

ORIGINAL ARTICLES

The automated detection of proliferative diabetic retinopathy using dual ensemble classification

R.A. Welikala¹, M.M. Fraz², T.H. Williamson³, S.A. Barman¹

1. School of Computing and Information Systems, Faculty of Science, Engineering and Computing, Kingston University, London, United Kingdom. 2. School of Electrical Engineering and Computer Science, National University of Sciences and Technology, Islamabad, Pakistan. 3. Ophthalmology Department, St Thomas' Hospital, London, United Kingdom

Correspondence: R.A. Welikala. Address: School of Computing and Information Systems, Faculty of Science, Engineering and Computing, Kingston University, Penrhyn Road, Kingston upon Thames, Surrey KT1 2EE, United Kingdom. Email: Roshan.Welikala@kingston.ac.uk

Received: May 11, 2015

Accepted: June 9, 2015

Online Published: June 17, 2015

DOI: 10.5430/ijdi.v2n2p72

URL: <http://dx.doi.org/10.5430/ijdi.v2n2p72>

Abstract

Objective: Diabetic retinopathy (DR) is a retinal vascular disease that is caused by complications of diabetes. Proliferative diabetic retinopathy (PDR) is the advanced stage of the disease which carries a high risk of severe visual impairment. This stage is characterized by the growth of abnormal new vessels. We aim to develop a method for the automated detection of new vessels from retinal images.

Methods: This method is based on a dual classification approach. Two vessel segmentation approaches are applied to create two separate binary vessel maps which each hold vital information. Local morphology, gradient and intensity features are measured using each binary vessel map to produce two separate 21-D feature vectors. Independent classification is performed for each feature vector using an ensemble system of bagged decision trees. These two independent outcomes are then combined to produce a final decision.

Results: Sensitivity and specificity results using a dataset of 60 images are 1.0000 and 0.9500 on a per image basis.

Conclusions: The described automated system is capable of detecting the presence of new vessels.

Keywords

Retinal images, Proliferative diabetic retinopathy, New vessels, Dual classification, Ensemble classification

1 Introduction

Diabetes is a condition that causes hyperglycemia (high blood sugar) resulting from defects in insulin secretion, insulin action, or both. Hyperglycemia can damage blood vessels, diabetic retinopathy (DR) is the resultant disorder affecting the retinal vasculature. DR can cause loss to vision and is recognized as the leading cause of blindness in the working age population^[1]. Early detection of the disease can allow for timely intervention^[2,3], therefore diabetic patients are required to attend regular retinal screening appointments. Manual assessment of the retinal images (see Figure 1) captured from these appointments can be a time consuming and costly task, therefore this is a field that would significantly benefit from the introduction of automated detection systems^[4].

Blood and fluid from damaged retinal blood vessel will leak on the retina and form features such as microaneurysms, hemorrhages, exudates, cotton wool spots and venous loops [5]. As DR progresses, the blockages and damage to blood vessels cause areas of retinal ischemia to develop. These areas of the retina send signals to the body to grow new blood vessels for nourishment. The growth of new vessels represents the most advanced stage of DR known as proliferative diabetic retinopathy (PDR). This stage poses a high risk of severe vision loss due to the fragile nature of new vessels making them prone to extensive bleeding [6]. Patients presenting PDR require an urgent referral to an ophthalmologist. The role of retinal screening programs is to not only detect patients presenting the onset of DR, but also to differentiate those patients presenting the advance stages of DR.

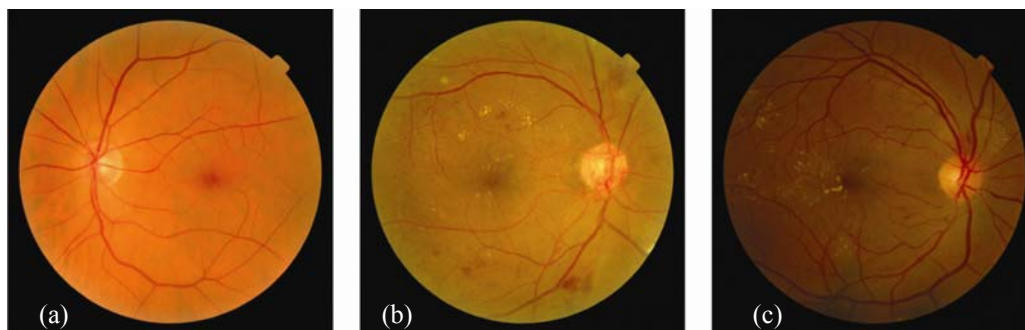


Figure 1. (a) Healthy retinal image. (b)-(c) Retinal images with DR

For clarity, it should be stated that the terms PDR and new vessels are used synonymously in this paper and throughout literature in this field. In reality it is not as simple as this. New vessels will always be present in a PDR case, although they may be obscured or completely hidden from view by features associated with new vessels such as pre-retinal hemorrhages, vitreous hemorrhages, fibrosis tissue and tractional retinal detachment. These make up a small percentages of cases, therefore the main focus should remain on new vessel detection.

New vessels can be categorized according to their location, new vessels at the optic disc (NVD) and new vessels elsewhere (NVE). They tend to be fine in calibre and more tortuous and convoluted than normal vessels. They start as loops or small networks and as they grow they form dense lacy networks which usually pass across the underlying veins and arteries [7]. New vessels grow away from the retinal surface, thus they tend to be out of the focal plane of the photograph and can appear blurry in appearance. Examples of new vessels as well as some other DR features (hemorrhages and exudates) are shown in Figure 2.

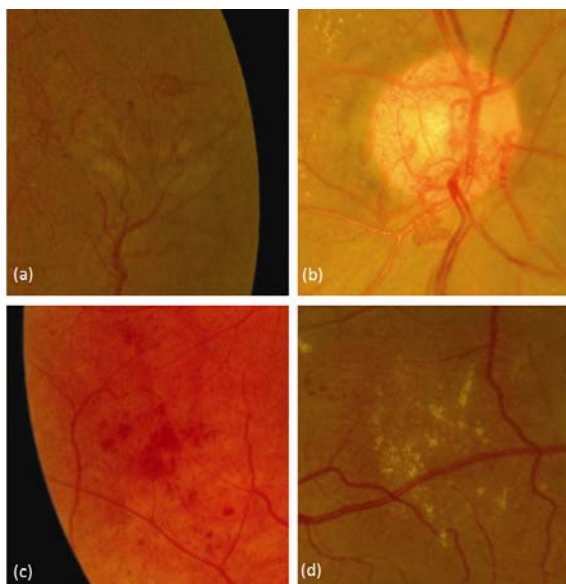


Figure 2. Zoom-in regions of retinal images: (a) NVE. (b) NVD. (c) Hemorrhages (dark lesions). (d) Exudates (bright lesions)

There are many studies investigating the automatic detection of DR focused on microaneurysm and haemorrhage detection^[8-12], and exudate detection^[13-16]. Only a small number of automated new vessel detection methods exist and these can be divided into two main categories. The first category is methods based on the inclusion of vessel segmentation. These methods analyze the morphology of the binary vessel map in search of abnormalities to its structure. The second category is methods based on analyzing texture from the images and thus avoids the difficulties that arise from vessel segmentation.

Vessel segmentation is the most mature area of research in the field of retinal image analysis. A review of vessel segmentation is provided by Fraz^[17]. The majority of these techniques strive to achieve a segmentation performance that would allow for accurate quantification of vessel calibre for cardiovascular studies. However, these techniques struggle to segment new vessels due to their fine calibre and irregular appearance. Ramlugun^[18] addressed this problem with a technique designed for small vessel extraction. The main contribution was the varying of the clip limit for contrast limited adaptive histogram equalization (CLAHE) to allow more contrast for small vessels. Most techniques also do not put enough emphasis on removing false responses caused by artefacts and other lesions. In particular false responses to exudates can cause large local densities on the segmented map and therefore can be mistaken for new vessels. B. Zhang^[19] proposed a matched filter with first-order derivative of the Gaussian and L. Zhang^[20] described a modified matched filter that used double sided thresholding, both methods were designed to reduce false responses to exudates.

The following methods include vessel segmentation prior to the described analysis methods for the detection of new vessels. Both Daxer^[21] and Karperien^[22] described the vasculature of the retina as a fractal and used the fractal dimension to measure its complexity to indicate the presence of new vessels. Hassan^[23] applied a local region based technique, where a scanning sub-window was used to measure the number of vessels and the area of vessels to indicate new vessels. Another local region based technique was proposed by Welikala^[24] which used five morphology features. This technique also included a step to discard the majority of normal vasculature and therefore simplified new vessel detection. Jelinek^[25] applied the derivatives of Gaussian wavelets to the vessel skeleton to extract several morphological based features to detect new vessels. A comprehensive set of 15 features was developed by Goatman^[26] to specifically detect NVD, this included the number of vessel segments, the mean vessel wall gradient and various tortuosity measures. Akram^[27] proposed a ten dimensional feature set based on morphological, intensity and gradient based values along with a multivariate m-Medoids based classifier.

The following methods describe the second category of new vessel detection methods, those based on extracting textural information. Frame^[28] used the grey level co-occurrence matrix (GLCM) to calculate statistical texture measures in order to identify irregular distributions of pixel intensities associated with neovascularization. Texture features from the run length matrix and the GLCM were calculated by Acharya^[29] to identify the stage of DR. Spectral texture analysis using multi-scale amplitude modulation frequency modulation (AM-FM) methods were proposed by Agurto^[30] to characterize different retinal structures, including new vessels. However, Agurto^[31] extended their work to include vessel segmentation and granulometry along with AM-FM to detect NVD. A new vessel detection method which combined statistical texture analysis, high order spectrum analysis and fractal analysis was proposed by Lee^[32]. However vessel segmentation was required for multi-fractal analysis. The performances of six different texture based methods for the detection of neovascularization were compared by Vatanparast^[33]. These include AM-FM, GLCM, Contourlet transform, Gabor filters and local binary patterns.

The application of a dual classification approach to independently process the binary maps from two different vessel segmentation methods^[34] was previously proposed by our research group. This included vessel segmentation by a standard line operator and a novel modified line operator. The former targets the accurate segmentation of new vessels and the latter (based on double sided thresholding) targets the reduction of false responses to non-vessel edges. The overall aim was to detect new vessels whilst reducing false responses caused by other retinal features.

This paper presents a new supervised method for the detection of new vessels by using an ensemble classifier of boot strapped decision trees combined with the dual classification approach^[34]. The feature vector has been expanded to include morphology, gradient and intensity based features. The classifier based on the boot strapped decision trees is a classic ensemble classifier, which has been broadly applied in many application areas of image analysis^[35], but has not been extensively utilized for detecting neovascularization. To the best of our knowledge, this is the first use of a decision trees based ensemble method for new vessel classification. An important feature of the bagged ensemble is that the classification accuracy can be estimated during the training phase, without supplying the classifier with test data. Moreover, the importance of each feature in classification can also be predicted during the training phase, which helps in identifying the most relevant features used in new vessel detection thus automatically reducing the dimensionality of the feature vector and boosting computational speed. Selecting the most relevant features also provides the potential to improve the classifier's performance. Furthermore, as compared to other supervised methodologies utilized in retinal image analysis, the ensemble classifier is fast in training and classification and comparatively requires less number of training samples.

2 Methodology

2.1 Framework

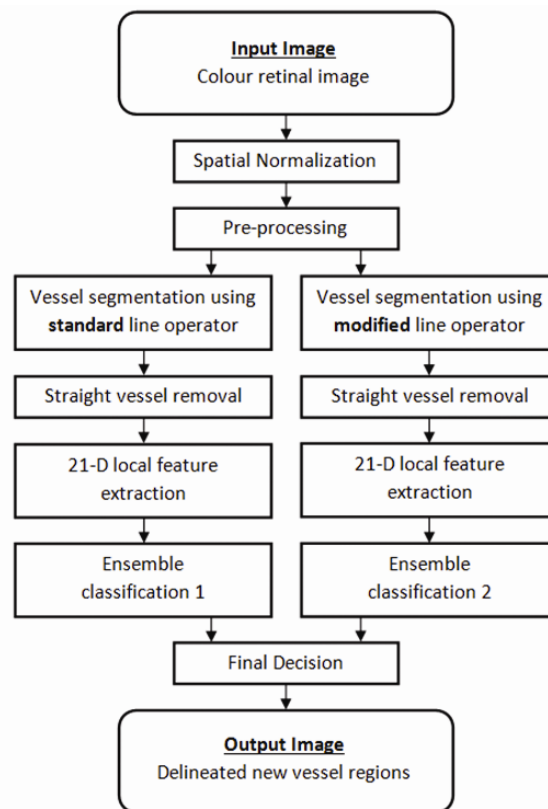


Figure 3. System architecture of the dual ensemble classification approach

Following alternations to the dual classification framework of Welikala *et al.*^[34], the architecture of the proposed method is shown in Figure 3. Following spatial normalization and pre-processing the system splits into two pathways as two different vessel segmentation methods were applied to create two binary vessel maps. Each pathway had its own feature set produced, using the same set of 21 local features. Independent classification was performed for each feature set using an ensemble classifier. The system produced a final decision by combining the two individual classification

outcomes in which regions of the retina were labelled as new vessels or non-new vessels. The initial steps of the framework are identical to that from Welikala *et al.* [34]. Therefore sections prior to 2.3 shall be kept concise, for a full explanation with all parameters listed see Welikala *et al.* [34].

2.2 Line operator/modified line operator

The line operator is a simple and effective method for vessel segmentation. The standard line operator is illustrated in Figure 4. The average grey-level of the pixels along a line passing through the target pixel was calculated for multiple orientations. The line strength of the pixel was obtained from the orientation with the largest value subtracted by the average grey-level of the similar orientated neighborhood. An empirically derived threshold was applied to the line strength image to produce the segmentation of the vessels. The modified line operator differed by being based on three line strength measures as opposed to one. This included measures that separately took into account the left and right side of the neighborhood. For full details and equations see Welikala *et al.* [34].

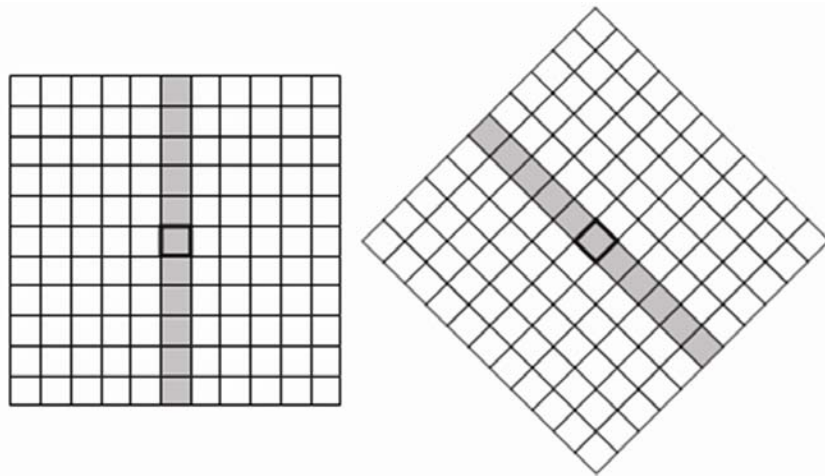


Figure 4. Standard line operator at different orientations

Figures 5(e)-(h) show that the standard line operator segments normal and new vessels with a high level of accuracy, however it also responds to non-vessel edges (mainly exudates/bright lesions and reflection artefacts). These false vessel detections often cause large local densities and large curvatures which are indistinguishable from new vessels. Figures 5(i)-(l) show the segmentation generated by the modified line operator has significantly reduced the false vessel detections to the extent that non vessel edges were distinguishable from new vessels. Unfortunately the segmentation of new vessels has worsened. This meant that new vessels were no longer so distinguishable from normal vasculature. Both of the segmentation methods showed disadvantages and therefore neither method alone was suitable for the detection of new vessels. However extraction of information from both maps could be used effectively and this lead to the development of the dual classification approach.

2.3 Feature extraction

The design of this method was aimed at the classification of image regions that contain new vessels. In a simple 4-D feature vector based on the description of new vessel regions that contain many vessel segments, which are closely packed and appear in multiple orientations was used [34]. However new vessels can be distinguished using other characteristics. New vessels tend to appear finer in calibre, shorter in length and possess a tortuous nature. New vessels also appear less homogeneous than normal vessels. Therefore, additional features were created based on these characteristics. This formed a 21-D feature vector (see Table 1) which contained morphology, intensity and gradient based features.

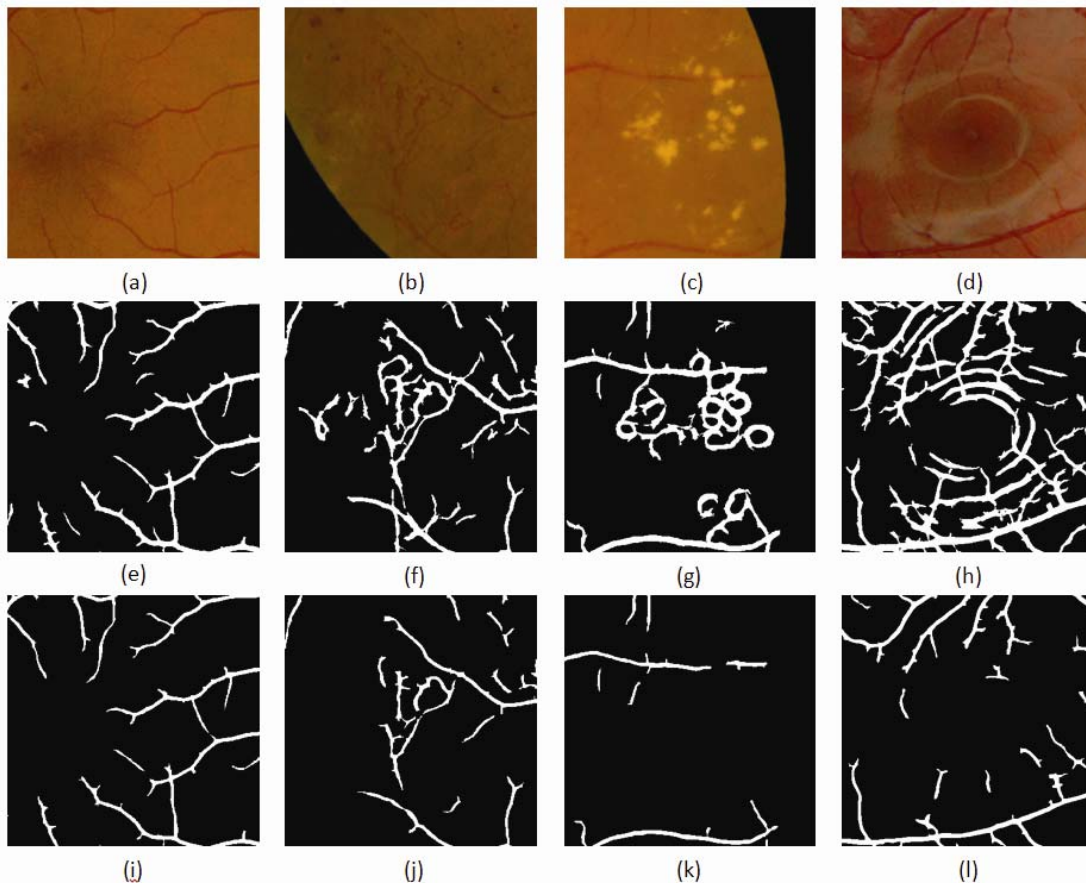


Figure 5. (a)-(d) Zoom-in regions of retinal images. (a) Normal vessels. (b) New vessels. (c) Exudates (bright lesions). (d) Reflection artefacts. (e)-(h) Segmentation of (a)-(d) using the standard line operator. (i)-(l) Segmentation of (a)-(d) using the modified line operator

Prior to measurements of features, the binary maps were simplified. This included straight vessel removal and the creation of vessel segments (single pixel in thickness), for full details see Welikala *et al.* [34]. A sub window of size 151×151 pixels was created in order to calculate local features. This sub window was scanned through the image and at each pixel position the 21-D feature vector was calculated. This same set of features was measured from each of the pathways to produce two separate feature vectors.

Table 1. Description of the features

Feature Number	Feature Name	Feature Description
1	Number of vessel pixels	The sum of all segment pixels within the sub window.
2	Number of vessel segments	The number of whole and partially included segments within the sub window.
3	Number of vessel orientations	The end points of a segment were connected by a straight line. The angle the line made with the x-axis that fell within the range -90° to 90° of the unit circle was calculated. The calculated angle was accordingly dropped into one of eight bins, each representing a range of angles. This was done for each segment within the sub window and the number of non-empty bins represented the number of orientations.
4	Vessel density	A segment was dilated with a disk structuring element with a radius of 20 pixels. The number of vessel pixels within the dilated area was divided by the number of pixels within the segment to give its vessel density. This was done for each segment within the sub window and the mean vessel density was calculated.

(Table 1 continued on page 78)

Table 1. (continued.)

Feature Number	Feature Name	Feature Description
5	Tortuosity 1	The tortuosity of each segment was calculated using the true length (measured with the chain code) divided by the Euclidean length. The mean tortuosity was calculated from all the segments within the sub window.
6	Tortuosity 2	The maximum tortuosity amongst all segments in the sub window.
7	Vessel length	The mean true length (in pixels) of all segments within the sub window.
8	Number of bifurcation points	The number of bifurcation points removed within the sub window when creating segments.
9	Grey level mean	The mean grey level of all segment pixels within the sub window.
10	Grey level coefficient of variation	The ratio of the standard deviation to the mean of the grey level of all segment pixels within the sub window.
11	Gradient mean	The mean gradient magnitude along all segment pixels within the sub window. Calculated using the Sobel gradient operator applied on the pre-processed image.
12	Gradient coefficient of variation	The ratio of the standard deviation to the mean of the gradient of all segment pixels within the sub window.
13	Line strength	The mean line strength of all segment pixels within the sub window.
14	Vessel width	Skeletonization correlates to vessel centre lines. The distance from the segment pixel to the closest boundary point of the vessel using the vessel map prior to skeletonization. This gives the half-width at that point which is then multiplied by 2 to achieve the full vessel width. The mean is calculated for all segment pixels within the sub window.
15	Mean vessel wall gradient	As for the vessel width above, the closest boundary point was assumed to be the vessel wall. The mean gradient magnitude along all vessel wall pixels within the sub window.
16	Vessel wall gradient coefficient of variation	The ratio of the standard deviation to the mean of the vessel wall gradient along all vessel wall pixels within the sub window.
17	Compactness	The full vessel map prior to skeletonization and straight vessel removal was used. Area and perimeter within the sub window were measured and used in the circularity equation, $C = 4\pi \cdot \text{area} / \text{perimeter}^2$.
18	Connectivity	The full vessel map prior to skeletonization and straight vessel removal was used. The vessel area was divided by the number of objects within the sub window.
19	Local grey level mean	The mean grey level of all pixels within the sub window using the pre-processed image.
20	Local grey level max	The maximum grey level amongst all pixels within the sub window using the pre-processed image.
21	Local grey level variation	The standard deviation of all pixels within the sub window using the pre-processed image.

Note. The term segment refers to the skeletonized vessel segments.

2.4 Dual ensemble classification

In this work two individual decisions have to be made. Therefore the two same sets of features measured from two different sources were kept separate and independent classification was performed on each. Classifier 1, associated with the feature set measured from the standard line operator approach, was intended to distinguish new vessels from normal vessels. Classifier 2, associated with the feature set measured from the modified line operator approach, was intended to distinguish new vessels from exudates and reflections. Combining the outcomes then removed the false new vessel responses that each classifier made. Alternatively, the conventional approach would be to combine all features into a single feature set. The classifier’s decision boundary would attempt to deal with making both decisions. However, such a boundary would compromise each of the decisions.

All features were normalized so that each feature had zero mean and unit standard deviation. Independent classification was performed for each of the two feature sets using an ensemble classifier of boot strapped decision trees (explained in section 2.5). Each pixel in the retinal image was characterized by a vector in 21-D feature space.

$$Fv(x,y) = [f_1(x,y), f_2(x,y), \dots, f_{21}(x,y)]$$

Each classifier independently labelled the candidate pixel as new vessels or non-new vessels when its representation in feature space $Fv(x, y)$ was known. The system produced a final decision by combining the two outcomes. The candidate pixel achieved a new vessel label only when both classifications agreed on its identity being new vessels, otherwise it achieved a non-new vessel label. When complete, all pixels labelled as new vessels were dilated with a structuring element the size of the sub window to illustrate the new vessel regions.

2.5 Ensemble classifier

The process of consulting multiple experts or seeking multiple opinions ahead of final decision making is almost our second nature. The wide-spread gains of such a procedure in automated decision making applications give rise to the ensemble classification framework.

In ensemble classification^[36], multiple classifiers and models are tactically generated and combined in order to give the solution to a machine learning problem; with a goal of obtaining better performance than could be obtained from any of the constituent classifiers/models. This process is used to abbreviate the likelihood of inadequate or unfortunate selection while improving the performance of the classifier. We use this strategy instinctively in our day to day activities, where we consider the opinion from several experts, evaluate and mingle their recommendations for establishing a well optimized and well-versed conclusion. In the same manner, the ensemble methods utilize multiple classifiers / models to accomplish gain in classification performance by mixing/aggregating the outcomes from several weak learners into one high-class classifier, with the goal of reducing the variance and amplifying the confidence in the decision. In this approach, we used the decision trees as the classification model and the results of these weak learners were combined using bootstrap aggregation also known as bagging.

Breiman's bagging^[37] is among the most primitive ensemble methods, which is most perceptive and simplest to implement, with surprisingly fine results. In bagging, the component classifiers (in this case the decision trees), are developed on the bootstrap replicas of the training dataset. This is done by a random selection of N training instances out of N with replacement, where N is the size of the training set. Majority voting is used to combine the responses of the individual component classifier (the decision trees). The decision of ensemble is the class chosen by most of the component classifiers. Traditionally, the component classifiers are of the same general form; for example, all hidden Markov model, all neural networks or all decision trees, which was the case in this work.

Given the original training set T , multiple sets of training data T_b are created, where $b = 1, 2, \dots, B$, by randomly sampling T with replacement. B is the number of component classifiers used in the ensemble system. On average, each training set T_b only contains two-thirds of the original samples. The bagging algorithm as explained by Polikar^[38] is illustrated below,

Inputs for bagging algorithm

- Training data $T = \{x_1, x_2, \dots, x_N\}$, $x_i \in X$, provided with correct class labels $w_i \in \Omega = \{w_1, \dots, w_C\}$. N is the size of training set.
- WeakLearn, (the weak learning algorithm).
- The Integer B specifies the total number of iterations.

Do $b = 1 \dots B$

1. Acquire the bootstrap sample T_b by randomly choosing N observations, with replacement, from the training set T .
2. Call the routine WeakLearn with T_b and obtain the hypothesis from the classifier

$$h_b : X \rightarrow \Omega.$$
3. Add h_b to the ensemble, E .

End Do Loop

Test the algorithm with simple Majority Voting

Given the unlabelled data instance z

1. Estimate the ensemble $E = \{h_1, \dots, h_B\}$ on z .
2. Suppose the classifier h_b gives vote to the class w_j as,

$$v_{b,j} = 1 \text{ or } 0 \quad (1) \text{ If } h_b \text{ picks class } w_j, (0) \text{ otherwise}$$
3. The total votes obtained by each of the class are,

$$V_j = \sum_{b=1}^B v_{b,j}, \quad j = 1, \dots, C.$$
4. The final classification/label is the class that gains the majority votes.

3 Experimental evaluation

3.1 Materials and evaluation

Evaluation was performed using images collected from the publicly available MESSIDOR retinal image database^[39] and the St Thomas' hospital ophthalmology department. A dataset of 60 images was created, 20 images with confirmed new vessels and a further 40 images without new vessels. The image data from each source were as follows:

- MESSIDOR: 5 new vessel images, 20 normal images and 20 images with other DR pathology (mainly bright lesions) and strong reflection artefacts. These images were acquired from a color video 3CCD camera on a Topcon TRC NW6 fundus camera with a 45 degree field of view (FOV) and an image resolution of 2240×1488 pixels.
- St Thomas' Hospital: 15 new vessel images acquired with a Nikon D80 digital SLR camera on a Topcon TRC NW6 fundus camera with a 45 degree FOV and an image resolution of 2896×1944 pixels. Ethical approval was obtained for the use of these images.

Images were spatially normalized using a technique proposed by Zhang^[40] along with bicubic interpolation and anti-aliasing. All images were required to be captured with the same FOV angle. Images were normalized to have a FOV width of 1379 pixels. Cropping was used to remove some of the surrounding black border to produce images of size 1479×1479 pixels.

The training data comprised of a specific selection of pixels chosen from the dataset. For convenience, a pixel can be referred to as an image patch as features were extracted using information from the local neighborhood contained within the sub window centred over the target pixel. An ophthalmologist labelled each image patch as either new vessels or non-new vessels. Separate training data was used for each classifier. Classifier 1 was trained with 50 new vessel patches and 50 normal vessel patches. Classifier 2 was trained with 50 new vessel patches and 50 patches made up of a variety of bright lesions, dark lesions and reflection artefacts. Note that each new vessel patch correlated to a pixel selected in the centre of a new vessel region.

Testing was performed across the whole of each retinal image, in terms of the classification process being performed at every pixel location. Splitting the data set to create separate training and testing sets was not suitable due to the limited size of the dataset. Instead, the leave-one-out cross validation method was applied. This meant the training data available to the classifiers was obtained from all the patches from all the images except those from the single test image, and this process was repeated for each image. The feature value normalization was also recalculated each time, leaving out the test image.

All pixels classified as positive by the system were dilated by the size of the sub window in order to represent the local neighborhood in which information had been extracted from. The effect of this was the delineation of the new vessel regions. However, the performance was assessed on a per image basis as this is more useful from a clinical point of view. An image simply achieved a new vessel label if it contained any delineated regions. Prior to this, an ophthalmologist labelled all images using the same labels as before but on a per image basis.

3.2 Ensemble classifier evaluation

An important feature of the bagged ensemble is that the classification accuracy can be estimated during the training phase without supplying the test data. Moreover, the importance of each feature in classification can also be predicted during the training phase. The trustworthy estimates of classification accuracy and the feature importance during the training of classifier is a smart feature of bagging.

3.2.1 Out-of-Bag classification error

In bagging, the component classifiers (in this case the decision trees), were developed on the bootstrap replicas of the training dataset. This was done by a random selection of N training instances out of N with replacement, where N is the size of the training set. The phenomenon of choosing the N out of N training instances with replacement leaves out 37% of instances on average for every component classifier in the ensemble. The left-overs were called the out-of-bag (OOB) observations and were used in the estimation of the predictive power of the ensemble. The OOB predicted responses were compared against the observed responses of all the training instances to estimate the average OOB error, which is an unbiased estimator of the true ensemble error. Figure 6 and 7 show the plot of the OOB classification error computed for the classifiers comprised of 100 decision trees for each of the classifications respectively. 25 and 30 trees were the optimal choices for classification 1 and classification 2 respectively as OOB classification error remained constant with the increase in the number of decision trees in the ensemble.

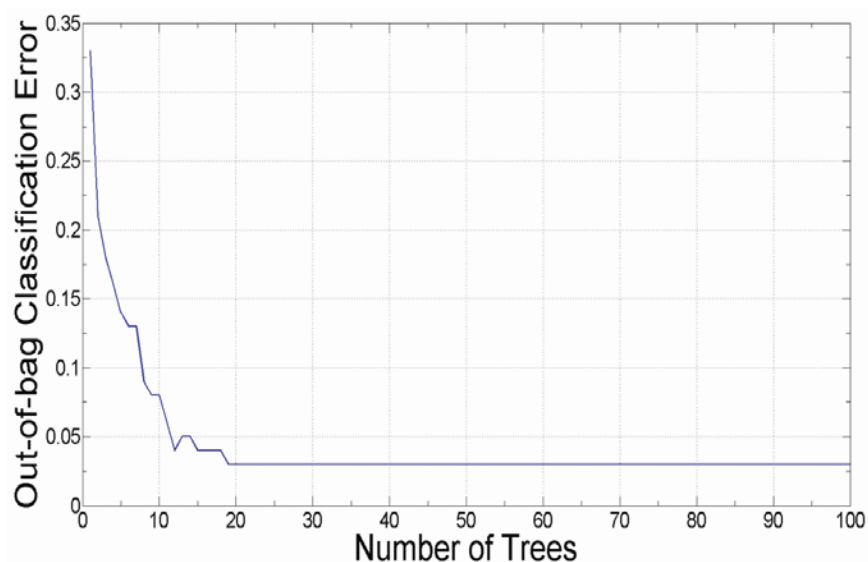


Figure 6. Out-of-Bag classification error for classification 1

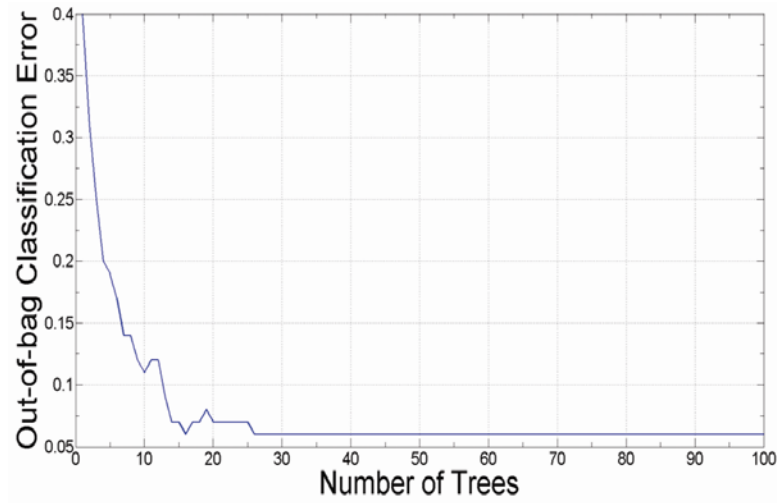


Figure 7. Out-of-Bag classification error for classification 2

3.2.2 Feature analysis

Initially we have used a feature vector consisting of twenty one features, which are summarized in Table 2, and detailed in section 2.3. For establishing the importance of each feature, the feature importance index was computed during the classifier training phase. The OOB observations were used to determine the significance of each individual feature from the feature vector in the classification. In order to predict the feature importance, the OOB data was randomly permuted across one variable and the increase in OOB error due to this permutation was estimated. This increase was directly proportional to the importance of the feature in the classification. The larger the increase, the more important the feature was in the classification.

Table 2. Summary of feature vector

Feature Number	Feature Name	Feature Number	Feature Name
1	Number of vessel pixels	12	Gradient coefficient of variation
2	Number of vessel segments	13	Line strength
3	Number of vessel orientations	14	Vessel width
4	Vessel density	15	Mean vessel wall gradient
5	Tortuosity 1	16	Vessel wall gradient coefficient of variation
6	Tortuosity 2	17	Compactness
7	Vessel length	18	Connectivity
8	Number of bifurcation points	19	Local grey level mean
9	Grey level mean	20	Local grey level max
10	Grey level coefficient of variation	21	Local grey level variation
11	Gradient mean		

Figure 8 and 9 shows the graphs for feature importance index (FII) calculated from classifiers created with 100 decision trees. The lowest ranked feature was removed (without return) from the feature vector and the system’s performance was evaluated. This was performed simultaneously for both feature vectors. This procedure continued until there was a drop off in the optimal operating point, the feature vector prior to this drop off was chosen. This resulted in 33% of the most significant features being kept for both feature vectors. The objective was to reduce the dimensionality of the feature vector which in turn decreases the computational cost of feature generation, classifier training and classification. Selecting the most relevant features also provides the potential to improve the classifier’s performance.

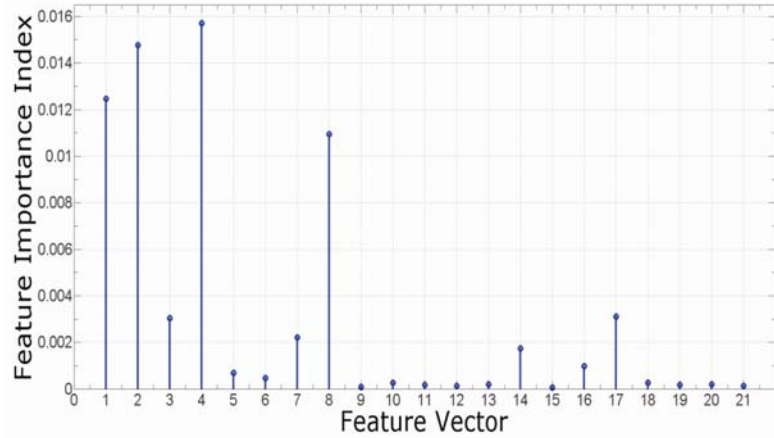


Figure 8. Feature Importance Index, classification 1

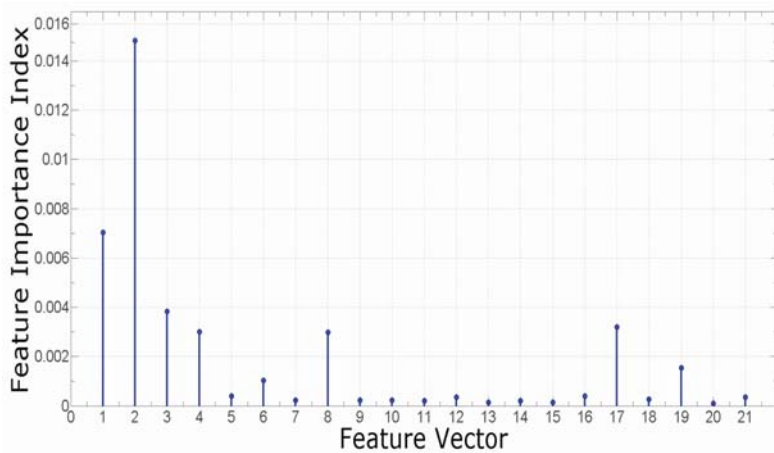


Figure 9. Feature Importance Index, classification 2

3.3 Performance measures

As mentioned above, the performance was assessed on a per image basis. An image was classified as either new vessels or non-new vessels. Consequently there are four outcomes, two classifications and two misclassifications which are defined in Table 3. The algorithm was evaluated in terms of sensitivity (SN), specificity (SP) and accuracy (Acc). These are often used in machine learning and are measures of the quality of binary classification. These metrics are defined in Table 4.

The use of the receiver operating characteristic (ROC) curve allows for the visualization of the performance of a binary classifier system, expressing the trade-off between increased detection and false alarms. This was created by plotting the true positive rate (SN) versus the false positive rate (1-SP) at various threshold levels of the probability score of the classifier.

With a dual classifier approach and therefore two probability scores, the creation of ROC curves was not a straight-forward task. This was resolved with the addition of a third axis to the conventional 2D ROC plot to accommodate for varying the threshold of the probability score of the additional classifier that arises in the dual classification approach. The resultant was a 3D ROC surface that explored all combinations of thresholds for the dual classification. Information from this 3D ROC surface was extracted to create a conventional 2D ROC curve. Full details of the procedure are described in Welikala *et al.* [34]. From this 2D plot the area under the curve (AUC) was extracted and used as a performance measure.

Maximum Acc is often used as the criteria to select the optimal operating point from the ROC curve. However, the algorithm could reach a SN of 100% at a high SP. This was important as from a clinical point of view a SN of 100% was

considered an essential requirement for assessment on a per image basis. Max Acc may not always equate to an operating point with a SN of 100%. Therefore an application specific performance measure was created, in which the operating point with the highest SP at a SN of 100% was selected as the optimal operating point.

Table 3. New vessel classification

	New vessels present	New vessels absent
New vessels detected	True positive (TP)	False positive (FP)
New vessels not detected	False negative (FN)	True negative (TN)

Table 4. Performance measures for new vessels detection

Measure	Description
SN	$TP/(TP+FN)$
SP	$TN/(TN+FP)$
Acc	$(TP+TN)/(TP+FP+TN+FN)$

3.4 Results

The ROC curve of the proposed system for evaluation on a per image basis is depicted in Figure 10. This represents the performance with the feature vector reduced in dimensionality. The features retained are shown in Table 5. The AUC value is 0.9505. The optimal operating point according to the application specific performance measure is a SN of 1.0000 and a SP of 0.9500. Table 6 shows these results along with the reported results from other new vessel detection methods. Figure 11 provides the manual delineation of new vessel regions marked by an ophthalmologist. Examples of classified images are given in Figure 12. Classified new vessel regions have also been indicated with a white boundary. Images containing any delineated regions are classified as new vessel images.

The MATLAB Code took 683 seconds on an Intel(R) core(TM)2 Quad CPU Q9300 at 2.5 GHz to process each image using the full 21-D feature vector. This was reduced to 498 seconds once the feature vector had been reduced in dimensionality.

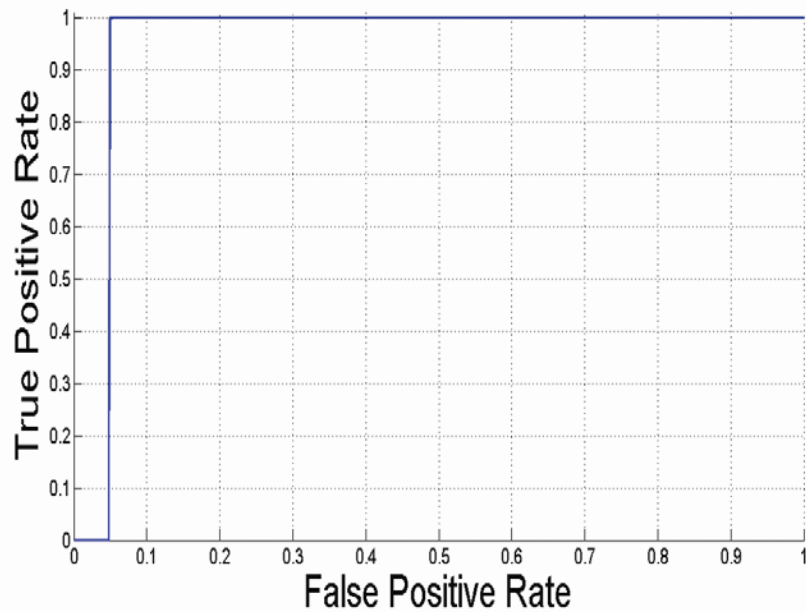


Figure 10. ROC curve for the performance of the proposed method (7 features)

Table 5. Selected features

Classification	Features
1	4, 2, 1, 8, 17, 3, 7
2	2, 1, 3, 17, 8, 4, 19

Table 6. Reported results for new vessel detection methods

Method	Year	Database	Level	AUC	Acc	SN	SP
Hassan ^[23]	2011	Combination	Pixel	0.7045	-	0.6390	0.8940
Welikala ^[24]	2013	Public	Image	-	-	1.0000	0.7000
Jelinek ^[25]	2007	Local	Image	0.9000	-	0.9400	0.8200
Goatman ^[26]	2011	Local	Image	0.9110	-	0.8420	0.8590
Akram ^[27]	2013	Public	Segment	0.9800	0.9800	0.9800	0.9700
Agurto ^[31]	2012	Local	Image	0.9400	-	0.9600	0.8300
Lee ^[32]	2013	Public	Image	0.9930	0.9850	0.9630	0.9910
Vatanparast ^[33]	2012	Local	Patch	-	-	0.9962	0.9661
Welikala ^[34]	2014	Combination	Image	0.9682	0.9333	1.0000	0.9000
Proposed (21 features)	2015	Combination	Image	0.9734	0.9667	1.0000	0.9500
Proposed (7 features)	2015	Combination	Image	0.9505	0.9667	1.0000	0.9500

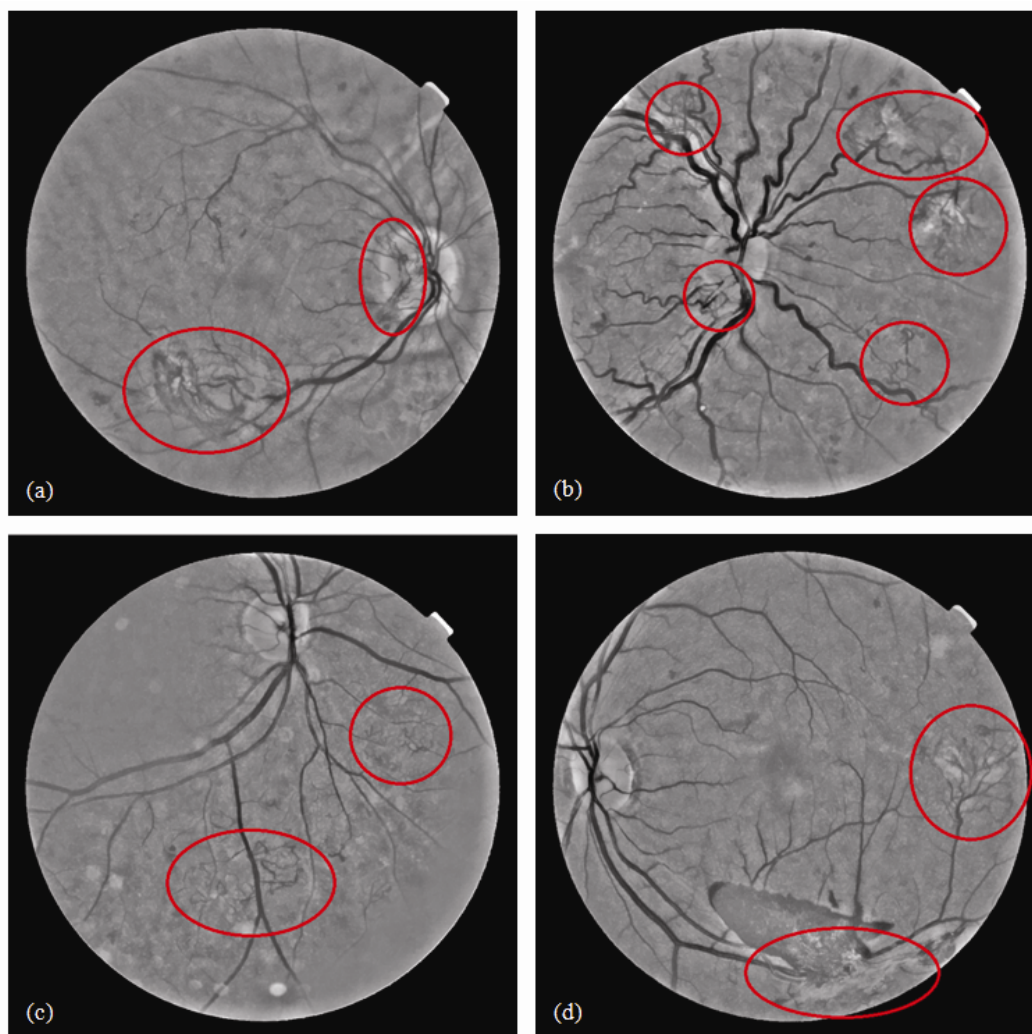


Figure 11. Location and delineation of new vessel regions marked by an ophthalmologist

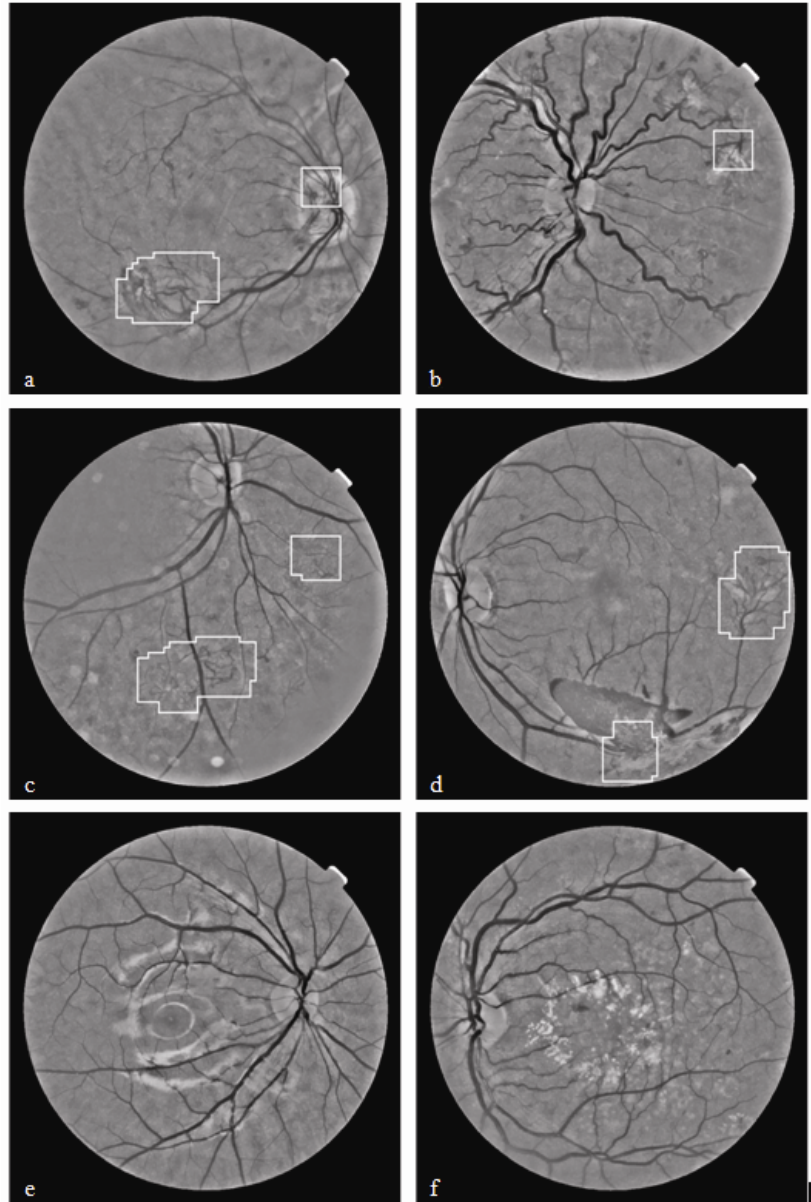


Figure 12. Results of the proposed system, per image basis (7 features). (a)-(d) True positive images, (e)-(f) true negative images.

4 Discussion and conclusion

In this paper, we have presented an effective new vessel detection method based on an ensemble classifier of boot strapped decision trees combined with dual classification. This also included the dimensionality reduction of a 21-D feature vector consisting of morphology, intensity and gradient based features.

The main contribution of this paper was the integration of ensemble classification which replaced the support vector machine classification used in Welikala *et al.* [34]. The ensemble method utilized multiple classifiers to accomplish gain in classification performance by mixing the outcomes from several weak learners into one high-class classifier. Decision trees were used as the classification model and the results of these weak learners were combined using bootstrap aggregation (bagging). An advantage of using the bagged ensemble was its ability to predict the importance of each feature during the training phase. This was useful to identify the most relevant features, allowing for the reduction of the feature vector dimensionality and thus increasing the computational speed of the system.

The results from table 6 show the proposed system performs better than Welikala *et al.* [34], achieving a SN of 1.0000 and a SP of 0.9500 compared to a SN of 1.0000 and a SP of 0.9000 on a per image basis. The AUC value also increases from 0.9682 to 0.9734. From the examples of classified images shown in Figure 12 it is clear that the algorithm responds well to a range of new vessel formations, including NVE, NVD, new vessels with associated fibrosis and obscure new vessels. Also evident is the algorithm's ability to avoid false responses despite the presence of bright lesions, dark lesions and reflections artefacts.

Feature removal using the feature importance index allowed for large reductions in the dimensionality of the feature vectors and therefore saved significantly on computational time. More than 66% of features were removed, with only 7 features being retained for each feature vector. These were features 4, 2, 1, 8, 17, 3, 7 for classification 1 and features 2, 1, 3, 17, 8, 4, 19 for classification 2 (see Table 5). The retained features are mainly morphology based, see table 2 for a summary of the features. However, on this occasion feature removal did not result in an increase to the classifier's performance, with the system's performance remaining the same at a SN of 1.0000 and a SP of 0.9500 (there was a slight drop in the AUC value to 0.9505).

The system puts no emphasis on correctly detecting all new vessels. Instead identifying any part of any new vessel region in the image is sufficient for the image to achieve a new vessel label. Figure 12(b) illustrates how there is no requirement to identify all new vessels in the image, with only one out of the five new vessel networks being identified. Figure 11(b) shows the same image with the location of all five new vessel networks marked by an ophthalmologist. Such an approach assists in ensuring a higher specificity is obtained.

It is evident from Figure 12 that the proposed method delineates new vessel regions to some extent, despite this it was not an objective to evaluate its performance in these terms. Striving to achieve a more complete delineation of new vessel regions would make the algorithm more susceptible to false positives. Therefore, our main concern is to identify whether new vessel regions are present in an image, as opposed to their accurate delineation. This suits the clinical requirements for identification of new vessels within a retinal screening program.

From Table 6 it is evident that the proposed method performs well in comparison to the other published methods. However, true comparisons are difficult to make as there exist no standard datasets for testing. Therefore, to assist in the development of PDR detection algorithms it is vital that a large new vessel data set becomes publicly available. Another difficulty in making comparisons is due to the variability in terms of their application. Jelinek [25] applied their methods on fluorescein images as opposed to conventional retinal images. Goatman [26] and Agurto [31] seek to only detect NVD.

Further developments of this method will involve the inclusion and evaluation of the boosting technique for comparison to our current methodology. Boosting [41] also creates an ensemble of classifiers by re-sampling the data, which is then combined by majority voting but it takes a different re-sampling approach than bagging. Our final goal is to develop an interactive retinal image analysis software tool focusing on early detection of PDR in screening programs. Prior to this our algorithm will need to be more stringently tested by using a larger new vessel data set. Therefore the aim is to continue to work in collaboration with local hospitals in expanding the data set, with the intent to publicly release.

In conclusion, this paper has demonstrated an automated system that is capable of detecting the presence of new vessels whilst reducing false responses to bright lesions, dark lesions and reflection artefacts. This involved the use of an ensemble classifier combined with dual classification.

Acknowledgments

This research was funded by the Faculty of Science, Engineering and Computing, Kingston University London.

References

- [1] Kocur I, Resnikoff S. Visual impairment and blindness in Europe and their prevention. *Br J Ophthalmol*. 2002; 86: 716-722. <http://dx.doi.org/10.1136/bjo.86.7.716>
- [2] De La Cruz JP, Gonzalez-Correa JA, Guerrero A. Pharmacological approach to diabetic retinopathy. *Diabetes Metab Res Rev*. 2004; 20: 91-113. <http://dx.doi.org/10.1002/dmrr.432>
- [3] Bek T, Erlandsen M. Visual prognosis after panretinal photocoagulation for proliferative diabetic retinopathy. *Acta Ophthalmol Scand*. 2006; 84: 16-20. <http://dx.doi.org/10.1111/j.1600-0420.2005.00574.x>
- [4] Klonoff DC, Schwartz DM. An economic analysis of interventions for diabetes. *Diabetes Care*. 2000; 23: 390-404. <http://dx.doi.org/10.2337/diacare.23.3.390>
- [5] Frank RN. Diabetic retinopathy. *Prog Retin Eye Res*. 1995; 14: 361-392. [http://dx.doi.org/10.1016/1350-9462\(94\)00011-4](http://dx.doi.org/10.1016/1350-9462(94)00011-4)
- [6] Kanski JJ, Bowling B. *Clinical Ophthalmology: A Systematic Approach*. Seventh edition. Elsevier Limited. 2011.
- [7] Dodson PM. *Oxford Diabetes Library: Diabetic Retinopathy*. Oxford University Press. 2008. <http://dx.doi.org/10.1093/med/9780199544967.001.0001>
- [8] Spencer T, Olson JA, McHardy KC, et al. An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. *Comput Biomed Res*. 1996; 29: 284-302. <http://dx.doi.org/10.1006/cbmr.1996.0021>
- [9] Niemeijer M, van Ginneken B, Staal J, et al. Automatic detection of red lesions in digital color fundus photographs. *IEEE Trans Med Imaging*. 2005; 24: 584-592. <http://dx.doi.org/10.1109/TMI.2005.843738>
- [10] Sopharak A, Uyyanonvara B, Barman SA. Simple hybrid method for fine microaneurysm detection from non-dilated diabetic retinopathy retinal images. *Comput Med Imaging Graph*. 2013; 37: 394-402. <http://dx.doi.org/10.1016/j.compmedimag.2013.05.005>
- [11] Tang L, Niemeijer M, Reinhardt JM, et al. Splat feature classification with application to retinal hemorrhage detection in fundus images. *IEEE Trans Med Imaging*. 2013; 32: 364-375. <http://dx.doi.org/10.1109/TMI.2012.2227119>
- [12] Zhang B, Wu X, You J, et al. Detection of microaneurysms using multi-scale correlation coefficients. *Pattern Recognit*. 2010; 43: 2237-2248. <http://dx.doi.org/10.1016/j.patcog.2009.12.017>
- [13] Walter T, Klein J, Massin P, et al. A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina. *IEEE Trans Med Imaging*. 2002; 21: 1236-1243. <http://dx.doi.org/10.1109/TMI.2002.806290>
- [14] Sopharak A, Uyyanonvara B, Barman SA. Automatic exudate detection from non-dilated diabetic retinopathy retinal images using fuzzy C-means clustering. *Sensors*. 2009; 9: 2148-2161. <http://dx.doi.org/10.3390/s90302148>
- [15] Sopharak A, Dailey MN, Uyyanonvara B, et al. Machine learning approach to automatic exudate detection in retinal images from diabetic patients. *J Mod Optic*. 2010; 57: 124-135. <http://dx.doi.org/10.1080/09500340903118517>
- [16] Osareh A, Mirmehdi M, Thomas B, et al. Automated identification of diabetic retinal exudates in digital colour images. *Br J Ophthalmol*. 2003; 87: 1220-1223. <http://dx.doi.org/10.1136/bjo.87.10.1220>
- [17] Fraz MM, Remagnino P, Hoppe A, et al. Blood vessel segmentation methodologies in retinal images - A survey. *Comput Methods Programs biomed*. 2012; 108: 407-433. <http://dx.doi.org/10.1016/j.cmpb.2012.03.009>
- [18] Ramlugun GS, Nagarajan VK, Chakraborty C. Small retinal vessels extraction towards proliferative diabetic retinopathy screening. *Expert Syst Appl*. 2012; 39: 1141-1146. <http://dx.doi.org/10.1016/j.eswa.2011.07.115>
- [19] Zhang B, Zhang L, Zhang L, et al. Retinal vessel extraction by matched filter with first-order derivative of Gaussian. *Comput Bio Med*. 2010; 40: 438-445. <http://dx.doi.org/10.1016/j.combiomed.2010.02.008>
- [20] Zhang L, Li Q, You J, et al. A modified matched filter with double-sided thresholding for screening proliferative diabetic retinopathy. *IEEE Trans Inf Technol Biomed*. 2009; 13: 528-534. <http://dx.doi.org/10.1109/TITB.2008.2007201>
- [21] Daxer A. Characterisation of the neovascularisation process in diabetic retinopathy by means of fractal geometry: diagnostic implications. *Graefes Arch Clin Exp Ophthalmol*. 1993; 231: 681-686. <http://dx.doi.org/10.1007/BF00919281>
- [22] Karperien A, Jelinek HF, Leandro JJG, et al. Automated detection of proliferative retinopathy in clinical practice. *Clin Ophthalmol*. 2008; 2: 109-122 PMID:19668394
- [23] Hassan SSA, Bong DBL, Premseenthil M. Detection of neovascularization in diabetic retinopathy. *J Digit imaging*. 2012; 25: 437-444. <http://dx.doi.org/10.1007/s10278-011-9418-6>
- [24] Welikala RA, Tah V, Williamson TH, et al. Differing matched filter responsivity for the detection of proliferative diabetic retinopathy. *Proceedings of the IASTED International Conference Signal Processing, Pattern Recognition and Applications*. 2013: 356-364. <http://dx.doi.org/10.2316/p.2013.798-096>

- [25] Jelinek HF, Cree MJ, Leandro JJ, et al, Luckie A. Automated segmentation of retinal blood vessels and identification of proliferative diabetic retinopathy. *J Opt Soc Am A Opt Image Sci Vis.* 2007; 24: 1448-1456. <http://dx.doi.org/10.1364/JOSAA.24.001448>
- [26] Goatman KA, Fleming AD, Philip S, et al. Detection of new vessels on the optic disc using retinal photographs. *IEEE Trans Med Imaging.* 2011; 30: 972-979. <http://dx.doi.org/10.1109/TMI.2010.2099236>
- [27] Akram MU, Khalid S, Tariq A, et al. Detection of neovascularization in retinal images using multivariate m-Medoids based classifier. *Comput Med Imaging Graph.* 2013; 37: 346-357. <http://dx.doi.org/10.1016/j.compmedimag.2013.06.008>
- [28] Frame AJ, Undrill PE, Olson JA, et al. Texture analysis of retinal neovascularization. *IEE Colloquium on Pattern Recognition (Digest No. 1997/018) (1997) 5/1-5/6.*
- [29] Acharya UR, Ng EYK, Tan J, et al. An integrated index for the identification of diabetic retinopathy stages using texture parameters. *J Med Syst.* 2012; 36: 2011-2020. <http://dx.doi.org/10.1007/s10916-011-9663-8>
- [30] Agurto C, Murray V, Barriga E, et al. Multiscale AM-FM methods for diabetic retinopathy lesion detection. *IEEE Trans Med Imaging.* 2010; 29: 502-512. <http://dx.doi.org/10.1109/TMI.2009.2037146>
- [31] Agurto C, Honggang Y, Murray V, et al. Detection of neovascularization in the optic disc using an AM-FM representation, granulometry, and vessel segmentation. *Annual International Conference of the IEEE, Engineering in Medicine and Biology Society (EMBC).* 2012: 4946-4949. <http://dx.doi.org/10.1109/embc.2012.6347102>
- [32] Lee J, Zee BCY, Li Q. Detection of neovascularization based on fractal and texture analysis with interaction effects in diabetic retinopathy. *PLoS One.* 2013; 8: e75699. <http://dx.doi.org/10.1371/journal.pone.0075699>
- [33] Vatanparast M, Harati A. A feasibility study on detection of Neovascularization in retinal color images using texture. *2nd International eConference on Computer and Knowledge Engineering.* 2012: 221-226.
- [34] Welikala RA, Dehmeshki J, Hoppe A, et al. Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification. *Comput Methods Programs Biomed.* 2014; 114: 247-261. <http://dx.doi.org/10.1016/j.cmpb.2014.02.010>
- [35] Fraz MM, Remagnino P, Hoppe A, et al. An ensemble classification based approach applied to retinal blood vessel segmentation. *IEEE Trans Biomed Eng.* 2012; 59: 2538-2548. <http://dx.doi.org/10.1109/TBME.2012.2205687>
- [36] Polikar R. Ensemble based systems in decision making. *Circuits and Systems Magazine, IEEE.* 2006; 6: 21-45. <http://dx.doi.org/10.1109/MCAS.2006.1688199>
- [37] Breiman L. Bagging predictors. *Machine Learning.* 1996; 24: 123-140. <http://dx.doi.org/10.1007/BF00058655>
- [38] Polikar R. Bootstrap - Inspired Techniques in Computation Intelligence. *Signal Processing Magazine, IEEE.* 2007; 24: 59-72. <http://dx.doi.org/10.1109/MSP.2007.4286565>
- [39] MESSIDOR [Internet]. Available from <http://messidor.crihan.fr>.
- [40] Zhang X, Thibault G, Decencière E, et al. Spatial normalization of eye fundus images. *IEEE International Symposium on Biomedical Imaging.* 2012.
- [41] Schapire RE. The boosting approach to machine learning: An overview. *Nonlinear Estimation and Classification.* 2003; 171: 149-171. http://dx.doi.org/10.1007/978-0-387-21579-2_9