An Automated System for Detecting and Measuring Nailfold Capillaries

Michael Berks, **Phil Tresadern**, **Graham Dinsdale**, **Andrea Murray**, **Tonia Moore**, **Ariane Herrick**, and **Chris Taylor**

University of Manchester, Manchester, United Kingdom

Abstract

Nailfold capillaroscopy is an established qualitative technique in the assessment of patients displaying Raynaud's phenomenon. We describe a fully automated system for extracting quantitative biomarkers from capillaroscopy images, using a layered machine learning approach. On an unseen set of 455 images, the system detects and locates individual capillaries as well as human experts, and makes measurements of vessel morphology that reveal statistically significant differences between patients with (relatively benign) primary Raynaud's phenomenon, and those with potentially life-threatening systemic sclerosis.

1 Introduction

Systemic sclerosis (SSc) is a connective tissue disorder which can lead to morbidity and mortality, often in young people – with a reported prevalence among adults of 250 per million [1]. Clinically, it results in fibrosis and microvascular abnormality, leading to ischaemic injury (e.g. ulceration, scarring, and gangrene), particularly in the fingers and toes. The commonest presenting feature is Raynaud's phenomenon (episodic colour change and pain in the fingers, usually in response to cold), but this is also a symptom of the more common, and relatively benign, primary Raynauds phenomenon (PR).

There is thus a clinical need to distinguish between PR and SSc-related Raynauds phenomenon. There is also a pressing need for quantitative biomarkers for monitoring SSc response to treatment, both clinically and in clinical trials, where existing endpoints (eg digital ulceration) are unreliable, leading to a limited evidence base [2, 3]. Nailfold capillaroscopy, a non-invasive technique for imaging capillaries at the base of the fingernails (see Figure 1), is already used clinically to assess the degree of microvascular abnormality, and has the potential to provide quantitative biomarkers for SSc. Standardised protocols for qualitative grading of nailfold images exist [4], but do not provide quantitative data. (Semi-)Manual measurements of capillary spacing, vessel width at the tops of loops (apices) and vessel tortuosity have been shown to have potential as quantitative biomarkers for SSc [5], but are too time-consuming and open to subjective factors for routine use. There is thus a clear rationale for developing automated methods for analysing nailfold images.

michael.berks@manchester.ac.uk.

In this paper we describe and evaluate a fully automated system for detecting and measuring capillaries in nailfold images, adopting a machine learning approach, and building on experience with existing semi-automated systems [5, 6]. Specifically, our contributions are: (a) a method for detecting vessels and estimating their width and orientation; (b) a method for generating candidate capillary apices; (c) a method for refining candidate capillary apices and measuring apical width and tortuosity; (d) a large-scale evaluation of performance, compared to human experts and; (e) initial results for classifying patients based on image level summary statistics of the automated capillary measurements.

In summary we show that our automated system is indistinguishable from human experts in detecting and locating capillary apices, and that there are statistically significant differences, at the population level, between automated capillary measurements for SSc patients and normal or PR subjects.

2 Nailfold Capillaroscopy Data

Our images were acquired by a capillaroscopy system in a tertiary referral centre for patients with SSc. Patients gave informed consent. High magnification, 768×576, 8-bit monochome pixel video frames were captured at a resolution of $1.25\mu m$ per pixel. Images captured along the nailbed were registered and compounded into a single mosaic (similarly to [7]) showing the whole nailfold (Figure 1).

In normal subjects, capillary loops are all similar, arranged regularly and approximately vertical. Current clinical practice confines attention to the distal row of capillaries, whose apices lie on a smooth, approximately horizontal line near the top of the image. This pattern is disrupted in SSc by structural damage to the microvasculature (Figure 1).

We have used a set of 990 mosaics, manually annotated as part of a separate clinical study, involving three subject groups: healthy controls (HC), subjects with PR, and patients with SSc. Each image was annotated independently by two expert observers, one of whom (Observer 1) was the same for every image, with the other drawn from a pool of experts. In each image, the observers attempted to mark the locations and apical widths of all distal capillaries. This is a challenging and subjective task for which perfect agreement is rare (see Section 6).

In a subset of 80 images, Observer 1 provided a precise demarcation of the inner and outer edges of the distal capillaries. Regions of interest around these capillaries were created, resulting in a set of 450 training RoIs with matching capillary masks, which we used for training learning algorithms (Section 3, Section 4, Section 5). The remaining images were split into a validation set of 456 images (104 HC, 83 PR, 269 SSc) used to locate capillaries and determine the distal row, and a test set of 455 images (104 HC, 83 PR, 268 SSc) used to evaluate performance (Section 6).

3 Vessel Detection and Characterisation

Nailfold images are challenging – capillaries are often very low contrast and of variable width ($10 - 300\mu m$) and appearance, whilst significant artefacts can be present. Although

previous work on automated capillaroscopy is limited, we can draw on the extensive literature on curvilinear structure detection in medical images (eg [8–10]. In particular, we adopt the well-established machine learning approach to predict, at each pixel, the probability that it belongs to a vessel (vesselness) [9], and the local orientation and width [10]. In the following sections we explain the features we used to describe local image structure, and the learning methods we used to obtain vesselness, orientation and width models.

3.1 Image Features

To characterise local structure at each training pixel, we use a feature vector of responses to symmetric (even) and asymmetric (odd) filters across scale and orientation. Specifically, we use a steerable filter bank [11] of directional second order derivatives of a Gaussian kernel [8] and their (approximate) Hilbert transforms so that responses at any orientation can be computed efficiently. From an initial scale $\sigma = 1$ pixel we compute even and odd responses at six angles $\theta_i = {i\pi}/{6}$ over five scales, in each case keeping σ fixed whilst downsampling the image by a factor of 2 in each direction (i.e. the coarsest scale is equivalent to $\sigma = 16$ at the original image resolution). For the coarser scales, we use bilinear interpolation to approximate the responses at the finest scale. At each pixel the responses at that pixel, along with its 8-connected neighbours, are concatenated into a feature vector without further manipulation. These features are designed to accommodate the wide range in size, shape and orientation of imaged capillaries.

3.2 Learning Vesselness, Orientation and Width Models

We treat vessel detection as a supervised classification problem, and orientation and width prediction as regression problems, all based on the features outlined in the previous section. We have used Random Forests [12], due to their ease of training, flexibility, relative robustness to over-fitting and strong performance in comparable learning tasks, but we do not believe the choice of learning method is critical.

For training, we used the 80 fully annotated images described in Section 2 each of which provides a binary vessel mask for training a vesselness classifier. To provide ground-truth for orientation and width regression, we skeletonised the binary masks, measured orientation and width at each centreline point, and propagated the measurements back to every point on the binary vessel mask using a simple nearest pixel interpolation. For orientation regression, we represented orientation as a unit vector in the complex plane, $t = \cos 2\theta + i \sin 2\theta$, doubling the angle θ to make orientation invariant to direction [13], and avoiding wraparound problems that arise if angle is used directly.

We trained three random forests (RFs), each containing 100 trees: one classifier on background versus labelled vessel pixels, and one regressor each for orientation and width. Although orientation and width are only defined for the vessel points, to ensure we generate unbiased, random predictions in the background of unseen images, we include background points with uniformly sampled random widths and orientations in training the regressors.

To make predictions in an unseen image we apply the separable basis filters, compute the interpolated, steered responses and extract feature vectors, before feeding the vectors through the trees in each forest and computing the mean, pooled over all leaf nodes, as the prediction output. For orientation, the unit vectors are converted back to radians. The result for each input image is a map of vesselness (V_v), orientation (V_θ) and width (V_w) (Figure 2(a–b)). Note that predictions of width and orientation made in this manner are significantly more accurate than analytic estimates, and, by reusing the same features as extracted for V_v, requires minimal additional computation.

4 Locating Candidate Capillary Apices

The maps V_v, V_θ and V_w provide an approximate segmentation of vessels and their low-level properties. However to extract meaningful measurements we need to detect and localise the apices of individual capillaries. Treating this as an object detection problem, we again adopt a learning approach in which each vessel pixel votes for the location of a nearby apex. Specifically, we use RF regression, trained using patches sampled near annotated capillary apices in the training data, to encode observed relationships between appearance and location [14].

To train the RF, we thin V_v by applying non-maximal suppression along the line normal to the estimated vessel orientation. We then extract 64×64 training patches, centred at local maxima in V_v and scaled and rotated according to the estimated width and orientation. If a patch contains a marked apex and is also centred at a vessel pixel in the original training masks, we label the patch as positive and record the offset to the apex; otherwise, we label the patch as negative. For each patch, a histogram of gradients (HoG) feature vector is formed by concatenating weighted histograms of gradient direction, computed for overlapping blocks in the patch [15]. A 100 tree RF classifier is then trained to distinguish between positive and negative patches, whilst a regression forest is trained to predict the offset associated with each positive patch.

For an unseen image, we extract a scale- and rotation-normalised patch about each local maximum in V_v (Figure 2(b)) and pass these through the classification forest. The output labels at each leaf node are pooled and averaged, forming $a(p)$, the probability that the patch contains an apex. The regression forest then predicts the offset to the nearest apex, and therefore the apex location in the image. A Gaussian kernel – centred at the apex, scaled by $\sigma = V_w(a)$ and weighted by $a(p)$ – is added to a vote map of apex locations A. Here the classifier plays a crucial role in weighting the regressor votes, so that only patches in which an apex is visible significantly contribute to A. The local maxima of A are the candidate locations for capillary apices (Figure 2(c)).

5 Refining Candidate Apices and Making Measurements

Given a set of candidate apex locations as training data, we can now train a new classifier that better discriminates between true apices and false positives, and between distal and non-distal apices. We do so in two ways: first, we reclassify each candidate, now using patches

located only at candidate apex locations; second, we exploit the fact that capillaries should lie in a single, approximately horizontal line across the image.

To train the new classifier, we used candidate apices computed for the set of 441 validation images. Each candidate (x_i, y_i) was labelled as positive only if it fell within a circle centred at a marked apex with diameter equal to the marked apical width. Of the 13,492 positive and 24,369 negative samples, the positive samples were further split into 9,735 distal and 3,757 non-distal capillaries. A 100 tree RF classifier was trained as before, using HoG features from scale- and rotation-normalised patches extracted about each sampled candidate from the original image (rather than V_v). The outputs of the trees were pooled to give a single appearance score, $c_i \in [0, 1]$, for each candidate, allowing us to define each candidate as the tuple of its location and score, (x_i, y_i, c_i). In addition to the final appearance score, the location of a candidate relative to the other candidates in any image can indicate its likelihood of being a distal row apex. Specifically, we assume the ideal distal row to be a smooth line running across the mosaic and passing through every true distal capillary apex. If we can estimate this line in an unseen image, we can use the vertical displacement of each candidate to the line to classify it as distal or nondistal.

To achieve this, we compute a density map of candidate locations by summing a Gaussian kernel centred at each discrete candidate location. Each kernel is weighted by the candidate's appearance score, c_i, so that strong candidates contribute more to the density. The density for each location in the mosaic can thus be computed as

$$D(u,v) = Z \sum_{i=1}^{N} c_i exp \left[-\frac{(u-x_i)^2}{2\sigma_x^2} \right] exp \left[-\frac{(v-y_i)^2}{2\sigma_y^2} \right] \quad (1)$$

where there are N candidates in the image, Z is a normalisation constant so that D sums to unity, and σ_x, σ_y are functions of the variances of x_i, y_i. Each apex's displacement from the distal row is then given by

$$d_i = y_i - arg \max_v D(x_i, v) \quad (2)$$

Using the observers' annotations, each candidate $C_i = (x_i, y_i, c_i, d_i)$ in the validation images is assigned a label $L_i \in [1, 2, 3]$ for false positives, distal capillaries and non-distal capillaries respectively. Figure 2(e) shows a kernel estimate of the class-conditional probability density $P(c_i, d_i | L_i)$. The class priors for each label type $P(L_i)$ can be estimated empirically from the data, allowing Bayes' rule to be used to compute the class probability, $P(L_i | c_i, d_i)$, of each candidate, given its appearance score and displacement.

Each candidate is rejected, or labelled as either a distal or non-distal capillary. For each kept candidate C_i we record its width $V_w(x_i, y_i)$, and use V_θ and V_v to compute the entropy of an orientation histogram of pixels connected to C_i as a measure of tortuosity. Finally, we compute capillary density as the mean distance between the distal apexes.

6 Results

We applied our detection method (O_3) to the set of 455 test images, and compared the selected capillaries to the annotations of the two human observers (O_1 and O_2). In total, 15,391 capillaries were selected by at least one of the three observers. The results illustrate the difficulty of the task. For example, of the 7047 distal capillaries (DCs) marked by O_1, only 56% were also marked by O_2, whilst 74% were marked by the automated system (O_3). Defining true positives via consensus between the experts, of the 3913 DCs marked by both O_1 and O_2, 84% were marked by O_3, which is a probably a workable performance level. Given the level of disagreement between the experts, false positives are even harder to assess. Of the 7913 DCs marked by O_3, 27% were marked by neither O_1 nor O_2. On the other hand, of the 7047 DCs marked by O_1, 17% were marked by neither O_2 nor O_3, and 44% were not marked by O_2.

Overall, the results suggest similar performance between O_1, O_2 and O_3. To test this formally, we considered the capillaries (both distal and non-distal) selected by each observer as truth and assessed how well each of the two observers detected these capillaries. Table 1 shows the precision, recall and F-measure for each pair of observers. To test the labelling of capillaries as distal or non-distal, we considered all capillaries jointly selected by each pair of observers, and computed accuracy and Cohen's kappa statistic to quantify agreement. Accuracy and kappa statistics were not computed for capillary detection because there is no defined true negative class for the task (this is effectively the space of all images not containing a capillary). In addition to the pairwise tests, we computed equivalent values for the software versus the consensus of the two humans.

Standard errors for each value were computed from 1,000 bootstrap samples of the test images. For the F-measures and Cohen's kappa statistics we computed the difference between the agreement of O_1 and O_2 and the average of their individual agreement with O_3. 95% confidence intervals of (-7.01, -3.90) and (-0.019, 0.12) respectively, suggest there was no significant reduction between the software's performance and either human on the two tasks (if anything the software showed greater agreement to the humans individually than the humans did to each other on the detection task).

Finally, we present initial results for disease status characterisation, based on the measurements extracted for the detected capillaries. Distributions of capillary measurements (capillary density, median width and median tortuosity) for the three groups HC (104 images), RP (83 images), and SSc (268 images) are shown in Figure 3. To test for differences between the distributions we used the non-parametric Wilcoxon rank sum test. Tests for all measurements between all groups showed significant differences at the 0.01 confidence level (at least), except for the comparison between RP and HC for vessel density – an expected result given that we would not expect the RP group to show any signs of capillary loss.

7 Conclusions

We have presented a fully automated system for measuring vessel morphology from nailfold capillaroscopy images. Evaluation on a large data set suggests that our system performs as well as experts in detecting vessels, and initial results for automated measurement suggest that we can detect significant differences between disease groups. Further work will involve refining apex detection, adding more sophisticated measurements, and combining measurements to make useful predictions for individuals.

Acknowledgements

References

1. Mayes M, Lacey J, Beebe-Dimmer J, Gillespie B, Cooper B, Laing T, Schottenfeld D. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large us population. Arthritis Rheum. 2003; 48:2246–55. [PubMed: 12905479]

2. Herrick AL. Contemporary management of raynaud's phenomenon and digital ischaemic complications. Current Opinion in Rheumatology. 2011 Nov.23:555–561. [PubMed: 21885977]

3. Herrick AL, Roberts C, Tracey A, Silman A, Anderson M, Goodfield M, McHugh N, Muir L, Denton CP. Lack of agreement between rheumatologists in defining digital ulceration in systemic sclerosis. Arthritis & Rheumatism. 2009; 60(3):878–882. [PubMed: 19248100]

4. Cutolo M, Pizzorni C, Secchi ME, Sulli A. Capillaroscopy. Best Practice and Research Clinical Rheumatology. 2008; 22(6):1093–1108. [PubMed: 19041079]

5. Murray AK, Moore TL, Manning JB, Taylor C, Griffiths CEM, Herrick AL. Non-invasive imaging techniques in the assessment of scleroderma spectrum disorders. Arthritis Rheum. 2009; 61(8): 1103–11. [PubMed: 19644893]

6. Paradowski M, Markowska-Kaczmar U, Kwasnicka H, Borysewicz K. KES. 2009; 2:151–158.

7. Anderson ME, Allen PD, Moore T, Hillier V, Taylor CJ, Herrick AL. Computerized nailfold video capillaroscopy – A new tool for assessment of raynauds phenomenon. J Rheumatology. 2005:841–848.

8. Staal J, Abràmoff MD, Niemeijer M, Viergever MA, Ginneken Bv. Ridge-based vessel segmentation in color images of the retina. IEEE Trans Med Imag. 2004 Apr; 23(4):501–509.

9. Soares JVB, Leandro JJG, C RM Jr, Jelinek HF, Cree MJ. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. IEEE Trans Med Imag. 2006 Sep; 25(9):1214–1222.

10. Berks, M.; Chen, Z.; Astley, S.; Taylor, C. Detecting and classifying linear structures in mammograms using random forests. Information Processing in Medical Imaging. Szkely, G.; Hahn, H., editors. Vol. 6801 of LNCS. Springer; Berlin Heidelberg: 2011. p. 510-524.

11. Freeman WT, Adelson EH. The design and use of steerable filters. IEEE Trans Pattern Anal Mach Intell. 1991 Sep; 13(9):891–906.

12. Breiman L. Random forests. Mach Learn. 2001; 45:5–32.

13. Mardia, KV.; Jupp, PE. Directional Statistics. Wiley; 2000.

14. Criminisi, A.; Shotton, J.; Robertson, D.; Konukoglu, E. Regression forests for efficient anatomy detection and localisation in CT studies. Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging. 2011. p. 106-117.

15. Dalal N, Triggs B. Histograms of oriented gradients for human detection. Proc IEEE Conf on Comp Vis and Patt Recog. 2005
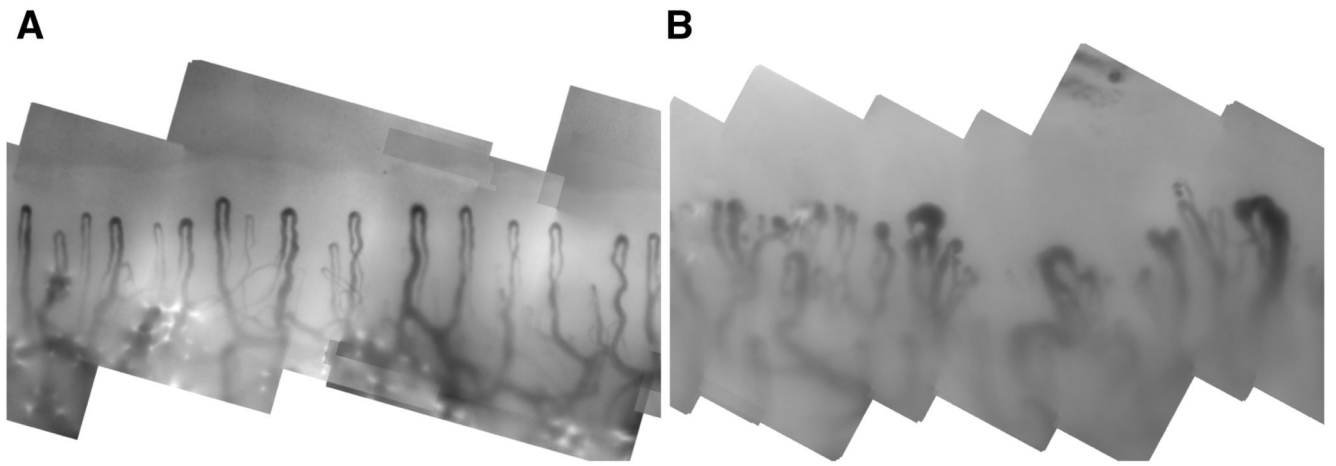
A **B**

Fig. 1.
Sections from two nailfold mosaics: one from a healthy control subject (left) and one from a patient with SSc (right) showing enlargement, distortion and irregular spacing of capillaries.
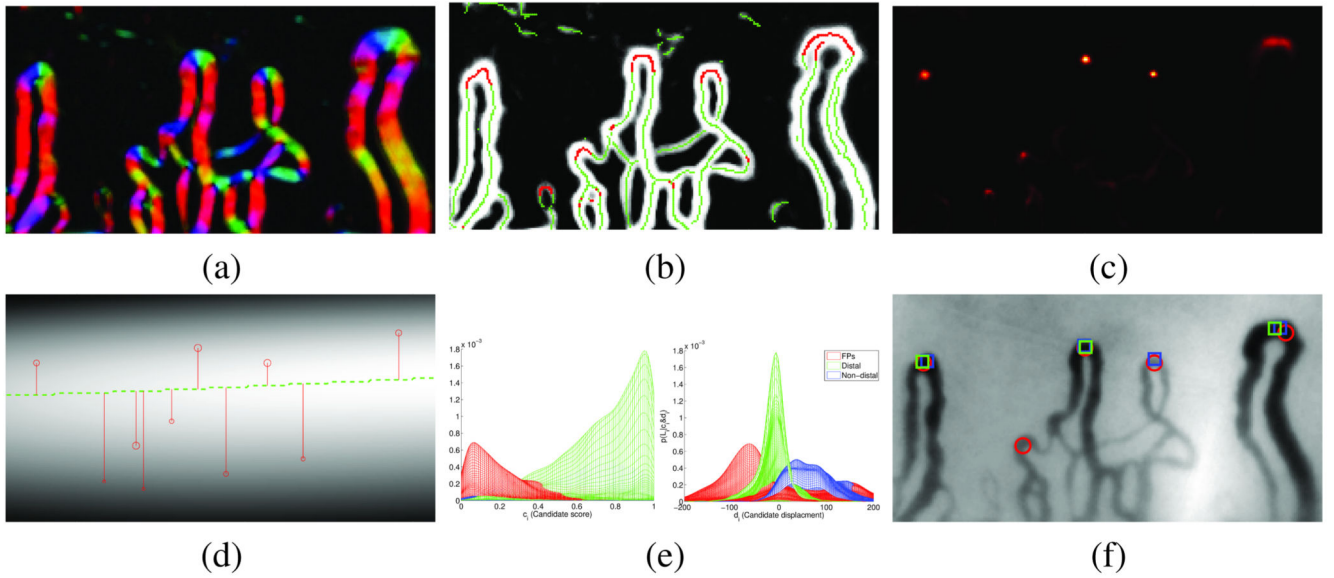
Fig. 2.
Detecting nailfold capillaries (a) Estimated vessel orientation V_θ, displayed using an hue/intensity color-map for angle/prediction confidence; (b) Vesselness V_v with oriented local maxima marked by green dots. Red dots have $a(q) > 0.5$ and vote for apex locations; (c) Vote map A of apex locations; (d) Estimated distal row (green line), from weighted kernel density of maxima in A. Candidate capillary locations are shown as red circles (scaled by candidate appearance score) with y displacements to estimated distal line; (e) Mesh plots of the joint class conditional distribution over candidate displacement and appearance score, viewed from score axis (left) and displacement axis (right): false positves (red), distal capillaries (green), non-distal capillaries (blue); (f) Capillaries selected by our method (red circle) and experts observers (blue and green squares), solid markers show distal capillaries.
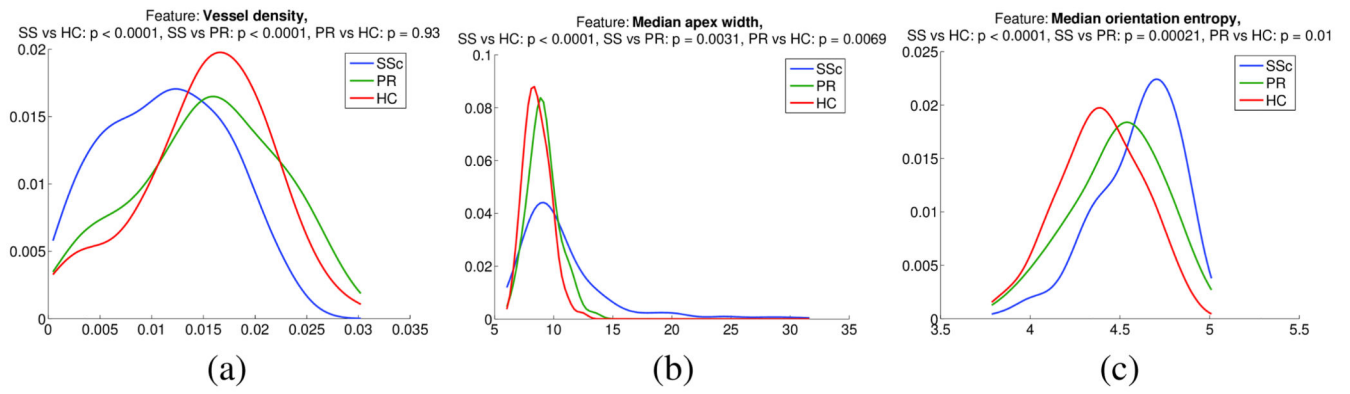
Fig. 3.
Distributions of capillary measurements by subject group: (a) Capillary density; (b) Median apical width; (c) Median capillary tortuosity.

Table 1

Pairwise agreement between observers. For each row O_i v O_j denotes the performance of observer i using j as ground truth. O_1, O_2 denotes consensus of the two human observers. Precision, recall and F-measures are given for vessel detection, classification accuracy and Cohen's Kappa statistic for labelling capillaries as distal or not.

Observers	Capillary detection			Distal v Non-distal	
	Precision	Recall	F-measure	Accuracy	Cohen's κ
O_2 v O_1	70.7 ± 1.3	46.2 ± 1.8	55.8 ± 1.4	89.5 ± 0.8	0.613 ± 0.025
O_3 v O_1	66.8 ± 1.1	59.6 ± 0.9	63.0 ± 0.8	90.8 ± 0.9	0.574 ± 0.028
O_3 v O_2	51.7 ± 1.9	70.5 ± 1.0	59.6 ± 1.3	87.2 ± 1.2	0.557 ± 0.028
O_3 v O_1, O_2	64.1 ± 1.8	80.9 ± 0.7	71.5 ± 1.1	93.6 ± 1.0	0.690 ± 0.034