Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Differentiation of arterioles from venules in mouse histology images using machine learning

J. Sachi Elkerton
Yiwen Xu
J. Geoffrey Pickering
Aaron D. Ward

Differentiation of arterioles from venules in mouse histology images using machine learning

J. Sachi Elkerton,^{a,b,*†} Yiwen Xu,^{a,b,c,*†} J. Geoffrey Pickering,^{a,c} and Aaron D. Ward^{a,b}

^aWestern University, Department of Medical Biophysics, 1151 Richmond Street, London, Ontario N6A 3K7, Canada

^bBaines Imaging Research Laboratory, London Regional Cancer Program, 800 Commissioners Road East, London, Ontario N6A 5W9, Canada

^cWestern University, Robarts Research Institute, 1151 Richmond Street, London, Ontario N6A 3K7, Canada

Abstract. Analysis and morphological comparison of the arteriolar and venular components of a microvascular network are essential to our understanding of multiple diseases affecting every organ system. We have developed and evaluated the first fully automatic software system for differentiation of arterioles from venules on high-resolution digital histology images of the mouse hind limb immunostained with smooth muscle α -actin. Classifiers trained on statistical and morphological features by supervised machine learning provided useful classification accuracy for differentiation of arterioles from venules, achieving an area under the receiver operating characteristic curve of 0.89. Feature selection was consistent across cross validation iterations, and a small set of two features was required to achieve the reported performance, suggesting the generalizability of the system. This system eliminates the need for laborious manual classification of the hundreds of microvessels occurring in a typical sample and paves the way for high-throughput analysis of the arteriolar and venular networks in the mouse. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.4.2.021104](https://doi.org/10.1117/1.JMI.4.2.021104)]

Keywords: vasculature quantification; whole slide analysis; feature analysis; machine learning; arteriole venule classification.

Paper 16130SSPRR received Jul. 1, 2016; accepted for publication Dec. 12, 2016; published online Feb. 28, 2017.

1 Introduction

The microvascular network consists of a hierarchical arrangement of arterioles, capillaries, and venules. Of these vessel categories, the arterioles and venules are wrapped by vascular smooth muscle cells. They serve vital functions delivering blood to and from the tissue to ensure oxygen and nutrient requirements are met. Assessment of the arteriolar and venular content, wall structure, and network organization is thus essential to our understanding of multiple disease scenarios. For example, in the eye, narrowing of retinal arterioles is related to diabetic macular ischemia,¹ and the arteriole-to-venule ratio is associated with hypertensive retinopathy.² In the brain, proper arteriolar function is essential to maintaining adequate cerebral blood flow; occlusion of even a single penetrating arteriole connecting the pial network to the subsurface network results in the death of a 0.5-mm cylinder of cortical tissue.³ In Alzheimer's disease, penetrating arterioles become tortuous, whereas the venules do not.⁴ Arterioles and venules may also have distinct genetic drivers during embryonic development of the cardiovascular system as well as angiogenesis in the adult.⁵ Importantly, however, histological analysis of the microvasculature is far more complex than for the macrovasculature. The small caliber vessels and their vast content carry a substantial risk of not reliably distinguishing arterioles from venules. Thus, it is critical to have approaches that visualize and quantify microvascular networks that can also distinguish, analyze, and compare arterioles and venules.

Brightfield histology provides a wealth of information about vessel wall structure, with a multitude of stains available to capture different structural aspects. For assessing the microvasculature,

whole-mount slide scanners for digitization of entire tissue cross sections at 0.25- μ m pixel size is particularly helpful.⁶ However, the volume of the acquired data makes comprehensive analysis very challenging; manual detection and measurement of vessels in tissue require a prohibitive amount of time. An automated microvessel detection and segmentation algorithm for histological mouse hind limb tissue has been previously reported,⁷ and other semiautomated approaches have been developed for this task.⁸ However, to the best of our knowledge, no automated method has yet been developed to classify arterioles and venules to allow for the separate analysis of arteriolar and venular structure. Such a system would eliminate the need for the laborious and observer-dependent manual classification performed in previous studies.^{9–11} Our interest is in understanding the cellular structure of the microvascular vessel wall both under normal circumstances and in the postischemic and regenerated microvasculature. Such a comparison may yield powerful insights into improving therapeutic regimes after stroke or myocardial infarction. However, in regenerated tissue, the arterioles and venules can be even more difficult to classify, partially due to the fact that newly generated arterioles often lack a fully developed vessel wall structure and can have a similar appearance on histology to venules.¹² This leads to a strong dependence on a skilled observer to make the distinction between arterioles and venules.

In this article, we addressed this issue by designing, implementing, and testing a software system for automatic classification of arterioles and venules on brightfield histology of the mouse hind limb smooth muscle. The system automatically detects and segments microvessels throughout the whole slide using our method described in Ref. 7. It then extracts texture and morphological features from all the vessels, performs feature selection to extract the most useful features for classification,

*Address all correspondence to: J. Sachi Elkerton, E-mail: jelkerto@uwo.ca; Yiwen Xu, E-mail: yxu225@uwo.ca

†Jointly first-authored paper.

and trains a classifier via supervised machine learning. We tested system performance using cross validation on a sample containing tissue samples from normal mice and from mice previously subjected to hind limb ischemia.

2 Materials and Methods

2.1 Materials

The experiments were conducted on tissue samples of the upper one-third tibialis anterior hind limb muscle from 10 wild-type C57BL/J6 mice (one tissue section per mouse). In five of the 10 mice, tissues were collected for 2 weeks after induction of hind limb ischemia by femoral artery excision; these samples were expected to contain regenerated vasculature, where the surgeries were performed under isoflurane anesthesia. The postischemic samples are referred to as the regenerated samples. This particular muscle segment was selected because of the consistent development of microvessels of various sizes following hind limb ischemia. The mice were perfused with saline post-mortem to remove the red blood cells from the vessel lumina and then perfusion fixed at physiological pressure with 4% paraformaldehyde. The tissues were processed and paraffin embedded after harvesting, then cut into 7-mm $\times$ 5-mm blocks and sectioned at 5 μ m. All experiments in this study were approved by the Animal Care and Veterinary Service Committee at the University of Western Ontario (protocol no. 2010-244) and were carried out in accordance with their requirements.

To visualize the smooth muscle distribution down to the level of the arterioles and venules, sections were immunostained with smooth muscle α -actin using a monoclonal antibody (DAKO, M0851) and bound primary antibody detected with horseradish peroxidase-conjugated secondary antibody and 3,3'-diaminobenzidine chromogen (DAB, Vector Laboratories, SK-4100). This marked the smooth muscle layer of the vasculature (resulting in the vessel walls being stained with a brown color) which is shown in Fig. 1(a). The tissue was then counter stained with hematoxylin, resulting in blue-stained nuclei. The immunostain was performed manually, and the samples were processed in two separate batches. The stained sections were then imaged with a ScanScope CS (Aperio Technologies, Vista, California) bright-field slide scanner, at 40 $\times$ objective magnification (achieved on this scanner using a 20 $\times$ objective with the 2 $\times$ magnification engaged), yielding a 0.25- μ m isotropic pixel size. This resulted

Fig. 1 Arteriole with the DAB stain (a) and the binary mask (b) output from the automated segmentation. Scale bar is 50 μ m.

in whole slide image dimensions ranging from 12,341 to 26,398 pixels in height and 9000 to 17,135 pixels in width.

One tissue section was analyzed from each mouse, where each tissue section was known to be either normal or regenerated postischemia, with vessels detected and segmented automatically using our previously reported method.⁷ Our segmentation algorithm resulted in a Dice similarity coefficient of 0.89 with no bias toward over- or undersegmentation and a detection sensitivity and specificity of 0.91 and 0.97, respectively, and the whole tissue section was automatically extracted from the whole slide scan. A sample vessel and its segmentation are shown in Fig. 1. Every vessel on the whole slide image was manually classified by the consensus of two observers (i.e., both observers had to be in agreement) into the classes: arteriole, venule, or unknown/other based on vessel size, thickness of the vessel walls, intensity of the vessel walls, and the relative location of the vessel (Fig. 2). Only the arteriole and venule image classes were used for this experiment, with a total of 192 arterioles and 127 venules. This included all vessel sections that the observers were confidently able to classify manually. Each of the 10 mice had a total of 7 to 66 vessels used from each tissue section.

2.2 Methods Overview

After automatic vessel detection and segmentation which defined the outer boundary of the vessel wall and the vessel lumen, features were extracted from these vessel walls. This included statistical, texture, and morphological features to use for supervised machine learning. Cross validation was used to test the system's ability to differentiate arterioles from venules, with the manual classifications used as the reference standard for comparison (Fig. 3).

2.3 Feature Extraction

First- and second-order feature computation was performed within each segmented vessel wall on 10 color channels separately: three red, green, and blue, three hue, saturation, and value (HSV), three Lab, and a DAB stain channel extracted by color deconvolution.^{13,14} The color images were decomposed into separate channels as they carry different information regarding the stains represented in the image. For example, HSV takes into the account the color spectrum, saturation of the color, and the brightness of the image, and Lab space includes all perceivable colors and approximates human perception of color differences. Twenty-two first-order features and 22 second-order gray-level cooccurrence matrix (GLCM) texture features¹⁵ were calculated on each color channel (Table 1). The GLCM calculations were performed in all four neighboring pixel directions, which were then averaged. Nineteen morphological features were also computed. In total, 459 features were computed for each vessel (Table 1).

2.4 Feature Selection

To select the optimal combination of features, forward feature selection was performed and evaluated using PRTools 5.2.1 (Delft Pattern Recognition Research, Delft, The Netherlands), with an inter-intraclass distance criterion.¹⁶ Other forms of feature selection were not explored with our limited sample size to avoid overfitting. It is important to note that feature selection was performed only on the training set at each iteration of

Fig. 2 Arterioles and venules are shown in the normal and regenerated tissues. The microvessels have distinguishable differences in the normal vasculature, where the smooth muscle layer is thicker and more pronounced on the arteriolar side and have a visibly darker stain. This difference is less apparent in the regenerated vasculature.

Fig. 3 Block diagram to demonstrate the method followed in this experiment for a LOMO cross validation.

cross validation. The entire dataset was not used to select a single set of features for supervised machine learning.

The number of features used for classification was chosen by completing the process for several different numbers of features and comparing the resulting errors and area under the receiver operating characteristic (ROC) curves (AUCs). It was desired to minimize error and maximize the AUC while keeping the set of features consistent through the cross validations.

2.5 Supervised Machine Learning

We tested the performance of three different machine learning classifiers using the selected features, performed using PRTools. The following classifiers were used: support vector machine (SVM), a random forest classifier (RFC), and a logistic linear classifier (LOGLC). The SVM calculates the separation by maximizing the margin between the support vectors of the two classes,¹⁷ trained with a linear kernel. In the random forest approach, multiple decision trees are created by sampling different feature and data subsets, with classification determined by the majority vote from all the trees.¹⁸ Fifty decision trees were trained for each classifier, each with data subsets chosen by bootstrapping. LOGLC is a linear classifier computed by maximum likelihood estimation using the logistic function.^{19,20}

To optimize the classifier performance, all the features were normalized. This was done by taking each individual feature, subtracting the mean, and dividing by the standard deviation. It is important to normalize the features because of classifiers that are sensitive to the scale of the features, such as the SVM.²¹ The other classifiers were unaffected by this normalization.

2.6 Validation

Two separate analyses were performed to validate the resulting classifiers. First, each classifier was tested using a leave-one-mouse-out (LOMO) cross validation (at each iteration, one mouse sample was left out). Feature selection and normalization were performed on the training set at each iteration, and the selected features were used to train the classifiers. The classification error rates, false positive rates (FPRs), false negative rates (FNR), and AUC were measured for each classifier. For this

Table 1 List of all the features extracted from each vessel image.

First-order features	Second-order features	Morphological/shape features
Mean $\frac{1}{N} \sum_i x_i$	GLCM contrast $\sum_{i,j} i - j ^2 g(i, j)$	Maximum diameter
Standard Deviation $\sqrt{\frac{\sum_i (x_i - \mu)^2}{N-1}}$	GLCM dissimilarity $\sum_{i,j} i - j g(i, j)$	Mean radial length
Energy $\sum_i x_i^2$	GLCM energy $\sum_{i,j} g(i, j)^2$	Spiculation: standard deviation of radial length
Median	GLCM entropy $-\sum_{i,j} g(i, j) \log_2 g(i, j)$	Aspect ratio
Maximum	GLCM autocorrelation $\sum_{i,j} i j g(i, j)$	Waviness $\frac{P_{\text{Vessel Hull}}}{P}$
Minimum	GLCM homogeneity 1 $\sum_{i,j} \frac{g(i, j)}{1 + i - j }$	Skeleton end point count
Range Maximum – Minimum	GLCM homogeneity 2 $\sum_{i,j} \frac{g(i, j)}{1 + i - j ^2}$	Normalized radial length mean
Mean absolute deviation $\frac{1}{N} \sum_i x_i - \mu $	GLCM correlation $\sum_{i,j} \frac{(i - \mu_i)(j - \mu_j) g(i, j)}{\sigma_i \sigma_j}$	Normalized radial length entropy $-\sum \frac{N_{\text{L}}}{N_{\text{L}}} \log_2 \left[\frac{N_{\text{L}}}{N_{\text{L}}} \right]$
Root mean square $\sqrt{\frac{1}{N} \sum_i x_i^2}$	GLCM cluster prominence $\sum_{i,j} [(i - \mu_i) + (j - \mu_j)]^4 g(i, j)$	Normalized Radial Length Sphericity $\frac{N_{\text{L}}}{\text{SD}(N_{\text{L}})}$
Kurtosis $\frac{1}{N} \frac{\sum_i (x_i - \mu)^2}{\sigma^4}$	GLCM cluster shade $\sum_{i,j} [(i - \mu_i) + (j - \mu_j)]^3 g(i, j)$	Normalized radial length roughness $\frac{\sqrt[4]{(N_{\text{L}} - N_{\text{L}})^4} - \sqrt[4]{(N_{\text{L}} - N_{\text{L}})^2}}{N_{\text{L}}}$
Skewness $\frac{1}{N} \frac{\sum_i (x_i - \mu)^3}{\sigma^3}$	GLCM cluster tendency $\sum_{i,j} [(i - \mu_i) + (j - \mu_j)]^2 g(i, j)$	Normalized radial length RATIO $\frac{\sum N_{\text{L}} - N_{\text{L}} }{\text{Number of boundary points} \times N_{\text{L}}}$
Variance $\frac{1}{N-1} \sum_i (x_i - \mu)^2$	GLCM variance $\sum_{i,j} (i - \mu)^2 g(i, j)$	Area box ratio $\frac{A_{\text{Vessel}}}{A_{\text{Box}}}$
Entropy $-\sum_i I P \log_2 I P$	GLCM max. probability $\max[g(i, j)]$	Area perimeter ratio $4\pi \frac{A_{\text{Vessel}}}{P_{\text{Vessel}}^2}$
Uniformity $\sum_i I P^2$	GLCM inverse variance $\sum_{i,j} \frac{g(i, j)}{ i - j ^2}$	Area convex hull ratio $\frac{A_{\text{Vessel}}}{A_{\text{Vessel Hull}}}$
1 percentile	GLCM sum average $\sum_{i=2}^N i g_{x+y}(i)$	Median wall thickness $\frac{\text{Median thickness}}{P_{\text{Lumen}}}$
10 percentile	GLCM sum entropy $-\sum_{i=2}^N \log_2 [g_{x+y}(i)] g_{x+y}(i)$	70 percentile wall thickness ratio $\frac{70 \text{ percentile thickness}}{P_{\text{Lumen}}}$
90 percentile	GLCM sum variance $\sum_{i=2}^N i^2 g_{x+y}(i)$	90 percentile wall thickness ratio $\frac{90 \text{ percentile thickness}}{P_{\text{Lumen}}}$
99 percentile	GLCM difference entropy $-\sum_{i=0}^{N-1} \log_2 [g_{x+y}(i)] g_{x+y}(i)$	Wall thickness SD $\frac{\text{Thickness SD}}{P_{\text{Lumen}}}$
Mean gradient value	GLCM information measure correlation 1 $\frac{H_{xy} - H_{xy1}}{\max\{H_X, H_Y\}}$	
Variance of gradient	GLCM information measure correlation 2 $\sqrt{1 - \exp[-2(H_X Y^2 - H_X Y)]}$	
Bimodality coefficient $\frac{\text{Skewness}^2 + 1}{\text{Kurtosis} + \frac{3(N-1)^2}{(N-2)(N-3)}}$	GLCM diagonal moment $\sum_{i,j} \sqrt{\frac{1}{2}} i - j g(i, j)$	
Median absolute deviation $\frac{1}{N} \sum_i x_i - \tilde{x} $	GLCM inverse difference moment $\sum_{i,j} \frac{g(i, j)}{1 + (i - j)^2}$	

$g(i, j)$ refers to the GLCM, where H_X and H_Y are entropies of g_x and g_y , $H_X Y = -\sum_{i,j} g_{ij} \log_2 g_{ij}$, $H_X Y 1 = -\sum_{i,j} g_{ij} \log_2 [g_x(i) g_y(j)]$, $H_X Y 2 = -\sum_i \sum_j g_x(i) g_y(j) \log_2 [g_x(i) g_y(j)]$, μ refers to the mean, $\tilde{x}$ refers to the median, σ refers to the standard deviation, and N refers to the number of pixels. A_x , P_x , and W_x refer to the area, perimeter, and average width, respectively, of structure x ; vessel refers to the vessel wall; box refers to the vessel's bounding box; vessel hull refers to the vessel's convex hull; and lumen refers to the inner vessel boundary. NRL is the normalized radial length. The GLCM neighborhood voxel distance was 0.25 μm averaged over all directions.

experiment, a classification as an arteriole was considered to be a positive classification.

In the second experiment, a single classifier was defined using an independent training set and tested on a separate testing set. Six mice were used for training (three normal, three regenerated), and four mice were used for testing (remaining two normal and two regenerated). Aggregation of values in the ROC curves by averaging in the first experiment may add a degree of freedom to the analysis; however, there is no such aggregation

in experiment 2. The performance of the single classifier was validated with the testing set with the same classification errors as the cross validation experiment.

2.7 Classifier Confidence

A confidence value was assigned to each vessel ranging from 0 to 1, where the confidence threshold of 0.5 determines the binary classification of each vessel. For the random forest,

the confidence was based on the proportion of positive votes from all the decision trees. For the LOGLC and SVM, the confidence was based on the distance from the classification boundary that was then scaled by a sigmoid function.¹⁶

To reduce the classification error rates, vessels with varying confidence levels were thresholded and removed. The lowest confidence level is 0.5, representing the situation where the vessel lies on the decision boundary in feature space, and the highest confidence level is 1.0. This enables the system to retain vessels that were classified with high confidence and to eliminate vessels with confidence values near the decision boundary which have a higher proportion of false positives. This shrinks the dataset, but lowers the error rates, and can be performed depending on the distribution of confidence values in a given sample.

2.8 Statistics

To determine statistical significance, a two-way ANOVA test was done. This was performed on the classification error, FNR, and FPR with the two factors of vessel type (normal and regenerated) and classifier type (RFC, LOGLC, and SVM). A separate ANOVA was done to also test for significant differences in errors between the classifier types and between classifiers trained using different numbers of features.

3 Results

3.1 Experiment 1

The differences between the three classifiers with respect to the number of features used for each classifier were reported to test the sensitivity of each classifier to the number of features. Significant differences in the error rates ($p < 0.05$) were found between the 1-feature and 2-feature classifiers; no other significant differences were found [Fig. 4(b)]. All classifiers were trained on features from forward feature selection and evaluated using a LOMO cross validation. The AUCs leveled off after 2 to 3 features at AUCs of >0.89 for the LOGLC, >0.89 for the SVM, and >0.84 for the RFC [Fig. 4(a)]. Diminishing returns on error rates were observed beyond two features, for which the LOGLC had an error rate of 17%, the SVM of 16%, and the RFC of 21% [Fig. 4(b)].

The sets of three features chosen from forward feature selection are shown in Fig. 5. The features chosen consistently at each of the 10 LOMO cross validations were area perimeter ratio (the ratio of the vessel wall area to the outer perimeter squared) and DAB skewness (skewness in the DAB channel). The third feature chosen was not consistent with every cross validation. The errors resulting from the 2-feature classifiers are discussed for the remainder of the results below.

The classification performance metrics using two features is provided in Table 2, where the metrics for normal and regenerated tissues are shown separately. From Table 2, all three classifiers performed similarly, with mean classification error rates between 17% and 21% and AUCs between 0.84 and 0.89. The ROC curves for each classifier type are shown in Fig. 6, where the curves were averaged over 10 cross validations. Each mouse sample was given equal weightings for the ROC curves. However, ROC curve averaging may add a degree of freedom to the analysis. By contrast, in the second experiment, only one ROC curve was computed using a single classifier and a larger testing set.

Fig. 4 AUC (a) and error rate (b) as a function of number of features used for classification. The error bars represent the standard error of the mean. All features were chosen using a forward feature selection, and classifiers were trained with a LOMO cross validation.

Fig. 5 Three features chosen from forward feature selection over the 10 cross validations.

The distribution of correctly and incorrectly classified vessels by the LOGLC, SVM, and RFC, respectively, for different levels of classifier confidence is shown in Fig. 7 using two features. The LOGLC has 48%, SVM has 48%, and RFC has 54% of the vessels being correctly classified with confidence >0.9 . The results from categorizing vessels based on varying confidence thresholds reduced the proportion of incorrectly classified vessels (Fig. 7), where no removal of vessels or no confidence threshold is denoted by “none.” The error rates of the remaining vessels after the removal of lower confidence vessels after a confidence threshold are shown in Table 3. As an example of how this confidence level was used, in Table 3, the results reported in the “RFC” row and the “0.9” column show that 39.81% of all the classified vessels had confidence thresholds <0.9 . Removing those vessels from the analysis resulted in a classification error of 10.42% using an RFC. By contrast, the none column shows the classification error when no confidence threshold was

Table 2 Performance metrics from cross validation using two features from forward feature selection (mean $\pm$ SD).

	All vessels		
	LOGLC	SVM	RFC
Error rate	0.17 $\pm$ 0.11	0.16 $\pm$ 0.10	0.21 $\pm$ 0.11
False negative rate	0.15 $\pm$ 0.24	0.14 $\pm$ 0.22	0.22 $\pm$ 0.27
False positive rate	0.26 $\pm$ 0.32	0.26 $\pm$ 0.32	0.25 $\pm$ 0.24
AUC	0.89	0.89	0.84
	Normal		
	LOGLC	SVM	RFC
Error rate	0.12 $\pm$ 0.11	0.11 $\pm$ 0.11	0.15 $\pm$ 0.10
False negative rate	0.04 $\pm$ 0.06	0.04 $\pm$ 0.05	0.12 $\pm$ 0.13
False positive rate	0.31 $\pm$ 0.41	0.31 $\pm$ 0.41	0.23 $\pm$ 0.16
AUC	0.94	0.93	0.90
	Regenerated		
	LOGLC	SVM	RFC
Error rate	0.21 $\pm$ 0.10	0.20 $\pm$ 0.08	0.29 $\pm$ 0.06
False negative rate	0.27 $\pm$ 0.31	0.25 $\pm$ 0.28	0.33 $\pm$ 0.36
False positive rate	0.21 $\pm$ 0.25	0.21 $\pm$ 0.25	0.33 $\pm$ 0.32
AUC	0.84	0.85	0.81

applied; this column shows the classification errors using all vessels regardless of their confidence values.

A two-way ANOVA test was performed to determine statistical significance between the different classifiers and the normal versus regenerated tissues. This was performed separately for the error rates, FNRs, and FPRs. There was found to be no

Fig. 6 ROC curves for each of the classifiers trained on two features. LOGLC AUC: 0.89, SVM AUC: 0.89, RFC AUC: 0.84. **Fig. 7** Histogram of the confidences from the LOGLC (a, b), SVM (c, d) and RFC (e, f) classification using two features with their respective confidence thresholds, where "none" indicates no confidence threshold was applied and all vessels were categorized.

statistical difference between the three classifiers ($p > 0.05$). There was statistical difference between the normal and regenerated tissue error rates and FNRs ($p < 0.05$) but not the FPRs ($p > 0.05$).

A two-way repeated-measures ANOVA was also performed on the different classifiers and between the errors from using two and three features. Again there was found to be statistical

Table 3 The classification errors from the 2-featured forward feature selection with a LOMO cross validation as the confidence threshold increases and vessels are eliminated.

		Confidence threshold				
		None	0.6	0.7	0.8	0.9
LOGLC	Classification error	17.24	15.86	11.63	8.60	4.94
	Fraction removed		9.09	19.12	30.72	49.22
SVM	Classification error	16.61	16.15	12.26	8.11	4.94
	Fraction removed		8.78	18.18	30.41	49.22
RFC	Classification error	24.14	19.86	18.53	14.29	10.42
	Fraction removed		11.60	18.81	27.59	39.81

differences neither between the three classifiers ($p > 0.05$) nor between the errors using two and three features ($p > 0.05$).

3.2 Experiment 2

A single classifier was defined using an independent training set (six mouse samples; three normal and three regenerated) and tested on a separate testing set (four mouse samples; two normal and two regenerated). In total, there were 206 vessels (84 normal and 122 regenerated) in the training set and 113 vessels (50 normal and 63 regenerated) in the testing set.

The AUCs leveled off after two features at AUCs of >0.90 for the LOGLC, >0.91 for the SVM, and >0.89 for the RFC [Fig. 8(a)]. Diminishing returns on error rates were observed beyond two features, for which the LOGLC had an error rate of 15%, the SVM of 17%, and the RFC of 21% [Fig. 8(b)].

The forward feature selection resulted in the DAB skewness and area perimeter ratio as the top two features. These features chosen were identical to those chosen in the LOMO cross validation.

The error metrics are shown in Table 4 using those two features, and AUC curves are shown in Fig. 9. The classifier confidence histograms for all the classifiers with confidence threshold levels are shown in Fig. 10.

4 Discussion

Both arterioles and venules contain smooth muscle cells, which are contractile cells that wrap the endothelial cells. We have established that differentiation of these two types of microvessels is feasible with the use of supervised machine learning. All the tested classifiers provided promising performance, with AUCs between 0.84 and 0.89 determined by cross validation

Fig. 8 AUC (a) and error rate (b) as a function of number of features used for classification. All features were chosen using a forward feature selection, and classifiers were trained with an independent training set.

Table 4 Performance metrics using two features from forward feature selection.

	All vessels		
	LOGLC	SVM	RFC
Error rate	0.15	0.17	0.21
False negative rate	0.15	0.15	0.24
False positive rate	0.15	0.20	0.17
AUC	0.91	0.92	0.89

	Normal		
	LOGLC	SVM	RFC
Error rate	0.06	0.10	0.16
False negative rate	0.08	0.08	0.23
False positive rate	0.04	0.12	0.08
AUC	0.97	0.95	0.92

	Regenerated		
	LOGLC	SVM	RFC
Error rate	0.22	0.22	0.25
False negative rate	0.20	0.20	0.24
False positive rate	0.27	0.27	0.27
AUC	0.85	0.84	0.87

Fig. 9 ROC curves for each of the classifiers from classifier trained on two features. LOGLC AUC: 0.91, SVM AUC: 0.92, RFC AUC: 0.89.

for our two-feathered forward feature selection, where the same two features were consistently chosen. Error rates, FNRs, and FPRs were consistent across classifiers, and only two features were required to achieve the reported performance. Variability in vessel size, shape, and staining was observed in our dataset. The upper row in Fig. 11 shows arterioles that were incorrectly

Fig. 10 Histogram of the confidences from the LOGLC (a,b), SVM (c, d) and RFC (e,f) classifications with their respective confidence thresholds.

classified as venules ($0.6 < \text{confidence} < 0.9$). The lower row shows correctly classified arterioles ($0.6 < \text{confidence} < 1.0$); note the variability in size, shape, and staining tolerated by the LOGLC.

The ability to accurately separate the arterioles from the venules is a key step in analyzing the vascular system. The arterioles control the blood flow directly leading to the capillaries, whereas venules are critical to the outflow of blood and tissue waste products. Not only does their function differ, but the manner in which they remodel during disease can differ. Accordingly, assessment of diseases characterized by ischemia, injury, and inflammation may be substantially enhanced by separating the two types of vessels.¹⁻⁵ Automatic separation allows for fast, high-volume analysis that is not user dependent. This will ultimately be included in a total automated procedure that will include detection and segmentation,⁷ 3-D reconstruction,⁶ and analysis.

The two features chosen during forward feature selection were consistent with each round of cross validation. This is promising and suggests generalizability of classifiers trained on these features. Although the error rates were slightly better in the 3-feature system, the inconsistent choice of the third feature across iterations makes this a less favorable solution. The two features chosen included a morphological feature, area perimeter ratio, and vessel wall intensity feature, DAB skewness, and they were both chosen at every round of cross validation. The utility of the area perimeter ratio feature is concordant with the fact that venules are in general thinner and have less vessel wall area compared with arterioles, which generally have thicker walls. It also may capture the amount of waviness of the vessel walls, where the venules in general have more curved walls than arterioles. The other selected feature that was chosen for every cross validation was the DAB skewness, which is concordant with the fact that in general, arterioles are more darkly stained than the venules since they have more smooth muscle. It is most likely the lack of dark DAB stain on many of the venules and the original DAB threshold used for segmentation that causes a

Fig. 11 Incorrectly (a)–(e) and correctly (f)–(j) classified arterioles by the LOGLC. Scale bar is 20 μm .

positive skewness on many of the venules. The performance based on the type of classifier used was evaluated. There was no significant difference between the errors from each of the classifiers ($p < 0.05$). This indicates that the error rates are independent of the classifier used.

The classification was performed on independent training and testing sets and demonstrated the consistency of the feature selection and classification processes. To maintain the generalizability of the classifier with two features, linear and RFCs were used to prevent overfitting to our current dataset. The features chosen were consistent with the features chosen during cross validation including one feature based on vessel shape, area perimeter ratio, and the other feature based on vessel stain content with regard to intensity in the form of DAB skewness. The error rates were also comparable with the rates when performing a LOMO cross validation. This suggests the possibility of using a single trained classifier that will be able to accurately separate the vasculature on unseen data.

We observed that in many instances of incorrect vessel classification, errors occurred due to staining variability [e.g., weak staining in Fig. 11(e)]. Lightly stained arterioles tended to be incorrectly classified as venules, and heavily stained venules incorrectly classified as arterioles. This suggests a need for additional preprocessing to normalize the images for the overall level of staining on each slide. Unusually thin-walled arterioles also caused false negatives, whereas thick-walled venules caused false positives. Many of these unusually thick or thin vessels came from the regenerated vasculature. Figures 11(a)–11(e) provides a qualitative illustration of the false negatives. There was significant difference between the normal and the regenerated vessel errors for the total error rates and the FNRs. This coheres with the observation that the vessels in the regenerated tissues have less distinct differences between the arterioles and the venules and are also more difficult for an operator to differentiate.

The distributions of confidences for each classifier are similar. Between the two experiments, the distributions were also similar, suggesting that the smaller training set seems representative of the entire dataset. Thresholding the confidences and eliminating the vessels are a possibility if more accurate classifications are desired. A subset of the vessels with lower confidences would remain unclassified [Figs. 7(b), 7(d), and 7(f)]. The goal is to replicate a human classifier, where a human will likely exclude several vessels based on uncertainty. However, there is a trade-off between accuracy of the classification and the number of vessels remaining for analysis, and this is dependent on the classifier and the distribution of confidences. In this experiment, thresholding the confidence values based on all the classifiers decreased the error rates (Table 3). However, the optimal threshold is dependent on the type of classifier used as the distribution of the vessel confidences varies. One needs to take into consideration when analyzing abnormal vasculature that the vessels of interest may lie in the range of lower confidences due to their structure (i.e., interesting discoveries may be found when looking at unusual vasculature occurring in the “low confidence” category). Once the algorithm has separated the vasculature into the three groups (arteriole, venule, and unknown), the user can then study the vessels in each category.

This work must be considered in the context of its strengths and limitations. Although we had a large sample size of 192 arterioles and 127 venules of varying sizes, shapes, and staining levels, all the tissues in this study were prepared in a single laboratory. Additional testing would be required to report on the

generalizability of our methods to tissues prepared in different laboratories, allowing for testing of various feature selection and classification techniques without overfitting to the current dataset. Even so, there was still a large staining variability between the mouse samples, including weakly stained vessels that may have been systematically under or over segmented. The technique used for histological slide preparation and immunohistochemistry staining is a standard method for visualizing the microvascular wall. However, the classifiers, as trained in our experiments, cannot be directly applied to images using other methods of vessel visualization, such as fluorescent staining. This study was also only done on the mouse hind limb tissue and cannot be directly generalized to other tissue types. Although we did not observe substantial improvement in classification performance when more than two features were used, this may have been due to a lack of normalization for staining variability. Our ongoing work includes staining normalization and a subsequent analysis of performance with the inclusion of additional features and more feature selection methods.

5 Conclusions

Distinguishing arterioles from venules is an unattained prerequisite to high-throughput, quantitative analysis of the microvascular network morphology. In addition, analysis and morphological comparison of arteriolar and venular subnetworks are essential to our understanding of multiple diseases affecting all vascularized organ systems. We have developed and evaluated the first fully automatic software system for differentiation of arterioles from venules on high-resolution digital histology images of the mouse hind limb. This system eliminates the need for laborious manual classification of the hundreds of microvessels occurring in a typical sample and paves the way for high-throughput analysis of the arteriolar and venular networks in the mouse.

Our software system based on supervised machine learning provided useful classification accuracy for differentiation of arterioles from venules on high-resolution digital histology images of the mouse hind limb immunostained with smooth muscle α -actin. Our system achieved an area under the ROC curve of 0.89. Feature selection was consistent across cross validation iterations, and a small set of two features was required to achieve the reported performance; this augurs well for the generalizability of this system. This system will enable scientists to conduct high-throughput studies of animal models, measuring and comparing arteriolar and venular networks to deepen our understanding of human disease so that improved treatments can be sought.

Disclosures

The authors have no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada, Cancer Care Ontario, and the Canadian Institutes of Health Research (FDN-143326). A.D. Ward holds a Cancer Care Ontario Research Chair in Cancer Imaging. J.G. Pickering holds the Heart and Stroke Foundation of Ontario/Barnett-Ivey Chair. Dr. Trevor Shepherd of the London Health Sciences Centre, London, Ontario, provided the Aperio Scanscope CS system.

References

1. G. Liew et al., "Diabetic macular ischaemia is associated with narrower retinal arterioles in patients with type 2 diabetes," *Acta Ophthalmol.* **93**(1), e45–e51 (2015).
2. M. Cavallari et al., "Novel method for automated analysis of retinal images: results in subjects with hypertensive retinopathy and CADASIL," *BioMed Res. Int.* **2015**, 752957 (2015).
3. A. Y. Shih et al., "Robust and fragile aspects of cortical blood flow in relation to the underlying angioarchitecture," *Microcirculation* **22**(3), 204–218 (2015).
4. A. Y. Lai et al., "Venular degeneration leads to vascular dysfunction in a transgenic model of Alzheimer's disease," *Brain* **138**(4), 1046–1058 (2015).
5. D. Shin et al., "Expression of ephrinB2 identifies a stable genetic difference between arterial and venous vascular smooth muscle as well as endothelial cells, and marks subsets of microvessels at sites of adult neovascularization," *Dev. Biol.* **230**(2), 139–150 (2001).
6. Y. Xu et al., "A method for 3D histopathology reconstruction supporting mouse microvasculature analysis," *PLoS One* **10**(5), e0126817 (2015).
7. Y. Xu et al., "Segmentation of digitized histological sections for vasculature quantification in the mouse hind limb," *Proc. SPIE* **9420**, 942003 (2015).
8. C. Taylor and R. Levenson, "Quantification of immunohistochemistry —issues concerning methods, utility and semiquantitative assessment II," *Histopathology* **49**(4), 411–424 (2006).
9. T. Nakagawa et al., "Diabetic endothelial nitric oxide synthase knockout mice develop advanced diabetic nephropathy," *J. Am. Soc. Nephrol.* **18**(2), 539–550 (2007).
10. F. Akhavein et al., "Decreased left ventricular function, myocarditis, and coronary arteriolar medial thickening following monocrotaline administration in adult rats," *J. Appl. Physiol.* **103**(1), 287–295 (2007).
11. T. R. Nurkiewicz et al., "Systemic microvascular dysfunction and inflammation after pulmonary particulate matter exposure," *Environ. Health Perspect.* **114**(3), 412–419 (2006).
12. M. J. Frontini et al., "Fibroblast growth factor 9 delivery during angiogenesis produces durable, vasoresponsive microvessels wrapped by smooth muscle cells," *Nat. Biotechnol.* **29**(5), 421–427 (2011).
13. A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol./Int. Acad. Cytol. Am. Soc. Cytol.* **23**(4), 291–299 (2001).
14. H. Yin et al., "Fibroblast growth factor 9 imparts hierarchy and vaso-reactivity to the microcirculation of renal tumors and suppresses metastases," *J. Biol. Chem.* **290**(36), 22127–22142 (2015).
15. R. W. Connors and C. A. Harlow, "A theoretical comparison of texture algorithms," *IEE Trans. Pattern Anal. Mach. Intell.* **PAMI-2**(3), 204–222 (1980).
16. R. Duin et al., *PRTools4. 1, A Matlab Toolbox for Pattern Recognition*, Delft University of Technology, Delft, The Netherlands (2007).
17. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**(3), 273–297 (1995).
18. L. Breiman, "Random forests," *Mach. Learn.* **45**(1), 5–32 (2001).
19. A. R. Webb, *Statistical Pattern Recognition*, John Wiley & Sons, New York (2003).
20. J. Anderson, "Logistic discrimination," in *Handbook of Statistics*, P. R. Krishnaiah and L. N. Kanal, Eds., Vol. **2**, North Holland Publishing Co., Amsterdam, North Holland (1982).
21. R. Herbrich and T. Graepel, "A PAC-Bayesian margin bound for linear classifiers: why SVMs work," *IEEE Trans. Inf. Theory* **48**(12), 3140–3150 (2002).

J. Sachi Elkerton is an undergraduate physics student at the University of Western Ontario in London, Canada. She is currently working in the Baines Imaging Research Laboratory at the London Regional Cancer Program.

Yiwen Xu is a PhD - MCISc candidate in the Department of Medical Biophysics at the University of Western Ontario in London, Canada.

Biographies for the other authors are not available.