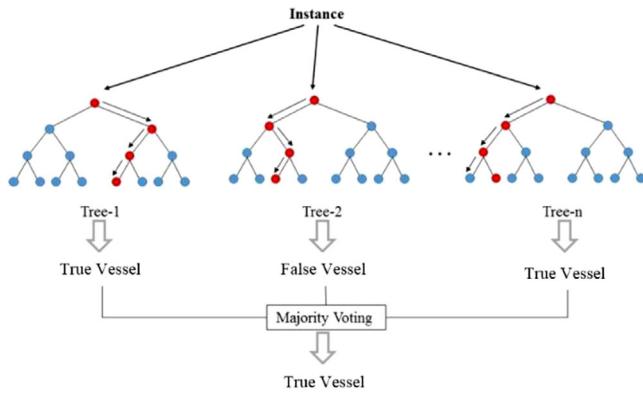


FIGURE 8 Random forest classifier to classify a pixel as a true vessel or not

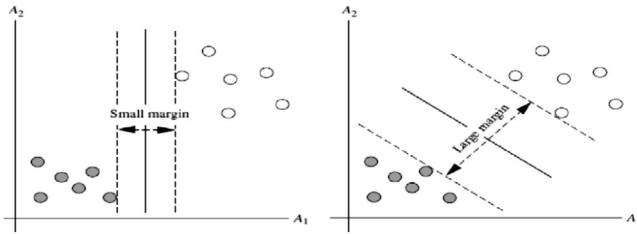


FIGURE 9 Optimal hyperplane in SVM

Descriptors are computed for all the pixels determined as vessels by Nguyen et al.'s method [9]. Ground truth labels of these pixels determined by experienced grader were made available while training the classifier. Once trained, it classified a given pixel as vessel or not vessel depending on its described feature vector.

3.2.2 | Support Vector Machine

Support vector machine (SVM) is usually a non-probabilistic linear classification (or regression) method that classifies tuples by constructing optimal linear separating hyperplane. Even though it is possible to find a number of hyperplanes to separate the data points into two-component, the optimal hyperplane always maximizes the margin of the training samples as shown in Figure 9 and is not too close to the data points which will be noise sensitive and test tuples might be misclassified.

Training

- If non-linear data, map data into high dimensional space using selected kernel function, thus data become linear.
- Select the Support Vectors from both classes (one class is type +1, another is type -1).
- Execute the training algorithm and obtain the parameters a_i from augmented support vectors.

- Find the weight vector w and offset b using parameters a_i s and augmented support vectors and construct the hyperplane using the linear discriminant function, $y = wx + b$.

Classification

- If non-linear SVM, map test data using same Kernel function into that same space.
- Find the appropriate class using weight vector w and (mapped) data comparing with offset b .

3.2.3 | Adaboost

Adaptive boosting, shortly known as Adaboost, is an iterative boosting approach introduced by Freund and Schapire. The major objective of this approach to put emphasis on the instances that are hard to classify. At first, every instance has the same weight that after every iteration increase if the instances are misclassified and decrease if the instances are correctly classified [26].

The Adaboost algorithm is explained below where inputs are D , a set of d class-labelled training tuples; k , the number of iterations; and a classification learning scheme. After the iterations are completed a composite model is created.

Method

- Initialize the weight of each tuple in D to $1/d$
- for $i = 1$ to k :
- sample D with replacement according to the tuple weights to obtain D_i
- use training set D_i to compute a model M_i
- calculate error(M_i) with the following equation:

$$\text{error } (M_i) = \sum_{j=1}^d w_j \times \text{err}(X_j)$$

- if $\text{error}(M_i) > 0.5$ then go back to step 3
- for each correctly classified tuple D_i , update it by multiplying the weight of the tuple by $\text{error}(M_i)/(1-\text{error}(M_i))$
- normalize the weight of each tuple
- endfor

To use the ensemble to classify tuple X , initialize the weight of each class to 0. For each classifier compute

$w_i = \log \frac{1-\text{error}(M_i)}{\text{error}(M_i)}$ and $= M_i(X)$, then add w_i to the weight of class c. Finally, the class with the highest weight will be return as the class of tuple X .

4 | EXPERIMENTS AND RESULTS

Experiments are conducted on publicly available colour fundus image datasets named DRIVE, STARE and CHASE_DB1. All of these datasets are developed to evaluate the performance of vessel segmentation methods and contain verified

vessel structure ground truth made by expert ophthalmologists. The proposed method has been compared with the state-of-the-art supervised and unsupervised methods. A summary of the datasets and evaluation protocol is provided below.

4.1 | Datasets

Among publicly available datasets for retinal fundus images, DRIVE, STARE and CHASE_DB1 are some of the significant ones.

4.1.1 | DRIVE

DRIVE dataset was collected from a diabetic retinopathy screening program in Netherlands [8]. Canon CR5 non-mydiatic 3 CCD camera with a 45° field of view (FOV) was used to acquire the images. The images are 768 × 584 pixels in dimensions, have 8 bits per colour plane and provided in jpeg. There are 40 images in total and among them, 33 do not have any pathology and the rest show early diabetic retinopathy symptoms. Images are provided in training and test sets. Each of the set contains 20 images. The training set contains four pathological images and the test set contains three pathological images. Vessels are segmented by two expert human graders.

4.1.2 | STARE

STARE [27] contains 20 labelled images, of which 10 have pathology and 10 are normal. There is no separate set of training and test images in the STARE dataset. The images are captured using Tapcon [27] TRV-50 fundus camera at a 35° FOV. The images are 605 × 700 pixels, and have 8 bits per colour channel. Two expert observers manually segmented the vessels in the images.

4.1.3 | CHASE_DB1

CHASE_DB1 contains a total of 28 fundus images acquired from multiethnic school children [28]. A hand-held Nidek NM-200-D fundus camera is used to capture the images at a 30° field of view FOV. The size of each image is 960 × 999 pixels. Two individual observers labelled the images. No specific pathology information is provided.

4.2 | Performance Measurement

Every pixel is classified either as vessel or non-vessel. There are four possible classification results: (1) true positive and (2) true negative (3) false positive and (4) false negative [29]. True positive (TP) refers to a pixel classified as a vessel in both in the

ground truth and the segmented image, while false positive (FP) refers to a pixel is classified as a vessel in the segmented image but it is recognized as a non-vessel in the ground truth. Sensitivity (SE), specificity (SP), accuracy and area under the curve (AUC), define mathematically as below, are used to evaluate the performance of the methods.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}.$$

The conventional AUC is calculated from a number of operating points [29]. Normally, AUC is used to evaluate a balanced data classification problem. But blood vessel segmentation is an unbalanced data classification problem as there are fewer vessel pixels than background pixels in a retinal image [29]. The performance of vessel segmentation can be evaluated by applying the following formula [29],

$$\text{AUC} = \frac{\text{Sensitivity} + \text{Specificity}}{2}.$$

4.3 | Results

4.3.1 | Qualitative Results

Figure 10 illustrates two images from DRIVE dataset, and other two from STARE and CHASE_DB1 datasets respectively. The first, second, third and fourth column show original fundus images, images generated after applying multi-scale line detection operation by Nguyen et al. [9], images generated by the proposed method with random forest classifier (RFC) with LHP-16 and mSURF respectively. Random Forest Classifier depicts the best average results with these two descriptors and they are certainly better than Nguyen et al.'s performance. Visual performance of any classifier with LHP-32 is almost identical to the same classifier and LHP-16 combination as shown in Figure 11, consequently, left from comparing with other methods in Figure 10.

4.3.2 | Quantitative Result

Table 1 illustrates performance of the proposed method's best two classifier and feature descriptor combination compared with the state-of-the-art methods in DRIVE and STARE datasets. We performed 10-fold cross validation to evaluate the performance of the proposed method.

As of the promise of the proposed method's to segment retinal vessels in present of pathological regions, it performs significantly better than other methods in terms of sensitivity and shows around 2% and 3% more sensitivity than second best

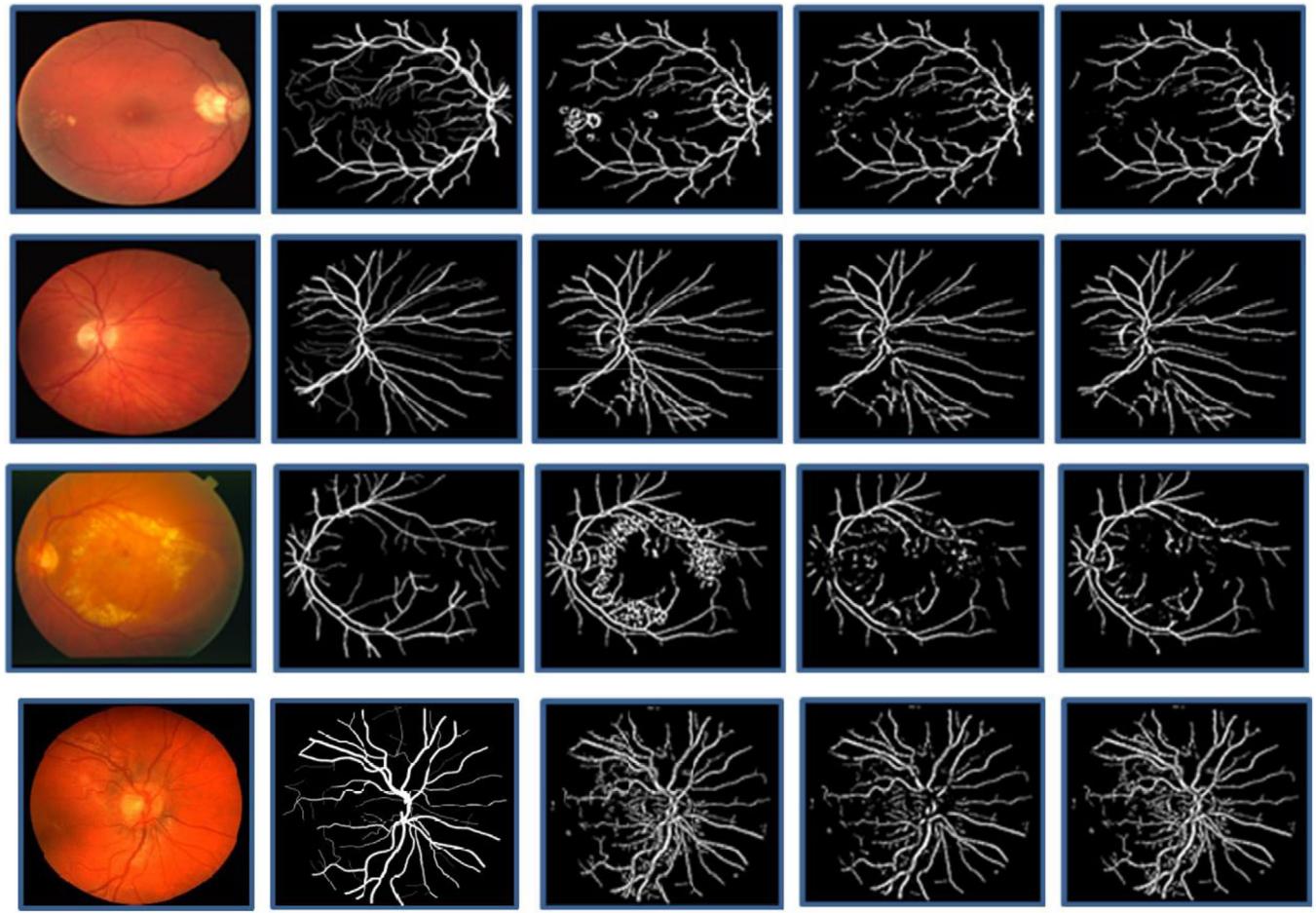


FIGURE 10 Fundus image, ground truth, multiscale segmentation, segmentation by RFC with LHP-16 and mSURF from left to right

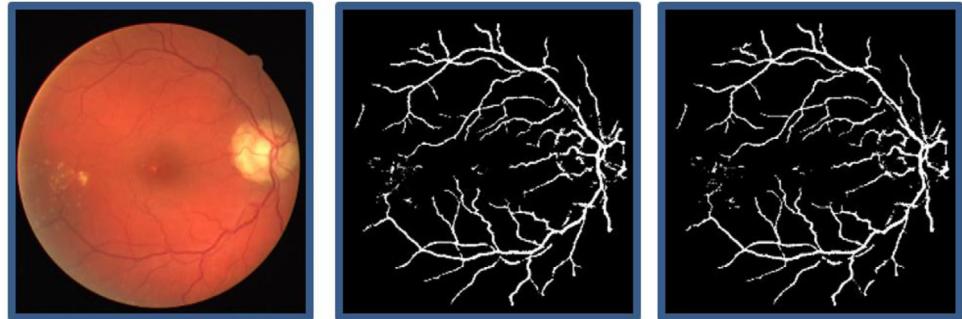


FIGURE 11 DRIVE image (08_test), segmented image by RFC with LHP-16 and LHP-32 respectively

method in DRIVE and STARE datasets respectively. The RFC and mSURF combination of the method depicts the best performance in terms accuracy among the state-of-the-art methods, whereas the SVM and LHP-16 stays on top in terms of the area under curve and sensitivity.

Performance evaluation on DRIVE

The descriptor's performance together with RFC as in Table 2, it appears that the performance of LHP-16 and LHP-32

was quite similar in accuracy, sensitivity, specificity and AUC. In addition, the accuracy of the model with the descriptors including mSURF was also identical and it was slightly more than 96%. The model illustrates 1% and 3% more sensitivity with LHP-16 and mSURF respectively than the sensitivity with LHP-32 which was 71.01%. Moreover, the RFC model with mSURF descriptor shows the best AUC value (85.88%) among the three descriptors. Though all three of the descriptors depicts almost similar performance, the performance of

TABLE 1 Comparison of performance with the state-of-the-art on DRIVE and STARE datasets

Methods	DRIVE				STARE			
	Accuracy	AUC	SE	SP	Accuracy	AUC	SE	SP
Supervised methods								
Lupascu et al. [30]	0.959	—	0.720	—	—	—	—	—
Marin et al. [15]	0.945	0.843	0.706	0.980	0.952	0.838	0.694	0.982
Roychowdhury et al. [31]	0.952	0.844	0.725	0.962	0.951	0.873	0.772	0.973
Jiang et al. [18]	0.959	0.848	0.712	0.983	0.965	0.881	0.782	0.979
Somoro et al. [16]	0.948	0.844	0.739	0.956	0.947	0.855	0.748	0.962
Unsupervised methods								
Annunziata et al. [32]	—	—	—	—	0.956	0.849	0.731	0.984
Zhao et al. [33]	0.948	—	0.735	0.979	0.951	—	0.719	0.977
Budai et al. [34]	0.957	0.816	0.644	0.987	0.938	0.781	0.580	0.982
Nguyen et al. [9]	0.941	—	—	—	0.932	—	—	—
Proposed								
RFC + mSURF	0.961	0.859	0.736	0.982	0.960	0.889	0.806	0.973
SVM + LHP-16	0.958	0.867	0.786	0.973	0.953	0.897	0.831	0.963

TABLE 2 Performance on DRIVE dataset

Classifiers	Descriptors	Accuracy	Sensitivity	Specificity	AUC
RFC	LHP-16	0.9609	0.7173	0.9835	0.8512
	LHP-32	0.9601	0.7101	0.9828	0.8465
	mSURF	0.9606	0.7355	0.9820	0.8588
SVM	LHP-16	0.9581	0.7578	0.9765	0.8672
	LHP-32	0.9575	0.7575	0.9756	0.8667
	mSURF	0.9521	0.6230	0.9826	0.8028
AdaBoost	LHP-16	0.9601	0.6846	0.9851	0.8586
	LHP-32	0.9603	0.6852	0.9853	0.8353
	mSURF	0.9511	0.5951	0.9832	0.7892

mSURF can be considered as a slightly better one. Regarding the SVM model, the LHP-16 and LHP-32 descriptors perform the same in all four criteria. However, these two descriptors outperform mSURF in accuracy, sensitivity and AUC and they perform at around 75.8% sensitivity and 86.7% AUC which are 13% and 6% more than mSURF in sensitivity and AUC respectively. mSURF performs slightly well only in specificity. Interestingly in AdaBoost, the descriptors depict almost identical performance to SVM. The accuracy, sensitivity, specificity and AUC of both LHP-16 and LHP-32 are around 96%, 68.5%, 98.5% and 85.5% respectively which is around 1%, 9%, 0.2% and 6% more than that of mSURF respectively. Overall, SVM with LHP-16 descriptor performs the best in the aspect of all four criteria and computation complexity. The standard deviation of the accuracy of the proposed method was respectively 0.003 (for RFC+mSURF) and 0.005 (for SVM and

TABLE 3 Performance on STARE dataset

Classifiers	Descriptors	Accuracy	Sensitivity	Specificity	AUC
RFC	LHP-16	0.9593	0.7866	0.9735	0.8801
	LHP-32	0.9587	0.7763	0.9725	0.8744
	mSURF	0.9601	0.8058	0.9726	0.8892
SVM	LHP-16	0.9531	0.8313	0.9628	0.8970
	LHP-32	0.9523	0.8284	0.9626	0.8955
	mSURF	0.9576	0.7253	0.9763	0.8508
AdaBoost	LHP-16	0.9584	0.7605	0.9752	0.8679
	LHP-32	0.9579	0.7521	0.9750	0.8636
	mSURF	0.9512	0.6557	0.9757	0.8157

LHP-16). For Nguyen et al.'s method the standard deviation was 0.005.

Performance evaluation on STARE

As in Table 3, where the performance of the three descriptors with RFC are summarized, LHP-16 and LHP-32 exhibit similar results as it was previously observed in DRIVE dataset. They show the accuracy, sensitivity, specificity and AUC around 95.85%, 78.2%, 97.3% and 87.7% respectively. Though the accuracy and specificity of the model with mSURF are not significantly different from the previous two, the sensitivity and AUC are around 1% and 2% higher than the previous two. Similarly, LHP-16 and LHP-32 with the SVM model perform at the accuracy, sensitivity, specificity and AUC of 95.3%, 83%, 96.25% and 89.5% respectively. However, for mSURF sensitivity drops by 10%, consequently, AUC drops by 4%.

TABLE 4 Performance on CHASE_BD1 dataset

Classifiers	Descriptors	Accuracy	Sensitivity	Specificity	AUC
RFC	LHP-16	0.9513	0.7418	0.9669	0.8544
	LHP-32	0.9498	0.7434	0.9653	0.8544
	mSURF	0.9555	0.7630	0.9693	0.8662
SVM	LHP-16	0.9401	0.7911	0.9516	0.8714
	LHP-32	0.9393	0.7912	0.9521	0.8717
	mSURF	0.9428	0.6172	0.9692	0.7932
AdaBoost	LHP-16	0.9539	0.6921	0.9731	0.8326
	LHP-32	0.9535	0.6862	0.9734	0.8298
	mSURF	0.9395	0.5796	0.9667	0.7732

TABLE 5 Comparison of performance on CHASE_DB1

Methods	Accuracy	Sensitivity	Specificity	AUC
Nguyen et al.	0.9433	0.7625	0.9602	0.8613
RFC+ mSURF	0.9555	0.7630	0.9693	0.8662
SVM + LHP-16	0.9401	0.7911	0.9516	0.8714

Regarding AdaBoost, the descriptors illustrate performance commensurate with previous observations, although the percentages are quite different. The accuracy of LHP-16, LHP-32 and mSURF with AdaBoost are 95.84%, 95.79% and 95.12% respectively. Their sensitivity is 76.09%, 75.24% and 65.62% respectively, yet their specificity stays around 97.5% while the AUC varies from 86.79% with LHP-16 to 81.57% with mSURF. The standard deviation of accuracy of the proposed method was respectively 0.011 (for RFC+mSURF) and 0.018 (for SVM + LHP-16). For Nguyen et al.'s [9] model the standard deviation was 0.017.

Performance evaluation on CHASE_DB1

As in Table 4, the accuracy, sensitivity, specificity, area under the curve of the method for CHASE_DB1 dataset for LHP-16 and LHP-32 are identical and they are around 95.1%, 74.3%, 96.69% and 85.44% with RFC, 94.01%, 79.12%, 95.21% and 87.15% with SVM, and 95.4%, 69%, 97.3% and 83% with AdaBoost respectively. Moreover, the accuracy of mSURF is 95.55%, 94.28% and 93.95 with RFC, SVM and AdaBoost respectively and the sensitivity drops from 76.3% to 57.96%, although the specificity remains a pretty average of over 96% in all three classifiers with mSURF. However, the AUC of the models with mSURF in CHASE_DB1 is the lowest of all datasets and they are 86.62%, 79.32% and 77.32% with RFC, SVM and AdaBoost respectively.

Interestingly, it appears that mSURF with RFC performs best in this dataset. As appeared in Table 5, the Random Forest Classifier and mSURF combination of the proposed method also performs better than Nguyen et al. [9] at every four aspects of the comparison. The standard deviation of accuracy of the proposed method was respectively 0.038 (for RFC+mSURF)

and 0.056 (for SVM + LHP-16). For Nguyen et al.'s [9] model the standard deviation was 0.058.

5 | DISCUSSIONS AND CONCLUSION

In this paper, we proposed an innovative approach to segment retinal blood vessels in colour fundus photographs that combines both supervised and unsupervised methods. In the unsupervised part, we have used the multi-scale line detector for a crude segmentation of blood vessels. The supervised technique then applied to remove falsely classified vessels from the initial segmentation. Two innovative descriptors named LHP and mSURF is proposed to describe keypoints in retinal imaging context. In principal mSURF descriptor has two benefits over normal SURF. Firstly, since the mSURF descriptor is computed on a scale rather than multiple scales, it is faster than SURF. Secondly, mSURF, in principle should be more discriminative than normal SURF, as it has the same length like SURF, and is able to compute more information about the patch for being focused on a single scale. In practical experiment, we found that mSURF gives about 5% improvement in computational speed over normal SURF, however, we did not observe any differences in the discrimination power between SURF and mSURF. LHP also relies on Haar like patterns, however, generates a binary descriptor. Three different classifiers named Random Forest, SVM and AdaBoost are independently trained to classify vessel and non-vessel pixels.

The proposed method has been compared against other methods in terms of sensitivity, specificity, accuracy and AUC. Three publicly available datasets that are purposed designed to evaluate the performance of vessel segmentation methods and namely DRIVE, STARE and CHASE_DB1 have been used for experimental evaluation. The proposed method has been found to outperform other state-of-the-art methods. On DRIVE dataset the proposed method shows an accuracy of 96.1%, which is about 2% better than the state-of-the-art Nguyen et al.'s method, about 1% better than state-of-the-art supervise methods. On STARE dataset the proposed method achieves an accuracy of 96% which is about 1–2% better than other methods in comparison. On CHASE_DB1 dataset also the proposed method is found to be 2% more accurate than state-of-the-art Nguyen et al.'s method. The proposed method shows better or at least similar sensitivity, specificity and AUC in comparison to other methods, which is equally applicable for all the datasets.

While generating LHP, there are an enormous number of feature combinations and patch decomposition strategies that are feasible. In this work, we have explored 2 different feature combinations and patch decomposition strategy. In future, we would like to explore, other decomposition strategies, and also want to employ machine to determine optical set of patterns [35].

REFERENCES

1. Habib, M. M., et al.: Detection of microaneurysms in retinal images using an ensemble classifier. Informatics in Medicine Unlocked 9, 44–57 (2017)

2. Saha, S. K., et al.: Color fundus image registration techniques and applications for automated analysis of diabetic retinopathy progression: A review. *Biomed. Signal Process. Control* 47, 288–302 (2019)
3. Saha, S. K., et al.: Performance evaluation of state-of-the-art local feature detectors and descriptors in the context of longitudinal registration of retinal images. *J. Med. Syst.* 42(4), 57 (2018)
4. Saha, S. K., et al.: A two-step approach for longitudinal registration of retinal images. *J. Med. Syst.* 40(12), 277 (2016)
5. Ghosh, T. K., et al.: Retinal blood vessel segmentation: A semi-supervised approach. In: *Iberian Conference on Pattern Recognition and Image Analysis*, pp. 98–107. Springer, Cham (2019)
6. Sayed, M. A., et al.: A semi-supervised approach to segment retinal blood vessels in color fundus photographs. In: *Conference on Artificial Intelligence in Medicine in Europe*, pp. 347–351. Springer, Cham (2019)
7. Fraz MM, et al.: Blood vessel segmentation methodologies in retinal images—a survey. *Comput. Methods Programs Biomed.* 108(1), 407–433 (2012)
8. DRIVE Homepage, <https://www.isi.uu.nl/Research/Databases/DRIVE/>. Accessed 18 December 18 2019
9. Nguyen, U. T., et al.: An effective retinal blood vessel segmentation method using multi-scale line detection. *Pattern Recognit.* 46(3), 703–715 (2013)
10. Wang, Y., et al.: Retinal vessel segmentation using multiwavelet kernels and multiscale hierarchical decomposition. *Pattern Recognit.* 46(8), 2117–2133 (2013)
11. Dash, J., Bhoi, N.: A thresholding based technique to extract retinal blood vessels from fundus images. *Future Computing and Informatics Journal* 2(2), 103–109 (2017)
12. Ricci, E., Perfetti, R.: Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Trans. Med. Imaging* 26(10), 1357–1365 (2007)
13. Yang, Y., et al.: An automatic hybrid method for retinal blood vessel extraction. *Int. J. Appl. Math. Comput. Sci.* 18(3), 399–407 (2008)
14. Wang, S., et al.: Hierarchical retinal blood vessel segmentation based on feature and ensemble learning. *Neurocomputing* 149, 708–717 (2015)
15. Marín, D., et al.: A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Trans. Med. Imaging* 30(1), 146–158 (2010)
16. Soomro, T. A., et al.: Strided U-Net model: Retinal vessels segmentation using dice loss. In: *2018 Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1—8. IEEE, Piscataway, NJ (2018)
17. Guo, S., et al.: BTS-DSN: Deeply supervised neural network with short connections for retinal vessel segmentation. *Int. J. Med. Inf.* 126, 105–113 (2019)
18. Jiang, Z., et al.: Retinal blood vessel segmentation using fully convolutional network with transfer learning. *Comput. Med. Imaging Graph.* 68, 1–15 (2018)
19. Yan, Z., et al.: A three-stage deep learning model for accurate retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics* 23(4), 1427–1436 (2018)
20. Rammy, S. A., et al.: Conditional patch-based generative adversarial network for retinal vessel segmentation. In: *2019 22nd International Multitopic Conference (INMIC)*, pp. 1–6. IEEE, Piscataway, NJ (2019)
21. Saha, S., Démoulin, V.: ALOHA: An efficient binary descriptor based on Haar features. In: *2012 19th IEEE International Conference on Image Processing*, pp. 2345–2348. IEEE, Piscataway, NJ (2012)
22. Calonder, M., et al.: BRIEF: Computing a local binary descriptor very fast. *IEEE Trans. Pattern Anal. Mach. Intell.* 34(7), 1281–1298 (2011)
23. Bay, H., et al.: Speeded-up robust features (SURF). *Comput. Vision Image Understanding* 110(3), 346–359 (2008)
24. Rust, C., et al.: A robust algorithm for optic disc segmentation and fovea detection in retinal fundus images. *Current Directions in Biomedical Engineering* 3(2), 533–537 (2017)
25. Breiman, L.: Random forests. *Machine Learning* 45(1), 5–32 (2001)
26. Quinlan, J. R.: Bagging, boosting, and C4. 5. In: *AAAI-96 Proceedings*, pp. 725–730. AAAI Press, Palo Alto, CA (1996)
27. STARE Homepage, <http://cecas.clemson.edu/ahover/stare/>. Accessed 29 November 2019
28. CHASE_DB1 Homepage, <https://blogs.kingston.ac.uk/retinal/chasedb1/>. Accessed 15 November 2019
29. Fan, Z., et al.: A hierarchical image matting model for blood vessel segmentation in fundus images. *IEEE Trans. Image Process.* 28(5), 2367–2377 (2018)
30. Lupascu, C. A., et al.: FABC: retinal vessel segmentation using AdaBoost. *IEEE Trans. Inf. Technol. Biomed.* 14(5), 1267–1274 (2010)
31. Roychowdhury, S., et al.: Blood vessel segmentation of fundus images by major vessel extraction and subimage classification. *IEEE Journal of Biomedical and Health Informatics* 19(3), 1118–1128 (2014)
32. Annunziata, R., et al.: Leveraging multiscale hessian-based enhancement with a novel exudate inpainting technique for retinal vessel segmentation. *IEEE Journal of Biomedical and Health Informatics* 20(4), 1129–1138 (2015)
33. Zhao, Y., et al.: Automated vessel segmentation using infinite perimeter active contour model with hybrid region information with application to retinal images. *IEEE Trans. Med. Imaging* 34(9), 1797–1807 (2015)
34. Budai, A., et al.: Robust vessel segmentation in fundus images. *Int. J. Biomed. Imaging* 2013, 1–11 (2013)
35. Saha S, Kanagasingam Y: Haar pattern based binary feature descriptor for retinal image registration. In: *2019 Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–6. IEEE, Piscataway, NJ (2019)

How to cite this article: Sayed A, Saha S, Rahaman GMA, Ghosh TK, Kanagasingam Y. An innovative approach for retinal blood vessel segmentation using mixture of supervised and unsupervised methods. *IET Image Process.* 2020;1-11.

<https://doi.org/10.1049/ijpr2.12018>