

RESEARCH ARTICLE

Open Access



# Retinal vascular tortuosity assessment: inter-intra expert analysis and correlation with computational measurements

Lucía Ramos<sup>1,2\*</sup>, Jorge Novo<sup>1,2</sup>, José Rouco<sup>1,2</sup>, Stephanie Romeo<sup>3</sup>, María D. Álvarez<sup>3</sup> and Marcos Ortega<sup>1,2</sup>

## Abstract

**Background:** The retinal vascular tortuosity can be a potential indicator of relevant vascular and non-vascular diseases. However, the lack of a precise and standard guide for the tortuosity evaluation hinders its use for diagnostic and treatment purposes. This work aims to advance in the standardization of the retinal vascular tortuosity as a clinical biomarker with diagnostic potential, allowing, thereby, the validation of objective computational measurements on the basis of the entire spectrum of the expert knowledge.

**Methods:** This paper describes a multi-expert validation process of the computational vascular tortuosity measurements of reference. A group of five experts, covering the different clinical profiles of an ophthalmological service, and a four-grade scale from non-tortuous to severe tortuosity as well as non-tortuous / tortuous and asymptomatic / symptomatic binary classifications are considered for the analysis of the the multi-expert validation procedure. The specialists rating process comprises two rounds involving all the experts and a joint round to establish consensual rates. The expert agreement is analyzed throughout the rating procedure and, then, the consensual rates are set as the reference to validate the prognostic performance of four computational tortuosity metrics of reference.

**Results:** The Kappa indexes for the intra-rater agreement analysis were obtained between 0.35 and 0.83 whereas for the inter-rater agreement in the asymptomatic / symptomatic classification were between 0.22 and 0.76. The Area Under the Curve (AUC) for each expert against the consensual rates were placed between 0.61 and 0.83 whereas the prognostic performance of the best objective tortuosity metric was 0.80.

**Conclusions:** There is a high inter and intra-rater variability, especially for the case of the four grade scale. The prognostic performance of the tortuosity measurements is close to the experts' performance, especially for Grisan measurement. However, there is a gap between the automatic effectiveness and the expert perception given the lack of clinical criteria in the computational measurements.

**Keywords:** Retinal circulation, Vascular tortuosity, Fundus images, Computer-aided diagnosis, Image analysis

## Background

In recent years, medical imaging has become crucial in the clinical decision-making process, playing an important role to improve the public health due to its ability to extract information for diagnosis and treatment purposes. The use of large databases for medical imaging

also implies the challenge of handling such amount of information in a reliable and useful way for the clinical expert. In addition to this, many medical imaging-based procedures present low repeatability, mainly due to the subjective appreciation of the analyzed data, the variability of the image conditions, or even the expert training for a specific task. Besides the subjectivity, the manual characterization of a large image dataset is a tedious and time-consuming task that inevitably leads to a decreasing performance over time for the same expert. In that sense, the use of computer-based systems that provide the image storage and analysis by a common repeatable procedure

\*Correspondence: [lramos@udc.es](mailto:lramos@udc.es)

<sup>1</sup>University of A Coruña, Department of Computer Science, Campus de Elviña, 15071 A Coruña, Spain

<sup>2</sup>CITIC-Research Center of Information and Communication Technologies, University of A Coruña, A Coruña, Spain

Full list of author information is available at the end of the article



allows ensuring an objective and reliable environment for the specialists, improving, thereby, the productivity and efficiency of the clinical performance.

In ophthalmology, retinal image analysis is an useful tool for the noninvasive diagnosis of many relevant diseases, such as hypertension, diabetes or atherosclerosis. Common symptoms of those pathologies include neo-vascularization, occurrence of pathological structures, or increased tortuosity that can be observed analyzing the vascular tree of the eye fundus. Given the importance of the eye fundus study, Sirius (System for the Integration of Retinal Images Understanding Services) was proposed in [1] as a computer aided diagnosis tool for the analysis of retinal images. It provides a framework for ophthalmologists or other experts to collaboratively work using retinal image-based applications in a distributed, fast and reliable environment. Sirius integrates several image processing algorithms structured as independent modules. One of the modules is in charge of the automatic arterio-venous ratio (AVR) calculation [2], a relevant biomarker to determine the vascular risk that is associated to diseases that affect the circulatory system such as hypertension. Another module localizes microaneurysms [3], which are small red points that appear in early stages of diabetic retinopathy. A third module is focused on measuring the vascular tortuosity of the blood vessels [4, 5], that is, how and how many times a vessel bends, complementary to the AVR parameter. It is a indicator for a number of vascular and nonvascular diseases such as diabetic retinopathy, cerebrovascular disease, stroke, and ischemic heart disease [6–9]. This module integrates four different metrics of tortuosity of reference [10–13].

The validation of Sirius modules against the manual evaluations performed by clinical experts is crucial to ensure a repeatable and reliable analysis of the biomedical parameters that are extracted from the retinal images. The AVR prognostic value, as computed in Sirius, has been clinically validated by Pose et al. [14]. The posterior validation of this module has been carried out in different real environments involving several health care systems [1]. Moreover, additional evaluations of Sirius vessel width measurement have been conducted in DRIVE and REVIEW databases [15, 16]. Regarding the tortuosity module, a preliminary validation over a set of retinal images previously classified as tortuous / non-tortuous has been presented by Sánchez et al. [4, 5].

Although retinal vascular tortuosity is underlying both vascular and systemic diseases, its manual characterization is affected by several limitations that still restrict its use to research purposes. Systematic reviews of retinal vessel tortuosity measures and clinical findings related to them conducted by Kalitzeos et al. [17] and Abdalla et al. [18] compile the main limitations for using the retinal vascular tortuosity as a clinical marker for diagnostic,

treatment and monitoring purposes. One of the main limitations is the lack of a precise and standard guide for the tortuosity assessment regarding the image acquisition, measurement location and consequent calculation. In the clinical practice, the manual characterization of the retinal vascular tortuosity is mostly based on clinical experience by identifying relative characteristics such as the dissimilarity to normal healthy vessels in terms of length, width or number of twists, among others, also evaluating changes in and around each vessel. Therefore, the grading is performed on a subjective scale resulting in a tedious and time-consuming task with a remarkable inter and intra expert variability. Another aspect stated in these reviews is that different diseases produce different tortuosity effects [9, 19, 20], so that the vascular tortuosity should be analyzed from each specific pathological point of view. Despite this, the absence of unified public datasets, the limited size of the existing ones or the differences in the segmentation techniques for extracting blood vessels and the medical state of the patients at the moment of screening, hinder the validation processes of available computational measurements. Additionally, most computational metrics are depending on one or two factors such as the curvature or the number of twists. However, the experts, based on their experience, consider additional parameters such as dilation, elongation, vessel calibers or branching angles [21, 22], among others, that are non incorporated in the current computational metrics of reference. The limitations extracted from these reviews indicate the necessity for standardizing the image acquisition, parameter calculation and analysis of the retinal vascular tortuosity in order to become more useful and reliable to support the clinical decision-making processes.

In the work herein described, a complete and exhaustive multi-expert validation procedure for the Sirius tortuosity module is proposed. This study aims, first, to lay the foundations for advancing the standardization of the retinal vascular tortuosity as a clinical biomarker with diagnostic potential. Once a consistent clinical criteria is established, the validation of the prognostic performance of objective computational measurements of reference is performed.

In order to cover the entire spectrum of the expert knowledge, the validation experiments included a group of five different experts with gradual levels of expertise that usually work in a ophthalmological service of the health care systems, from the head of the service to resident physicians. The manual rating was performed on the basis of a four-grade qualitative scale from non-tortuous to severe tortuosity, being complemented with non-tortuous / tortuous and asymptomatic / symptomatic binary classifications. A rating procedure divided in several rounds was designed in order to set a consensual ground-truth and the extraction of uniform criteria. To this end, first, the five experts rated separately the whole

dataset in a blind process. In order to gain consensus, the discrepancies were analyzed followed by a second rating round that was carried out by each expert. Finally, a joint session involving all the experts was held to set total consensual rates. Therefore, the expert agreement was analyzed throughout the rating procedure and, then, the consensual rates were set as reference to compare the individual manual and automatic measurements. This way, the prognostic performances of the tortuosity metrics presented in [4] were evaluated in relation to the experts performance.

This paper is organized as follows: “**Materials and methods**” section describes the designed dataset, the details of the automatic tortuosity metrics and the procedure for the multi-expert validation. Next, Section **Results** exposes all the conducted experiments and Section **Discussion** discusses the obtained results and the constraints and potential of the tortuosity characterization. Finally, “**Conclusions**” section presents the conclusions and possible future work.

## Materials and methods

### Dataset and rating procedure

Given the association of the retinal vascular tortuosity with diabetes and, specifically, diabetic retinopathy [23–25], fundus images from diabetic patients were found representative for this study. Although vascular tortuosity is underlying more pathologies, different diseases present different tortuosity effects [18]. Given the limited cohort of patients, in order to incorporate homogeneous data, the dataset is limited to diabetic patients ranging from none, mild, moderate, severe or proliferative diabetic retinopathy in a balanced distribution. Therefore, the designed dataset consists of 60 fundus images varying from non visible anomalies in the vessels to severe tortuosity.

In order to analyze the intra and inter-rater variability, a group of five experts belonging to different levels of an ophthalmological service, from the head of the service to resident physicians, manually rated these images. Therefore, the manual characterization of the retinal vascular tortuosity covers the entire spectrum of the expert knowledge. In the clinical practice, ophthalmologists commonly evaluate the degree of a retinal blood vessel tortuosity by considering changes in and around that vessel. These signs are mostly based on clinician experience and knowledge since there is no standard guide for the tortuosity evaluation. Given the lack of standard, the experts, on the basis of their experience, decided to use initially a qualitative four-grade scale which comprises none, mild, moderate, and severe tortuosity degree. Throughout the rating procedure, the experts self-instructed themselves and jointly decided the use of two binary classifications obtained by grouping the grades according to their association with

meaningful clinical conditions. In one hand, a classification to discriminate between no sign of tortuosity and any tortuosity level from mild to severe is considered. It is equivalent to the tortuous / non-tortuous classification used in previous validation experiments [4, 5]. On the other hand, with the knowledge that mild tortuosity is asymptomatic whereas moderate and severe tortuosity can lead to significant risks [22], the classification asymptomatic / symptomatic is defined by grouping none and mild grades in one class and moderate and severe grades in the other class. Before and during the rating procedure, the experts were only instructed with the explicit indication of sticking to the evaluation of vascular tortuosity, abstracting from other clinical findings in the fundus images that could bias the manual rating. For the same reason, the information related to the patient medical state was not available for the experts. The rating procedure was carried out through the progressive following steps:

- A first rating round ( $R_1$ ) using the four-grade scale involving all the experts separately, in a totally blind process.
- A meeting with the experts to discuss the discrepancies and clarify the criteria for the next round. In order to have a control expert that preserves the initial criteria, the expert  $E_5$  did not attend to this meeting.
- A second rating round ( $R_2$ ) for each expert using asymptomatic / symptomatic classification, since after the meeting it is selected as more relevant for the clinical practice.
- A joint round ( $R_c$ ) to set unified consensual rates for the asymptomatic / symptomatic classification.

Table 1 summarizes the manual rates that were provided by the set of experts  $E = \{E_1, E_2, \dots, E_5\}$  for the whole dataset using the four-grade scale in the initial blind rating round  $R_1$ , grouped by grading.

In addition to this grading, these rates were grouped according to the tortuous / non-tortuous and the asymptomatic / symptomatic binary classifications. After the meeting with the experts for analyzing these rates, the binary classification asymptomatic / symptomatic was set as the most relevant for the clinical practice due to the

**Table 1** Retinal vessel tortuosity rated by 5 experts in the first round using a four-grade scale

Grading	$E_{1R_1}$	$E_{2R_1}$	$E_{3R_1}$	$E_{4R_1}$	$E_{5R_1}$
0:none	34	10	21	20	31
1:mild	19	29	23	25	18
2:moderate	7	20	16	12	11
3:severe	0	1	0	3	0

**Table 2** Retinal vessel tortuosity rated by 5 different experts in  $R_1$  and  $R_2$  using a classification asymptomatic / symptomatic

		E1	E2	E3	E4	E5
R1	0:none-mild	53	39	44	45	49
	1:moderate-severe	7	21	16	15	11
R2	0:none-mild	37	39	34	27	50
	1:moderate-severe	23	21	26	33	10

need for medical treatment in symptomatic cases. Table 2 summarizes the rates for the asymptomatic / symptomatic classification obtained by grouping the labels of  $R_1$  and from the manual rates of  $R_2$ .

In order to extract a global expert assessment and discard some isolated dissenting cases, the most voted rates for each image are analyzed in the rounds  $R_1$ ,  $R_2$ , as well as the set formed by the union of  $R_1$  and  $R_2$ . Therefore, for each image,  $V_{R1}$  is set to the most voted rate in the round  $R_1$ ,  $V_{R2}$  is the most voted rate in the round  $R_2$  and  $V_{R1R2}$  is the most voted rate in the set including all the rates in rounds  $R_1$  and  $R_2$ . Finally, in order to clarify the debatable cases and set unified consensual rates, a joint  $R_c$  was carried out involving the five experts. Table 3 summarizes the rates from  $V_{R1}$ ,  $V_{R2}$ ,  $V_{R1R2}$  and  $R_c$ .

**Automatic measurement of the retinal vessel tortuosity**

Retinal blood vessels are normally straight or slightly and gradually curved. However, vascular diseases can cause tortuosity in its structure, defined as a non-smooth appearance of the vessel course. Tortuosity may affect to a small region or involve the entire retinal vascular tree. Figure 1 shows representative examples of retinal images with non-tortuous and tortuous blood vessels.

Based on the computational metrics of reference, all the vessels composing the retinal vascular tree are involved in the tortuosity computation. Therefore, given a color retinal image (see Fig. 2a), the first step to automate the tortuosity measurement consists of the extraction of the arterio-venous tree, and then, its division into the composing vessels. Then, a tortuosity value is computed for each of these vessels, and finally, the tortuosity values corresponding to all the vessels are integrated in a total tortuosity value that is associated with the whole retina.

To this end, the retinal vessels are extracted, first, from an algorithm based on the crease extraction [26, 27]. This algorithm consists in detecting the blood vessels from the ridges or valleys in the retinal image, this is, regions that

**Table 3** Retinal vessel tortuosity rates from  $V_{R1}$ ,  $V_{R2}$ ,  $V_{R1R2}$  and  $R_c$  using the asymptomatic / symptomatic classification

	$V_{R1}$	$V_{R2}$	$V_{R1R2}$	$R_c$
0:none-mild	52	37	45	44
1:moderate-severe	8	23	15	16

form an extreme and tubular level on the neighborhood. Therefore, the Multi Local Set of Extrinsic Curvature enhanced by the Structure Tensor (MLSEC-ST) operator is applied to detect the vessels from the ridge lines. Then, a thinning process is performed to extract the centerline of a maximum of  $1px$  width for each vessel [28]. After this, an edge tracking algorithm is applied to decompose the vessel tree into its constituent vessels. Finally, the vessel point coordinates are locally smoothed in order to minimize the discrete effect of the pixel representation. The resulting vessel segments are used for the tortuosity computation (see Fig. 2b).

Using the vessel segments as reference, four different metrics for measuring the tortuosity of retinal vessels were considered:

**First approach - Hart et al. [10]**

The first approach, proposed by Hart et al. [10], is the simplest and most widely used measure. It computes the tortuosity of a vessel by examining how long the curve is relative to its chord length, as follows:

$$\tau_{Hart} = \frac{L_c}{L_x} - 1 \tag{1}$$

where  $L_c$  is the arc length or curve length obtained by counting all the points from the start of the vessel to its end, and  $L_x$  is the length of the underlying chord, that is, the euclidean distance between the two end points of the vessel.

**Second approach - Grisan et al. [11]**

Another tortuosity metric has been proposed by Grisan et al. [11]. This approach subdivides each vessel in  $n$  segments of constant-sign curvature and then combines the evaluation of such segments and their number as follows:

$$\tau_{Grisan} = \frac{n - 1}{L_c} \sum_{i=1}^n \left[ \frac{L_{csi}}{L_{xsi}} - 1 \right] \tag{2}$$

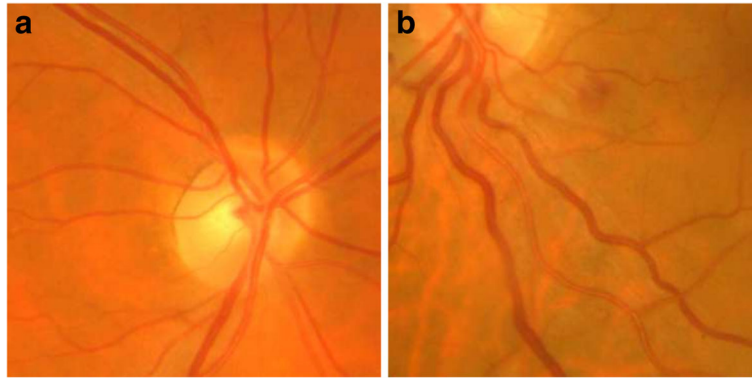
where  $L_c$  corresponds to the arc length of the vessel whereas  $L_{csi}$  and  $L_{xsi}$  represent the arc length and the chord length of each subsegment. This metric integrates the information about how many times a vessel changes convexity so that a higher number of subsegments implies higher tortuosity.

**Third approach - Trucco et al. [12]**

The third approach implements the proposal of Trucco et al. [12], that consists of a measure only depending on the vessel skeleton curvature. This metric is a generalized version of the curvature based metrics presented in Hart et al. [10]. It is defined by:

$$\tau_{Trucco} = \left( \sum_j |k_s(j)|^p \right)^{\frac{1}{p}} \tag{3}$$





**Fig. 1** Retinal images **a** Non-tortuous and **b** tortuous blood vessels

where  $p$  is a strictly positive integer and  $k_s(j)$  is the curvature at the  $j$ th point of the vessel  $s$ , defined as follows:

$$k_s(j) = \frac{x'(j)y''(j) - x''(j)y'(j)}{[y'(j)^2 + x'(j)^2]^{3/2}}, \tag{4}$$

The curvature measures the variation of the slope of the line tangent to the curve at each point along the segment. A significant difference in slope between the point and its surrounding neighbors implies a high curvature.

**Fourth approach - Onkaew et al. [13]**

Finally, the system implements the metric proposed by Onkaew et al. [13], that uses the number of points where the curvature changes its sign, but this curvature is calculated from a improved chain-code algorithm. This method labels each point using its relative position in relation to the previous point, traveling all the intermediate points along the vessel. The tortuosity metric is defined as follows:

$$\tau_{Onkaew} = \frac{n-1}{n} \frac{1}{L_c} \sum_{i=1}^n K(p_i, k) \tag{5}$$

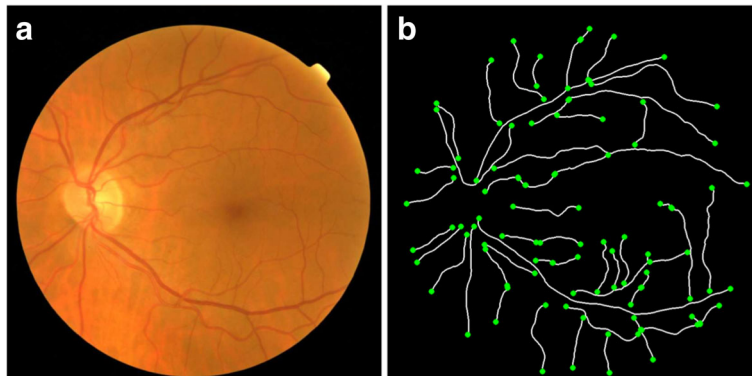
where  $L_c$  corresponds to the arc length of the vessel,  $n$  is the number of subsegments composing the vessel and  $K(p_i, k)$  is the curvature at each point computed by using the chain-code algorithm.

**Integration of the vessel tortuosity values**

The proposed metrics allow the tortuosity calculation from each particular vessel. Once the tortuosity values are computed for each vessel composing the vascular tree, these values are integrated in a total tortuosity measurement using their weighted average. This way, the computed total tortuosity is associated as a single score to the whole retina. Therefore, using the compositionality property of the proposed measures in (1), (2), (3) and (5), each vessel contributes inversely proportional to its arc length [10]. This weighted additivity is defined as:

$$\tau(c1, c2) = \frac{[L_{c1}\tau_{c1} + L_{c2}\tau_{c2}]}{L_{c1} + L_{c2}} \tag{6}$$

where  $L_{ci}$  is the arc length of the vessel  $ci$  and  $\tau_{ci}$ , the tortuosity value for that vessel. Therefore, the resulting tortuosity is within the range of the constituent vessels [11]. These metrics have a dimension of  $1/length$  and thus



**Fig. 2** Arterio-venous tree segmentation. **a** Original retinal image. **b** Ends of each vessel of the segmented vascular tree

may be interpreted as a tortuosity density, allowing the comparison between retinal images at different scales.

**Expert agreement analysis**

An overall comparison including all the manual rates for the whole dataset was performed in order to evaluate the expert agreement and set a reference for validating the performance of the computational approaches. To this end, the percentages of retinal images with full consensus and with four or at least three expert coincidences were extracted for the four-grade scale and for the tortuous / non-tortuous and asymptomatic / symptomatic binary classifications in rounds  $R_1$  and  $R_2$ .

Then, an extended analysis was performed for the asymptomatic / symptomatic classification since the discussion between experts throughout the rating process concluded that this classification is more relevant for the clinical practice. For this purpose, Cohen-Kappa indexes [29] were computed between each pair of experts in rounds  $R_1$  and  $R_2$  in order to evaluate the intra and inter-rater agreement for this classification. Moreover, the agreement in relation to  $V_{R1}$ ,  $V_{R2}$ ,  $V_{R1R2}$ , and the consensus session  $R_c$  was also analyzed.

**Multi-expert validation analysis**

In order to evaluate the prognostic performance of a tortuosity metric, a ROC (Receiver Operating Characteristic) analysis was carried out by using the asymptomatic / symptomatic classification as target prediction. To this end, ROC curves were built from the reciprocal relation between sensitivity and specificity calculated for all the

possible threshold values in the automatic metrics [30]. Thus, tortuosity metrics can be evaluated against each of the expert predictions. This same ROC analysis can be used to evaluate the performance of each expert against the others, but in this case, only one point of the ROC space is obtained.

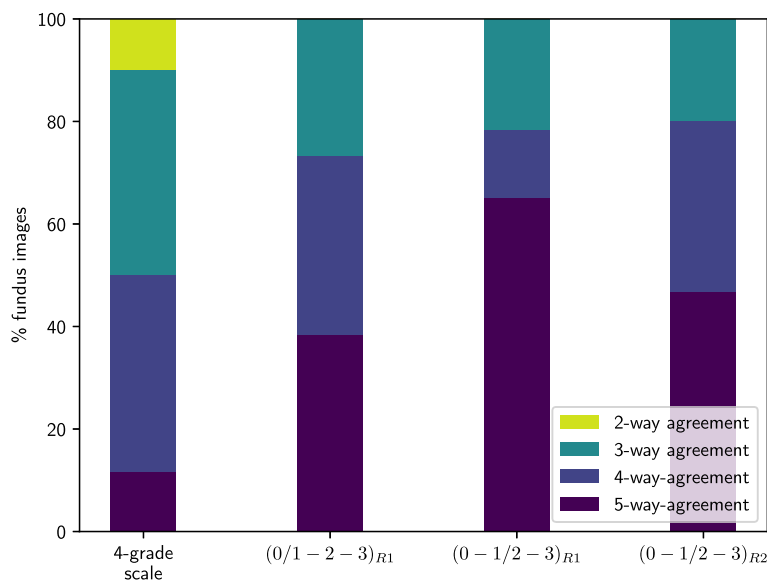
With the aim of setting a consistent ground-truth that includes all the unified criteria that were extracted throughout the rating process, the consensual rates in  $R_c$  are set as the reference to evaluate the performance of the different tortuosity metrics. Therefore,  $\overline{ROC}(Ei_{Rj}, R_c)$  represent the point in the space ROC for expert  $i$  in round  $j$  respect to  $R_c$ . Similarly,  $\overline{ROC}(\tau_m, R_c)$  corresponds to the ROC curve for tortuosity metric  $\tau_m$  in relation to  $R_c$ . Comparing the AUC values of the equivalent curves allow to evaluate if the prognostic performance is similar to the experts performance.

**Results**

**Expert agreement results**

An overall comparison was performed among all the manual rates in the whole dataset for the four-grade scale and also for the non-tortuous / tortuous and asymptomatic / symptomatic classifications in rounds  $R_1$  and  $R_2$ . Figure 3 shows the percentages of retinal images with full consensus among all the experts, the percentages where there are four coincidences, and the percentages where, at least, three experts agree in their labels.

Complementarily, Table 4 shows the Cohen-Kappa indexes between each pair of experts in rounds  $R_1$  and  $R_2$  and between each expert and  $V_{R1}$ ,  $V_{R2}$ ,  $V_{R1R2}$ , and



**Fig. 3** Overall comparison among all the manual rates in the whole dataset. Expert agreement for four-grade, (0 / 1-2-3) and (0-1 / 2-3) scales in rating rounds  $R_1$  and  $R_2$

**Table 4** Inter-intra expert agreement analysis

Cohen-Kappa	$E_{2R_1}$	$E_{3R_1}$	$E_{4R_1}$	$E_{5R_1}$	$E_{1R_2}$	$E_{2R_2}$	$E_{3R_2}$	$E_{4R_2}$	$E_{5R_2}$	$V_{R_1}$	$V_{R_2}$	$V_{R_1R_2}$	$R_c$
<b>E1<sub>R1</sub></b>	0.39	0.53	0.57	0.61	<b>0.35</b>	0.39	0.22	0.20	0.66	0.92	0.35	0.57	0.53
<b>E2<sub>R1</sub></b>		0.50	0.53	0.42	0.43	<b>0.56</b>	0.34	0.35	0.38	0.44	0.43	0.53	0.42
<b>E3<sub>R1</sub></b>			0.43	0.48	0.44	0.50	<b>0.50</b>	0.33	0.52	0.59	0.51	0.70	0.74
<b>E4<sub>R1</sub></b>				0.51	0.55	0.61	0.39	<b>0.43</b>	0.45	0.63	0.55	0.64	0.70
<b>E5<sub>R1</sub></b>					0.37	0.51	0.31	0.25	<b>0.83</b>	0.69	0.45	0.71	0.57
<b>E1<sub>R2</sub></b>						0.50	0.76	0.48	0.23	0.40	0.86	0.62	0.66
<b>E2<sub>R2</sub></b>							0.48	0.48	0.46	0.44	0.64	0.69	0.73
<b>E3<sub>R2</sub></b>								0.57	0.34	0.26	0.83	0.54	0.57
<b>E4<sub>R2</sub></b>									0.28	0.22	0.61	0.43	0.46
<b>E5<sub>R2</sub></b>										0.74	0.49	0.65	0.52
<b>V<sub>R1</sub></b>											0.40	0.63	0.59
<b>V<sub>R2</sub></b>												0.70	0.74
<b>V<sub>R1R2</sub></b>													0.87

Cohen-Kappa indexes for intra and inter-rater agreement as well as between each expert and  $V_{R_1}$ ,  $V_{R_2}$ ,  $V_{R_1R_2}$  and  $R_c$  in rounds  $R_1$  and  $R_2$  for asymptomatic / symptomatic classification. The indexes that are highlighted in bold correspond to the intra-rater agreement between rounds  $R_1$  and  $R_2$

the consensus session  $R_c$  for asymptomatic / symptomatic classification. A standard guideline for interpreting these indexes [31] assumes slight agreement from 0.0 to 0.20, fair agreement from 0.21 to 0.40, moderate agreement from 0.41 to 0.60, substantial agreement from 0.61 to 0.80 and almost perfect or perfect agreement for values greater than 0.81. The Cohen-Kappa indexes that are highlighted in bold correspond to the intra-rater agreement between rounds  $R_1$  and  $R_2$ . The rates in  $R_c$  are selected as the reference to validate the prognostic performance of the tortuosity automatic measurements since they represent the consensual criteria concluded throughout the rating process.

**Multi-expert validation results**

The consensual rates from  $R_c$  are used for validating the prognostic performance of tortuosity measurements. To this end, the points in the ROC space  $ROC(Ei_{Rj}, R_c)$  for  $i = 1..5, j = 1, 2$  are computed as shown in Fig. 4 represented by triangular and square marks. In the same way, the ROC curves  $ROC(\tau_m, R_c)$  for each tortuosity metric in relation to  $R_c$  are built. Moreover, the intra-expert reliability, represented as circular marks, is also considered.

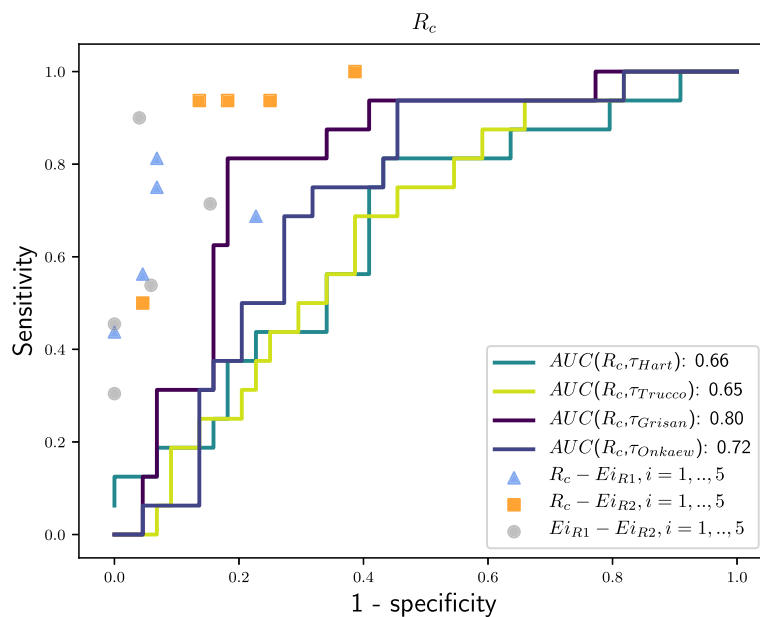
**Discussion**

The results extracted from the overall comparison among the manual rates show that there is a high inter-rater variability, especially for the four-grade scale. Regarding the binary classifications, the experts agree with higher rates in the discrimination between asymptomatic / symptomatic than between non-tortuous / tortuous retinal images. In the rating round  $R_2$ , the percentage of images with full consensus decreased mostly due to the rates of

$E_5$ , the control expert that did not attend to the meeting to discuss the discrepancies and keep its initial criteria, indicating, thereby, the utility and suitability of the meeting. Hence, there is a slight increment in the percentage of images where at least four expert converge since the discussion allowed to unify criteria and gain consensus.

Regarding the Cohen-Kappa indexes, they show low or fair agreement for the four experts who attended the session to clarify the discrepancies found in  $R_1$ , since after the meeting, they change their criteria for the second round  $R_2$ . However,  $E_5$ , the control expert who was not involved in that session, made a similar rating in both rounds, given its criteria was not influenced and modified, presenting, thereby, a high intra-rater agreement. According to the data showed in Table 2, in the round  $R_1$ , the experts were more conservative for asymptomatic cases whereas in the round  $R_2$  the sensitivity for symptomatic cases increased. The change in the criteria is mainly due to the fact that initial rates corresponds to a global assessment of the whole retina, mostly focused on the main vessels, nevertheless, the expert meeting for analyzing the discrepancies led to a more local analysis taking into account each specific vessel during the round  $R_2$ . The criteria refinement is also reflected in the low index between  $V_{R_1}$  and  $V_{R_2}$ . However, the rates obtained by combining  $R_1$  and  $R_2$  are quite close to the consensual rates in  $R_c$  since  $V_{R_1R_2}$  represents the majority inclination comprising the conservative criteria based in the global perception followed in  $R_1$  as well as the analysis of specific vessels considered in  $R_2$ .

With respect to the objective tortuosity measurements, Fig. 4 shows that the prognostic performance is below, at different distances, of the experts performance. As detailed before, the analyzed computational metrics



**Fig. 4** Prognostic performance of retinal vascular tortuosity measurements. ROC curves to evaluate prognostic performance of automatic tortuosity measurements in relation to  $R_c$

incorporate parameters as amplitude, number of twists or curvature of the retinal blood vessels, depending on each case. The results show that the best performance is provided by the metrics which integrate the information about how many times a vessel changes its convexity. In particular, the metric that reached the best score was the Grisan proposal, followed by the Onkaew proposal, given they combine the number of segments with constant convexity within a vessel with the evaluation of such segments. However, Hart and Trucco proposals analyze each vessel globally, regardless of whether it has a constant sign or presents twists.

#### Tortuosity characterization. Constraints and potential

The assessment of the retinal vascular tortuosity is affected by several factors that prevent its use for diagnostic and treatment purposes. Thus, the lack of precise and standard guides for tortuosity characterization leads to a remarkable disagreement among the experts. In this sense, the multi-expert validation process throughout a rating procedure in several stages is raised in order to lay the foundations for advancing the standardization of the retinal vascular tortuosity as a potential indicator for diagnostic purposes.

Besides the subjective appreciation of tortuosity signs, the manual characterization is also depending on the experience of the rater. In order to cover the entire spectrum of the expert knowledge, a group of five clinicians belonging to different levels of an ophthalmological service was considered for the rating procedure. In particular,

they cover the head of the service, experienced clinicians with different levels of expertise and also the participation of resident physicians. This way, the manual characterization of the retinal vascular tortuosity incorporates assessments at different levels of expertise and medical profiles. In order to avoid biased rates, the information related to the patient medical state was not known by the experts at the time of the manual rating. The rating procedure was performed individually in a totally blind process in which the experts were only instructed with the explicit indication of sticking to the evaluation of the vascular tortuosity, abstracting from other clinical findings that could bias the rating.

Despite this, there are also limitations related to the availability of normative data due to the absence of unified public datasets. Moreover, even the available datasets, public or private, present limitations in terms of type and size. This, along with the lack of a standard regarding the computational algorithms used for extracting the blood vessels or the location of the tortuosity measurements, hinder the validation processes of the available computational methods. Furthermore, different diseases produce different tortuosity effects, so that the vascular tortuosity should be analyzed from each specific pathological point of view. In this work, given the association of the retinal vascular tortuosity with diabetes and, more specifically, diabetic retinopathy, diabetic patients were found representative for this study. Although vascular tortuosity is underlying more pathologies, the dataset was limited to diabetic patients in order to analyze a representative



cohort of homogeneous data. With respect to the type and size of the retinal images, the implemented methods allow a high degree of normalization in the computed tortuosity values, independently of the acquisition procedure. Regarding the location or zone of the vessels involved in the tortuosity computation, this analysis is based on the metrics of reference in the literature [10–13], included in the Sirius framework [1]. According to these metrics, the vascular tree is extracted by means of a consolidated computational methodology [26] for that purpose, being all the vessels composing the vascular tree involved in the global tortuosity computation.

Regarding the prognostic performance of the computational metrics, despite the acceptable results of some of the metrics, all of them remain at a distance of the experts performance. The metrics of reference generally use mathematical properties depending on one or two factors such as curvature or number of twists. However, the experts, based on their experience, analyze a larger set of properties being, therefore, differentiated of the computational metrics.

## Conclusions

The retinal vascular tortuosity constitutes a potential indicator of relevant vascular and non-vascular diseases, so that a reliable quantitative measurement would be a potential biomarker for early detection and disease prevention. However, there is no a precise and standard definition of the vascular tortuosity, and consequently, its manual characterization is a subjective task with a high variability. This work is raised with the aim of establishing the basis for advancing in the standardization of the retinal vascular tortuosity as a clinical marker with diagnostic potential allowing, thereby, the robust validation of computational measurements to ensure an objective and reliable environment for the retinal experts. For this purpose, a multi-expert validation procedure is presented in order to assess the prognostic performance of the computational calculation of the vascular tortuosity following the main referenced strategies, included in the Sirius framework. The presented validation included the participation of a group of five different experts and considered a four-grade scale from non-tortuous to severe tortuosity as well as non-tortuous / tortuous and asymptomatic / symptomatic binary classifications. The rating procedure comprised 2 rating rounds in which each expert manually rated the whole dataset and a posterior final joint consensus session where the debatable cases were discussed to reach a global agreement.

For the multi-expert validation procedure, firstly, the expert agreement was analyzed along the different rating rounds. The intra and inter-rater reliability were computed and the discrepancies were discussed involving

the experts in order to clarify the criteria and extract additional information from their clinical perception. This rating process allowed to gain consensus among the experts and get consensual rates comprising all the unified criteria extracted throughout the rating process. Therefore, the consensual rates were set as a reference for validating the computational tortuosity metrics that were included in this analysis. Once a consolidated clinical criteria was established, the prognostic performance of the computational measurements was compared to the experts performance, allowing a robust validation of the strengths and limitations of the different tortuosity metrics of reference.

The multi-expert validation provided acceptable results, especially regarding the Grisan proposal. However, all of the considered computational measurements remain at a distance of the experts performance. The analyzed metrics use mathematical properties to define the degree of tortuosity according to one or two factors such as the amplitude, the curvature or the number of twists, depending on each case. Despite that, the experts, based on their experience, analyze additional parameters such as the neovascularization, the vessel caliber or the distinction between arteries and veins that are not incorporated, at the moment, in the existing computational metrics, thereby causing differences between the automated effectiveness and the expert perception. The results extracted from this work demonstrate that the metrics of reference do not provide a full representation of the expert perception so that additional parameters should be incorporated in the computational metrics in order to have a more accurate and reliable tortuosity assessment. Thus, future work in this research line includes the integration of additional properties in new computational proposals that could approach the performance of the computational metrics to the knowledge of the expert clinicians.

## Abbreviations

AUC: Area under curve; AVR: Arterio-venous ratio; ROC: Receiver operating characteristic; Sirius: System for the integration of retinal images understanding services

## Acknowledgements

Not applicable.

## Funding

This research has been economically supported by the Instituto de Salud Carlos III, Government of Spain and FEDER funds of the European Union through the PI14/02161 and the DTS15/00153 research projects and by the Ministerio de Economía y Competitividad, Government of Spain through the DPI2015-69948-R research project. Also, this work has received financial support from the European Union (European Regional Development Fund - ERDF) and the Xunta de Galicia, Centro singular de investigación de Galicia accreditation 2016-2019, Ref. ED431G/01; and Grupos de Referencia Competitiva, Ref. ED431C 2016-047.

## Availability of data and materials

The datasets used and/or analysed during the current study are not publicly available given they are part of private data from a clinical service but are available from the corresponding author on reasonable request.

**Authors' contributions**

LR, JN, JR and MO made substantial contributions to conception and design of the retinal vascular tortuosity assessment, analysis of data and interpretation of the results. SR and MA were involved in the acquisition of the dataset, the rating procedure and the process of establishing a consensual clinical criteria to evaluate the retinal vascular tortuosity. LR performed the experiments, drafted the manuscript and integrated critical feedback from the other authors. JN has been involved in drafting the manuscript and revising it critically for relevant content. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

The "Comité de Ética da Investigación de A Coruña-Ferrol" committee belonging to the "Rede Galega de Comités de Ética da Investigación" attached to the regional government "Secretaría Xeral Técnica da Consellería de Sanidade da Xunta de Galicia" approved this study (Ref. 2015/501), which was conducted in accordance with the tenets of the Helsinki Declaration. This study was carried out retrospectively on existing data that have previously been anonymized.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>University of A Coruña, Department of Computer Science, Campus de Elviña, 15071 A Coruña, Spain. <sup>2</sup>CITIC-Research Center of Information and Communication Technologies, University of A Coruña, A Coruña, Spain. <sup>3</sup>Servizo de Oftalmoloxía, Complexo Hospitalario Universitario de Ferrol, Ferrol, A coruña, Spain.

Received: 25 May 2018 Accepted: 30 October 2018

Published online: 20 November 2018

**References**

- Ortega M, Barreira N, Novo J, Penedo MG, Pose-Reino A, Gómez-Ulla F. Sirius: A web-based system for retinal image analysis. *IJ Med Inform.* 2010;79(10):722–32. <https://doi.org/10.1016/j.ijmedinf.2010.07.005>.
- Vázquez SG, Cancela B, Barreira N, Penedo MG, Rodríguez-Blanco M, Pena Seijo M, de Tuero GC, Barceló MA, Saez M. Improving retinal artery and vein classification by means of a minimal path approach. *Mach Vis Appl.* 2013;24(5):919–30. <https://doi.org/10.1007/s00138-012-0442-4>.
- Barreira N, Penedo MG, González S, Ramos L, Cancela B, González A. Automatic analysis of the microaneurysm turnover in a web-based framework for retinal analysis. *Image Anal Model Ophthalmol.* 2014;111–128.
- Sánchez L, Novo J, Fernandez A, Barja JM. Unified methodology for evaluating vessel tree tortuosity metrics in eye fundus images. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, Porto, Portugal, June 20–22, 2013. Porto: IEE; 2013. p. 528–9. <https://doi.org/10.1109/CBMS.2013.6627862>.
- Sánchez L, Barreira N, Penedo MG, de Tuero GC. Computer aided diagnosis system for retinal analysis: Automatic assessment of the vascular tortuosity. In: Innovation in Medicine and Healthcare 2014, Proceedings of the Second KES International Conference on Innovation in Medicine and Healthcare, InMed 2014. San Sebastian: IOS Press; 2014. p. 55–64. <https://doi.org/10.3233/978-1-61499-474-9-55>.
- Witt N, Wong TY, Hughes AD, Chaturvedi N, Klein BE, Evans R, McNamara M, Thom SAM, Klein R. Abnormalities of retinal microvascular structure and risk of mortality from ischemic heart disease and stroke. *Hypertension.* 2006;47(5):975–81. <https://doi.org/10.1161/01.HYP.0000216717.72048.6c>. <http://hyper.ahajournals.org/content/47/5/975.full.pdf>.
- Dougherty G, Johnson MJ, Wiers MD. Measurement of retinal vascular tortuosity and its application to retinal pathologies. *Medical & Biological Engineering & Computing.* 2009;48(1):87. <https://doi.org/10.1007/s11517-009-0559-4>.
- Cheung CY-I, Zheng Y, Hsu W, Lee ML, Lau QP, Mitchell P, Wang JJ, Klein R, Wong TY. Retinal vascular tortuosity, blood pressure, and cardiovascular risk factors. *Ophthalmology.* 2011;118(5):812–18. <https://doi.org/10.1016/j.ophtha.2010.08.045>.
- Sasongko MB, Wong TY, Donaghue KC, Cheung N, Jenkins AJ, Benitez-Aguirre P, Wang JJ. Retinal arteriolar tortuosity is associated with retinopathy and early kidney dysfunction in type 1 diabetes. *Am J Ophthalmol.* 2012;153(1):176–183. <https://doi.org/10.1016/j.ajo.2011.06.005>.
- Hart WE, Goldbaum MH, Côté B, Kube P, Nelson MR. Measurement and classification of retinal vascular tortuosity. *IJ Med Inform.* 1999;53(2–3):239–52. [https://doi.org/10.1016/S1386-5056\(98\)00163-4](https://doi.org/10.1016/S1386-5056(98)00163-4).
- Grisan E, Foracchia M, Ruggeri A. A novel method for the automatic grading of retinal vessel tortuosity. *IEEE Trans Med Imaging.* 2008;27(3):310–9. <https://doi.org/10.1109/TMI.2007.904657>.
- Trucco E, Azegrouz H, Dhillon B. Modeling the tortuosity of retinal vessels: Does caliber play a role?. *IEEE Trans Biomed Eng.* 2010;57(9):2239–47. <https://doi.org/10.1109/TBME.2010.2050771>.
- Onkaew D, Turior R, Uyyanonvara B, Akinori N, Sinthanayothin C. Automatic retinal vessel tortuosity measurement using curvature of improved chain code. In: International Conference on Electrical, Control and Computer Engineering 2011 (InECCE). 2011. p. 183–6. <https://doi.org/10.1109/INECCCE.2011.5953872>.
- Pose-Reino A, Rodríguez-Fernández M, Hayik B, Gómez-Ulla F, Carrera-Nouche MJ, Gude-Sampedro F, Estévez-Nuñez J, Méndez-Naya I. Regression of alterations in retinal microcirculation following treatment for arterial hypertension. *J Clin Hypertens.* 2006;8:590–5. <https://doi.org/10.1111/j.1524-6175.2006.05476.x>.
- Barreira N, Ortega M, Rouco J, Penedo MG, Pose-Reino A, Mariño C. Semi-automatic procedure for the computation of the arteriovenous ratio in retinal images. *J Comput Vision Biomech.* 2010;3(2):135–47.
- Vázquez SG, Barreira N, Penedo MG, Pena-Seijo M, Gómez-Ulla F. Evaluation of sirius retinal vessel width measurement in review dataset. In: Proc. of the 26th IEEE International Symposium on Computer-Based Medical Systems. 2013. p. 71–6. <https://doi.org/10.1109/CBMS.2013.6627767>.
- Kalitzos AA, Lip GYH, Heitmar R. Retinal vessel tortuosity measures and their applications. *Exp Eye Res.* 2013;106:40–6. <https://doi.org/10.1016/j.exer.2012.10.015>.
- Abdalla M, Hunter A, Al-Diri B. Quantifying retinal blood vessels' tortuosity — review. In: 2015 Science and Information Conference (SAI). London: IEEE; 2015. p. 687–93. <https://doi.org/10.1109/SAI.2015.7237216>.
- Mohsenin A, Adelman RA, Mohsenin V. Retinal vascular tortuosity in obstructive sleep apnea. *Clin Ophthalmol.* 2013;7:787–92. <https://doi.org/10.2147/OPHTH.S41795>.
- Weiler DL, Engelke CB, Moore ALO, Harrison WW. Arteriole tortuosity associated with diabetic retinopathy and cholesterol. *Optom Vis Sci.* 2014;92(3):384–91. <https://doi.org/10.1097/OPX.0000000000000484>.
- Ikram MK, Cheung CY, Lorenzi M, Klein R, Jones TLZ, Wong TY. Retinal vascular caliber as a biomarker for diabetes microvascular complications. *Diabetes Care.* 2013;36(3):750–9. <https://doi.org/10.2337/dc12-1554>.
- Han H. Twisted blood vessels: Symptoms, etiology and biomechanical mechanisms. *J Vasc Res.* 2012;49:185–97. <https://doi.org/10.1159/000335123>.
- Sasongko MB, Wong TY, Nguyen TT, Cheung CY, Shaw JE, Wang JJ. Retinal vascular tortuosity in persons with diabetes and diabetic retinopathy. *Diabetologia.* 2011;54(9):2409–16. <https://doi.org/10.1007/s00125-011-2200-y>.
- Sasongko MB, Wong TY, Nguyen TT, Cheung CY, Shaw JE, Kawasaki R, Lamoureux EL, Wang JJ. Retinal vessel tortuosity and its relation to traditional and novel vascular risk markers in persons with diabetes. *Curr Eye Res.* 2016;41(4):551–7. <https://doi.org/10.3109/02713683.2015.1034371>.
- Lim LS, Chee ML, Cheung CY, Wong TY. Retinal vessel geometry and the incidence and progression of diabetic retinopathy. *Investig Ophthalmol Vis Sci.* 2017;58(6):200–5. <https://doi.org/10.1167/iov.17-21699>.
- López AM, Lloret D, Serrat J, Villanueva JJ. Multilocal creaseness based on the level-set extrinsic curvature. *Comp Vision Image Underst.* 2000;77(2):111–44. <https://doi.org/10.1006/cviu.1999.0812>.

27. Ortega M, Penedo MG, Rouco J, Barreira N, Carreira MJ. Retinal verification using a feature points-based biometric pattern. *EURASIP J Adv Sig Proc.* 2009;2009. <https://doi.org/10.1155/2009/235746>.
28. Zhang TY, Suen CY. A fast parallel algorithm for thinning digital patterns. *Commun ACM.* 1984;27(3):236–9. <https://doi.org/10.1145/357994.358023>.
29. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas.* 1960;20(1):37. <https://doi.org/10.1177/001316446002000104>.
30. Fawcett T. An introduction to roc analysis. *Pattern Recogn Lett.* 2006;27(8):861–74. <https://doi.org/10.1016/j.patrec.2005.10.010>.
31. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–74. <https://doi.org/10.2307/2529310>.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

