Computer-aided diagnosis of retinopathy in retinal fundus images of preterm infants via quantification of vascular tortuosity

Faraz Oloumi
Rangaraj M. Rangayyan
Anna L. Ells

Computer-aided diagnosis of retinopathy in retinal fundus images of preterm infants via quantification of vascular tortuosity

Faraz Oloumi,^{a,*} Rangaraj M. Rangayyan,^a and Anna L. Ellis^{a,b}

^aUniversity of Calgary, Department of Electrical and Computer Engineering, Schulich School of Engineering, 2500 University Drive N.W., Calgary, AB T2N 1N4, Canada

^bUniversity of Calgary, Division of Ophthalmology, Department of Surgery, Cumming School of Medicine, 2500 University Drive N.W., Calgary, AB T2N 1N4, Canada

Abstract. Retinopathy of prematurity (ROP), a disorder of the retina occurring in preterm infants, is the leading cause of preventable childhood blindness. An active phase of ROP that requires treatment is associated with the presence of plus disease, which is diagnosed clinically in a qualitative manner by visual assessment of the existence of a certain level of increase in the thickness and tortuosity of retinal vessels. The present study performs computer-aided diagnosis (CAD) of plus disease via quantitative measurement of tortuosity in retinal fundus images of preterm infants. Digital image processing techniques were developed for the detection of retinal vessels and measurement of their tortuosity. The total lengths of abnormally tortuous vessels in each quadrant and the entire image were then computed. A minimum-length diagnostic-decision-making criterion was developed to assess the diagnostic sensitivity and specificity of the values obtained. The area (A_z) under the receiver operating characteristic curve was used to assess the overall diagnostic accuracy of the methods. Using a set of 19 retinal fundus images of preterm infants with plus disease and 91 without plus disease, the proposed methods provided an overall diagnostic accuracy of $A_z = 0.98$. Using the total length of all abnormally tortuous vessel segments in an image, our techniques are capable of CAD of plus disease with high accuracy without the need for manual selection of vessels to analyze. The proposed methods may be used in a clinical or teleophthalmological setting. © 2016 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JMI.3.4.044505]

Keywords: computer-aided diagnosis; feature extraction; plus disease; retinal fundus image; retinopathy of prematurity; retinal vessel tortuosity.

Paper 16154R received Jul. 27, 2016; accepted for publication Nov. 18, 2016; published online Dec. 15, 2016.

1 Introduction

1.1 Retinopathy of Prematurity

Retinopathy of prematurity (ROP) is a retinal disorder that occurs in approximately 70% to 80% of preterm infants who are screened based on low birth weight (BW) and gestational age (GA).^{1,2} ROP is the leading cause of preventable childhood blindness.^{3–5} Preventive treatment in the form of laser photocoagulation may be effective only if ROP is diagnosed at its initial advancing stages.³ The need for treatment of ROP is warranted mainly based on the presence of plus disease,^{4,5} which is almost always present in acute ROP requiring treatment.^{3,6–8}

1.2 Plus Disease

Plus disease is clinically defined by the assessment of the presence of certain level of increase in tortuosity and width of retinal vessels by qualitative comparison to a standard photograph introduced in the original publication that classified ROP by a committee of international experts,⁹ that was later revisited by a larger committee.⁵ Such a procedure is inherently subjective, the limitations of which were confirmed by Chiang et al.⁶ in a study designed to assess the interexpert agreement of plus disease diagnosis based on RetCam retinal fundus images: the experts agreed on the diagnosis of only 12% of the images (4/34).

Furthermore, the standard photograph is believed to be atypical because it shows more vascular dilation and less tortuosity as compared to most cases with plus disease.³ Also, the image possesses a narrow field of view (FOV), and does not reveal possible tortuous vessels in the periphery of the retina, which have recently been shown to be more correlated to the presence of plus disease.^{10–12} The narrow FOV of the standard photograph is believed to represent the FOV of a funduscope when centered on the optic nerve head (ONH). However, to the best of the authors' knowledge, such a limit to the area of analysis has never been justified. In fact, it has been shown recently that experts do consider assessment of peripheral vessels in the diagnosis of plus disease.^{10,11} In addition to the mentioned changes to the vasculature, a change in the openness of the major temporal arcade (MTA), the thickest branch of venules, has also been observed as a sequela of ROP as well as an indicator of compromised structural integrity of the macular region.^{5,13–16} Such factors and limitations indicated the need for computer-aided methods to quantify and analyze the changes in the architecture of retinal blood vessels, such as tortuosity related to the presence of plus disease.

In recent years, there has been an increase in attempts for automated detection, segmentation, and quantification of various retinal features, such as the ONH, fovea, and the vasculature, as evidence by studies such as QUARTZ,¹⁷ ARIA,¹⁸ and VAMPIRE.¹⁹ There

*Address all correspondence to: Faraz Oloumi, E-mail: oloumif@ucalgary.ca

have been various semiautomated computer-aided procedures designed to perform diagnosis of plus disease by quantification of changes in the tortuosity of blood vessels.^{7,8,10,11,20-26} These studies have demonstrated that methods for computer-aided diagnosis (CAD) are capable of discriminating among cases with and without plus disease as accurately as experts. However, to the best of our knowledge, all studies that have performed diagnosis of plus disease via quantification of vascular tortuosity have employed methods that involve manual marking of vessel segments to be analyzed, or manual selection and correction of parts of automatically detected vessels to include only the desired or selected vessels for further analysis; such a procedure may not be feasible in a clinical or teleophthalmological setting due to lack of time and/or resources. Furthermore, all mentioned studies limited the area of analysis to the posterior of the retina; however, peripheral-vessel tortuosity has been shown to be more correlated to the presence of plus disease as compared to posterior-vessel tortuosity.¹⁰⁻¹²

The main focus of the present work is on the development of image processing methods for detection and quantitative analysis of abnormally tortuous vessels in retinal fundus images as well as an approach to use measures of tortuosity for diagnosis of plus disease. In addition to the results of quantification of tortuosity as a single diagnostic feature, the results of our previous related studies on the thickness²⁷ and the openness of the MTA¹⁴ are used to perform feature selection and pattern classification.

1.3 Quantification of Vascular Tortuosity

Although there have been several measures of tortuosity proposed in the literature, tortuosity does not have a specific mathematical definition. The methods in the literature can typically be divided into three categories: length-to-chord (LTC) measures, curvature-based measures, and angle-based measures. Measures of tortuosity are typically obtained using a skeletal representation of the vasculature, which is either obtained using image analysis algorithms in an automated or semiautomated environment or drafted manually using a specially designed graphical user interface (GUI). To the best of the authors' knowledge, all of the methods presented in the literature for measurement of tortuosity in retinal images of preterm infants for assessment of plus disease have used manual segmentation or selection of vessels. The following paragraphs explore 12 measures of tortuosity in the literature; for a more thorough review of the available methods for measurement of tortuosity in retinal images, please refer to the review article by Abdalla et al.²⁸

The simplest and one of the first formal definitions of tortuosity is the LTC measure, which is obtained as the ratio of the geodesic length, or arc-length, of a line segment (a part of a vascular skeleton in the context of this work) to the length of the straight line connecting the segment's end points (chord).^{7,24,29} The drawback of the LTC measure is that it may not account for possible changes in the curvature of a segment (twists and turns); as a result, an arched and a sinusoidal segment with the same geodesic and chord lengths may lead to the same tortuosity values, which is undesirable. Various studies have proposed modified definitions of the LTC measure in order to overcome this inherent limitation.

Wilson et al.³⁰ proposed a multiscale approach to the LTC measure, in which a given vessel segment was divided into two subsegments using the perpendicular bisector of the chord of the

segment (a line that is normal to the chord and divides it into two equal parts); the point of intersection of the perpendicular bisector with the vessel separates it into two subsegments. Each resulting subsegment was divided using the same method until a minimum specified length was reached. Tortuosity was then defined as the level of increase in total chord length after the final subdivision as compared to the initial chord length. It is expected that a tortuous segment would lead to a large increase in the value of the total chord length, providing a larger multi-scale LTC measure.

Hart et al.³¹ originally proposed the use of curvature to define several measures of tortuosity using skeletonized retinal vessel segments, including the total curvature and total squared curvature measures. Dougherty et al.³² fitted a polynomial spline to the traced center-line of a vessel segment and used the curvature of the spline to compute a measure of tortuosity.

Bolón-Canedo et al.³³ started with manually annotated maps of the vascular tree that were broken into separate segments by removing the branching points. Cubic splines were then fitted to the separated vessel segments and sample points from the fitted models were obtained and used for further analysis. A measure of tortuosity was defined as the second-order rate of change (acceleration) of the sampled points on a segment. Bolón-Canedo et al. did not specify the sampling rate on the cubic spline fits.

Using curvature, Grisan et al.³⁴ separated a given vessel segment into curved and linear parts. Each half of a linear subsegment was then assigned to its previous and next curved subsegments. Grisan et al. defined a measure of tortuosity for the entire segment as the sum of the LTC measure of each subsegment normalized by the true length of the entire segment and finally weighted based on the number of subsegments in the segment (number of sign changes in the curvature). The authors noted that their measure would provide misleading values if a subsegment does not contain any nonlinear parts.

The angle-based tortuosity measures are derived by first defining a set of vectors of fixed length connecting various points on a given segment and then using the angles formed at the tip and tail of two consecutively connected vectors to obtain a measure of tortuosity.

Gelman et al.²⁰ defined tortuosity as the sum of all angles (SOA) among a set of vectors, as previously explained, normalized by the total length of the segment. However, Gelman et al. did not specify the length of the vectors used to obtain the angles. The authors noted that the proposed methods would not work on a segment without any branching points.

Makkapati and Ravi³⁵ proposed a modified definition of the SOA measure by defining the angle for a given skeleton pixel using the two angular bisectors of three consecutive vectors normalized by the sum of the lengths of each subsegment. However, they did not specify the length of each vector.

Instead of defining an angle between successive vectors, Lisowska et al.³⁶ used the absolute change in the slope of the vectors and defined a measure of tortuosity for a segment as the sum of the slope differences.

By defining the angle of a given pixel based on the coordinates of the current and previous pixels, Poletti et al.³⁷ proposed an angle-based measure of tortuosity. The tortuosity measure was computed as the sum of the squared angle changes, divided by the total length of the vessel segment, where the local change in angle was defined as the difference in the inverse tangent of the angle of the current and previous pixels. It should be noted that such a definition of the angle at a pixel based on its

8-connected neighboring pixels will only provide angle measures with increments that are integral multiples of 45 deg.

The tortuosity measures that use the SOA approach are dependent on the length of the vectors used on a vessel segment to define the angles; short vector lengths may result in large angle values, whereas long vector lengths may miss small local variations. Ideal representation of the skeleton (or center-lines) of vessels is a difficult task that is prone to discretization errors as well as inaccuracies in detection and segmentation of vessels; defining the orientation of a given vessel pixel using the coordinates of such a skeletal representation of vessels is also prone to errors.

Several studies^{7,8,10,11,20-26} have measured vessel tortuosity in retinal fundus images of premature infants and correlated their findings to the presence of plus disease in the images. The protocols used and the results obtained in these studies as compared to the same in this work are presented in Sec. 3.3.

We propose image-processing methods that do not require manual selection or correction of vessels to be analyzed and do not limit the area of analysis to only the posterior of the retina. We define an angle-variation-based tortuosity (AVT) measure^{38,39} that considers the precise dominant orientation of vessel at each pixel and not the discrete and limited variations in a skeletal representation of a vessel. We also define diagnostic-decision-making criteria that consider the clinical definition of plus disease with regard to tortuosity as well as a practical approach for objective and clear presentation of the information related to abnormally tortuous vessels to an ophthalmologist. The proposed analysis of vascular tortuosity is performed in three regions: the posterior, the periphery, and the entire FOV of the image.

The present paper is an updated, revised, and vastly expanded version of the related preceding work published in a conference proceeding,³⁹ in terms of introduction, review of the state-of-the-art methods, the diagnostic-decision-making criterion, bootstrapping, receiver operating characteristic (ROC) analysis, analyzing the entire length of tortuous vessels in three regions of the retina, correlation of image features to patient features, feature selection, pattern classification, results,

comparative analysis, illustrations, discussion, limitations, and future work.

2 Materials and Methods

2.1 Database of Fundus Images of the Retina

The telemedicine for ROP in Calgary (TROPIC) database⁴⁰ is a private collection of retinal fundus images of preterm infants that is used in the present work for CAD of plus disease. Written consent was obtained from the parents of the patients to capture and use the images. The tenets of the declaration of Helsinki were followed while compiling the database. The images of the TROPIC database were captured using the RetCam II camera equipped with a wide-angle ROP lens (130 deg). The RetCam II images have a size of 480 × 640 pixels and are considered to possess a poor spatial resolution of about 30 μm per pixel.⁴¹

In total, 110 images from 41 preterm patients (16 females and 25 males) were selected from the database for the present study. In most cases, there are five different images available for each eye of each patient from each visit, representing different retinal fields to provide collectively an almost complete photographic documentation of the retina. In each case, the image with the highest visibility of the entire vasculature was chosen. Images were not selected based on overall quality and/or vessel-to-background contrast. Nineteen of the 110 selected images are from patients diagnosed with plus disease (stages 2 and 3 of ROP) and 91 images show no signs of plus disease (stages 0, 1, 2, and 3 ROP). At most, two images from the same patient were included for the same stage of ROP (one image from each eye). Multiple images of the same eye from the same patient were included only if the ROP stages were different at the time of imaging. All diagnoses were performed by a retinal specialist (A. L. Ells) at the time of clinical examination using a funduscope and by considering the patient's clinical records. It is of importance to note that the diagnosis was not based solely on the RetCam images of the patient. Patients corresponding to 90 of the 110 images were diagnosed with no ROP or with stages

Table 1 Upper portion: BW, GA, and CA of the patients. Lower portion: computed total abnormally tortuous-vessel length in the posterior, periphery, and the entire FOV. The p -values, area under the ROC curve (A_z), and 95% ($\alpha = 0.025$) asymmetric confidence interval (CI_a) in the discrimination of 19 cases with plus disease versus 91 cases without plus disease using the total abnormally tortuous-vessel length in each region as well as the BW, GA, and CA of the patients are provided.

Parameter	Without plus, mean $\pm$ STD ($n = 91$)	With plus, mean $\pm$ STD ($n = 19$)	p -value	A_z , CI_a
BW (g)	818.00 $\pm$ 210.78	815.89 $\pm$ 203.71	0.968	0.51, [0.367, 0.654]
GA (weeks)	26.73 $\pm$ 1.88	24.95 $\pm$ 1.77	<0.001***	0.82, [0.690, 0.904]
CA (days)	71.05 $\pm$ 23.67	69.84 $\pm$ 13.00	0.831	0.50, [0.390, 0.604]
Posterior tortuous-vessel length (mm)	0.53 $\pm$ 0.77	4.62 $\pm$ 2.65	<0.001***	0.88, [0.712, 0.978]
Peripheral tortuous-vessel length (mm)	0.95 $\pm$ 1.34	11.91 $\pm$ 9.85	<0.001***	0.94, [0.859, 0.987]
Total tortuous-vessel length (mm)	1.48 $\pm$ 1.64	16.53 $\pm$ 10.56	<0.001***	0.98, [0.910, 0.997]

*** $p < 0.001$.

1 or 2 ROP (30 images per category), and patients corresponding to 20 images were diagnosed with stage 3 ROP.

The upper portion of Table 1 provides the mean, standard deviation (STD), and the associated p -values for the BW, GA, and chronological age (CA) of the patients. Note that GA + CA = postmenstrual age. The TROPIC database was used in our previous studies on the openness¹⁴ and thickness²⁷ of the MTA in the presence of plus disease.

A training set of 10 images, including 5 without and 5 with signs of plus disease, was also formed. All abnormally tortuous vessel segments in the five training images corresponding to plus disease were manually identified by the retinal specialist. Note that abnormal tortuosity includes vessels that are not sufficiently tortuous (preplus) for plus diagnosis, but are not normal either. The images of the training set are mutually independent of the test set described in the previous paragraph, even though they are drawn from the same population of patients, i.e., the training set images either show a different stage of ROP or were taken at another visit, even if showing the same stage of ROP as in those visits from which the test set was collected.

2.2 Detection and Segmentation of the Retinal Vascular Skeleton

Detection and segmentation of retinal vessels are prerequisite steps for the analysis of retinal features. In this work, Gabor filters, which are sinusoidally modulated Gaussian functions tuned to a certain bandwidth (thickness, τ), elongated (l), and rotated over a range of orientations (K), were employed for detection of retinal vessels. The K -number of oriented filters are spaced evenly over the range $[-\pi/2, \pi/2]$.⁴² The frequency response of the main Gabor kernel oriented at $-\pi/2$ is defined as

$$G(u, v) = \frac{1}{2} (\exp\{-2\pi^2[\sigma_x^2(u + f_o)^2 + \sigma_y^2v^2]\} + \exp\{-2\pi^2[\sigma_x^2(u - f_o)^2 + \sigma_y^2v^2]\}), \quad (1)$$

where $\sigma_x = \tau/(2\sqrt{2 \ln 2})$ and $\sigma_y = l\sigma_x$ are the STDs of the Gaussian function in the x and y directions, respectively, and $f_o = 1/\tau$ indicates the frequency of the modulating sinusoid. The vessel-strength (Gabor-magnitude response) image is a gray-scale representation of the detected vessels where the higher the intensity value of a pixel, the higher the probability of the pixel belonging to a vessel. For more details on Gabor filters and comparison to various other vessel detection methods, please refer to the related publications.⁴³⁻⁴⁴ The vessel-strength image was used to obtain a skeletal representation of the vasculature in this work. The vessel-orientation (Gabor-angle response) image, which provides the dominant vessel orientation at each pixel, was used to compute the measure of AVT. The steps required in the application of Gabor filters to retinal fundus images of preterm infants and obtaining a skeletal representation of the vasculature are stated below.

1. Apply the K Gabor filters to the inverted green-channel image and obtain the Gabor-magnitude-response image using the maximum response at each pixel over the K orientations and the Gabor-angle-response image as the corresponding orientation (angle) of the filter with the maximum response at each pixel.
2. Normalize the gray-scale values of the Gabor-magnitude-response image to the range $[0, 1]$.

3. Binarize the normalized Gabor-magnitude-response image using a threshold slider via a GUI.
4. Remove segments shorter than a maximum length (in pixels) as specified by the user via the GUI using the morphological area-open procedure.
5. Skeletonize the binary image by considering 8-connectedness for all vessel segments.
6. Remove an elliptical area representing the ONH from the skeleton image, centered at the center of the ONH.
7. Remove spurs that are 5 pixels or shorter in length.
8. Detect branching points on the skeleton.
9. Remove an area of size 3×3 pixels centered at previously detected branching points to separate all vessel segments.
10. Remove remaining segments that are 7 pixels or shorter in length.

The width and height of the ellipse representing the ONH area were set based on the average and STD of ONH width (ONHW) and ONH height (ONHH) in preterm infants⁴¹ as 44 and 60 pixels (mean plus two times the STD in each case), respectively. The approximate location of the center of the ONH was previously marked by the user in a related study.¹⁴

All 110 test images were randomly presented to the user (F. Oloumi) via the GUI, one at a time, without revealing the associated diagnosis and processed based on the above steps. The parameters of Gabor filters were set empirically as $\tau = 7$ pixels, $l = 1$, and $K = 45$ for all images. The two binarization parameters in steps 3 and 4 were selected separately for each image with the objective of maximizing the number of segmented retinal vessels obtained while minimizing the artifacts due to choroidal vessels.

2.3 Detection of the Linear Parts of Vessel Segments

A key challenge in the computation of a measure of tortuosity of a vessel segment is to determine whether parts of the segment are linear (with little to no change in orientation) and to remove such parts from further analysis. In the present work, the vessel-orientation (Gabor-angle response) information at each pixel for each identified segment along with the median absolute deviation (MAD)⁴⁵ measure were used to analyze and separate the vessel segments into linear and nonlinear subsegments. The MAD measure is defined as $\text{median}[|X - \text{median}(X)|]$, where X represents the values under analysis and $|\cdot|$ represents the absolute value.⁴⁵ Various steps of the algorithm are presented below.

1. Detect all disconnected vessel segments obtained in the previous step and number them.
2. Determine the sequential order of the skeleton pixels as follows:
 - a. Ensure that the segment under analysis has exactly two end points.
 - b. Set the starting point at one of the end points.
 - c. Move sequentially one pixel at a time in a 3×3 -pixel neighborhood along the segment.

- d. Record the position (spatial index) of each traversed pixel and its associated Gabor-angle response in a vectorial format.
- e. Repeat steps (c)–(d) until the other end point is reached.
3. Compute the MAD measure in a window of length 7 pixels applied to the sequenced Gabor-angle responses.
4. Mark a pixel on the segment under analysis as being linear only if all MAD measures in a 7-pixel-long window centered at the current pixel are zero.
5. Obtain skeleton image of only linear segments as indicated by the MAD measure.
6. Obtain skeleton image of nonlinear vessel segments by subtracting the image of linear segments from the original skeleton image.

The length of the analysis window was empirically set to 7 pixels. At the end of this procedure, two skeleton images of the vasculature are available: one image representing the linear segments, which will be excluded from further analysis, and one image representing the segments that are considered to be nonlinear. The following sections provide details of the method for derivation of the AVT measure as applied to the nonlinear retinal vessel segments.

2.4 Quantification of Vascular Tortuosity

Given a nonlinear vessel segment, a local-tortuosity index (LTI)³⁸ based on the sequenced vessel orientation (Gabor-angle response) information was obtained at every pixel as

$$\text{LTI}(p) = \{ |\sin[\phi(p) - \phi(p-1)]| + |\sin[\phi(p) - \phi(p+1)]| \} / 2, \quad (2)$$

where p , $p-1$, and $p+1$ are the current, previous, and next pixels, respectively. This formulation ensures that LTI has a theoretical maximum value of 1 and a minimum of 0. The AVT measure for a given nonlinear vessel segment is defined as the average of the LTI values along the segment as

$$\text{AVT} = \frac{1}{P} \sum_{p=1}^P \text{LTI}(p), \quad (3)$$

where P is the total number of pixels in a given segment. The AVT measure is normalized so that it provides a maximum value of 1 and a minimum of 0 for each vessel segment. By definition, $\text{AVT} = 0$ for a straight line.

The training set of images (see Sec. 2.1) was used to obtain a suitable threshold for the AVT measure by comparing the AVT values for the abnormally tortuous vessel segments, as marked by the retinal specialist, to the AVT values for vessels with normal levels of curvature (variation in the orientation). Based on this empirical analysis, a threshold of $t = 0.07$ was determined in order to discriminate between abnormally tortuous and normal vessel segments. Each nonlinear vessel segment with $\text{AVT} > t$ was automatically marked as being abnormally tortuous and as normal if $\text{AVT} \leq t$.

2.5 Diagnostic Decision Making

Clinical diagnosis of plus disease requires the presence of sufficient increase in tortuosity and thickness, as compared to the standard photograph, in at least two quadrants of the image. Since tortuosity is not formally defined, numerical representations of its quantitative measurement, such as the AVT, may not be directly meaningful to an ophthalmologist. Furthermore, providing one average or maximum measure of tortuosity for an entire retinal fundus image, or even for each quadrant of an image, may be misleading if the tortuous segments are not sufficiently long, i.e., two highly tortuous but short segments in two different quadrants should not lead to a positive diagnosis. In addition, averaging the AVT measure over all segments in an image or per quadrant of the image may lead to a washout effect, which is undesirable.

In this work, we assess the diagnostic accuracy of a single diagnostic-decision-making threshold that follows the clinical definition of plus diagnosis concerning vessel tortuosity (based on each quadrant of the image), called the minimum-length criterion. Furthermore, overall diagnostic significance of the total lengths of abnormally tortuous vessels in a given image, in the posterior, periphery, and the entire FOV, were considered as separate diagnostic features without the use of a single threshold; a sliding-threshold method was used to determine the overall diagnostic accuracy of each feature as discussed in the following section. Note that the results of the sliding-threshold method may also be used to obtain a different classification threshold for the minimum-tortuous-vessel length for the entire image as opposed to each quadrant to be used in a future study that does not employ the same set of images. The posterior section of the image was defined as a circle centered at the center of the ONH with radius $= 2 \times \text{ONHH}$, and the peripheral section was taken as the entire FOV minus the posterior area.

In the present work, the total length of tortuous vessels in each quadrant was computed by dividing the skeleton image of the automatically labeled abnormally tortuous vessels into four quadrants with reference to the center of the ONH. All segments in each quadrant were then identified and their associated chain-code⁴⁶ representations were obtained. The length of each segment was computed as the number of even codes plus $\sqrt{2}$ multiplied by the number of odd codes. Such an approach ensures that the diagonal length of a pixel is taken as $\sqrt{2}$ pixels. The true length of a segment was obtained by multiplying the total length of a segment (in pixels) with the spatial resolution of the TROPIC images ($30 \mu\text{m}/\text{pixel}$).

A suitable minimum-tortuous-vessel-length was obtained as the diagnostic-decision-making threshold of plus disease by analyzing the length of the abnormally tortuous vessels, over the entire FOV in the training set, as marked by the retinal specialist. A diagnosis (classification) of plus disease is made in the present work if an image contains at least 2.5 mm of abnormally tortuous vessels in each of at least two quadrants, or at least 5 mm of abnormally tortuous vessels in any one quadrant. It was not possible to obtain separate minimum-tortuous-vessel-length thresholds for the posterior and the periphery as the abnormally tortuous vessels were marked over the entire FOV in the training set.

2.6 Evaluation of the Diagnostic Performance

Using the minimum-tortuous-vessel-length criterion, a single set of sensitivity (true-positive rate) and specificity (true-negative rate) values was obtained to characterize the diagnostic accuracy

of the proposed methods at a single operating point over the entire FOV. To assess the overall diagnostic performance of the total length of abnormally tortuous vessels in the posterior, periphery, and the entire FOV of the image at multiple operating points, ROC analysis was performed using the ROCKIT software.⁴⁷ The values of the area under the ROC curve, A_z , as well as their asymmetric 95% confidence intervals (CI_a) were obtained from ROCKIT. The A_z value provided by ROCKIT is based on the area under the binormal fit to the ROC curve, which may vary slightly as compared to the area under a discrete version of the ROC curve (AUC).

In addition to the diagnostic criteria mentioned in the previous section, it is possible to select the most suitable threshold for a minimum-tortuous-vessel-length criterion for the entire image using the point on the ROC curve that provides the best trade off between sensitivity and specificity values. The ROC curve provides the sensitivity and specificity of the diagnostic feature used by varying the threshold for classification and plotting the obtained sensitivity values against $(1 - \text{specificity})$ values. The most suitable operating point is often taken as the point on the ROC curve that is closest to the point $(0, 1)$.

Considering the limited number of cases in both classes, the bootstrap method^{48,49} was implemented for further statistical analysis of the results. The bootstrap is a method used to analyze the level of confidence in an estimate of a parameter of a population such as the mean, when there are a small number of data points available. The bootstrapping method resamples a population (a class) by randomly selecting a larger number of cases from the given population, with replacement. This random-selection procedure is repeated hundreds or thousands of times and the associated confidence intervals of the results are obtained.

In the present work, both of the classes with and without plus disease were resampled 150 times with replacement and the associated mean parameter of each class and the AUC values were obtained. The resampling procedure was repeated 500 times and the associated symmetric 95% confidence intervals (CI_s), calculated by assuming a normal distribution for the obtained mean and AUC values, were obtained. An in-house ROC analysis program was used to perform the bootstrapping procedure, since ROCKIT does not allow for batch processing of large amounts of data. The in-house ROC analysis program provides the AUC value and not the A_z value based on the binormal estimation of the curve as in ROCKIT. As a result, there could be small differences among the values obtained using the two programs, as discussed in Sec. 3.

2.7 Feature Selection and Pattern Classification

When dealing with a set of features to perform CAD, it would be of interest to determine whether a subset of the computed features will yield a more precise level of discrimination than a single feature. When combined together through pattern classification methods, combinations of various features may provide complementary information and lead to higher diagnostic accuracy than any individual feature on its own. Pattern classification methods of naïve Bayes,⁵⁰ logistic regression (LR),⁵¹ and multilayer perceptrons (MLPs)⁵⁰ were employed in this work for such analysis.

Supervised pattern classification methods require training. Considering the limited number of available cases in this work (110 cases from the TROPIC database) the k -fold cross-validation method was used to perform the training and testing steps. Such analysis divides the dataset into k parts (folds), and at each

iteration, one fold is kept aside while the rest of the data are used to train the chosen classifier. Next, the trained classifier is applied to the fold kept aside;^{50,52} the procedure is repeated over all folds and averaged measures of performance are obtained.

The feature selection technique of the wrapper method that employs supervised learning along with classification algorithms, such as those explained above, was used to select the most discriminatory combination of features based on the given set of data.⁵² Two search methods of exhaustive⁵⁰ and best-first⁵² were employed to obtain the set of features that provided the highest level of discrimination between cases with and without plus disease. In all instances, $k = 11$ folds (10 images each) were used to perform cross-validation. The features that were selected more than half the time ($\geq k/2$ folds) using the proposed setup were used for pattern classification using the same classifier as with the wrapper method.

In this work, the three tortuosity features obtained in various regions of the image were combined with the features obtained as part of our previous related studies including the thickness of the MTA²⁷ (one feature) and openness of the MTA (five features),¹⁴ for a total of nine features, to perform feature selection and pattern classification. Please refer to the associated references for the specific details of how the features of thickness²⁷ and openness of the MTA¹⁴ were quantified.

3 Results

Using a Lenovo Thinkpad T510 equipped with an Intel Core i7 (hyperthreaded-dual-core) 2.67-GHz processor, 8 GB of DDR 3 RAM, running 64-bit Windows 10 Professional, and using 64-bit MATLAB® R2013 software, on the average, it took approximately 45 s to process and analyze each image using the proposed methods.

Figure 1 shows the results of applying the proposed methods to an image without any signs of plus disease from the TROPIC database. No segment was found to be abnormally tortuous in all four quadrants based on the predetermined AVT threshold.

Figure 2 shows the same steps for an image of a patient diagnosed with plus disease. The total sum of the lengths of abnormally tortuous vessels in the first to fourth quadrants of the image were computed to be 11.75, 4.20, 1.99, and 1.42 mm, respectively.

Figure 3(a) shows the steps involved in statistical analysis of the vessel-orientation information in order to break apart an extracted vessel skeleton segment into its linear and nonlinear parts. The analysis is shown only for an extracted segment of the superior temporal arcade (STA) [(b) of the same figure] corresponding to the image in Fig. 2. The resulting subsegments shown in (c) are considered to be linear and those shown in (d) are considered to be nonlinear. Using the AVT threshold of 0.07, only the right-most nonlinear subsegment (the longest of the three subsegments) in Fig. 3(d) was determined to be abnormally tortuous, with $AVT = 0.15$; this subsegment is shown in red in Fig. 2(h) along with other abnormally tortuous subsegments.

Using the minimum-tortuous-vessel-length threshold for the entire FOV derived in the present work (see Sec. 2.5), sensitivity = 0.89 (17/19) and specificity = 0.99 (90/91) were obtained leading to a classification accuracy = 0.97 (107/110). Note that the same minimum-tortuous-vessel-length threshold may not be applicable to the results obtained over the posterior and the periphery of the retina.

By treating the patient attributes of BW, GA, and CA as separate features, the same analysis was performed to assess the

Fig. 1 (a) Image 4002 of the TROPIC database showing no signs of plus disease. (b) The vessel orientation for every third pixel on vessels for a portion of (a) is shown by the blue needles. (c) The vessel-strength image obtained using the green channel of (a). (d) The skeleton image after thresholding (c) at 0.14 of the normalized intensity value and removal of 8-connected segments having fewer than 100 pixels each. (e) The skeleton image after removal of the ONH area and the branching points. (f) The skeleton image of the linear vessel subsegments obtained using the MAD measure. (g) Skeleton image of the nonlinear vessel subsegments obtained by subtracting (f) from (e). (h) Color-coded skeleton image distinguishing the abnormally tortuous vessel segments, if any, in red and the normal vessel segments in green, using the AVT threshold of 0.07; in this case, no segment was found to be abnormally tortuous. Images in (d)–(h) have been morphologically dilated using a disk of radius one pixel for better visual representation in the figure.

Fig. 2 (a) Image 2903 of the TROPIC database that shows signs of plus disease. (b) and (c) The same as the caption in Figs. 1(b) and 1(c). (d) The skeleton image after thresholding the vessel-strength image at 0.045 of the normalized intensity value and removal of 8-connected segments having fewer than 100 pixels each. (e)–(g) The same as the caption in Figs. 1(e)–1(g). (h) Color-coded skeleton image distinguishing the abnormally tortuous vessel segments in red, and the normal vessel segments in green, using the AVT threshold of 0.07.

diagnostic power of these attributes. The results (also presented in the upper portion of Table 1) indicate that the difference in mean GA of the patients with plus disease as compared to those without plus disease is highly statistically significant, and also provides a high value of $A_z = 0.82$ in discrimination between the two classes.

The Pearson correlation coefficients indicating the relationships between the total length of abnormally tortuous vessels in the image and the three patient attributes of BW, GA, and CA as well as their associated levels of statistical significance were obtained. The results indicated that there is little to no correlation⁵³ between total tortuous-vessel length and the patient

Fig. 3 (a) The vessel orientation along a sequenced skeleton segment, its associated normalized MAD measure obtained over a 7-pixel-long window, and the detected linear samples are plotted against the sequence index. (b) The extracted STA segment [in the first quadrant of the image in Fig. 2(e)] on which the statistical analysis shown in (a) is performed. The result of segmenting the skeleton in (b) into (c) linear and (d) nonlinear subsegments. The AVT measures computed for the three nonlinear subsegments in (d) of the figure are 0.05, 0.04, and 0.15, from left to right, respectively.

attributes of BW and CA. The total tortuous-vessel length and GA showed higher correlation coefficient values, yet were not found to be meaningfully correlated to each another.

3.1 ROC Analysis

The lower part of Table 1 shows the results of statistical and diagnostic analysis of the total abnormally tortuous vessel length using 91 cases without plus disease and 19 cases with plus disease over the posterior, periphery, and the entire FOV of the image. The results indicate extremely statistically significant differences (p -value < 0.001) between the means of the total abnormally tortuous vessel lengths for cases with and without plus disease over the entire FOV, and an excellent performance in classification and diagnosis of plus disease with $A_z = 0.98$. The same measure provides high values of $A_z = 0.88$ and 0.94 over the posterior and periphery, respectively. Figure 4 shows the three ROC curves obtained based on analysis of the total length of abnormally tortuous vessels in the posterior, periphery, and the entire FOV.

Table 2 presents the results of applying the bootstrapping approach as explained in Sec. 2.6. The results indicate similar average AUC and mean values as compared to the results in the lower portion of Table 1. However, the 95% confidence intervals are much narrower, indicating that there is a high level of confidence in the estimated values for the average tortuous-vessel lengths for cases with and without plus disease, as well as in discrimination between the two classes using ROC analysis. Based on the results of bootstrapping, it may be concluded that the original set of values obtained (Table 1) can be considered to be representative of the population distribution to which they belong. As mentioned in Sec. 2.6, the ROC analysis procedure used for the bootstrapping approach provides the AUC values as opposed to the A_z values provided by ROCKIT; hence, there is a small difference (0.01) between the A_z values in Table 1 and the AUC values in Table 2.

Analysis of the total tortuous-vessel length related to images from the same eye of the same patient that progressed to plus disease and a higher stage of ROP, on average over the three patients and four imaging instances available in the TROPIC

Fig. 4 ROC curves presenting the diagnostic accuracy of the total length of tortuous vessels obtained using the proposed methods to distinguish between cases without and with plus disease over the posterior ($A_z = 0.88$), periphery ($A_z = 0.94$), and the entire FOV ($A_z = 0.98$). A potential optimal operating point is the point on the ROC curve, associated with the entire FOV, that is closest to the point (0, 1), with sensitivity = 0.93 and specificity = 0.97 in the present case. The ROC curves associated with the posterior and peripheral vessel lengths provide potential optimal operating points with sensitivity = 0.80 and 0.86, as well as specificity = 0.90 and 0.89, respectively. The trade off between sensitivity and specificity must be determined based on the clinical application.

database, showed an increase of 10.82 mm in the total length of abnormally tortuous vessels detected.

3.2 Feature Selection and Pattern Classification

Table 3 presents the results of feature selection using the wrapper method and cross-fold validation. In all instances, the tortuous-vessel length over the entire FOV was selected in 100% of the folds; only the naïve-Bayes classifier selected other features as well, including the openness parameter of the single-parabolic model and the two TAA measures.

Pattern classification was performed using the results of feature selection via the naïve-Bayes model using the two search methods. However, in both instances, the results of pattern classification did not yield higher A_z values as compared to using the total tortuous-vessel length alone ($A_z = 0.98$). The pattern

classifier based on the naïve-Bayes model achieved the same sensitivity or specificity as compared to classification using the minimum-vessel-length criterion with total tortuous-vessel length, at the expense of lower specificity or sensitivity values, respectively.

The measure of the thickness of the MTA was not selected in any of the analyses. Various other methods were tested for feature selection and classification, but the results are not provided as they were the same as stated above; all methods always included the total tortuous-vessel length with at least 90% selection rate during cross validation. None of the methods that selected any additional features along with total tortuous-vessel length led to improved diagnostic results in terms of A_z values.

3.3 Comparative Analysis

To the best of our knowledge, 11 studies^{7,8,10,11,20–26} have performed CAD of plus disease based on quantification of vessel tortuosity; however, the approaches to analysis, diagnosis, and validation of the results have not been standardized across all of the studies and show inconsistencies. All of the mentioned studies, except the works of Ataer-Cansizoglu et al.¹⁰ and Campbell et al.,¹¹ have limited the area of analysis to the posterior, defined as a circle centered at the center of the ONH with radius $\approx 2 \times \text{ONHH}$. Some of the studies have obtained separate measures for arterioles and venules and some have obtained combined measures for both types of vessels. All studies have used manual markings of the center of the ONH, threshold values provided by the user for binarization, and manual selection and correction of the vessel segments to be analyzed. Table 4 presents comparative analysis of the results of the mentioned studies. The main differences among these studies are in the approach to diagnosis and validation of the results including the conditions that warrant a positive diagnosis of plus disease.

The studies conducted by Gelman et al.,^{7,20} Koreen et al.,⁸ and Thyparampil et al.²³ all used the software package RISA.⁵⁵ RISA provides two different measures of tortuosity (the LTC and SOA) and analyzes venules and arterioles separately. Only the best diagnostic results, as provided by these works, are reported in Table 4. Two of the studies^{20,23} listed above only provided A_z values, whereas the other two studies^{7,8} provided A_z as well as a single set of sensitivity and specificity values. However, the sensitivity and specificity values were determined by finding the intersection point between the sensitivity and specificity curves, plotted separately as functions of the ratio of the number of correctly identified vessels to the actual tortuosity measures. Such a measure classifies each single vessel. RISA requires the vessel segment under analysis to have at least one branching point. Manual correction of the detected branching

Table 2 Values of the mean of each class, the mean of the AUC, and their associated 95% ($\alpha = 0.025$) symmetric confidence intervals, CI_s , obtained by randomly selecting 150 cases with and without plus disease, repeated 500 times for total tortuous-vessel lengths in the posterior, the peripheral, and entire FOV. The confidence intervals were obtained by assuming a normal distribution for the AUC and class-mean values. In the last column, M^{***} means that, out of the 500 trials, the results indicated statistically extremely significant differences in M trials with $p \leq 0.001$.

Parameter	Without plus, mean (SE)	With plus, mean (SE)	AUC, mean (CI_s)	Statistical significance
Posterior tortuous-vessel length (mm)	0.533, (0.003)	4.623, (0.012)	0.905, [0.904, 0.907]	500***
Peripheral tortuous-vessel length (mm)	0.950, (0.006)	11.822, (0.044)	0.942, [0.941, 0.943]	500***
Total tortuous-vessel length (mm)	1.487, (0.008)	16.485, (0.050)	0.970, [0.970, 0.971]	500***

Table 3 Results of feature selection using various pattern classification methods employed via the wrapper method and cross-fold validation. In all instances, 11 folds were used during validation and a feature was included if it was selected in at least 50% of the folds.

Classifier	Selected feature(s)	
	Best-first	Exhaustive
LR	{Tortuosity _{total} }	{Tortuosity _{total} }
Naïve Bayes	{Tortuosity _{total} , a _{MTA} }	{Tortuosity _{total} , Tortuosity _{posterior} , TAA _{r=60} , TAA _{r=120} }
MLP	{Tortuosity _{total} }	{Tortuosity _{total} }

points and manual input regarding whether the selected vessels are arterioles or venules were also required. In all studies, the arteriolar and venular tortuosity measures were obtained separately. Although the studies of Gelman et al.⁷ and Koreen et al.⁸ provided results for various combinations of venular/arteriolar tortuosity/dilation measures, none of the mentioned studies provided diagnostic results for combinations of only the arteriolar and venular tortuosity measures. The values in Table 4 as provided by Thyparampil et al.²³ were obtained using venular

tortuosity, whereas the other three studies^{7,8,20} obtained higher diagnostic accuracy using arteriolar tortuosity.

Studies by Wallace et al.,²⁵ Johnston et al.,⁵⁴ Zhao et al.,²⁶ and Kiely et al.²² all used the ROPTool software that obtains a measure of tortuosity based on the LTC measure. Only the study by Johnston et al.⁵⁴ provided separate and combined measures of tortuosity for venules and arterioles for each quadrant of the image. The other three studies combined the tortuosity measures obtained from both venules and arterioles by taking the maximum tortuosity value in each quadrant as the overall measure. In all studies, diagnosis of plus disease was performed on each quadrant of the image. Studies by Wallace et al.²⁵ and Johnston et al.⁵⁴ also provided per-image diagnosis using the quadrant-based results; if at least two quadrants of an image showed a sufficient level of tortuosity, the image was considered to represent a case with plus disease. However, Johnston et al.⁵⁴ did not specify the required level of tortuosity. Furthermore, it is unclear how such a binary per-image diagnosis could be used to perform ROC analysis and obtain an A_z value as in the work of Johnston et al.⁵⁴ Only Wallace et al.²⁵ provided separate counts of the number of images with and without plus disease. Zhao et al.²⁶ provided only an A_z value; the threshold obtained based on this study was used in the work of Kiely et al.²² to obtain measures of sensitivity and specificity, as well as an overall A_z value. Johnston et al.⁵⁴ did not provide sensitivity and

Table 4 Comparative analysis of the state-of-the-art studies available in the literature on CAD of plus disease. If a particular parameter or result was not specified in a study, it is denoted by NP (not provided). Sen. and Spe. stand for sensitivity and specificity, respectively.

Publication	Vessel detection	Selection of vessels to analyze	Analysis of venules and arterioles	# of cases without plus and with plus	A_z	Sen. and Spe. (threshold)	Sen. and Spe. (ROC)
Gelman et al. ²⁰	Manual	Manual	Separate	20 and 12	0.91 ^a	NP	NP
Gelman et al. ⁷	Manual	Manual	Separate	21 and 13	0.82 ^a	NP	0.76 and 0.76 ^{a,d}
Koreen et al. ⁸	Manual	Manual	Separate	14 and 6	0.96 ^a	NP	1.00 and 0.85 ^{a,d}
Thyparampil et al. ²³	Manual	Manual	Separate	61 and 13	0.95 ^b	NP	NP
Wallace et al. ²⁵	Manual	Manual	Combined	9 and 11	NP	1.00 and 0.78	NP
Zhao et al. ²⁶	Manual	Manual	Combined	20 ^h	0.91 ^c	NP	NP
Johnston et al. ⁵⁴	Manual	Manual	Combined	77 ^h	0.97 ^g	NP	NP
Kiely et al. ²²	Manual	Manual	Combined	92 ^h	0.95 ^c	1.00 and 0.66 ^c	0.91 and 0.86 ^c
Wallace et al. ²⁴	Manual	Manual	Combined	11 and 5	NP	0.82 and 0.80	NP
Ataer-Cansizoglu et al. ¹⁰	Manual	Semiautomatic ^j	Combined	63 ⁱ and 14	NP	NP	NP
Heneghan et al. ²¹	Semiautomatic ^f	Manual	Combined	12 and 11	NP	NP	0.82 and 0.75 ^e
Present work	Semiautomatic ^f	Automatic	Combined	91 and 19	0.98	0.89 and 0.99	0.93 and 0.97

^aDiagnosis performed on each selected arteriole, not the image.

^bDiagnosis performed on each selected venule, not the image.

^cDiagnosis performed based on the maximum tortuosity value of selected vessels for each quadrant, not the image.

^dValue obtained at the point of intersection of the sensitivity and specificity plots of diagnosis of each vessel, not using ROC analysis.

^eDiagnosis performed using tortuosity and thickness measures combined.

^fRequires two parameters to be set by the user.

^gROC analysis performed using binary diagnostic results (Johnston et al.⁵⁴ did not describe how such analysis was performed).

^hDiagnosed per image quadrant.

ⁱIncludes preplus cases as well as cases without plus disease.

^jRequires a manually drafted vessel map.

specificity values based on the per-image diagnostic results and concluded that combining arterioles and venules provides similar or better diagnostic results as compared to considering arteriolar tortuosity alone.

Ataer-Cansizoglu et al.¹⁰ used hand-drawn annotations of blood vessels in 77 retinal fundus images of preterm infants to perform CAD of plus disease. Forty-seven images did not show any signs of plus disease, 16 were marked as showing increase in tortuosity and thickness that were not sufficient enough for plus diagnosis (designated as preplus), and 14 images showed signs of plus disease. The study was limited to images of sufficient quality and contrast. The diagnostic decision was reached by extracting the acceleration feature³³ of both vessel types (arterioles and venules combined) and training and testing the data assuming a Gaussian-mixture model via cross-fold validation. The authors performed three-class diagnosis, quantified the results in terms of accuracy, but did not provide values of sensitivity, specificity, or A_z . Ataer-Cansizoglu et al.¹⁰ varied the range of the area of analysis of tortuous vessels over circles with radii = $1 \times$ to $6 \times$ ONHW centered at the center of the ONH. The authors concluded that the larger the area of analysis, i.e., the farther away from the center of the ONH, the higher the level of diagnostic accuracy of plus disease when considering vascular tortuosity with an accuracy = 0.95. In a related study by Campbell et al.,¹¹ the same set of 77 images were assessed by 11 experts and a new reference diagnosis was obtained for each image. Similar to the original work, using the acceleration feature Campbell et al. obtained the highest accuracy (0.95) when considering the tortuosity of all types of vessels in a circular area with radius = $6 \times$ ONHW.

Heneghan et al.²¹ obtained the average thickness and tortuosity values for an image by combining measurements from both arterioles and venules. Sensitivity and specificity values in the diagnosis of threshold disease were then obtained using the combination of the two measures. The study did not provide A_z values or diagnostic results using only the tortuosity measurements.

The study conducted by Wallace et al.²⁴ is the only study that is almost directly comparable to the present work. Wallace et al.²⁴ used an LTC-based measure of tortuosity obtained from the standard photograph (venules and arterioles combined) and if an image had equal or greater tortuosity measures in at least two quadrants, the image was diagnosed as a case with plus disease. The study did not provide A_z values.

4 Discussion

Our proposed diagnostic-decision-making criteria combine the clinical definition of plus disease with respect to tortuosity⁵ with a practical understanding of the characteristics of a sufficiently tortuous vessel. A direct comparison between the results of the present study and those listed in Table 4 may not be appropriate since the diagnostic and evaluative criteria used in the studies have been different, as described in the previous section. Such limitations indicate the need for consistent protocols and criteria to be set in CAD of plus disease.

The clinical diagnostic definition of plus disease provides a diagnosis for the entire image (or eye of patient) based on at least two quadrants having vessels with abnormal tortuosity and dilation.⁵ Therefore, diagnosis of each single vessel or each quadrant (based on a single maximum or average value) may not qualify as diagnosis of plus disease, as plus-diagnosis is performed on a per-eye (image) basis.

All studies that have performed diagnosis of plus disease via quantification of vessel tortuosity have included only specific desired vessels for further analysis via manual marking of vessel segments, or manual selection and/or correction of parts of automatically detected vessels. Such manual marking of images may not be practical in a clinical or teleophthalmological setting due to lack of resources and time, and may introduce operator-dependent error and bias. The methods proposed in the present work are capable of distinguishing tortuous vessels in a given image without requiring any manual selection and/or correction.

Studies that have performed CAD of plus disease using tortuosity have provided a single measure of tortuosity for each selected vessel segment,^{7,8,20,23} the entire image,^{10,11,21} or for each quadrant of the image.^{22,24–26,54} However, as previously mentioned, because tortuosity is not formally defined, numerical representations of its quantitative measurement (per vessel, quadrant, or image) may not be meaningful to an ophthalmologist. Use of the AVT measure first to identify abnormally tortuous vessel segments and then to obtain the total length of all of such segments is more practical from a CAD point-of-view and more meaningful to an ophthalmologist.

Separate analysis of arterioles and venules may not be necessary for the diagnosis of plus disease. It has been observed that the distinction between arterioles and venules becomes difficult in the presence of zone I disease;⁵⁶ even experts cannot distinguish between arterioles and venules in retinal images of preterm infants about 20% of the time.^{54,57} As shown by various studies,^{10,11,21,22,24–26,54} combining the measures of venular and arteriolar tortuosity provides similar or better diagnostic results as compared to considering arteriolar or venular tortuosity alone, respectively. Furthermore, based on the findings in the present work, it is questionable whether peripheral-venular tortuosity is less important than posterior-arteriolar tortuosity (see Fig. 2 as an example).

As previously mentioned, even though the standard photograph of plus disease has been used for clinical diagnosis, the image is believed to be atypical because it shows more vascular dilation and less tortuosity as compared to most clinical cases with plus disease.³ Second, the image possesses a narrow FOV and does not reveal possible tortuous vessels in the periphery of the retina. All studies mentioned in Sec. 3.3, except the work of Ataer-Cansizoglu et al.¹⁰ and Campbell et al.,¹¹ limited the area of analysis to the posterior of the retina. However, peripheral-vessel tortuosity has been shown to be more correlated to the presence of plus disease as compared to posterior-vessel tortuosity.^{10–12} Indeed, as presented in Table 1, considering the total length of abnormally tortuous vessels over the entire FOV, as compared to considering only the posterior or the periphery of the retina, provides higher diagnostic accuracy. The methods proposed in the present work are capable of detecting tortuous vessels regardless of location.

Retinal images of preterm infants vary substantially in terms of pigmentation, contrast, and quality. All of the other studies mentioned above have selected only images with high vessel-to-background contrast and quality in order to more precisely observe the variations in the vasculature. In real-world applications (clinical or teleophthalmological), it may not be possible to obtain only high-quality and sharp pediatric retinal fundus images due various factors including the infant's movement. In the present work, we have used the largest database of retinal images of preterm infants in the literature (see Table 4), which

is not limited to images with high quality and contrast. Despite some limitations in the thresholding step, our methods are robust to be able to diagnose plus disease with high accuracy regardless of the quality of the image.

Based on the results provided in Table 1, the patient attributes of BW and CA do not show any statistically significant differences in the mean for the cases without plus disease as compared to the plus cases, whereas the difference in the mean GA for the two classes is statistically extremely significant. GA shows a high power of discrimination between cases with and without plus disease with $A_z = 0.82$. Although a statistically extremely significant correlation ($p < 0.001$) was found between GA and the total length of tortuous vessels, the results do not indicate a meaningful level of correlation between the two parameters. These results could indicate that plus disease may not be a developmental process and may be more probable in underdeveloped retinas, i.e., the lower the GA, the higher the probability of occurrence of plus disease.

The results of feature selection and pattern classification indicate that the total tortuous-vessel length over the entire FOV alone can lead to the best CAD performance in the detection of plus disease with the database used in this work (TROPIC). Larger databases with a large range of image characteristics and the need for robust performance may call for the use of multiple diagnostic features.

5 Limitations and Future Work

It has been observed that changes in vascular tortuosity that occur in the presence of plus disease are dynamic²³. Indeed, as mentioned in Sec. 3, using the proposed methods, an average increase of about 11 mm in total length of abnormally tortuous vessels was detected in the eyes of three patients that progressed to plus disease. However, given the limited number of such cases in the present work, further longitudinal analysis of this aspect with more cases would be of interest.

In the present work, the clinical diagnoses using indirect ophthalmoscopy along with consideration of the patients' records have been taken as the reference, not diagnoses based only on the RetCam images. However, since in a tele-ophthalmological setting RetCam images are employed for diagnosis, and given the fact that there exists interexpert variability in the diagnosis of plus disease via RetCam images,⁶ inclusion of more cases with and without plus disease as well as multiple three-class reference diagnoses (normal, pre-plus, and plus) of the same cases provided by several experts is needed to further strengthen the results and the statistical analysis.

Another limitation of the present work is that the proposed methods are not fully automated, with the vessel binarization step, the initial removal of small segments, and marking of the center of the ONH being the manual steps in the process (the same is true for all of the studies listed in Table 4). No single automated thresholding method provides consistent results for binarization of all images due to the variable nature of the retinal images of preterm infants, including varying pigmentation, blurring, and low vessel-to-background contrast. A combination of the results of multiple automated thresholding methods may lead to better binarization results.^{43,58,59} The method of Rangayyan et al.⁶⁰ will be adapted in the future to automatically detect the center of the ONH.

6 Conclusion

In comparison to our pilot studies on vascular tortuosity related to plus disease,^{38,39} the use of the MAD measure to detect and remove linear parts of vessel segments significantly improves the accuracy of the obtained tortuosity measure with an overall diagnostic accuracy of $A_z = 0.98$ using the total length of abnormally tortuous vessels in a given image. The use of a training set of images to obtain thresholds for the AVT measure and to define the minimum-tortuous-vessel-length criterion is advantageous and does not bias the final findings since the training and test sets are mutually independent. Furthermore, the minimum-tortuous-vessel-length threshold obtained from ROC analysis in this work could be used as a threshold for plus diagnosis and applied to other datasets of retinal fundus images of preterm infants.

There are many inconsistencies in the criteria used for diagnosis and evaluation of the results of quantification of vascular tortuosity and CAD of plus disease in the various studies in the literature; there is a need for standardized and consistent protocols to be used in CAD of plus disease based on the findings of the current study as well as those mentioned in Table 4.

Although the clinical diagnostic definition of plus-related tortuosity requires the presence of abnormally tortuous vessels in the posterior of the retina, based on the results presented in this work, it may be argued that considering the total length of all abnormally tortuous vessels in the entire image, as compared to only the posterior or the periphery, may be a more accurate and quantitatively comprehensible indicator of the presence of plus disease that could lead to higher diagnostic accuracy. Further analysis and testing of this observation, using the minimum tortuous-vessel-length threshold obtained in this work with the TROPIC database, with independent databases of retinal fundus images of preterm infants, including multiple expert diagnoses per image, would be of interest.

The methods presented in this work are capable of distinguishing abnormally tortuous vessels and provide high accuracy in the diagnosis of plus disease. The developed programs may be used as a tool for screening ROP in preterm infants.

Disclosures

The authors have no conflicts of interest to report.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada. We thank April Ingram (Ells Retina Center) for help with the TROPIC images and Dr. Eliana Silva de Almeida (Federal University of Alagoas, Brazil) for help with statistical analysis. A preliminary and shorter version of this work was presented at the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, held August 25–29, 2015, in Milan, Italy.³⁹

References

1. Early Treatment for Retinopathy of Prematurity Cooperative Group, "The incidence and course of retinopathy of prematurity: findings from the early treatment for retinopathy of prematurity study," *Pediatrics* **116**(1), 15–23 (2005).
2. D. J. Gunn, D. W. Cartwright, and G. A. Gole, "Incidence of retinopathy of prematurity in extremely premature infants over an 18-year period," *Clin. Exp. Ophthalmol.* **40**, 93–99 (2012).
3. B. V. Davitt and D. K. Wallace, "Plus disease," *Surv. Ophthalmol.* **54**, 663–670 (2009).

4. Early Treatment for Retinopathy of Prematurity Cooperative Group, "Revised indications for the treatment of retinopathy of prematurity: results of the early treatment for retinopathy of prematurity randomized trial," *Arch. Ophthalmol.* **121**, 1684–1696 (2003).
5. International Committee for the Classification of Retinopathy of Prematurity, "The international classification of retinopathy of prematurity revisited," *Arch. Ophthalmol.* **123**, 991–999 (2005).
6. M. F. Chiang et al., "Interexpert agreement of plus disease diagnosis in retinopathy of prematurity," *Arch. Ophthalmol.* **125**(7), 875–880 (2007).
7. R. Gelman et al., "Plus disease in retinopathy of prematurity: pilot study of computer-based and expert diagnosis," *J. Am. Assoc. Pediatr. Ophthalmol. Strabismus* **11**(6), 532–540 (2007).
8. S. Koreen et al., "Evaluation of a computer-based system for plus disease diagnosis in retinopathy of prematurity," *Ophthalmology* **114**(12), e59–e67 (2007).
9. International Committee for the Classification of Retinopathy of Prematurity, "An international classification of retinopathy of prematurity," *Int. Ophthalmol.* **8**, 3–10 (1985).
10. E. Ataer-Cansizoglu et al., "Computer-based image analysis for plus disease diagnosis in retinopathy of prematurity: performance of the 'i-ROP' system and image features associated with expert diagnosis," *Transl. Vision Sci. Technol.* **4**, 5–12 (2015).
11. J. P. Campbell et al., "Expert diagnosis of plus disease in retinopathy of prematurity from computer-based image analysis," *JAMA Ophthalmol.* E1–E7 (2016).
12. K. M. Keck et al., "Plus disease diagnosis in retinopathy of prematurity: vascular tortuosity as a function of distance from optic disk," *Retina* **33**(8), 1700–1707 (2013).
13. J. S. Grewal and A. R. Fielder, "Sequelae of retinopathy of prematurity," *Pediatr. Child Health* **22**(1), 19–24 (2011).
14. F. Oloumi, R. M. Rangayyan, and A. L. Ells, "Quantification of the changes in the openness of the major temporal arcade in retinal fundus images of preterm infants with plus disease," *Invest. Ophthalmol. Visual Sci.* **55**, 6728–6735 (2014).
15. C. M. Wilson et al., "The temporal retinal vessel angle and infants born preterm," *Br. J. Ophthalmol.* **90**, 702–704 (2006).
16. K. Wong et al., "The temporal and nasal retinal arteriolar and venular angles in preterm infants," *Br. J. Ophthalmol.* **95**(12), 1723–1727 (2011).
17. M. M. Fraz et al., "QUARTZ: quantitative analysis of retinal vessel topology and size—an automated system for quantification of retinal vessels morphology," *Expert Syst. Appl.* **42**(20), 7221–7234 (2015).
18. ARIA: Retinal Vessel Analysis with MATLAB, 2012, <http://petebankhead.github.io/ARIA/> (December 2016)
19. A. Perez-Rovira et al., "VAMPIRE: vessel assessment and measurement platform for images of the retina," in *35th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pp. 3391–3394 (2011).
20. R. Gelman et al., "Diagnosis of plus disease in retinopathy of prematurity using retinal image multiscale analysis," *Invest. Ophthalmol. Visual Sci.* **46**(12), 4734–4738 (2005).
21. C. Heneghan et al., "Characterization of changes in blood vessels width and tortuosity in retinopathy of prematurity using image analysis," *Med. Image Anal.* **6**(1), 407–429 (2002).
22. A. E. Kiely et al., "Computer-assisted measurement of retinal vascular width and tortuosity in retinopathy of prematurity," *Arch. Ophthalmol.* **128**, 847–852 (2010).
23. J. P. Thyparampil et al., "Plus disease in retinopathy of prematurity: quantitative analysis of vascular change," *Am. J. Ophthalmol.* **150**(4), 468–475 (2010).
24. D. K. Wallace et al., "Computer-automated quantification of plus disease in retinopathy of prematurity," *J. Am. Assoc. Pediatr. Ophthalmol. Strabismus* **7**, 126–130 (2003).
25. D. K. Wallace, Z. Zhao, and S. F. Freedman, "A pilot study using ROPTool to quantify plus disease in retinopathy of prematurity," *J. Am. Assoc. Pediatr. Ophthalmol. Strabismus* **11**, 381–387 (2007).
26. Z. Zhao et al., "A tool for computer-aided diagnosis of retinopathy of prematurity," *Proc. SPIE* **6915**, 69152W (2008).
27. F. Oloumi et al., "Computer-aided diagnosis of plus disease via measurement of vessel thickness in retinal fundus images of preterm infants," *Comput. Med. Biol.* **66**, 316–329 (2015).
28. M. Abdalla, A. Hunter, and B. Al-Diri, "Quantifying retinal blood vessels' tortuosity—review," in *Science and Information Conf. (SAI)*, pp. 687–693 (2015).
29. W. Lotmar, A. Freiburghaus, and D. Bracher, "Measurement of vessel tortuosity on fundus photographs," *Albrecht von Graefes Archiv für klinische und Experimentelle Ophthalmologie* **211**(1), 49–57 (1979).
30. C. M. Wilson et al., "Computerized analysis of retinal vessel width and tortuosity in premature infants," *Invest. Ophthalmol. Visual Sci.* **49**(1), 3577–3585 (2008).
31. W. E. Hart et al., "Measurement and classification of retinal vascular tortuosity," *Int. J. Med. Inf.* **53**(2–3), 239–252 (1999).
32. G. Dougherty, M. J. Johnson, and M. D. Wiers, "Measurement of retinal vascular tortuosity and its application to retinal pathologies," *Med. Biol. Eng. Comput.* **48**, 87–95 (2010).
33. V. Bolón-Canedo et al., "A GMM-based feature extraction technique for the automated diagnosis of retinopathy of prematurity," in *IEEE 12th Int. Symp. on Biomedical Imaging (ISBI)*, 1498–1501 (2015).
34. E. Grisan, M. Foracchia, and A. Ruggeri, "A novel method for the automatic grading of retinal vessel tortuosity," *IEEE Trans. Med. Imaging* **27**, 310–319 (2008).
35. V. V. Makkapati and V. V. C. Ravi, "Computation of tortuosity of two dimensional vessels," in *Eighth Int. Conf. on Advances in Pattern Recognition (ICAPR)*, pp. 1–4, IEEE, Kolkata, India (2015).
36. A. Lisowska et al., "An experimental assessment of five indices of retinal vessel tortuosity with the RET-TORT public dataset," in *36th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS)*, Chicago, Illinois, pp. 5414–5417 (2014).
37. E. Poletti, E. Grisan, and A. Ruggeri, "Image-level tortuosity estimation in wide-field retinal images from infants with retinopathy of prematurity," in *Engineering in Medicine and Biology Society, 34th Annual Int. Conf. of the IEEE*, San Diego, California, pp. 4958–4961 (2012).
38. F. Oloumi, R. M. Rangayyan, and A. L. Ells, "Assessment of vessel tortuosity in retinal images of preterm infants," in *Engineering in Medicine and Biology Society (EMBS), 36th Annual Int. Conf. of the IEEE*, Chicago, Illinois, pp. 5410–5413 (2014).
39. F. Oloumi, R. M. Rangayyan, and A. L. Ells, "Computer-aided diagnosis of plus disease in retinal fundus images of preterm infants via measurement of vessel tortuosity," in *Engineering in Medicine and Biology Society (EMBS), 37th Annual Int. Conf. of the IEEE*, Milan, Italy (2015).
40. P. L. Hildebrand, A. L. Ells, and A. D. Ingram, "The impact of telemedicine integration on resource use in the evaluation ROP... analysis of the telemedicine for ROP in Calgary (TROPIC) database," *Invest. Ophthalmol. Visual Sci.* **50**, 3151 (2009).
41. D. J. D. Silva et al., "Optic disk size and optic disk-to-fovea distance in preterm and full-term infants," *Invest. Ophthalmol. Visual Sci.* **47**(11), 4683–4686 (2006).
42. R. M. Rangayyan et al., "Detection of blood vessels in the retina with multiscale Gabor filters," *J. Electron. Imaging* **17**, 023018 (2008).
43. F. Oloumi et al., "Detection of blood vessels in retinal fundus images (invited paper)," *Comput. Sci. J. Moldova* **22**(2), 155–185 (2014).
44. F. Oloumi, R. M. Rangayyan, and A. L. Ells, *Digital Image Processing for Ophthalmology: Detection and Modeling of the Retinal Vascular Architecture*, Morgan & Claypool, California, USA (2014).
45. T. Pham-Gia and T. Hung, "The mean and median absolute deviations," *Math. Comput. Modell.* **34**(7–8), 921–936 (2001).
46. H. Freeman, "On the encoding of arbitrary geometric configurations," *IRE Trans. Electron. Comput.* **EC-10**, 260–268 (1961).
47. ROCKIT Metz ROC Software, "Radiology," uchicago.edu/page/metz-roc-software.
48. B. Efron, "Bootstrap methods: another look at the jackknife," *Ann. Stat.* **7**(1), 1–26 (1979).
49. A. M. Zoubir and B. Boashash, "The bootstrap and its application in signal processing," *IEEE Signal Process. Mag.* **15**, 56–76 (1998).
50. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., Wiley, New York (2001).
51. J. Neter et al., *Applied Linear Statistical Models*, 4th ed., Irwin, Chicago, Illinois (1990).
52. R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif. Intell.* **97**(1), 273–324 (1997).

53. T. Colton, "Regression and correlation," in *Statistics in Medicine*, pp. 189–217, Little, Brown and Company, Boston, Massachusetts (1974).
54. S. C. Johnston et al., "Tortuosity of arterioles and venules in quantifying plus disease," *J. Am. Assoc. Pediatr. Ophthalmol. Strabismus* **13**(2), 181–185 (2009).
55. M. E. Martinez-Pérez et al., "Retinal vascular tree morphology: a semi-automatic quantification," *IEEE Trans. Biomed. Eng.* **49**(8), 912–917 (2002).
56. J. T. Flynn and T. Chan-Ling, "Retinopathy of prematurity: two distinct mechanisms that underlie zone 1 and zone 2 disease," *Am. J. Ophthalmol.* **142**(1), 46–59.e2 (2006).
57. C. M. Wilson et al., "Digital image analysis in retinopathy of prematurity: a comparison of vessel selection methods," *J. Am. Assoc. Pediatr. Ophthalmol. Strabismus* **16**(3), 223–228 (2011).
58. I. Cruz-Aceves et al., "Automatic segmentation of coronary arteries using Gabor filters and thresholding based on multiobjective optimization," *Biomed. Signal Process. Control* **25**, 76–85 (2016).
59. R. Medina-Carnicer et al., "Determining hysteresis thresholds for edge detection by combining the advantages and disadvantages of thresholding methods," *IEEE Trans. Image Process.* **19**, 165–173 (2010).
60. R. M. Rangayyan et al., "Detection of the optic nerve head in fundus images of the retina with Gabor filters and phase portrait analysis," *J. Digital Imaging* **23**, 438–453 (2010).

Faraz Oloumi received his BSc, MSc, and PhD degrees in electrical and computer engineering, in 2009, 2011, and 2015, respectively, from the University of Calgary, Calgary, Alberta, Canada. He is currently the founder and director of Aurteen Inc., a research company developing biomedical image analysis software for screening and computer-aided diagnoses of various diseases that manifest themselves in fundus images of the retina. His current interests are biomedical image processing, CAD, artificial intelligence, deep learning, pattern analysis, and graphical user interface design.

Rangaraj M. Rangayyan is a professor Emeritus of electrical and computer engineering at the University of Calgary, Canada. His research areas are biomedical signal and image analysis and CAD. He is a fellow of the IEEE, SPIE, American Institute for Medical and Biological Engineering, Society for Imaging Informatics in Medicine, Canadian Medical and Biological Engineering Society, and Canadian Academy of Engineering, and was recognized with the 2013 Outstanding Engineer Medal by IEEE Canada.

Anna L. Ells is an ophthalmologist, with dual fellowships in pediatric ophthalmology and medical retina. She has a combined academic, hospital-based practice, and private practice. Her research focuses on ROP, global prevention of blindness in children, and teleophthalmological approaches to ROP. She has international expertise and has published extensively in peer-reviewed journals.