

# SCIENTIFIC REPORTS



OPEN

## Automated Detection of Vessel Abnormalities on Fluorescein Angiogram in Malarial Retinopathy

Received: 19 November 2014

Accepted: 18 May 2015

Published: 08 June 2015

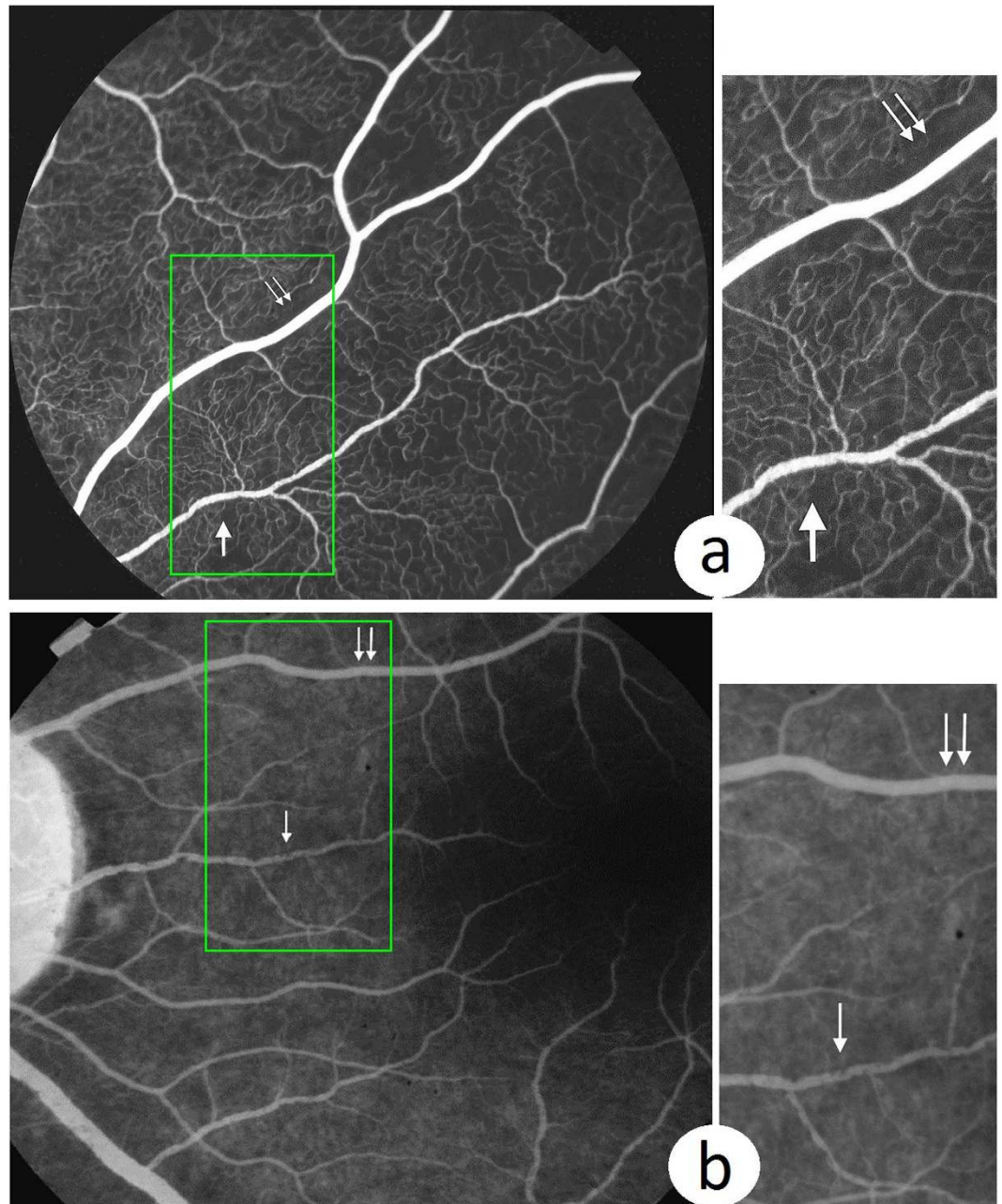
Yitian Zhao<sup>1,2</sup>, Ian J. C. MacCormick<sup>2,3</sup>, David G. Parry<sup>4</sup>, Nicholas A. V. Beare<sup>2,4</sup>, Simon P. Harding<sup>2,4</sup> & Yalin Zheng<sup>2,4</sup>

The detection and assessment of intravascular filling defects is important, because they may represent a process central to cerebral malaria pathogenesis: neurovascular sequestration. We have developed and validated a framework that can automatically detect intravascular filling defects in fluorescein angiogram images. It first employs a state-of-the-art segmentation approach to extract the vessels from images and then divide them into individual segments by geometrical analysis. A feature vector based on the intensity and shape of saliency maps is generated to represent the level of abnormality of each vessel segment. An AdaBoost classifier with weighted cost coefficient is trained to classify the vessel segments into normal and abnormal categories. To demonstrate its effectiveness, we apply this framework to 6,358 vessel segments in images from 10 patients with malarial retinopathy. The test sensitivity, specificity, accuracy, and area under curve (AUC) are 74.7%, 73.5%, 74.1% and 74.2% respectively when compared to the reference standard of human expert manual annotations. This performance is comparable to the agreement that we find between human observers of intravascular filling defects. Our method will be a powerful new tool for studying malarial retinopathy.

Cerebral malaria (CM) is a major cause of death and disability, especially in children in sub-Saharan Africa. CM is characterised by sequestration of parasitised erythrocytes in cerebral vessels<sup>1</sup>, but despite much research the mechanisms by which the intravascular malaria parasite causes coma and death remain unclear<sup>2,3</sup>. Malarial retinopathy (MR) has been identified as an important clinical sign in the diagnosis and prognosis of cerebral malaria<sup>4</sup>. The retina and brain are affected in similar ways in CM, and so the photographic features of MR are likely to give further valuable information about CM disease process, diagnosis, treatment and prognosis<sup>2</sup>.

Intravascular filling defects (IVFD) are a feature of MR that can be observed in fluorescein angiogram (FA) images. IVFD may represent sequestration of parasitised erythrocytes in the microvasculature<sup>5</sup>. Sequestration is the pathological hallmark of cerebral malaria<sup>6,7</sup>, but as yet, it has only been possible to quantify it histopathologically at post mortem. IVFD can be seen in large and small venules, arterioles and capillaries, but appear to be most prominent in venules. As shown in Fig. 1, the appearance ranges from mottling and slight irregularities of the vessel wall, to more obvious lesions that look as if small bites have been taken from the vessel<sup>4,8</sup>. Cerebral and retinal sequestration is always seen in fatal cases of CM with MR<sup>2,9</sup>, and the histopathological appearance of sequestration is similar to IVFD<sup>2,10</sup>. Moreover, IVFD often resolve the day after treatment with anti-malarial drugs is started (personal observation). This is consistent with resolution of sequestration and clinical recovery. It is plausible that IVFD represent this fundamental pathological process, and this lesion merits further investigation.

<sup>1</sup>School of Mechatronics Engineering, Beijing Institute of Technology, Beijing, China. <sup>2</sup>Department of Eye and Vision Science, University of Liverpool, Liverpool, UK. <sup>3</sup>Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Blantyre, Malawi. <sup>4</sup>St. Pauls Eye Unit, Royal Liverpool University Hospital, Liverpool, UK. Correspondence and requests for materials should be addressed to Y.Z. (email: yalin.zheng@liverpool.ac.uk)



**Figure 1. Two example fluorescein angiography images illustrating the appearances of IVFDs.** Vessels with IVFD are shown by single arrows. Vessels without IVFD, in the same image, are shown by double arrows. (a) Example 1: Intensity of mature parasitized red blood cells in vessels with IVFD is significantly different from normal vessels. (b) Example 2: Edges of vessels with IVFDs become unsmooth, the diameter is changed dramatically when compared to normal vessels. The images on the right are the zoom-in view of the regions enclosed by the green box within the original image on the left respectively.

Previous studies of FA in severe malaria were based on semi-quantitative data based on human observation<sup>11</sup>. Manual grading is often time consuming, is subject to observer variations<sup>12</sup>, and may not adequately capture important details such as the precise extent or location of lesions. By overcoming some of the limitations of manual grading, automated detection of IVFD may help to illuminate CM disease mechanisms. Unlike retinal haemorrhages, and capillary non-perfusion<sup>13</sup>, to the best of our knowledge automated quantification of IVFD has not yet been attempted. We address this by presenting a framework for automated detection of IVFD, with the aim of quantifying an under-researched retinal feature that has plausible links to the fundamental disease process involved in cerebral malaria.

Automated vessel analysis is an active research area in the field of medical imaging<sup>14</sup>. The primary effort has been focused on automated vessel segmentation, as evidenced by extensive reviews<sup>15,16</sup>, and

quantitative measurements of vessel geometry such as arteriovenous ratio (AVR), tortuosity, and fractal number<sup>14</sup>. There are few works on automatic vasculature analysis in FA<sup>17,18</sup>, and within this literature, the detection of discrete vessel abnormalities involving specific sections of the vessel wall has received little, if any, attention. Only one study addresses the related objective of detecting arteriolar narrowing in color fundus photography<sup>19</sup>. In their work<sup>19</sup>, a density analysis method is first used to detect the vessels, then connectivity analysis is performed to establish vessel trees, and finally arterioles are separated from venules by analysing vessel colour and width so as to assess arteriolar narrowing. This method had a sensitivity of about 75%.

We propose a new framework for automated detection of IVFD. Essentially we have formulated the problem in terms of image classification, where the objective is to train a classifier to determine if a vessel segment is normal or not based on a set of features that represent each segment. Throughout this paper, a vessel segment is defined as a connected segment of the detected vasculature between junctions or bifurcations, or a segment containing only one endpoint. The proposed framework will address three major challenges: 1) Accurate, efficient and reliable detection of vessels; 2) The process of deriving the features that are most discriminative and able to separate normal and abnormal vessels. 3) A classifier with good performance has to be identified and trained properly.

Our framework includes graph cut-based vessel segmentation, vessel geometry analysis, saliency map generation, and ensemble classification by AdaBoost (details of these technical components are described in the methods, below).

*Saliency* is a predictor of object regions which attract human attention. It indicates the relative importance of visual features and is closely related to characteristics of human perception and processing of visual stimuli<sup>20–22</sup>. Saliency originates from visual uniqueness, unpredictability, rarity, or surprise, and is often attributed to variations in image attributes like colour, gradient, edges, and boundaries<sup>23</sup>. Saliency in 2D images is the perceptual quality that makes an object, person, or pixel stand out relative to its neighbours, and that captures our attention<sup>22</sup>. Estimated saliency maps are widely used in many computer vision applications including object of interest image segmentation<sup>24</sup>, object recognition<sup>25</sup>, and so on. A pixel is salient if its appearance is unusual, considering the context of neighbouring pixels - one always looks at a pixel within its surrounding patch rather than simply observing a pixel in isolation. We define saliency in terms of information content: a key-point corresponds to a particular image location within a structure with a low probability of occurrence (i.e. high information content). Many saliency detection approaches for 2D images exist. They have a similar structure, computing several features in parallel and then fusing their values in a representation which is usually called a *saliency map*. The most general model of saliency detection is described by Itti and Koch<sup>21</sup>. Other existing saliency detection methods for feature determination can be divided into four classes: pixel-based methods<sup>21,26–30</sup>; region-based methods<sup>22,23,31</sup>; frequency-based methods<sup>32–35</sup>; parameter learning-based methods<sup>36–38</sup>.

In the case of IVFD, there is a contrast between the normally smooth vessel wall and individual discrete lesions that appear to protrude into the vessel lumen (Fig. 1(a)). These lesions may be defined as salient regions. Similarly, in the vessels affected by IVFD, some sections of the diameters or curvatures of vessel walls may be significantly different from neighbouring vessels or even other segments of the same vessel (Fig. 1(b)), such vessel edges may also be determined as salient features. These observations prompted us to use vessel intensity and shape saliency maps, and combine them to generate a combined saliency map.

## Results

In this section we describe the dataset used, evaluation metrics, experiments performed to evaluate the effects of various parameters, and the experimental results.

**Dataset.** Our automated framework was evaluated against a dataset containing 6,358 vessel segments (3,033 abnormal segments) from 10 retinal FA images with a size of  $3008 \times 1960$  pixels. These images were taken in the children with CM admitted to the Malaria Research Project Ward, Department of Paediatrics, Queen Elizabeth Central Hospital, Blantyre, Malawi. All subjects had signs of MR on admission. Ethical approval for retinal examination and imaging was given by committees in Blantyre and at collaborating institutions. Consent was given by the parents/guardians of subjects before examination and imaging. The tenets of the Declaration of Helsinki were adhered to. 50-degree images were taken after pupil dilation with Tropicamide 1% and Phenylephrine 2.5%, using a Topcon 50-EX optical unit (Topcon, Tokyo, Japan) and Nikon E1-H digital camera. Manual annotation of IVFD is extremely time consuming even aided by computer programs, it takes over an hour per image. Therefore, only 10 representative cases were selected for the evaluation of IVFD detection. We intentionally chose images that display a range of IVFD severity to create this dataset. This selection was made by ophthalmologists and professional graders who have been leading concurrent development of a protocol for manual grading of IVFD and other retinal features in cerebral malaria. Although the number of subjects is relatively small, we feel that these images represent a fair range of this spectrum.

Human expert graders used a systematic approach to label vessels as abnormal or normal in terms of IVFDs aided by an in-house Matlab program version 2013a (Mathworks, Natick, CA). During the process, the original and an overlay image of the original with centrelines of vessels highlighted in yellow were displayed side by side. Observers were asked to select abnormal and normal vessel segments in

turn by clicking on the vessel segment of interest. The selected abnormal segments were then highlighted in red while normal ones in green. In order to assess the detection performance of the framework on vessel segments with different diameters, the observers were asked to look at the peri-capillary vessels, small vessels or large vessels separately. Following our in-house FA grading workbook, we define capillaries as the smallest vessels visible on a well-focussed angiogram. A post-capillary venule is formed by the confluence of two or more capillaries, and extends up to the point where it is joined by a second post-capillary venule or other larger venular segment. Small venules are defined as any section of vein between the edge of the post-capillary venule complex up to the point of confluence with another vessel of similar or larger calibre. Large venules extend from the point where two small venules converge to the edge of the optic disc.

Three experienced observers in grading MR images were involved in the grading. A professional grader (DGP) and an ophthalmologist (IJCM) labelled the vessels using the same software and following the same guidelines in a masked pattern. The grading results by DGP were reviewed together by a senior ophthalmologist familiar with IVFDs (SPH) and the consensus between them was used as the final reference standard. When human graders were uncertain whether IVFDs were present or absent, vessels were left unlabelled and are not analyzed in this study.

**Evaluation Metrics.** Four commonly-used metrics were employed to evaluate the performance of the program in terms of vessel segment: sensitivity, specificity, accuracy, and the area under a receiver operating characteristic curve *AUC*. Sensitivity is a measure of effectiveness in identifying abnormal vessel segments while specificity performs the same function for normal vessel segments. Accuracy indicates the overall classification performance. *AUC* has the ability to reflect the trade-offs between the sensitivity and specificity in particular in the case of imbalanced data classification. These metrics are defined as follows:

$$\text{sensitivity (Se)} = \frac{tp}{tp + fn}, \quad (1)$$

$$\text{specificity (Sp)} = \frac{tn}{tn + fp}, \quad (2)$$

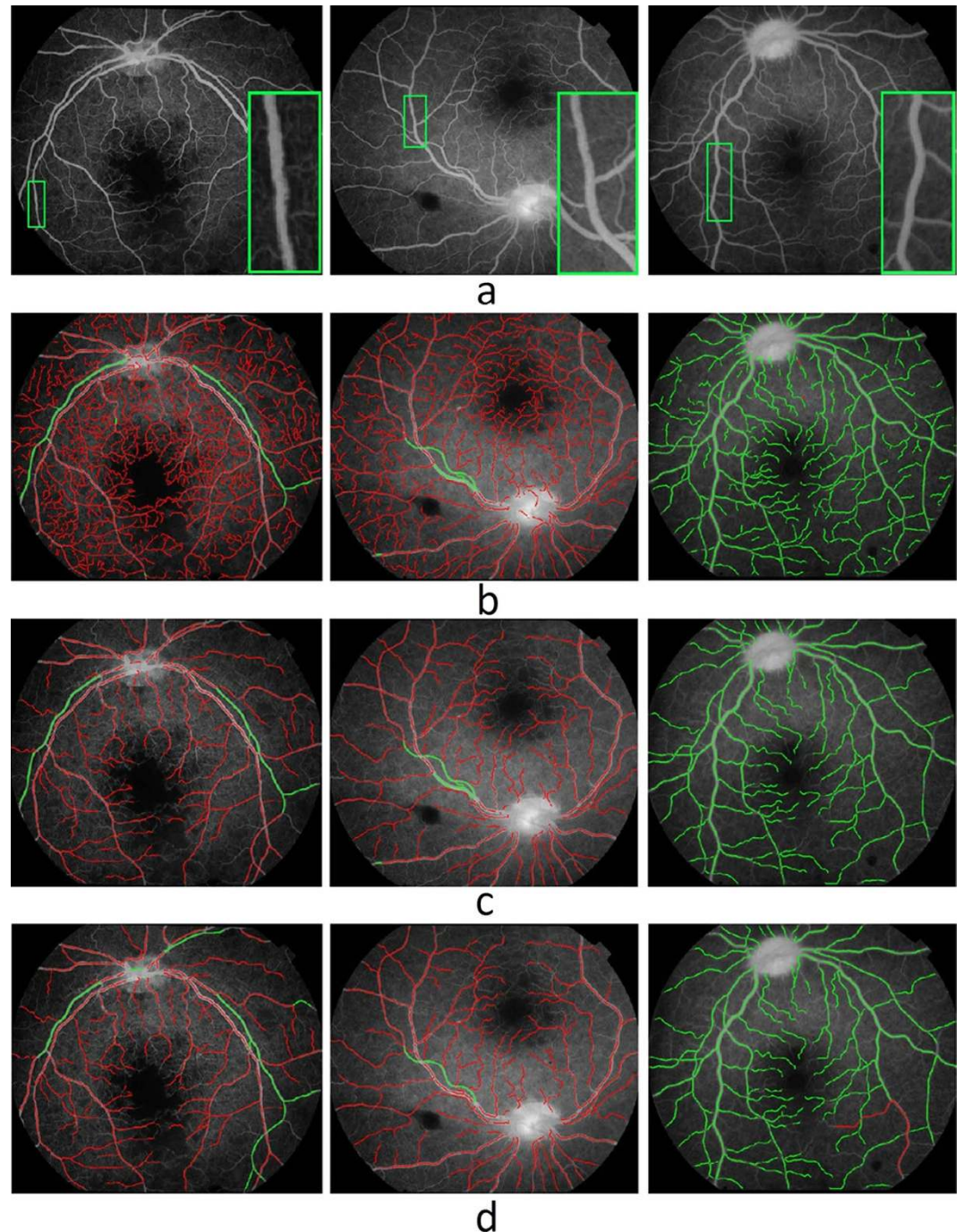
$$\text{accuracy (Acc)} = \frac{tp + tn}{tp + fp + tn + fn}, \quad (3)$$

$$\text{AUC} = \frac{Se + Sp}{2}. \quad (4)$$

where *tp*, *tn*, *fp* and *fn* indicate the true positive (the number of correctly identified abnormal vessel segments), true negative (the number of correctly identified normal vessel segments), false positive (the number of incorrectly identified abnormal vessel segments), and false negative (the number of incorrectly identified normal vessel segments), respectively. In particular *AUC* is calculated as suggested by Hong *et al.*<sup>39</sup>. An *AUC* of 1.0 means that the classifier distinguishes class examples perfectly.

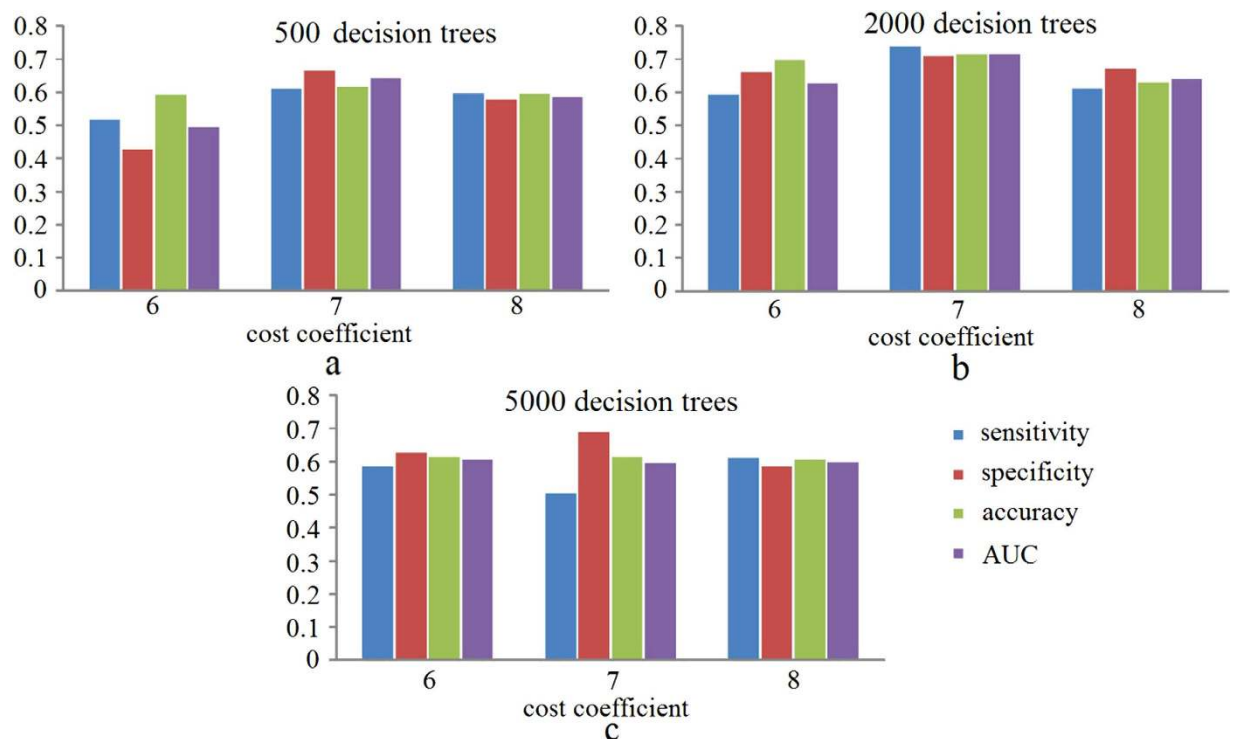
**Experiment Settings.** The 10 images in the dataset were randomly separated into a training set (8 images) and testing set (2 images). The training set was used to train and validate models while the testing set for evaluating the performance of the final model. An image-wise partition strategy was chosen in order to avoid possible overfitting, which could be introduced by a segment-wise partition strategy. With a segment-wise partition strategy, a classifier trained and tested on vessel segments from the same images may provide surprisingly good results on the training images, but perform poorly on new images. We applied repeated leave-one-out cross validation (LOOCV) to the training set for parameter optimization (or model selection)<sup>40</sup>. In brief, of the 8 images in the training set, 7 images were used to train a model while the remaining image was retained as the validation data for testing the model trained. The process was repeated 8 times with each single portion (image) used exactly once as the validation data. The LOOCV was then repeated five times on different random splits of the dataset, and the mean values of sensitivity, specificity, accuracy and *AUC* were used for comparisons of different parameter settings. The range tested for the number of trees was 500, 1000, 2000, 5000 and 10,000 while the range for the cost coefficient was 2 to 8 with an interval of 2. The 'optimal' values of the class weights and number of trees found from the repeated LOOCV were used to train the whole training set to obtain the final model. The performance of the final model was determined by applying it to the testing set. Sub-analysis on the performance of the final model for detection of vessel segments at different types was also performed.

**Experimental Results.** Figure 2 shows the results of the proposed automated abnormal vessel detection framework on 3 FA images where the normal vessels are illustrated in green colour whilst the abnormal vessels are in red colour. As we can see from Fig. 2(b), our method has classified all the vessel



**Figure 2.** Comparison of abnormal vessel detection between the automated framework and manual annotations (abnormal vessels are highlighted in red and normal vessels in green). (a) Three example fluorescein angiography images. The inset in each image shows the zoom-in view of the region enclosed by the green box within that image. (b) Detection results on all the segmented vessels by the framework. (c) Detection results on the gradable vessels only by the human observers. (d) Manual annotations from a consensus of two observers was used as the reference standard.

segments segmented by our vessel segmentation method into normal and abnormal segments respectively. However, there were a number of thin vessels that were ungradable for the human observers due to poor contrast. In this work, only the vessels labelled by human observers were considered for the purpose of comparison. Comparing results from our automated method (Fig. 2(c)) with those of human observer's (Fig. 2(d)), it can be seen that the results are visually very similar either in the case of lots of abnormal vessels contained images (Fig. 2 left and middle column) or fewer abnormal contained image (Fig. 2 right column).



**Figure 3.** Evaluation results in terms of sensitivity, specificity, accuracy, and area under the curve (AUC), under different combinations of decision trees and cost coefficients. (a) Results using 500 trees; (b) Results using 2000 trees; (c) Results using 5000 trees. On each plot, from left to right are the results with cost coefficient of 6, 7, and 8 respectively.

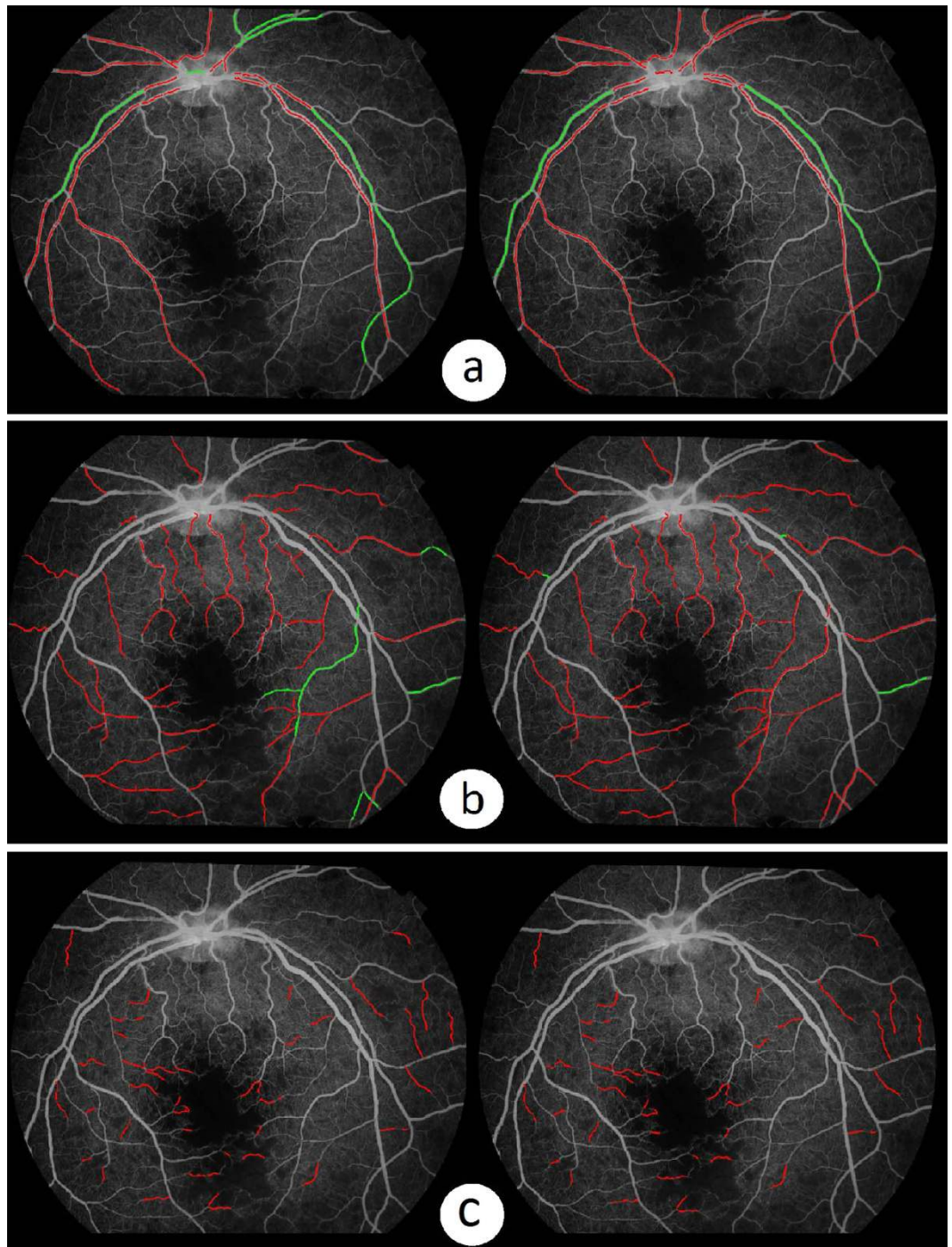
	sensitivity	specificity	accuracy	AUC
overall	0.747	0.735	0.741	0.742
large	0.732	0.713	0.729	0.728
small	0.765	0.782	0.751	0.776
peri-capillary	0.743	0.709	0.743	0.723

**Table 1.** Detection performance of the proposed framework for all the vessel segments, and for vessel segments with different types (large, small and peri-capillary vessels).

Figure 3 shows the classification performance for different values of the cost coefficient and different numbers of decision trees. It is clear that the classifier with 2000 trees and a cost coefficient of 8 was most effective in the detection of IVFD. With this set of parameters we were able to train a final model and apply it to the test set to obtain the evaluation results of the proposed method with the consensus annotations. Table 1 shows that the evaluation results in terms of sensitivity, specificity, accuracy, and AUC are 0.747, 0.735, 0.741, and 0.742, respectively. In addition, The overall inter-observer agreement for IJCM and DGP was found to be  $\kappa = 0.424$  ( $p < 0.001$ ) implying good agreement. The  $\kappa$  value for the framework and DGP is 0.555 ( $p < 0.001$ ).

In order to provide clinicians with more information about abnormalities in vessel segments, we also evaluated the performance on large, small and peri-capillary vessels separately. Figure 4 shows the results on one image by the program and the expert annotation side by side and Fig. 4(a–c) show the results on large, small and peri-capillary vessels respectively. The results for these three vessel types in terms of sensitivity, specificity, accuracy, and AUC were also presented in Table 1. Overall, the proposed abnormal vessel detection for the vessels from small vessel has the highest performance, which achieve sensitivity of 0.765, specificity of 0.782, accuracy of 0.751, and AUC of 0.776.

**Discussion and Conclusions.** We have developed a novel abnormal vessel detection framework to identify IVFD – a neurovascular sign that may represent an important part of CM pathogenesis. The framework comprises four major components: vessel segmentation, analysis of vessel geometry, salient feature generation, and vessel classification. Our evaluation of this framework yielded results that are



**Figure 4.** Illustration of the detection performance of the framework on vessels with different sizes (Left column: Manual annotations. Right column: Automated annotations). From top to bottom, (a) Results on large vessels. (b) Results on small vessels. (c) Results on peri-capillary vessels.

comparable to expert human observers. While much work has been done to develop tools to measure retinal vessel geometry, to the best of our knowledge this is the first report of automated analysis of discrete retinal vessel abnormalities.

Our method demonstrated satisfactory overall performance: sensitivity of 74.7%, specificity of 73.5%, and accuracy of 74.1%. In terms of vessel type-wise analysis, the framework achieved a sensitivity of 76.4%, specificity of 79.1%, and accuracy of 75.9% on small vessel. These results are consistent with the fact that there are relatively few large vessels, compared to smaller vessels. Unfortunately peri-capillary vessels were not typically photographed with sufficient quality for analysis to be accurate.

These promising results largely rely on our novel adaptation of the concept of salient features to the field of medical image analysis. In psychological terms, saliency is a predictor of visual object regions that attract human attention. Saliency indicates the relative importance of components of our visual world, and is closely involved in perception and processing of visual stimuli. In computational terms, saliency refers to a region or object that stands out from its neighbours or background. In this paper, we represented IVFD as salient regions on the background of the retinal image. IVFD can be thought of as minute vessel regions that have different diameter, curvature, or contrast to neighbouring regions. These features of IVFD are in line with the definition of saliency in computer vision field: the salient region is one that is significantly different from nearby regions in terms of contrast or shape.

Another highlight of our approach is the use of weighted ensemble classification method to deal with imbalanced data. This is very important as the proportion of abnormal to normal vessel segments in a retinal image is often skewed. A weighted classification strategy appears to be an appropriate way to penalize misclassification errors for each class differently. Furthermore, an ensemble classification technique will usually provide better performance compared to single classifiers. We chose weighted AdaBoost for this specific application because of its simplicity, efficiency and robustness against potential problem of overfitting. Other classification methods, such as weighted-SVM<sup>41</sup>, could also be used.

Automated analysis of retinal images is an important objective in medical research. The main emphasis has been on analysis of colour fundus photographs rather than FA, and on quantifying vessel geometry rather than identifying particular vessel segments affected by focal lesions. As a result the problem of detecting discrete vessel abnormalities is relatively unexplored. Achieving high performance in automated lesion detection is a challenging task. In our experience, there are many different factors that could compromise performance. First of all, there is often a very large variation in brightness, contrast, and artefact across images. This makes it difficult to have universal criteria to define the abnormalities. Secondly, IVFD can be difficult to grade, even for expert human graders. It is possible that an automated technique such as ours might provide more accurate detection than the current human expert reference standard.

Development of this framework is motivated by medical demands for a tool to measure the number of abnormal vessels in retinal FA images, and our method should allow better estimation of associations between MR and clinical outcome in patients with CM. This work is ongoing. The flexibility of this framework suggests it might be suitable for detecting abnormal vessel segments in other retinal or neurovascular diseases that involve discrete vascular lesions.

In conclusion, we have proposed and evaluated an innovative abnormal vessel detection framework to support the study of malaria retinopathy, and our experimental results have demonstrated its effectiveness. It has potential to be further developed as a useful tool for fast accurate and objective assessment for a range of retinal diseases.

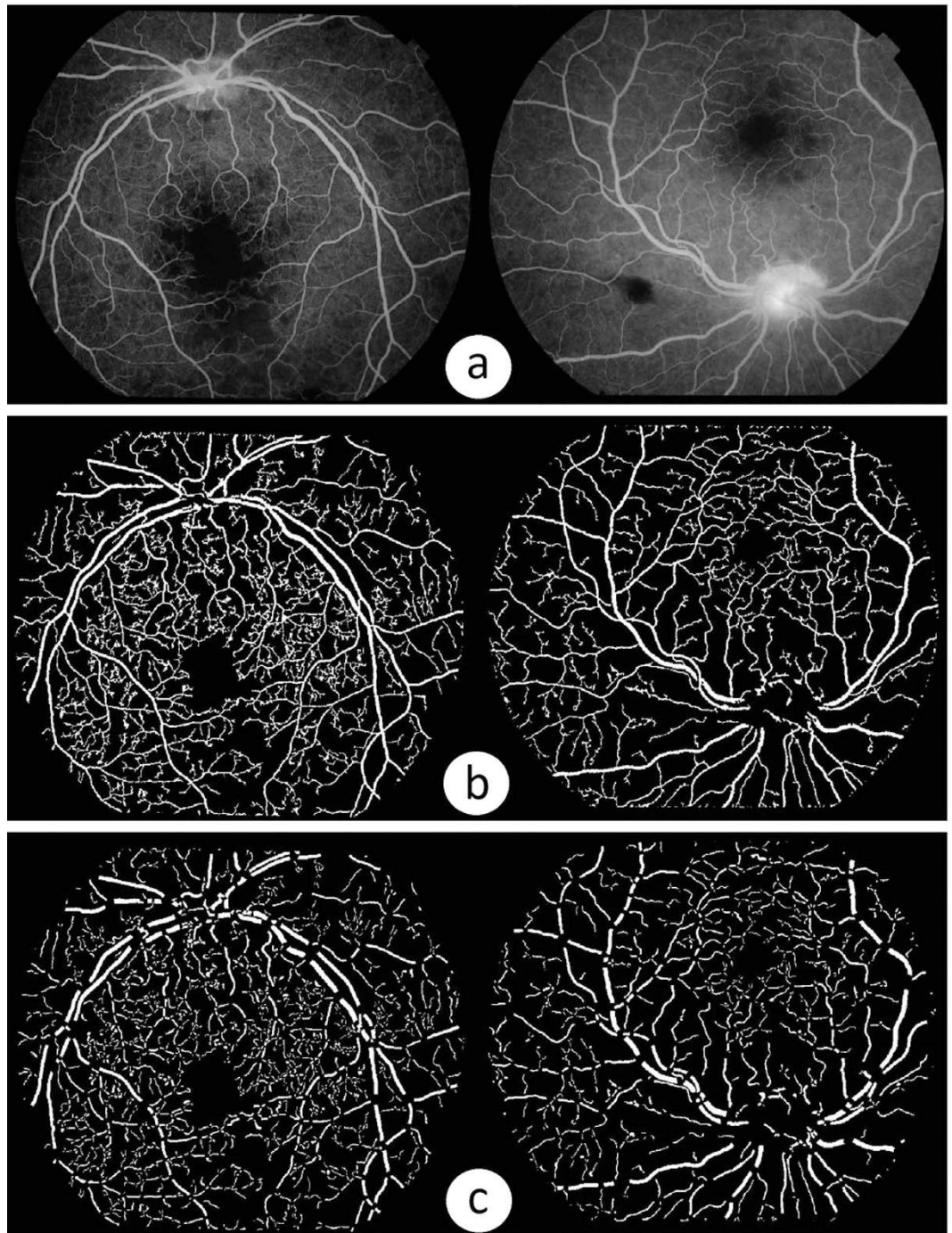
## Methods

In this section the proposed automated IVFD detection framework is described in detail.

**Vessel Segmentation.** The automated detection of blood vessels is a prerequisite in the development of automated system for the analysis of vessels. For this work, we adopted a state-of-the-art segmentation technique for its good accuracy and efficiency<sup>42</sup>. This technique is built on local phase enhancement and graph cut method. Local phase-based vessel enhancement is employed to enhance vessel-like structures in an image to form a ‘vesselness map’. As suggested by the name, this filter uses local phase information in the image to enhance vessel-like structures. Compared to the conventional intensity-based filters, this filter is invariant to intensity inhomogeneity within the image and also capable of producing more accurate enhancement results for vessels with different widths, even at the bifurcations or end of vessels. The vessels are segmented by applying a graph-cut based Chan-Vese (CV) model to the vesselness map for its computational efficiency. This model<sup>43</sup> as a region-based active contour model, segments the image into two regions (objects and background) by minimizing an energy for smooth boundary and low intra-region intensity variance. In this work we use the optimal parameter values as suggested by the original paper. In particular, for the graph-cut segmentation model, initialisation is achieved automatically by applying a threshold with an empirically chosen value of 0.5 to the vesselness map (afterwards ‘1’ denotes vessel pixel while ‘0’ background). Effects of different threshold values have been evaluated and it seems that the final results are not sensitive to it. Figure 5(a) shows two original example FA images, and their segmentation results are illustrated on Fig. 5(b).

**Geometric Analysis of Vessels.** Following the vessel segmentation step, geometrical analysis of the segmented vessels is performed in order to split the vasculature into individual segments for further processing. The morphological thinning algorithm is first applied to the segmented vessel trees in order to estimate the centre line and diameter of vessel segments: the exterior pixels from the segmented vessels are removed iteratively by using the thinning algorithm, and obtaining a new binary image containing connected lines of ‘on’ pixels locating along the vessel centres. The centerlines are refined by using a least-squared cubic spline technique in order to obtain smoother trajectories<sup>44</sup>. Branch points (>2 neighbours) are removed so as to divide the centerlines of the vascular tree into individual portions where each portion corresponds a vessel segment. Segments with a short centreline length (<10 pixels)





**Figure 5. Illustration of vessel segmentation and geometric analysis results.** (a) Two example fluorescein angiography images. (b) Vessel segmentation results. (c) Vessel segments after removing branch pixels from images in the second row. In (b) and (c) pixels in white denote vessels.

are eliminated to improve the speed of the later processing. Guided from the centerline location of each segment, individual vessel segments will be isolated from the original segmentation result by removing the branch points and their neighbour pixels. Figure 5(c) demonstrates the vessel segments produced after removing the branch pixels and pixels around them of Fig. 5(b).

The vessel diameters of each segment are estimated by using the distance transform of the inverted binary segmented image as suggested by<sup>45</sup>. It uses the Euclidean distance of each vessel pixel from the closest non-vessel pixel, and thus, doubling the maximum values of the distance transform along the thinned centerlines provides an estimate of the diameter of every vessel segment at its widest point. Bankhead *et al.* has demonstrated that this method can provide good width estimation results at locations in the middle of vessel segments<sup>45</sup>. It seems to us that this method may suffer at the two ends of

vessel segments due to the complex geometry. In order to avoid this problem, only diameters at locations 5 pixels away from branch (or end) pixels are considered for the subsequent analysis. A segment will be removed if its centreline contains fewer pixels than its estimated mean diameter. After this process, each segment will be indexed for subsequent analysis.

**Feature Generation.** To classify the vessel segments detected in the previous step as normal or abnormal, a set of features need to be derived to represent each vessel segment so as to form an input vector for the classifier to be used. In this work, for each segment, a total of 21 features including intensity and shape saliency maps are generated.

**Intensity-based Vessel Saliency.** Let  $w(x) \in V$  to be the viable local representation as a patch that represents pixel  $x$ , and  $V$  indicates all the vessel segments. The average vessel diameter of our dataset is around 5 pixels, so the size of the patch is set as  $3 \times 3$  in this work, where  $x$  is the centre of the patch. The patches can be seen as samples of a multivariate probability function (PDF). A number of methods to estimate an unknown multivariate PDF with a sufficient number of samples have been introduced in the literature. The kernel density estimator (KDE) is chosen in this paper. The KDE is appropriate since it is non-parametric, which will allow to estimate any PDF. Therefore, the probability of a patch  $w(y)$  can be defined as

$$p(w(y)) = \frac{1}{Nh} \sum_{x \in V} K\left(\frac{d(w(y), w(x))}{h}\right), \quad (5)$$

where  $d$  is a distance function that will be discussed later,  $K$  is a kernel,  $h$  is a smoothing parameter, and  $N$  represents the number of pixels. KDE method is capable to blur the contribution of each sample  $x$  by spreading it to a certain area in vessel segments with a certain shape<sup>46</sup>, which is defined by  $K$ . The multivariate distribution will have higher probability if the patches are in dense areas. From our experience, the most commonly used and the most appropriate kernel is Gaussian function with zero mean and standard deviation  $\sigma_k$ . Using a Gaussian kernel, Eq. (5) is rewritten as

$$p'(w(y)) = \frac{1}{N\Gamma} \sum_{x \in V} e^{-\frac{d^2(w(y), w(x))}{2\sigma_k^2}}, \quad (6)$$

The estimated probabilities are taken from an actual PDF by setting a proper constant  $\Gamma$ .  $\sigma = 0.2$  is chose to substitute  $h$ . After determining the probability of the patches, the intensity-based saliency measure can be defined as follows:

$$SI(y) = -\log\left(\frac{1}{N\Gamma} \sum_{x \in V} e^{-\frac{d^2(w(y), w(x))}{2\sigma_k^2}}\right). \quad (7)$$

In our application, the intensity-based saliency finally will be normalized into range  $[0,1]$ .  $d$  is estimated by relative average distance. The relative distance is used in case the distribution of the data is not uniform, and the distance metric mainly focuses on the relationships between neighboring points. Let a patch set  $W$  in a vessel contains  $n$  patches  $w_1, w_2, \dots, w_n$ . The relative average distance of a pair of patches  $w(x), w(y) \in W$  is defined as follows:

$$d(w(y), w(x)) = \frac{\|w(x) - w(y)\|_2}{ave_{w(k) \subset W}(\|w(x) - w(k)\|_2)} \quad (8)$$

The  $ave_{w(k) \subset W}(\|w(x) - w(k)\|_2)$  are the average Euclidean distance between  $w(x)$  and other patches  $w(k)$  belonging to  $W$  respectively. For two sets of points/pixels with similar neighboring relationships but different densities (i.e., similar relative density), the absolute distances between corresponding points differ dramatically from each other, but the relative distances are in general similar. This is an advantage of the relative distance in reflecting the relative density of points and relative scale of the objects.

**Shape-based Vessel Saliency.** Let  $u$  be the diameter of  $p_1$  and  $p'_1$ , and  $v$  be the diameter of  $p_2$  and  $p'_2$ . Let  $c_1$  and  $c_2$  be the centre points of these two diameters, and their coordinates  $c_1 = (x_u, y_u)$  and  $c_2 = (x_v, y_v)$ . Denotes  $u(p_1, p'_1), v(p_2, p'_2)$  are two random diameters of a given vessel, where  $(p_1, p'_1)$  and  $(p_2, p'_2)$  are the edge points on the vessel. The dissimilarity values of diameter  $u$  due to  $v$  in terms of length is given by  $\mathcal{L}(u, v)$ , where,

$$\mathcal{L}(u, v) = \frac{\|\mathcal{L}_u - \mathcal{L}_v\|}{\sqrt{(x_u - x_v)^2 + (y_u - y_v)^2}}, \quad (9)$$

where  $\mathcal{L}$  is diameter length, and  $\|\cdot\|$  calculates the Euclidean distance of centre points of diameter of  $u$  and  $v$  in the vessel. The dissimilarity values of the orientation of each centreline pixels  $c_1$  and  $c_2$  is calculated as

$$\Theta(u, v) = \frac{\|\Theta_u - \Theta_v\|}{\sqrt{(\mathbf{x}_u - \mathbf{x}_v)^2 + (\mathbf{y}_u - \mathbf{y}_v)^2}}, \quad (10)$$

where  $\Theta$  is the orientation of each pixel located on the centreline. After all the dissimilarity values in terms of diameter length and orientation, they are fused as weighted values  $\mathcal{W} = \sqrt{\mathcal{L}^2 + \Theta^2}$ . A dissimilarity measure between a pair of diameters may be given as:

$$diss(u, v) = \frac{\|\mathcal{W}(u) - \mathcal{W}(v)\|}{1 + h \cdot d_{position}(u, v)} \quad (11)$$

where  $h$  is the control parameter, and  $h=3$  in our implementation. Let  $d_{position}(u, v)$  be the Euclidean distance between the centre points  $c_u$  and  $c_v$  of the two diameters of  $u$  and  $v$ .

We need to compute a distinctness value for each diameter, given the dissimilarity values calculated above. Diameter  $u$  is considered salient when it is highly dissimilar to other vertices, i.e., when  $diss(u, v)$  is high:  $\forall v$ . The saliency value of  $u$  is defined as

$$SS(u) = 1 - \exp\left\{-\sum_{u=1}^U diss(u, v)\right\} \quad (12)$$

where  $U$  is the total number of the diameter in a given vessel. However, in practice, to evaluate the uniqueness of a diameters, there is no need to incorporate its dissimilarity to all the other diameters. If the most similar diameters (low dissimilarity diameters) are significantly different from diameter  $u$ , then clearly all diameters are also highly different from diameter  $u$ . Therefore, for diameter  $u$ , we search for the  $M$  most similar diameter according to the dissimilarity values, and define the diameters set as  $\{Q_m\}_{m=1}^M$ . Hence, the saliency value of diameter  $u$  can be rewritten as

$$SS(u) = 1 - \exp\left\{-\frac{1}{M} \sum_{m=1}^M diss(u, Q_m)\right\}. \quad (13)$$

In practice,  $M$  is the number of diameters whose dissimilarity value are higher than the average dissimilarity value. Similarly, the shape-based saliency values are also normalized into [0,1].

After obtaining the saliency values for each pixel of vessel and vessel centreline, the shape-based saliency and intensity-based saliency are simply combined together into a final saliency map SM ( $SM = SI + SS$ ), as shown as Fig. 6(b), the blue colour indicates the most salient regions, and the red colour shows the least salient regions. Two example images were selected: one with many abnormal vessel contained (top image of Fig. 6), and one with few abnormalities (bottom image of Fig. 6). It is clear that the salient regions in the top image are relatively more than the salient regions from the bottom one.

According to the pixel number of abnormal regions of each vessel in the final saliency map, the abnormality rate  $R(\mathbf{v})$  for each vessel is calculated as:

$$R(\mathbf{v}) = \frac{\text{the number of pixels belonging to abnormal regions of vessel } \mathbf{v}}{\text{the total number of pixels belonging to vessel } \mathbf{v}}. \quad (14)$$

Figure 6(c) illustrates the saliency map after the thresholding process applied: the vessel regions where their saliency values is larger than an empirically defined threshold value of 0.65 will be set to 1 (abnormal), otherwise will be decided to 0 (normal).

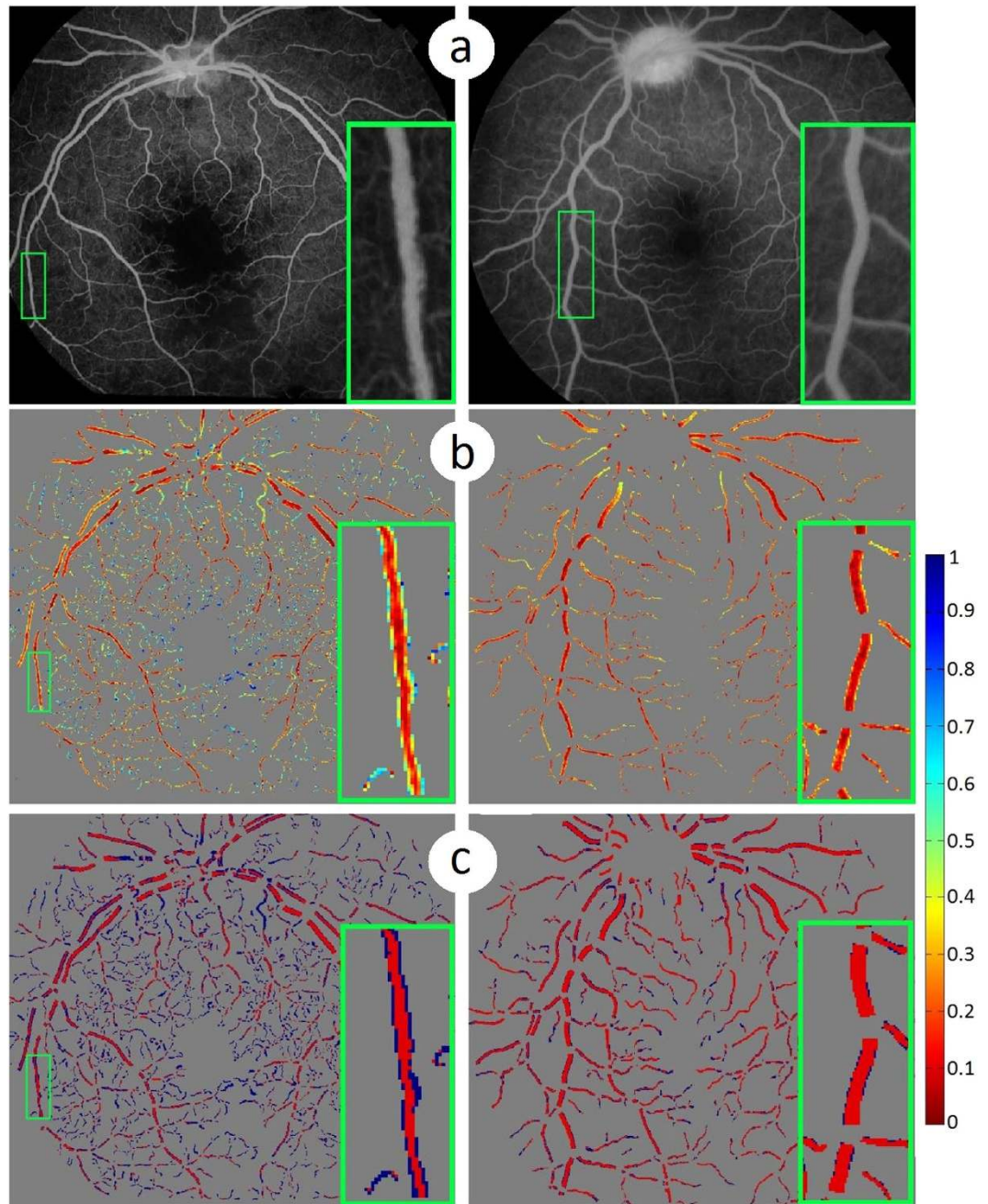
**Feature Vector.** Based on the saliency maps derived above, a feature vector of 21 features is derived for each vessel segment. These features are listed below:

*Feature 1–4.* Mean, standard deviation, entropy and the sum of gradient magnitude of the intensity-based saliency SI within the segment.

*Feature 5–8.* Mean, standard deviation, entropy and the sum of gradient magnitude of the shape-based saliency SS within the vessel segment.

*Feature 9.* Saliency-based abnormality rate  $R$  of the vessel segment.

*Other Feature 10–21.* Curvatures of the edges and the centerline of the vessel. Mean, standard deviation and entropy values of orientations of the two edges and the centerline.



**Figure 6. Illustration of saliency maps.** (a) Two example fluorescein angiography images. (b) Saliency maps of each individual vessel segments. (c) Vessels are divided into salient and non-salient regions after applying thresholding process to images in the second row respectively. Blue colour indicates the most salient regions while red colour shows the least salient regions for images in the second and third rows. The inset in each image shows the zoom-in view of the region enclosed by the green box within that image.

**AdaBoost Classification.** In this work we have used the AdaBoost classifier<sup>47</sup> with weighted cost coefficient classifier for the purpose of classification task. AdaBoost works by building a stronger and more powerful classifier from lots of smaller weak classifiers. We used a decision tree as the weak classifier<sup>47</sup>. The weak classifiers are generated sequentially in order to decrease the estimation error of the previous weak classifier<sup>48</sup>. Although various classification techniques have been proposed, such as artificial neural networks, support vector machine (SVM), decision trees, the choice of classifier is dependent on the complexity of that specific application and the nature of the data. The reasons for our choice of weighted AdaBoost are three-fold. First, AdaBoost is relatively simple, easy to train and less susceptible to over-fitting than other classifiers. As such it usually provides relatively good performance for most classification problems<sup>48</sup>. Second, as an ensemble classifier it can be more effective than a single classifier

in many cases, though this depends on the statistical properties of the data being analysed. Third, different weights can be easily introduced to tackle challenging classification problems. A weighted AdaBoost classifier has two parameters (class weights and number of trees) that have to be optimized in order to achieve the best classification performance. As described in Section Experiments, these are determined by the repeated LOOCV.

## References

- Milner, R. in *Cerebral Angiogenesis 15th edn*, Vol. 1135 (Springer, 2014).
- MacCormick, I. *et al.* Cerebral malaria in children: using the retina to study the brain. *Brain* **137**, 2119–2142 (2014).
- World Health Organization (WHO). Severe malaria. *Trop Med Int Heal* **19**, 7–131 (2014).
- Beare, N., Lewallen, S., Taylor, T. & Molyneux, M. Redefining cerebral malaria by including malarial retinopathy. *Future Microbiol* **6**, 349–355 (2011).
- Beare, N., Glover, S. & Molyneux, M. Malarial retinopathy in cerebral malaria. *Am J Trop Med Hyg* **80**, 171 (2009).
- White, N., Turner, G., Day, N. & Dondorp, A. Lethal malaria: marchiafava and bignami were right. *J Infect Dis* **208**, 192–198 (2013).
- Taylor T. *et al.* Differentiating the pathologies of cerebral malaria by postmortem parasite counts. *Nat Med* **10**, 143–145 (2004).
- Beare, N., Harding, S., Taylor, T., Lewallen, S. & Molyneux, M. E. Perfusion abnormalities in children with cerebral malaria and malarial retinopathy. *J Infect Dis* **199**, 263–271 (2009).
- Barrera, V. *et al.* Severity of retinopathy parallels the degree of parasite sequestration in eye and brain in Malawian children with fatal cerebral malaria. *J Infect Dis* DOI: 10.1093/infdis/jiu592 (2014).
- Lewallen, S. *et al.* Clinical-histopathological correlation of the abnormal retinal vessels in cerebral malaria. *Arch Ophthalmol* **118**, 924–928 (2000).
- Beare, N. *et al.* Prognostic significance and course of retinopathy in children with severe malaria. *Arch Ophthalmol* **122**, 1141–1147 (2004).
- Beare, N. *et al.* Inter-observer concordance in grading retinopathy in cerebral malaria. *Ann Trop Med Parasitol* **96**, 105–108 (2002).
- Zheng, Y. *et al.* A comprehensive texture segmentation framework for segmentation of capillary non-perfusion regions in fundus fluorescein angiograms. *PLoS ONE* **9**, e93624 (2014).
- Patton, N. *et al.* Retinal image analysis: Concepts, applications and potential. *Prog Retin Eye Res* **25**, 99–127 (2006).
- Fraz, M. *et al.* Blood vessel segmentation methodologies in retinal images - A survey. *Comput Methods Programs Biomed* **108**, 407–433 (2012).
- Niemeijer, M. *et al.* Comparative study of retinal vessel segmentation methods on a new publicly available database. *SPIE Med Imag* **5370**, 648–656 (2004).
- Vargas-Cañas, R. & Liatsis, P. Retinal vessel segmentation in fluorescein angiograms using a committee of local experts. *Conf Proc Syst Sig and Image Proc (IWSSIP)* **2012**, 240–243, (2012).
- Perez-Rovira, A., Zutis, K., Hubschman, J. & Trucco, E. Improving vessel segmentation in ultra-wide field-of-view retinal fluorescein angiograms. *Conf Proc IEEE Eng Med Biol Soc* **2011**, 2614–2617, (2011).
- Hatanaka, Y. *et al.* Automated detection algorithm for arteriolar narrowing on fundus images. *Conf Proc IEEE Eng Med Biol Soc* **2005**, 286–289, (2005).
- Kadir, T. & Brady, M. Saliency, scale and image description. *Iny J Comput Vision* **45**, 83–105, (2001).
- Itti, L., Koch, C. & Niebur, E. A model of Saliency-Based visual attention for rapid scene analysis. *IEEE Trans Pattern Anal Mach Intell* **20**, 1254–1259 (1998).
- Achanta, R., Hemami, S., Estrada, F. & Susstrunk, S. Frequency-tuned salient region detection. *Conf Proc IEEE Computer Vis Pattern Recogn* **2009**, 1597–1604 (2009).
- Cheng, M. *et al.* Global contrast based salient region detection. *Conf Proc IEEE Computer Vis Pattern Recogn* **2011**, 409–416 (2011).
- Ko, B. & Nam, J. Object-of-interest image segmentation based on human attention and semantic region clustering. *J Opt Soc Am A* **23**, 2462–2470 (2006).
- Rutishauser, U., Walther, D., Koch, C. & Perona, P. Is bottom-up attention useful for object recognition. *Conf Proc IEEE Computer Vis Pattern Recogn* **2004**, 37–44 (2004).
- Koch, C. & Ullman, C. Shifts in selective visual attention: towards the underlying neural circuitry. *Hum Neurobiol* **4**, 219–227 (1985).
- Clark, J. & Ferrier, N. Modal control of an attentive vision system. *Conf Proc IEEE Intl Conf Comput Vis* **1988**, 514–523 (1988).
- Milanesi, R. Detecting salient regions in an image: from biological evidence to computer implementation. *PhD thesis*, University of Geneva (1993).
- Bruce, N. & Tsotsos, J. Saliency based on information maximization. *Adv Neural Inf Process Syst* **18**, 155–162 (2006).
- Gao, D., Mahadevan, V. & Vasconcelos, N. On the plausibility of the discriminant center-surround hypothesis for visual saliency. *J Vis* **8**, 1–18 (2008).
- Aziz, M. & Mertsching, B. Fast and robust generation of feature maps for region-based visual attention. *IEEE Trans Image Process* **17**, 633–644 (2008).
- Hou, X. & Zhang, L. Saliency detection: A spectral residual approach. *Conf Proc IEEE Computer Vis Pattern Recogn* **2007**, 1–8 (2007).
- Wang, Z. & Li, B. A two-stage approach to saliency detection in images. *Conf Proc IEEE Intl Conf Acoust Speech and Signal Proc* **2008**, 965–968 (2008).
- Guo, C. & Zhang, L. Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform. *Hum Neurobiol* **2008**, 1–8 (2008).
- Cui, X., Liu, Q. & Metaxas, D. Temporal spectral residual: fast motion saliency detection. *Conf Proc ACM Intl Conf Multimedia* **2009**, 617–620 (2009).
- Kienzle, W., Wichmann, F. A., Franz, M. O. & Scholkopf, B. A nonparametric approach to bottom-up visual saliency. *Adv Neural Inf Process Syst* **19**, 689–696 (2006).
- Seo, H. & Milanfar, P. Static and space-time visual saliency detection by self-resemblance. *J Vis* **9**, 1–27 (2009).
- Goferman, S., Zelnik-Manor, L. & Tal, A. Context-aware saliency detection. *IEEE Trans Pattern Anal Mach Intell* **34**, 1915–1926 (2012).
- Hong, X., Chen, S. & Harris, C. A kernel-based two-class classifier for imbalanced data sets. *IEEE Trans Neural Netw* **18**, 28–41 (2007).
- Hastie, T. in *The elements of statistical learning: Data mining, inference, and prediction 2nd edn* (eds Hastie, T. *et al.*), Ch. 7, 219–260 (Springer, 2009).

41. Zheng, Y., Yang, X. & Beddoe, G. Reduction of false positives in polyp detection using weighted support vector machines. *Conf Proc IEEE Eng Med Biol Soc* **2007**, 4433–4436 (2007).
42. Zhao, Y. *et al.* Retinal Vessel segmentation: An efficient graph cut approach with Retinex and local phase. *PLoS ONE* **10**, e0122332 (2015).
43. Daněk, O., Matula, P., Maška, M. & Kozubek, M. Smooth Chan-Vese segmentation via graph cuts. *Pattern Recogn Lett* **33**, 1405–1410 (2012).
44. Lee, E. Choosing nodes in parametric curve interpolation. *Comput Aided Des* **21**, 363–370 (1989).
45. Bankhead, P., Scholfield, C., N., McGeown & J. G., Curtis, T. M. Fast retinal vessel detection and measurement using wavelets and edge location refinement. *PLoS ONE* **7**, e32435 (2012).
46. Martins, P., Carvalho, P. & Gatta, C. Context-aware keypoint extraction for robust image representation. *British Machin Vis Conf (BMVC)* **2012**, 1–12 (2012).
47. Freund, Y. & Schapire, R. A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* **55**, 119–139 (1997).
48. Hijazi, M., Coenen, F. & Zheng, Y. Data mining techniques for the screening of age-related macular degeneration. *Know-Based Syst* **29**, 83–92 (2012).

## Acknowledgments

This work was supported by the Wellcome Trust (grant number 092668/Z/10/Z). We thank Dr Yonghuai Liu at the Department of Computer Science, Aberystwyth University for valuable suggestions.

## Author Contributions

Conceived and designed the experiments: Y. Zhao, Y. Zheng. Performed the experiments: Y. Zhao. Analyzed the data: Y. Zhao, Y. Zheng. Contributed reagents/materials/analysis tools: D.G.P., S.P.H., I.J.C.M. Wrote the paper: Y. Zhao, Y. Zheng, I.J.C.M. Shared helpful suggestions during the implementation of method: I.J.C.M., N.A.V.B., S.P.H., D.G.P.

## Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Zhao, Y. *et al.* Automated Detection of Vessel Abnormalities on Fluorescein Angiogram in Malarial Retinopathy. *Sci. Rep.* **5**, 11154; doi: 10.1038/srep11154 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>