Proliferative Diabetic Retinopathy Characterization based on Fractal Features: Evaluation on a Publicly Available Data Set

José Ignacio Orlando,[1, 2, a)] Karel van Keer,[3] João Barbosa Breda,[3] Hugo Luis Manterola,[1, 2]
Matthew B. Blaschko,[4] and Alejandro Clausse[1, 2, 5]
[1)] *Pladema Institute, UNCPBA, Gral. Pinto 399, Tandil, Argentina*
[2)] *Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Argentina*
[3)] *Department of Ophthalmology, UZ Leuven, Leuven, Belgium*
[4)] *ESAT-PSI, KU Leuven, Kasteelpark Arenberg 10, B-3001, Leuven, Belgium*
[5)] *Comisión Nacional de Energía Atómica, CNEA, Argentina*

(Dated: 5 October 2017)

Purpose: Diabetic retinopathy (DR) is one of the most widespread causes of preventable blindness in the world. The most dangerous stage of this condition is proliferative DR (PDR), in which the risk of vision loss is high and treatments are less effective. Fractal features of the retinal vasculature have been previously explored as potential biomarkers of DR, yet the current literature is inconclusive with respect to their correlation with PDR. In this study we experimentally assess their discrimination ability to recognize PDR cases.
Methods: A statistical analysis of the viability of using three reference fractal characterization schemes–namely box, information and correlation dimensions–to identify patients with PDR is presented. These descriptors are also evaluated as input features for training ℓ_1 and ℓ_2 regularized logistic regression classifiers, to estimate their performance.
Results: Our results on MESSIDOR, a public data set of 1200 fundus photographs, indicate that patients with PDR are more likely to exhibit a higher fractal dimension than healthy subjects or patients with mild levels of DR ($p \leq 1.3 \times 10^{-2}$). Moreover, a supervised classifier trained with both fractal measurements and red lesion based features reports an area under the ROC curve of 0.93 for PDR screening and 0.96 for detecting patients with optic disc neovascularizations.
Conclusions: The fractal dimension of the vasculature increases with the level of DR. Furthermore, PDR screening using multiscale fractal measurements is more feasible than using their derived fractal dimensions. Code and further resources are provided at `https://github.com/ignaciorlando/fundus-fractal-analysis`.

Keywords: Proliferative diabetic retinopathy, Fundus imaging, Fractal dimension, Machine learning

I. INTRODUCTION

Diabetes mellitus (DM) is a highly prevalent metabolic disorder characterized by insulin deficiency and/or an inappropriate response to insulin.[1] The resulting hyperglycemia, through protein glycation, oxidative stress and other mechanisms, can lead to serious (micro) vascular and neurological complications. The number of people suffering from DM is expected to increase from 366 million in 2011 to 552 million by 2030,[2] with two thirds of the cases occurring in low to middle-income countries, where access to an ophthalmological consultation is scarcer.[3]

One of the major microvascular consequences of DM is Diabetic Retinopathy (DR), which is one of the leading causes of preventable blindness in the World.[4] As high glucose levels weaken the vessel walls, microaneurysms[5] develop on the vascular surface. When these pathological regions break, blood and/or lipids leak into the retinal layers producing hemorrhages, exudates and swelling.[1] At the same time, several biomechanical changes also lead to vessel occlusion, which means that some retinal areas do not receive enough oxygen. The body reacts to this condition by releasing angiogenic factors that lead to the generation of new vessels to bypass the damaged ones which grow in a fibrous scaffold. This new fibrovascular proliferation defines proliferative DR (PDR). These weak new vessels might also break and leak not only to the retina but also to the vitreous humor. At this stage patients have a higher risk of suffering retinal detachment, hemovitreous, and other retinal changes that lead to irreversible vision loss.[6]

Whereas end-stage DR is the leading cause of irreversible blindness among working-age adults in the world,[4] the disease is completely asymptomatic in its early stages. However, early and adequate correction of hyperglycemia and concomitant cardiovascular risk factor significantly reduces the risk of developing advanced DR in the first place, justifying the need for screening programs aimed at detecting DR in early stages of the disease. Identifying subjects with PDR, though, is essential to start other more urgent treatments such as laser photocoagulation or anti-VGEF injections to prevent further bleeding and/or neovascularization. Automated systems for computer-aided diagnosis of DR, in general, or PDR, in particular, can contribute not only to improving the efficiency of screening campaigns but also to reducing the inter-expert variability.[7] In general, these systems are aided by the automated computation

---

[a)] Electronic mail: jiorlando@conicet.gov.ar

TABLE I: Summary of statistical studies describing the fractal geometry of DR.

Study	Conclusions
Daxer, 1993[8]	**High fractal dimension** observed on **PDR patients**.
Cheung *et al.*, 2009[9]	**High fractal dimension** observed on **early DR** patients.
Grauslund *et al.*, 2010[10]	**Low fractal dimension** observed on **PDR** patients.
Aliahmad *et al.*, 2014[11]	**High fractal dimension** observed on **early DR** patients.
Mudigonda *et al.*, 2015[12]	**High fractal dimension** observed on **PDR** patients.
Huang *et al.*, 2016[13]	**No statistically significant differences** observed in the fractal dimension of patients with **different DR grades**.

TABLE II: Summary of existing methods for PDR detection. AUC = area under the ROC curve. Se = sensitivity. Sp = specificity.

Study	Methodology	Num. images	Results
Welikala *et al.*, 2014[17]	Vascular features and dual classification	60 images (from MESSIDOR and a private data set)	AUC = 0.9682, $Se = 1$, $Sp = 0.9$
Welikala *et al.*, 2015[18]	Feature selection based on genetic algorithms	60 images (from MESSIDOR and a private data set)	AUC = 0.9914, $Se = 1$, $Sp = 0.9750$
Roychowdhury *et al.*, 2016[19]	Texture, structural and intensity features	57 images (from STARE and a private data set)	AUC = 0.8291, $Se = 0.864$, $Sp = 0.76$
Gupta *et al.*, 2016[20]	Texture features	776 images (from STARE, MESSIDOR, HRF and a private data set)	AUC = 0.9597, $Se = 0.922$, $Sp = 0.83$

of pathology-related radiomics, which are used to train a classifier.[6]

Fractal dimensions have been extensively analyzed in the literature as potential descriptors of pathological changes in the retinal vasculature (Table I). However, the inconsistencies in their conclusions have suggested that they are sensitive to imaging conditions or characterization protocols.[13,14] In particular, Huang *et al.*[13,15] observed that fractal dimensions are unstable with respect to factors such as the capture devices, the segmentation method used to retrieve the vessels and the regions where the fractal properties are analyzed. However, such a study is guided by the labels provided in the data set MESSIDOR,[16] which were assigned following a diagnostic criterion in which both pre-PDR and PDR cases are grouped in the same category (Section II). Nevertheless, the most evident changes in the distribution of the retinal vessels are expected to occur on the proliferative cases as a consequence of abnormal vasculogenesis (Fig. 1). On the other hand, the study was performed in terms of the absolute values of the fractal features,[13] without combining them with any machine learning technique, which is the usual practice in developing DR screening systems.[6]

On the other side, only a few methods for automated PDR screening can be found in the literature and most of them are focused on segmenting areas with abnormal vasculogenesis (Table II). In general, all these approaches rely on a training phase from patch-level annotations, which requires an intensive effort to obtain. As a consequence, these methods are usually evaluated on relatively small data sets with a few images.

This article presents an extensive study of three different fractal dimensions and multiscale measurements, obtained from vessel segmentations and their skeletonizations, for characterizing patients with PDR. For reproducibility purposes, the entire study is conducted on the publicly available data set of fundus images MESSIDOR.[16] To this end, the images of MESSIDOR were relabeled by two experts to separate pre-PDR and PDR cases. The fractal dimensions are statistically evaluated for different DR grades, first, reporting a growing tendency for higher levels of retinopathy. Moreover, fractal dimensions retrieved from skeletonizations showed a better ability to discriminate PDR cases compared with those extracted from segmentations. Additionally, we investigate if the typical strategy used for estimating fractal dimensions from multiscale measurements affects their true discrimination ability. To this end, ℓ_1 and ℓ_2 regularized logistic regression classifiers are learned from these features for PDR screening. The empirical results show that the estimation of the fractal dimension as the slope of a fitted line actually reduces the discrimination ability of the original measurements. Furthermore, performance is significantly improved when using the raw measurements as features. It was also observed that the combination of fractal measurements with red lesion based features achieves high performance for PDR screening. To the best of our knowledge, this study is the first one reporting results obtained using multiscale fractal measurements for training a PDR screening method. On the other hand, instead of using patch-level categories, our approach is trained from the more economical image-level annotations. Furthermore, the evaluation is performed on a larger set of images compared to existing approaches (Table II), with a more realistic proportion of true positive and negative samples.[7]

II. MATERIALS

All experiments were carried out on the publicly available data set of retinal images MESSIDOR,[16] which is a benchmark set for evaluating computer-assisted methods for DR assessment. It comprises 1200 eye color fundus photographs provided by 3 medical institutions in France, captured using a Topcon TRC NW6 (Topcon, Japan) device. The original file format is TIF, and 3 different image resolutions can be found in the data set: 1440 × 960, 2240 × 1488 and 2304 × 1536. In our study, all images were standardized to an approximate resolution of 909 × 909 pixels so that the vessel calibre relatively matches that of the DRIVE data set.[21] This is required for segmenting the vasculature, as detailed in Section III A. Images are

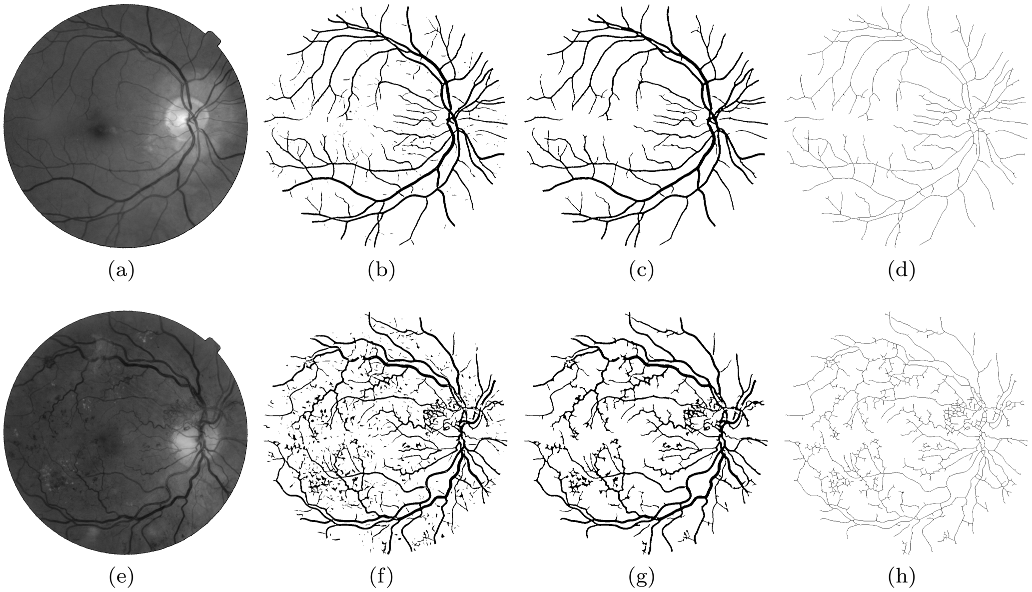


FIG. 1: Retinal vasculature segmentation and postprocessing. First row: non-proliferative case. Second row: proliferative case. From left to right: green band of the original color image (a, e), vessel segmentation (b, f), postprocessing removing spurius detections (c, g) and skeletonized vasculature (d, h).

labeled in 4 DR grades–R0, R1, R2 and R3–following the diagnostic criterion indicated at the top of Table III.

The R3 category originally provided in the set corresponds to both pre-proliferative and proliferative cases. As the purpose of this work is to estimate the ability of fractal features to identify patients suffering from abnormal vasculogenesis, images belonging to R3 were relabeled by two ophthalmologists, assigning the R3 grade to pre-proliferative cases (without neovessels) and the R4 label to PDR subjects (with neovessels) (second part of Table III). Neovascularization was labelled as neovessels in the optic disc and neovessels elsewhere in the retina, to follow the clinically used grading of neovascularization. A total number of 59 images with neovascularization (15 of them with neovessels in the optic disc area) were identified. The first human grader was used in all the experiments as the ground truth reference, while the second observer was used to estimate the inter-expert variability, as made in previous studies.[21] All these labels are made publicly available in the project website for reproducibility purposes.

III. METHODS

Fractal objects are structures that are characterized by *self-similarity*, which means that a pattern persists at multiple scales.[13] Branching structures such as roots or lightning bolts are natural examples of these objects.[22] The property of self-similarity is generally quantified by means of fractional dimensions. Retinal vasculature is known to constitute a fractal object as well, which means it can be characterized using standard fractal dimensions.[23] In order to study such properties, it is necessary to first segment the vascular structure from an image. In this study, we made use of fundus photographs, which are projective color images of the inner surface of the human eye that can be obtained non-invasively at a relatively economical cost.[6] Given a fundus photograph, the retinal vasculature is segmented and skeletonized as explained in Section III A. Afterwards, a fractal analysis procedure based on the box-counting method[24] is performed to retrieve different fractal dimensions and multiscale measurements (Section III B). Finally, two different regularized logistic regression classifiers (Section III C) are trained to assess these features for proliferative DR detection.

A. Vessel segmentation and skeletonization

Fractal characterization schemes are usually based on raw binary maps obtained from a manual or an automated vessel segmentation strategy, or using their skeletonizations. Fig. 1 presents examples of segmentations and skeletonizations obtained from two images sampled from the MESSIDOR data set.[16] Fractal features computed from vessel segmentations are expected to characterize not only their distribution patterns but also the relative caliber of the vascular segments. On the contrary, those features obtained from vessel skeletonizations are more likely to describe only the general organization of the vascular tree.

In the present study the retinal vessels are recovered by means of a fully-connected conditional random field model learned with a structured output support vector machine.[25,26] In particular, we follow the same configuration reported in Orlando *et al.*,[27] which resulted in a more connected representation of the general vascular structure. As a preprocessing step, the green channel of the RGB images is taken as it is the one that exhibits the best contrast between vessels and the remaining struc-

TABLE III: DR grades in the MESSIDOR data set. Top: diagnostic criterion provided in the set. Bottom: redefined criterion used in this study. N: number of lesions. Subindices: MA = microaneurysms, HE = hemorrhages and NV= neovascularization.

Grade	Criteria	Images	Prop.
R0	$(N_{\mathrm{MA}} = 0)$ AND $(N_{\mathrm{HE}} = 0)$	546	45.50%
R1	$(0 < N_{\mathrm{MA}} \leq 5)$ AND $(N_{\mathrm{HE}} = 0)$	153	12.75%
R2	$(5 < N_{\mathrm{MA}} < 15)$ AND $(0 < N_{\mathrm{HE}} < 5)$ AND $(N_{\mathrm{NV}} = 0)$	247	20.58%
Original diagnostic criterion			
R3	$(N_{\mathrm{MA}} \geq 15)$ OR $(N_{\mathrm{HE}} \geq 5)$ OR $(N_{\mathrm{NV}} > 0)$	254	21.17%
Redefined diagnostic criterion			
R3	$(N_{\mathrm{MA}} \geq 15)$ OR $(N_{\mathrm{HE}} \geq 5)$ AND $(N_{\mathrm{NV}} = 0)$	195	16.25%
R4	$(N_{\mathrm{MA}} \geq 15)$ OR $(N_{\mathrm{HE}} \geq 5)$ AND $(N_{\mathrm{NV}} > 0)$	59	4.92%

tures of the fundus. Afterwards, a series of pixel level features are computed.[27] For vessel skeletonization we apply an iterative thinning algorithm commonly used for this task.[28]

An additional postprocessing stage is introduced after segmentation to remove spurious detections that usually occur in highly pathological cases due to the presence of red lesions, scars of laser photocoagulation treatments or dark streaks between bright lesions (Fig. 1(f)). This approach consists of three stages. First, all binary connected components with less than 100 pixels are removed from the segmentation. A morphological closing[28] with a disk structuring element of 2-pixels radius is applied afterwards to connect isolated elements with the main vascular structure, and to remove holes in the main arteries due to the central reflex. Finally, all connected components with less than 200 pixels are removed. The resulting segmentations (Figs. 2(c) and 1(g)) exhibit less spurius structures and a more accurate representation of the main vascular structure, while preserving the typical features of the pathological areas like neovascularizations (Fig. 2).

B. Fractal dimensions and measurements

Fractal dimensions are metrics that quantify the self-similarity of fractal objects.[13] Loosely speaking, the self-similarity property can be formally described as:

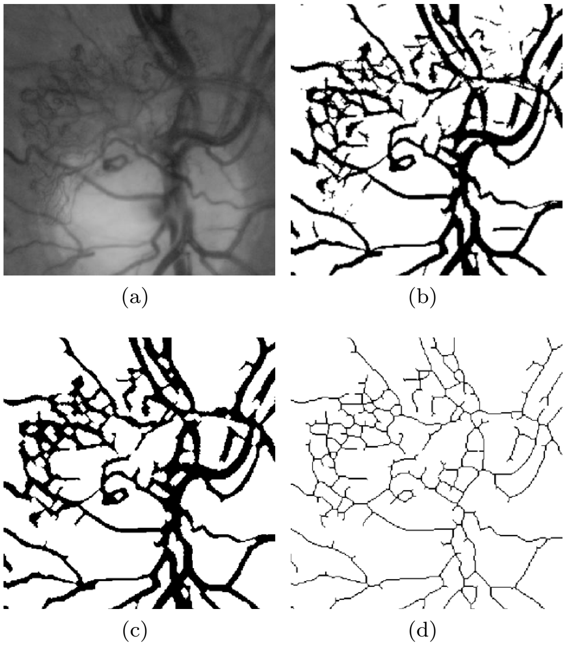
$$N(r) \propto r^{-D}, \tag{1}$$

where $N(r)$ is an appropriate quantification of the pattern (from now on called *measurement*) at a given scale r, and D is the so-called fractal dimension. D characterizes the similarity observed in the object when the scale r is increased or decreased, and knowing $N(r)$ from Eq. (1) it can be obtained as:

$$D = -\lim_{r \to 0} \frac{\log N(r)}{\log r}. \tag{2}$$

If the self-similarity property is assumed to approximately hold, then it is expected that computing the measurement N at two different scales r_n and r_{n-1} yields an estimate of D as follows:

$$D \simeq -\frac{\log N(r_n) - \log N(r_{n-1})}{\log r_n - \log r_{n-1}}. \tag{3}$$


(a) (b)

(c) (d)

FIG. 2: Vessel segmentation on areas with neovascularizations. (a) Green band of the original image. (b) Blood vessel segmentation. (c) Refined segmentation. (d) Skeletonized area.

Based on this relationship between measurements and scales, one of the simplest strategies to estimate D is to apply the box-counting method.[24] It consists of analyzing the measurement N as obtained from multiscale regular grids superimposed over the input object. More specifically, a binary input image I is divided in different squares of side-length r_n–with n being a given resolution– and the measurements are taken from the boxes that overlap with the object to be characterized. This process is repeated until r reaches a minimum scale, and a set of measurements over a wide range of scales is obtained. Finally, a log-log plot is produced to represent the changes in N with the grid resolution, and the fractal dimension D can be estimated as the slope of the regression line that fits those points.[13]

In fundus image analysis, different fractal dimensions have been applied to characterize the structural patterns of the retinal vasculature.[9,29] In this work we use three different estimators of fractal dimension–namely box, information and correlation dimensions–and their corresponding fractal measurements.

1. Box dimension.

The box dimension D_B, also known as the capacity or the Minkowski-Bouligand dimension,[22] is one of the simplest approaches for fractal characterization. It was first introduced by Liebovitch *et al.*,[24] and is defined as the number $N(r)$ of boxes with side-length r that are needed to cover an object that grows following $(1/r)^{D_B}$ as $r \to 0$:

$$D_B = \lim_{r\to 0} \frac{\log N(r)}{\log 1/r}. \tag{4}$$

where $N(r)$ represents the number of squared boxes with side-length r which overlap with the binary structure under analysis.

2. Information dimension.

The information dimension D_I[13] is defined as:

$$D_I = \lim_{r\to 0} \frac{H(r)}{\log 1/r} \tag{5}$$

where $H(r)$ is the sum of the Shannon entropy of each of the cells in the grid:

$$H(r) = -\sum_{i=1}^{n_r} p_i(r) \log p_i(r) \tag{6}$$

Here, n_r is the total number of cells of side-length r in the image, $p_i(r) = q_i(r)/M$ represents the probability of finding a part of the binary object in the i-th cell of side-length r, $q_i(r)$ is the number of pixels of the object contained in the i-th cell, and M is the total number of pixels of the object.

3. Correlation dimension.

The correlation dimension D_C is relatively similar to the box dimension, defining N as the average number of points contained in a box of size r. This is usually approximated by the probability density:[13]

$$C(r) = \frac{1}{{n_r}^2} \sum_{i=1,j=1,i\neq j}^{n_r} \Theta\left(r - ||\mathbf{p}_i - \mathbf{p}_j||\right) \simeq \sum_{i=1}^{n_r} p_i^2(r), \tag{7}$$

where $\Theta(x)$ is the Heaviside step function, $\mathbf{p}_i$ is the position of the i-th pixel belonging to the fractal object, and $p_i(r)$ is the probability of finding an object in the i-th cell as defined in Section III B 2. Then, the correlation dimension of a binary object is obtained as:

$$D_C = \lim_{r\to 0} \frac{\log C(r)}{\log 1/r}. \tag{8}$$

C. Regularized logistic regression

Proliferative DR detection can be modeled as a binary classification task that can be solved by means of a supervised machine learning approach. In particular, in this study the procedure will be guided by regularized logistic regression classifiers with two different regularizers,[30,31] and both fractal dimensions and measurements will be explored as potential features for training these classifiers.

Let S be a training set composed of n training instances $\{(\mathbf{x}_i, y_i), i = 1, ..., n\}$. Each $\mathbf{x}_i \in \mathbb{R}^d$ is a d-dimensional feature vector extracted from the i-th image, and $y_i \in \{-1, +1\}$ corresponds to its class label ($+1$ if proliferative DR or -1 if healthy or non-proliferative DR). Features in $\mathbf{x}_i$ are standardized to zero mean and unit variance before training the classifier.

Logistic regression can be written as a regularized risk minimization with logistic loss.[27] The objective function to minimize is given by the following expression:

$$\hat{\beta} = \operatorname*{argmin}_{\beta} \lambda\Omega(\beta) + \sum_{i=1}^{n} \log\left(1 + e^{-y_i\langle\beta,\mathbf{x}_i\rangle}\right) \tag{9}$$

where β is a coefficient vector of the linear discriminant function, $\lambda \geq 0$ is a scalar parameter controlling the degree of regularization by the regularizer $\Omega : \mathbb{R}^d \mapsto \mathbb{R}_+$ and $\langle\cdot,\cdot\rangle$ is the canonical inner product in $\mathbb{R}^d$. Two different regularizers were used as Ω, the $\ell_1 = \|\cdot\|_1$ and $\ell_2 = \|\cdot\|_2$ norms.[30] ℓ_1 imposes the sum of the absolute values of the parameters β to be small, resulting in sparse parameter vectors.[31] This setting makes this regularizer suitable for simultaneous learning and feature selection. We are interested in this specific ability as we want to identify relevant scales when using fractal measurements as features. By contrast, ℓ_2 regularization might not be effective when only a few features are relevant. Finally, the likelihood of a given image with feature vector $\mathbf{x}_i$ being proliferative DR is obtained by applying:

$$p(y_i = 1|\mathbf{x}_i) = \frac{1}{1 + e^{-\langle\beta,x_i\rangle}}. \tag{10}$$

IV. EXPERIMENTAL SETUP

To estimate the multiscale fractal measurements, all the vessel masks and their skeletonizations are extended using zero padding up to a regular resolution of 1024 × 1024 pixels. Thus, as the analyzed scales are powers of 2, a vector of 10 ($= \log_2 1024$) different measurements is obtained for each image.

Since MESSIDOR is not divided into training and test sets, we evaluated each classifier using k-fold cross-validation,[32] with $k = 5$. For this purpose, the data set is randomly divided into k folds, preserving almost the same proportion of positive and negative samples. At each iteration, $k - 1$ folds are used as a training set for learning the classifier, and the remaining one is used for testing. A validation set comprising a randomly sample of 30% of the training set samples is used at each

TABLE IV: Correlation analysis between different fractal dimensions, for each fractal object (top) and between fractal objects (bottom). ρ stands for the Pearson correlation coefficient.

Fractal dimensions	**Fractal object**	
	Vessel segmentation	Skeletonized vasculature
(D_B, D_I)	$\rho = 0.9766$	$\rho = 0.9908$
(D_B, D_C)	$\rho = 0.9451$	$\rho = 0.9768$
(D_I, D_C)	$\rho = 0.9897$	$\rho = 0.9954$
Fractal dimension	**Correlation between fractal objects**	
D_B	$\rho = 0.9109$	
D_I	$\rho = 0.9284$	
D_C	$\rho = 0.9048$	

iteration for model selection. The best value for the regularization parameter $\lambda = 10^i, i \in \{-10, -9, ..., 0, ..., 10\}$ is selected by maximizing the area under the receiver-operating characteristic (ROC) curve as measured on the validation set. The best model is then applied on the i-th test fold, and the process is repeated for each fold until $k = i$. Then, the overall performance of the classifier is estimated as the mean area under the ROC curve, averaged for all the k test folds.[33]

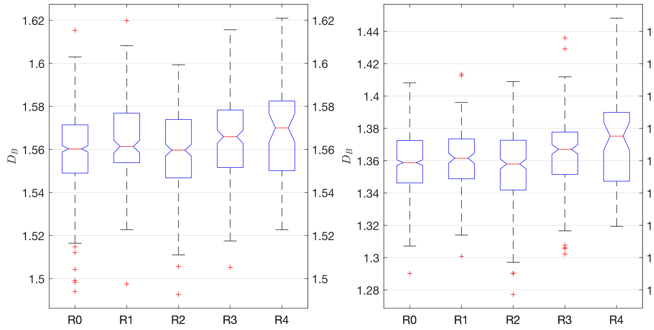
V. RESULTS

A. Statistical analysis

A first experiment was performed on the entire data set, using all fractal dimensions, to analyze their pairwise linear correlation. Two different comparisons were made: one contrasting fractal dimensions retrieved from the same object, and a separate one comparing the same fractal dimension but obtained from the different structures. The Pearson correlation coefficient ρ was used to quantify the relationship between variables. The results are presented in Table IV. Independently of the fractal object to be characterized, ρ values higher than 0.9 are always observed. The highest linear correlation occurs between D_I and D_C, while the lowest one is between D_B and D_C. Moreover, the correlation between fractal dimensions extracted from different objects are lower, yet high enough to assume that the features are strongly associated to each other.

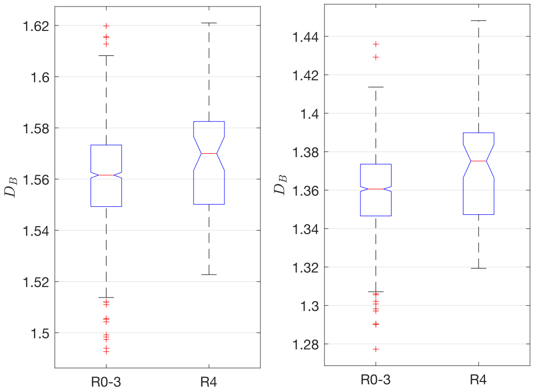
The distribution of the D_B fractal dimension computed from vessel segmentations and skeletonizations, and for each DR grade, are represented with boxplots in Figure 3. The remaining fractal dimensions are not included as they are highly correlated with D_B and have similar distributions.

To assess the level of significance of the differences in the inter-group distributions, a non-parametric Kolmogorov-Smirnov test ($\alpha = 0.05$) was performed for every combination of DR grades. ANOVA, Kruskal-Wallis and Wilcoxon sign rank tests were discarded because the variances of each grade are not homogeneous[34,35] (Levene test, $p < 0.013$). In general, it is observed that the differences between patients with PDR (R4) and those with mild (R1-2) or no DR (R0) are statistically significant, specially when the fractal features are computed from skeletonizations ($p < 7.1 \times 10^{-4}$). Furthermore, if R3 is contrasted with R4, the differences are also significant ($p < 1.9 \times 10^{-2}$). On the other hand, the fractal dimensions are extracted from the vessel segmentations, most of the differences are also statistical significant ($p < 1.2 \times 10^{-2}$). The differences in D_B and D_C between R1 and R4, and in all the fractal dimensions between R3 and R4, are not significant, though ($p \geq 0.07$ and $p > 0.33$, respectively).



(a) Vessel segmentations. (b) Vessel skeletonizations.

FIG. 3: Distribution of D_B per risk level as obtained from (a) vessel segmentations and (b) skeletonized vessels.



(a) Vessel segmentations. (b) Vessel skeletonizations.

FIG. 4: Distribution of D_B fractal dimensions for healthy and non proliferative DR (R0-3) vs. proliferative DR (R4), as obtained from (a) vessel segmentations and (b) skeletonized vessels.

Figure 4 compares the distribution of D_B for the R4 label with respect to all the other DR grades grouped together, to analyze the reliability of fractal dimensions to identify patients with PDR. It can be seen that the differences between groups are statistically significant independently of the fractal objects to be characterized ($p \leq 1.3 \times 10^{-2}$ for vessel segmentations, $p \leq 9.7 \times 10^{-5}$ for skeletonizations).

Finally, to evaluate the contribution of each fractal dimension for detecting PDR, their raw responses were thresholded at different values to construct ROC curves, and the area under each of the curves (AUCs) was used to quantitatively measure their performance. Two dif-

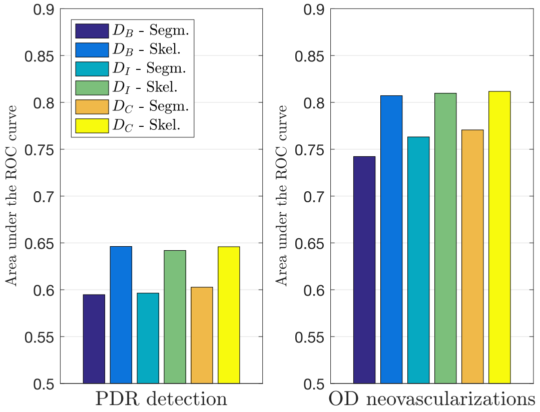


FIG. 5: AUC values obtained using each fractal dimension as a score to detect proliferative DR. Left: PDR screening (R0-3 vs. R4). Right: detection of patients with neovessels in the OD area.

ferent experiments were performed for this purpose: one estimating their ability for detecting all types of neovascularizations (R4 vs. all the other grades), and a separate one evaluating their ability to identify patients with neovessels only within the optic disc (OD) area. The resulting AUC values are depicted as bar charts in Figure 5. Despite the high correlation observed between fractal dimensions extracted from vessels and skeletonizations, higher AUC values are obtained with the latter. For PDR detection, the best performance is achieved using the D_B (AUC= 0.6462), while D_I outperformed all the other approaches for detecting optic disc neovascularizations (AUC= 0.8118).

B. Fractal dimensions and measurements as features for PDR screening

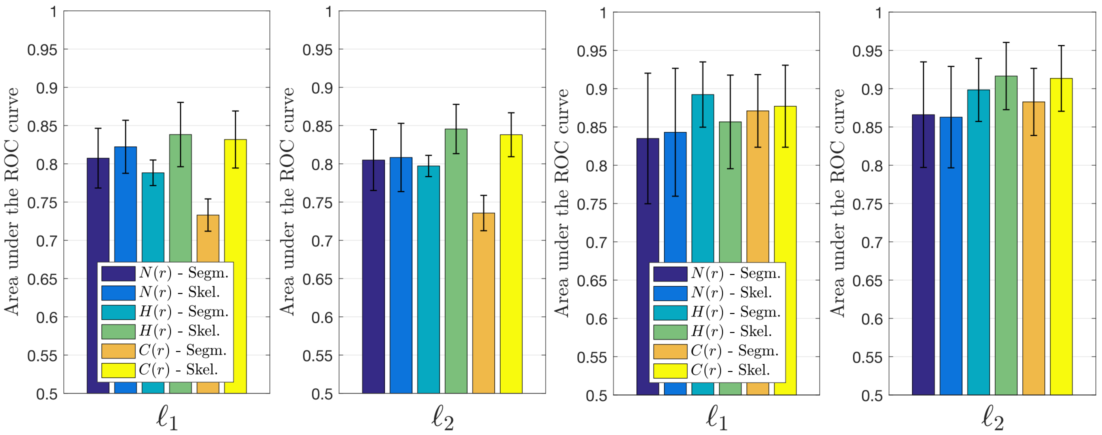
As previously indicated in Section III, fractal dimensions are normally obtained as the slope of a regression line that best fit a given set of multiscale fractal measurements. In order to study whether the linear regression step affects the discrimination ability of the original measurements, a first experiment was conducted training ℓ_1 and ℓ_2 regularized logistic regression classifiers using these features computed from segmentations or skeletonizations. Figure 6 depicts with bar charts the average AUC values obtained using 5-fold cross-validation for each specific configuration, for PDR detection and for detecting OD neovascularizations. Error bars correspond to the standard error through folds. In general, it is possible to see that the multiscale measurements reported higher performance than the fractal dimensions derived from them, with skeletonizations being more effective than segmentations for most of the comparisons. Comparing ℓ_1 and ℓ_2 regularization, the latter gives better results, which suggest that all the scales provide valuable information to the classifier. The best discriminating feature for both PDR screening and detecting OD neovascularizations, is $H(r)$ obtained from skeletonizations (AUC= 0.8455 and 0.9165, respectively).

Finally, fractal dimensions and/or measurements were evaluated as features for training classifier for PDR screening and for detecting patients with OD neovascularizations. In particular, we trained ℓ_1 and ℓ_2 regularized logistic regression classifiers with all the dimensions and measurements extracted from segmentations and skeletonizations. Additionally, we incorporated the maximum red lesion probability (obtained with our method for red lesion detection[5]) as a feature, to further assess if fractal features improve their performance when combined with it. Finally, the combination of the red lesion based feature and the best fractal measurement observed in Figure 6 (i.e. $H(r)$ from the skeletonized vasculature) was also evaluated. 5-fold cross-validation was used in all the experiments, as described in Section IV. Figure 7 presents the average ROC curves obtained for each configuration, including the mean AUC values and their standard deviations, and the second human observer sensitivity and specificity. The AUC values are also summarized in Table V. The best configurations reported AUCs of 0.93 and 0.96 for PDR screening and detecting patients with OD neovascularizations, respectively. It is worth noting that the results of our method for PDR detection are quantitatively similar to the performance of the second observer.

VI. DISCUSSION

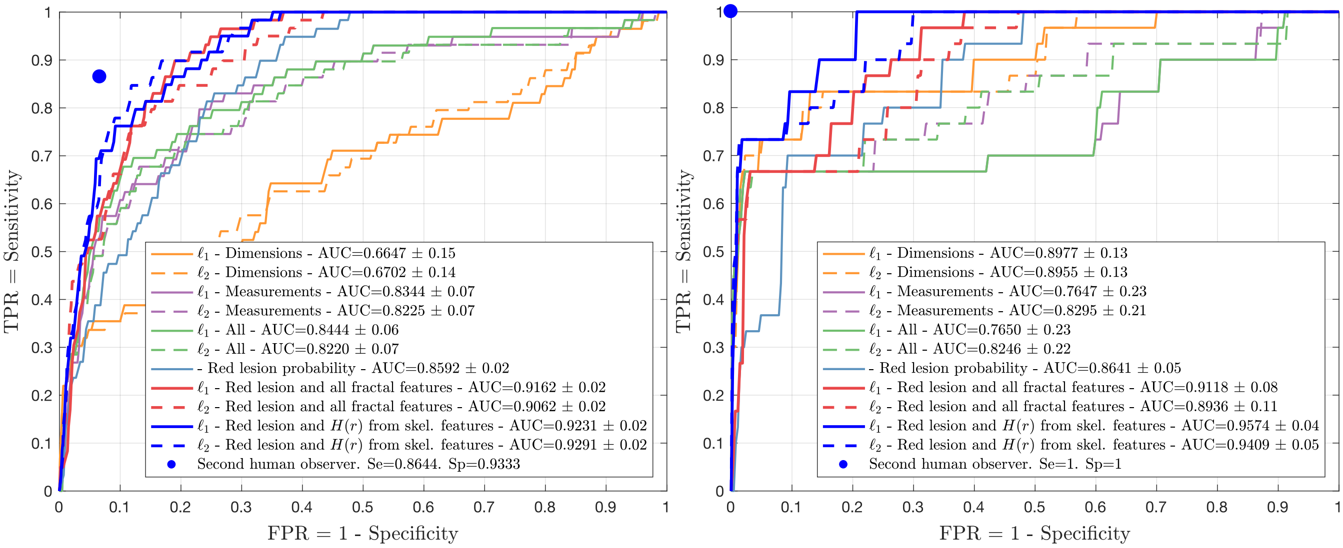
It was experimentally observed that, for low DR grades, fractal dimensions are distributed almost equally (Fig. 3). This result is in line with previous studies,[13] and is supported by the fact that red lesions, which are the earliest signs of the disease, are not expected to affect the overall distribution of the retinal vasculature. However, when PDR cases are separated from the original R3 grade in the data set, an increased fractal dimension value is more likely to be obtained, independently of the type of dimension considered. If healthy, mild and preproliferative subjects are grouped into the same category, the differences are more evident (Fig. 5), which in principle indicate that fractal dimension is a good feature for PDR screening.

Despite the fact that a high linear correlation is observed among different types of fractal dimensions (Table IV), the AUC values reported by the features extracted from the skeletonized vasculature are higher than those obtained using the segmentations (Fig. 5) for both PDR screening and for detecting OD neovascularizations. This can be explained by the fact that areas with agglomerations of abnormal vessels are better characterized by skeletonizations, because this postprocessing method does not take into account the caliber of the vessels. Another interesting result is that the multiscale measurements used as features for logistic regression produced higher AUC values than their associated fractal dimensions, as seen when comparing Figs. 5 and 6. This characteristic is a first indication that the linear regression model used for estimating the dimension from the fractal measurements weakens the discrimination ability of the latter by filtering valuable information. On the other hand, the regularization parameter slightly affects the



(a) PDR detection (b) OD neovascularizations

FIG. 6: Mean area under the ROC curves (AUCs) obtained by ℓ_1 and ℓ_2 regularized logistic regression classifiers, trained using multiscale fractal measurements for (a) PDR detection and (b) detection of OD neovascularizations, using 5-fold cross-validation. Error bars represent the standard error.



(a) PDR detection. (b) OD neovascularizations.

FIG. 7: Mean ROC curves obtained by ℓ_1 and ℓ_2 regularized logistic regression classifiers trained using different combinations of features, for (a) proliferative DR detection and (b) detection of optic disc (OD) neovascularizations.

TABLE V: Mean AUC values (± standard deviation) obtained for proliferative DR (PDR) detection and for detecting OD neovascularizations.

Method		**Detection task**	
Features	**Reg.**	**PDR detection**	**OD neovascularization**
Fractal dimensions	ℓ_1	0.6647 ± 0.14	0.8977 ± 0.13
	ℓ_2	0.6702 ± 0.14	0.8955 ± 0.13
Fractal measurements	ℓ_1	0.8344 ± 0.07	0.7647 ± 0.23
	ℓ_2	0.8225 ± 0.07	0.8295 ± 0.21
Fractal dimensions and measurements	ℓ_1	0.8444 ± 0.06	0.7650 ± 0.23
	ℓ_2	0.8220 ± 0.07	0.8246 ± 0.22
Max. red lesion probability	ℓ_1 / ℓ_2	0.8592 ± 0.02	0.8641 ± 0.05
Max. red lesion probability and all the fractal features	ℓ_1	0.9162 ± 0.02	0.9118 ± 0.08
	ℓ_2	0.9062 ± 0.02	0.8936 ± 0.11
Max. red lesion probability and $H(r)$	ℓ_1	0.9231 ± 0.02	**0.9574 ± 0.04**
	ℓ_2	**0.9291 ± 0.02**	0.9409 ± 0.05

results (Fig 6), with the ℓ_2 norm reporting better performance in most of the experiments. This suggests that, in principle, all the scales are partially relevant for the analysis. In any case, the best fractal measurement for both PDR and OD neovascularization detection is the $H(r)$ obtained from skeletonizations.

Combining all the fractal dimensions into the same feature vector for PDR detection did not seem to be effective, as seen in the low AUC values reported in Fig. 7. The combination of fractal measurements, by contrast, yields a significant improvement in results, with a corresponding reduction in the variability between folds. The best performance was achieved by the ℓ_1 regularizer in this case, which means that combining all the measurements is suboptimal. Adding the fractal dimensions to this feature set slightly improves performance if the ℓ_1 regularizer is used, as well. The maximum probability of the red lesion detector[5] achieved a similar AUC value, although with lower sensitivities under the low FPR regime. This property is a consequence of the inability of this feature to characterize the presence of neovascularizations. However, results are improved when the fractal features were added. Finally, it is worth noting that the highest performance is achieved by combining the red lesion based feature and $H(r)$ from the skeletonized vasculature. Under this setting, the ℓ_2 norm is the best regularizer, which is consistent with the results observed in Fig 6. In particular, this approach yielded 90% sensitivity for a specificity value of 83%, with the lowest standard deviation through folds.

A similar evaluation can be made for detecting patients with OD neovascularizations. In this case, the combination of all the fractal dimensions reported higher performance than the combination of all the fractal measurements, suggesting that the interaction between measurements does not provide valuable information for this task. Furthermore, the red lesion based feature yielded less accurate results than the combination of fractal dimensions. Nevertheless, the combination of $H(r)$ with such a feature also resulted in better performance, reporting 100% sensitivity for a specificity value of 80%. In this case, the ℓ_1 regularizer is more appropriate, suggesting than certain scales are avoided when the red lesion feature is considered.

A comparison with the state of the art in PDR detection is difficult as most of the available tools were tested on different data sets (Table II). For example, the works by Welikala *et al.*[17,18] reported higher AUC values than our method, although using smaller test sets containing only a few images from MESSIDOR and others obtained from private data sets. Similarly, Roychowdhury *et al.*[19] reported an AUC value of 0.8291 for PDR detection on a smaller data set made with a combination of 40 images from STARE[36] and another 17 images extracted from a private data set. In the present study we used the whole set of 1200 images provided in MESSIDOR in order to produce results as general as possible.[31] Even under this challenging but realistic setting, our combination of fractal features reported higher AUC values than other existing approaches.

A potential limitation regards the grading of neovascularizations. Neovessels were identified by looking into fundus images, without access to angiographic leakage information, which would confirm the diagnosis. However, experienced graders were responsible for this task, which makes it more likely to be accurate. Furthermore, the relative agreement between the two human observers (Figure 7) suggests a certain consensus regarding the images that correspond to PDR cases, which supports the reliability of the labels. Further investigation of inter-expert agreement is an interesting area of future research. On the other hand, the accuracy of the vessel segmentation is essential to ensure a proper fractal characterization. In our experiments, we observed that the method usually fails when the segmentation contains spurious segments that were not suppressed by the preprocessing method. This occurs when the images in MESSIDOR contain large hemorrhages covering large areas of the images or the illumination is uneven. In other data sets, it can also happen when the quality of the images is poor. As fundus photographs are non-invasive and their acquisition is relatively economical, low quality images could be recaptured to mitigate this issue and to ensure a proper input to the system.

VII. CONCLUSIONS

An extensive study of the reliability of using fractal dimensions to identify patients with proliferative DR (PDR) was presented. A statistical analysis of their distribution through different DR grades indicated that patients with PDR are more likely to have an increased fractal dimension, which is associated with the presence of neovascularizations. The statistically significant differences between R4 and the remaining labels indicate that fractal properties can be robust enough to characterize PDR cases with respect to non-proliferative and healthy subjects. Furthermore, the original multiscale fractal measurements used for the computation of the fractal dimensions were assessed as input features for training logistic regression classifiers. It was observed that this approach yields better results for PDR screening and for detecting patients with optic disc neovascularization than using the raw fractal dimensions. This setting indicates that the linear combination of fractal measurements performs better than the traditional linear regression model for PDR detection. Finally, state of the art performance is obtained for both classification tasks when integrating fractal measurements with a red lesion based feature. These settings should be taken into account when designing automated methods for DR grading, as they might potentially improve results. In conclusion, fractal properties are shown to be effective radiomics for characterizing neovacularization processes. Finally, an open source implementation of our experimental setup is made available online for the sake of contributing to the reproducibility of our experiments.[37]

ACKNOWLEDGEMENTS

This work is partially funded by ANPCyT PICT 2014-1730 and 2016-0116, PICT start-up 2015-0006, NVIDIA Hardware Grant, Internal Funds KU Leuven and FP7-

MC-CIG 334380.

CONFLICT OF INTEREST

All the authors have no relevant conflicts of interest to disclose.

[1] N. Cheung, P. Mitchell, and T. Y. Wong, "Diabetic retinopathy," The Lancet **376**, 124 – 136 (2010).

[2] D. R. Whiting, L. Guariguata, C. Weil, and J. Shaw, "IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030," Diabetes Research and Clinical Practice **94**, 311–321 (2011).

[3] F. B. Hu, "Globalization of diabetes," Diabetes care **34**, 1249–1257 (2011).

[4] E. Prokofyeva and E. Zrenner, "Epidemiology of major eye diseases leading to blindness in Europe: A literature review," Ophthalmic Research **47**, 171–188 (2012).

[5] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M. B. Blaschko, "Learning to detect red lesions in fundus photographs: An ensemble approach based on deep learning," arXiv preprint arXiv:1706.03008 (2017).

[6] M. D. Abràmoff, M. K. Garvin, and M. Sonka, "Retinal imaging and image analysis," IEEE Reviews in Biomedical Engineering **3**, 169–208 (2010).

[7] M. D. Abràmoff and M. Niemeijer, "Mass screening of diabetic retinopathy using automated methods," in *Teleophthalmology in Preventive Medicine* (Springer, 2015) pp. 41–50.

[8] A. Daxer, "The fractal geometry of proliferative diabetic retinopathy: implications for the diagnosis and the process of retinal vasculogenesis," Current eye research **12**, 1103–1109 (1993).

[9] N. Cheung, K. C. Donaghue, G. Liew, S. L. Rogers, J. J. Wang, S.-W. Lim, A. J. Jenkins, W. Hsu, M. L. Lee, and T. Y. Wong, "Quantitative assessment of early diabetic retinopathy using fractal analysis," Diabetes Care **32**, 106–110 (2009).

[10] J. Grauslund, A. Green, R. Kawasaki, L. Hodgson, A. K. Sjølie, and T. Y. Wong, "Retinal vascular fractals and microvascular and macrovascular complications in type 1 diabetes," Ophthalmology **117**, 1400–1405 (2010).

[11] B. Aliahmad, D. K. Kumar, M. G. Sarossy, and R. Jain, "Relationship between diabetes and grayscale fractal dimensions of retinal vasculature in the indian population," BMC ophthalmology **14**, 152 (2014).

[12] S. Mudigonda, F. Oloumi, K. M. Katta, and R. M. Rangayyan, "Fractal analysis of neovascularization due to diabetic retinopathy in retinal fundus images," in *E-Health and Bioengineering Conference (EHB), 2015* (IEEE, 2015) pp. 1–4.

[13] F. Huang, B. Dashtbozorg, J. Zhang, E. Bekkers, S. Abbasi-Sureshjani, T. T. Berendschot, and B. M. ter Haar Romeny, "Reliability of using retinal vascular fractal dimension as a biomarker in the diabetic retinopathy detection," Journal of Ophthalmology **2016** (2016).

[14] A. Wainwright, G. Liew, G. Burlutsky, E. Rochtchina, Y. P. Zhang, W. Hsu, J. M. Lee, T. Y. Wong, P. Mitchell, and J. J. Wang, "Effect of image quality, color, and format on the measurement of retinal vascular fractal dimension," Investigative ophthalmology & visual science **51**, 5525–5529 (2010).

[15] *Proceedings of the Ophthalmic Medical Image Analysis Second International Workshop (OMIA '15) Held in Conjunction with MICCAI* (Springer, 2015).

[16] E. Decencière, X. Zhang, G. Cazuguel, B. Laÿ, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, *et al.*, "Feedback on a publicly distributed image database: the Messidor database," Image Analysis & Stereology **33**, 231–234 (2014).

[17] R. Welikala, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, T. H. Williamson, and S. Barman, "Automated detection of proliferative diabetic retinopathy using a modified line operator and dual classification," Computer methods and programs in biomedicine **114**, 247–261 (2014).

[18] R. Welikala, M. M. Fraz, J. Dehmeshki, A. Hoppe, V. Tah, S. Mann, T. H. Williamson, and S. A. Barman, "Genetic algorithm based feature selection combined with dual classification for the automated detection of proliferative diabetic retinopathy," Computerized Medical Imaging and Graphics **43**, 64–77 (2015).

[19] S. Roychowdhury, D. D. Koozekanani, and K. K. Parhi, "Automated detection of neovascularization for proliferative diabetic retinopathy screening," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the* (IEEE, 2016) pp. 1300–1303.

[20] G. Gupta, S. Kulasekaran, K. Ram, N. Joshi, M. Sivaprakasam, and R. Gandhi, "Local characterization of neovascularization and identification of proliferative diabetic retinopathy in retinal fundus images," Computerized Medical Imaging and Graphics **55**, 124–132 (2017).

[21] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, M. D. Abramoff, *et al.*, "Comparative study of retinal vessel segmentation methods on a new publicly available database," in *Medical Imaging 2004* (International Society for Optics and Photonics, 2004) pp. 648–656.

[22] K. Falconer, *Fractal geometry: mathematical foundations and applications* (John Wiley & Sons, 2004).

[23] B. R. Masters, "Fractal analysis of the vascular tree in the human retina," Annual Review of Biomedical Engineering **6**, 427–452 (2004).

[24] L. S. Liebovitch and T. Toth, "A fast algorithm to determine fractal dimensions by box counting," physics Letters A **141**, 386–390 (1989).

[25] J. I. Orlando and M. B. Blaschko, "Learning fully-connected CRFs for blood vessel segmentation in retinal images," in *MICCAI 2014, LNCS*, Vol. 8149, edited by P. Golland, C. Barillot, J. Hornegger, and R. Howe (Springer, 2014) pp. 634–641.

[26] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," IEEE Transactions on Biomedical Engineering **64**, 16–27 (2017).

[27] J. I. Orlando, E. Prokofyeva, M. del Fresno, and M. B. Blaschko, "Convolutional neural network transfer for automated glaucoma identification," in *12th International Symposium on Medical Information Processing and Analysis* (International Society for Optics and Photonics, 2017) pp. 101600U–101600U.

[28] R. M. Haralick and L. G. Shapiro, *Computer and robot vision* (Addison-Wesley Longman Publishing Co., Inc., 1991).

[29] C. Y.-l. Cheung, E. Lamoureux, M. K. Ikram, M. B. Sasongko, J. Ding, Y. Zheng, P. Mitchell, J. J. Wang, and T. Y. Wong, "Retinal vascular geometry in asian persons with diabetes and retinopathy," Journal of Diabetes Science and Technology **6**, 595–605 (2012).

[30] A. Y. Ng, "Feature selection, l 1 vs. l 2 regularization, and rotational invariance," in *Proceedings of the twenty-first international conference on Machine learning* (ACM, 2004) p. 78.

[31] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, Vol. 1 (Springer, 2009).

[32] L. Lo Vercio, J. I. Orlando, M. del Fresno, and I. Larrabide, "Assessment of image features for vessel wall segmentation in intravascular ultrasound images," International Journal of Computer Assisted Radiology and Surgery **11**, 1397–1407 (2016).

[33] T. Fawcett, "An introduction to ROC analysis," Pattern Recognition Letters **27**, 861–874 (2006).

[34] D. W. Zimmerman, "Inflation of type i error rates by unequal variances associated with parametric, nonparametric, and rank-transformation tests," Psicologica **25** (2004).

[35] G. W. Corder and D. I. Foreman, *Nonparametric statistics: A step-by-step approach* (John Wiley & Sons, 2014).

[36] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," IEEE Transactions on Medical Imaging **19**, 203–210 (2000).

[37] `https://github.com/ignaciorlando/fundus-fractal-analysis`.