Red Lesion Detection Using Dynamic Shape Features for Diabetic Retinopathy Screening

Lama Seoud*, Thomas Hurtut, Jihed Chelbi, Farida Cheriet, and J. M. Pierre Langlois

Abstract—**The development of an automatic telemedicine system for computer-aided screening and grading of diabetic retinopathy depends on reliable detection of retinal lesions in fundus images. In this paper, a novel method for automatic detection of both microaneurysms and hemorrhages in color fundus images is described and validated. The main contribution is a new set of shape features, called Dynamic Shape Features, that do not require precise segmentation of the regions to be classified. These features represent the evolution of the shape during image flooding and allow to discriminate between lesions and vessel segments. The method is validated per-lesion and per-image using six databases, four of which are publicly available. It proves to be robust with respect to variability in image resolution, quality and acquisition system. On the Retinopathy Online Challenge's database, the method achieves a FROC score of 0.420 which ranks it fourth. On the Messidor database, when detecting images with diabetic retinopathy, the proposed method achieves an area under the ROC curve of 0.899, comparable to the score of human experts, and it outperforms state-of-the-art approaches.**

Index Terms—**Computer aided diagnostic, diabetic retinopathy, fundus IMAGING, lesion detection, retina, screening.**

I. INTRODUCTION

DIABETIC retinopathy (DR) is a complication of diabetes that can lead to impairment of vision and even blindness. It is the most common cause of blindness in the working-age population [1]. One out of three diabetic person presents signs of DR [2] and one out of ten suffers from its most severe and vision-threatening forms [3]. DR can be managed using available treatments, which are effective if diagnosed early. Since DR is asymptomatic until late in the disease process, regular eye fundus examination is necessary to monitor any changes in the retina.

With the increasing prevalence of diabetes and the aging population, it is expected that, in 2025, 333 millions diabetic patients worldwide will require retinal examination each year [4]. Considering the limited number of ophthalmologists, there is

an urgent need for automation in the screening process in order to cover the large diabetic population while reducing the clinical burden on retina specialists. Automation can be achieved at two levels: first, in detecting cases with DR, and, second, in grading these cases. Indeed, the identification of the severity level, through DR grading, allows more appropriate and consistent referral to treatment centers [5].

Our research focuses on the development of an automatic telemedicine system for computer-aided screening and grading of DR. Since computer analysis cannot replace the clinician, the system aims at identifying fundus images with suspected lesions and at sorting them by severity. Then, the annotated images are sent to a human expert for review, starting with the suspected most severe cases. Such an automatic system can help to reduce the specialist's burden and examination time, with the additional advantages of objectivity and reproducibility. Moreover, it can help to rapidly identify the most severe cases and to focus clinical resources on the cases that need more urgent and specific attention.

A computer-aided screening and grading system relies on the automatic detection of lesions. Fundus images with DR exhibit red lesions, such as microaneurysms (MA) and hemorrhages (HE), and bright lesions, such as exudates and cotton-wool spots. In this paper, we will focus only on MA and HE which are among the very first manifestations of DR.

Several methods have been developed for the automatic detection of red lesions in color fundus images. Most of them [6]–[15] focus solely on the detection of MAs. Because of their fairly uniform circular shape and limited size range, MAs can be detected using morphological operations such as diameter closing [11] and top-hat transformation using a linear structuring element at various orientations [6]–[10]. The goal here is to distinguish MAs from elongated structures. Another approach is to use a priori shape knowledge and to perform a convolution with a double ring filter [13] or through template matching with multiscale Gaussian kernels [12], [14], [15]. Contrary to vascular segments, which are directional, MAs indeed show a Gaussian-like peak in all directions.

Even though MAs are among the first signs of DR, HEs are also highly valuable for DR screening and useful for grading. In fact, retinal HEs are the result of MAs starting to leak into the retinal layers, indicating a more severe level of DR. According to the most common DR severity scale [5], their presence and number indicate either a moderate or a severe non-proliferative DR. HEs come in different types, such as "dot", "blot" and "flame" [16]. Dot HEs and MAs are difficult to distinguish from one another on fundus images, thus dot HEs are usually referred to as MAs. A flame HE corresponds to blood
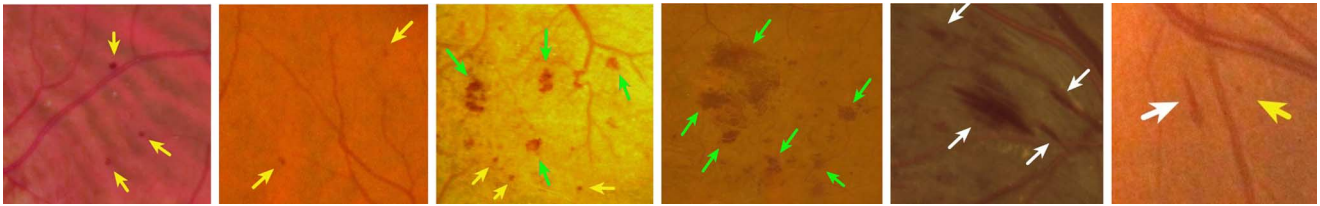
Fig. 1. Portions of different fundus images with red lesions. The yellow, green and white arrows point respectively to MAs/dot HEs, blot HEs and flame HEs.

leaking into the nerve fiber layer. Its shape, more elongated, follows the structure of the nerve fibers. A blot HE corresponds to blood leaking deeper in the retinal layer. It appears larger than a dot HE, and its borders are irregular, leading to various shapes. Fig. 1 shows examples of these lesions types.

Because HEs present a wide variety of shapes, template matching methods, such as the ones developed for MA detection, are difficult to adapt to this task. In fact, no single template can match all the possible sizes and shapes of HEs. A common methodology adopted in the literature for combined MA and HE detection consists in identifying all dark-colored structures in the image, mainly through a thresholding, combined with adapted preprocessing [16], [17], and then in removing the vessels from the resulting set of candidates. Vessel detection is performed using either a multilayer perceptron [16] or multiscale morphological closing [17]. Unfortunately, the major limitation to this approach is that most of the false positives at the vessel segmentation step are actually lesions. After their removal along with the detected vessels, these lesions are lost and not retrieved in subsequent processing.

In this paper, we propose a method for the detection of both MAs and HEs that does not require prior vessel segmentation. We consider a supervised classification scheme to discriminate between lesions and other structures like vessel segments and background noise. After image preprocessing, candidate regions are identified. Features are then extracted and used to classify each candidate. This approach is common in the literature. However, in previous work [7], [10], the candidate regions are first segmented; then, color and shape features are extracted and used for classification. Nevertheless, the reliabitily of the color and shape features depends greatly on the accuracy of the segmentation step. High accuracy is difficult to obtain in the case of diffuse HEs, particularly in low resolution images. A difference of a single pixel can have a significant impact on the circularity measure, for example, especially for small candidate lesions.

Our major contribution is a new set of shape features that do not require precise segmentation of the candidates. We consider every regional minimum as a candidate. Since the boundaries of the minima do not necessarily correspond to the edges of the structures of interest, we propose to extract shape features through the process of morphological image flooding. The general idea behind this approach lies in the physical phenomenon of blood leaking from (as opposed to blood flowing in) the vessels. In the case of a lesion, the local minimum represents the focal point from which the blood is leaking gradually, in a more or less isotropic manner, depending on whether the lesion is an MA or an HE. This can be represented as nested layers of progressively higher intensities: as the intensity threshold

increases, each evolving layer encompasses those found previously. The difference with a vessel segment is that the layers evolve more anisotropically in the latter case, following the vessel's orientation, and, at some intensity threshold, start merging with other vessel segments. This novel set of features, called Dynamic Shape Features (DSF), was briefly introduced in a preliminary study [18].

Moreover, because our goal is to use the proposed lesion detection method in a telemedicine context, the method must deal well with variability in image resolution and quality, in addition to the inherent variation in the appearance of the retina. To evaluate the performance of our proposed method in such conditions, we conducted extensive validation, using six independent image databases varying mostly in terms of image resolution and acquisition system. This validation, conducted partly on publicly available datasets, allows us to directly compare our method to the state-of-the-art.

The paper is organized as follows. In Section II, we provide a detailed description of the methodology and of the DSFs. Section III describes the experimental validation setup. Section IV reports the results which are then discussed in Section V.

II. PROPOSED METHOD

The proposed method takes as input a color fundus image together with the binary mask of its region of interest (ROI). The ROI is the circular area surrounded by a black background. It outputs a probability color map for red lesion detection. The method comprises six steps. First, spatial calibration is applied to support different image resolutions. Second, the input image is preprocessed via smoothing and normalization. Third, the optic disc (OD) is automatically detected, to discard this area from the lesion detection. Fourth, candidate regions corresponding to potential lesions, are identified in the preprocessed image, based on their intensity and contrast. Fifth, the DSF together with color features are extracted for each candidate. Sixth, candidates are classified according to their probability of being actual red lesions. Each of these steps is detailed in the following subsections.

A. Spatial Calibration

To adapt to different image resolutions, we use a spatial calibration method introduced in [19]. Images are not resized. Rather, the diameter D of the ROI (after removal of the dark background) is taken as a size invariant. This hypothesis is reasonable since most of the images for DR screening are acquired with a field of view (FOV) of 45°. D is used to set
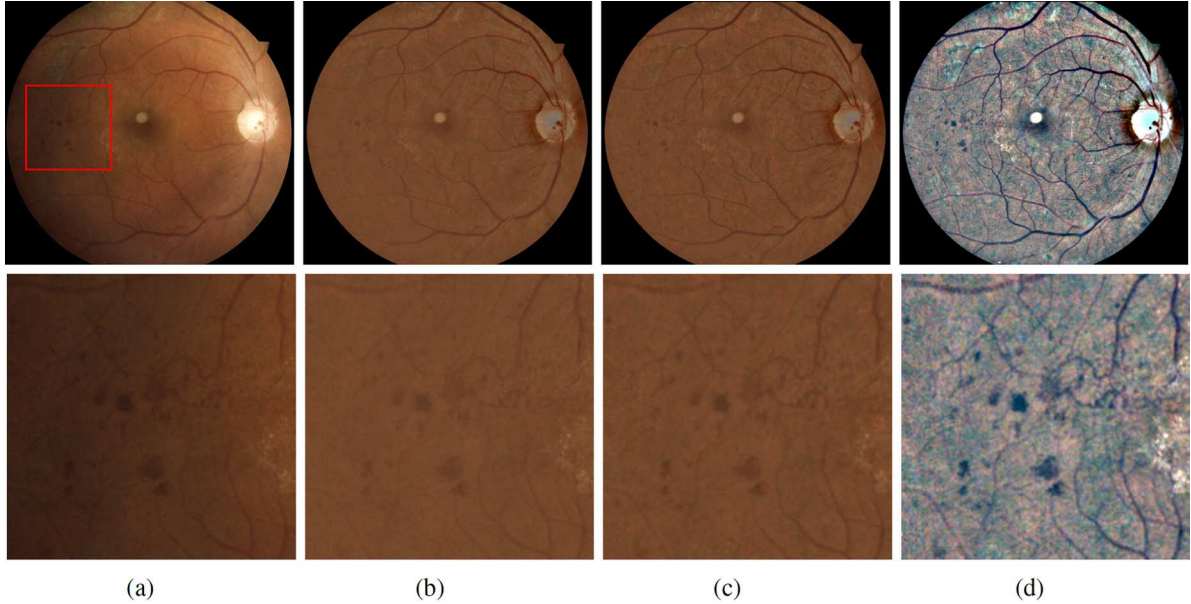
Fig. 2.   Image preprocessing steps. (a) Original image I; (b) illumination equalization; (c) adaptive contrast equalization; (d) color normalization. The lower row shows a detailed area of the image after each step. (a) I (b) I_{ie} (c) I_{ce} (d) I_p.

the kernel sizes of the different filters in the proposed method. Three size parameters are used in our method:

- d_1 is the average radius of the OD;
- d_2 is the size of the smallest MA;
- d_3 is the size of the largest HE.

In the case of eye fundus images obtained with a FOV of $45°$, we have experimentally set these parameters to $d_1 = D/10$; $d_2 = D/360$ and $d_3 = D/28$.

B. Image Preprocessing

The illumination of the retina is often nonuniform, leading to local luminosity and contrast variation. Lesions may be hardly visible in areas of poor contrast and/or low brightness. Moreover, in a telemedicine context, images are variable in terms of color and quality. Consequently, preprocessing steps are required to address these issues. The successive steps are detailed hereafter and illustrated in Fig. 2.

1) Illumination Equalization: To overcome the vignetting effect, the illumination equalization method in [20] is used:

$$I_{ie} = I + \mu - I * h_{M1}. \tag{1}$$

A large mean filter (h_{M1}) of diameter d_1 is applied to each color component of the original image I in order to estimate its illumination. Then, the resulting color image is subtracted from the original one to correct for potential shade variations. Finally, the average intensity μ of the original channel is added to keep the same color range as in the original image.

2) Denoising: A small mean filter (h_{M2}) of diameter d_2 is applied to each color channel of the resulting image I_{ie} in order to attenuate the noise resulting from the acquisition and compression steps without smoothing the lesions.

3) Adaptive Contrast Equalization: The contrast drift is approximated using the local standard deviation computed for

each pixel in a neighborhood of diameter d_1, for each color channel (I_{std}). Areas with low standard deviation indicate either low contrast or smooth background. To enhance low contrast areas, we sharpen the details in these specific regions using (2) for each color channel separately:

$$I_{ce} = I_{dn} + \frac{1}{I_{std}}(I_{dn} * (1 - h_{M3})). \tag{2}$$

Local image details are thereby added to the denoised image, weighted by the inverse of the contrast drift. The details are obtained using a high pass filter, derived from a mean filter (h_{M3}) of diameter d_3. The previous denoising step prevents undesirable noise sharpening.

4) Color Normalization: Color normalization is necessary in order to obtain images with a standardized color range. We perform, in each color channel of I_{ce}, histogram stretching and clipping in the range $\mu \pm 3\sigma$, where μ and σ are the mean and standard deviation of the color channel in the ROI. The resulting preprocessed image I_p is shown in Fig. 2(d).

C. Optic Disc Removal

The OD is a significant source of false positives in red lesion detection [18], [21]; therefore its removal is a necessary step.

Starting from the preprocessed image, we first use an entropy-based approach [22] to estimate the location of the OD's center. Basically, the OD is located in a high intensity region where the vessels have maximal directional entropy. A subsequent optimization step then estimates the OD's radius and refines its position. This consists in convolving a multi-scale ring-shaped matched filter to the image in a sub-ROI centered on the first estimation of the OD's center, of radius equal to a third of the ROI's radius. The radius and position of the matched filter that minimizes the convolution are selected as the OD's final radius and center position.

D. Candidate Extraction

Since blood vessels and dark lesions have the highest contrast in the green channel [11], the latter is extracted from the preprocessed image and is denoted G_p. The red and blue channels are used later to extract color features.

In the green channel, MAs and HEs appear as structures with local minimal intensity. A brute force approach would be to extract all the regional minima in G_p [15]. A regional minimum is a group of connected pixels of constant intensity, such that all the adjacent pixels have strictly higher intensities [23]. Unfortunately, this method is highly sensitive to noise. Depending on the smoothness of the image, the number of regional minima can thus be very large.

To overcome this limitation, we adopt the dynamics transformation [24] which rates regional minima according to their local contrast. Noisy minima usually have lower contrast than red lesions. In a topographic representation of G_p, the dynamic of a minimum is computed as the difference in intensity between the given minimum and the brightest points of the paths reaching a minimum of lower intensity. The main advantage of this definition is that the resulting contrast measurement is independent of the size and shape of the regional minimum.

Using this transformation, we can select the minima by thresholding the resulting contrast image. At this point, we would like to discard from the set of candidates as many local minima corresponding to noise as possible. In order to estimate the noise's intensity, we compute the local standard deviation in a neighborhood of the size of the papilla and consider the lowest standard deviation inside the ROI, which would correspond to a region in the retinal background with minimal signal intensity. This noise estimation is denoted σ_{noise}.

Finally, a selected minimum should have an intensity lower than the mean intensity in G_p to be considered a candidate region. This is supported by the fact that we are looking for red lesions, which are darker than the retinal background.

Contrast and illumination equalization gain importance at this point. Without these preprocessing steps, global contrast and intensity thresholding would be difficult to achieve.

In addition, all candidates whose distance to the OD's center is smaller than the OD's radius are removed from the set of candidates and not considered any further.

E. Dynamic Shape Features

Among the candidates, several regions correspond to non-lesions, such as vessel segments and remaining noise in the retinal background. To discriminate between these false positives and true lesions, an original set of features, the DSFs, mainly based on shape information, is proposed.

In a topographic representation of G_p, each candidate corresponds (by analogy) to a water source, denoted S_j. Morphological flooding is applied to G_p starting from the lowest water source and ending when the retinal background is reached. It is indeed hypothesized that when the flooding reaches the retinal background intensity, the catchment basins degenerate and no longer contextually represent a red lesion. The last flooding level, denoted I_{stop}, is set experimentally as described in Section III-C.

At each flooding level i, pixels that are adjacent to a water source S_j and lower than the flooding level i are added to the catchment basin of S_j, denoted $B_i^{S_j}$. When two basins merge, they start sharing the same pixels and thus the same attributes. We implement the image flooding using hierarchical queues [25].

At each flooding level i, for each candidate S_j, six shape attributes are computed on the catchment basin $B_i^{S_j}$:

- Relative area ($Rarea$): number of pixels in $B_i^{S_j}$, divided by the total number of pixels in the ROI.
- Elongation ($Elong$): $1 - W/L$ with W and L the width and length, respectively, of the bounding box of $B_i^{S_j}$ oriented along its major axis.
- Eccentricity (Ecc): $\sqrt{(L^2 - W^2)/L^2}$ with W and L the width and length, respectively, of the bounding box of $B_i^{S_j}$ oriented along its major axis.
- Circularity ($Circ$): ratio of the area of $B_i^{S_j}$ over its squared perimeter and multiplied by 4π.
- Rectangularity ($Rect$): ratio of the area of $B_i^{S_j}$ over the area of its bounding box oriented along its major axis.
- Solidity (Sol): ratio of the area of $B_i^{S_j}$ over the area of its convex hull.

Six curves are obtained for each candidate, one per shape attribute: $Rarea(i)$, $Elong(i)$, $Ecc(i)$, $Circ(i)$, $Rect(i)$ and $Sol(i)$ as a function of the flooding level i. A linear least-squares regression is performed on each curve. The order K of the polynomial linear fit is set experimentally as described in Section III-C. Depending on the order of the fit, the number of regression parameters is equal to $K + 1$. These parameters and the root mean squared error of the fit constitute the first $K + 2$ features. The last 2 features are the mean and median values over the flooding levels. In total, $6 \times (K + 4)$ DSFs are computed for each candidate.

Fig. 3 illustrates the DSF computed for 4 candidates (S_1 to S_4), taken from an image with signs of DR (Fig. 3(a)). Candidates S_1 and S_4 correspond to vessel segments, S_1 being on a relatively large segment while S_4 is part of a third order segment close to the fovea. Candidate S_2 corresponds to a medium HE, and candidate S_3 to a MA. Fig. 3(b) illustrates the catchment basins at each flooding level, starting from the candidate region and rising progressively toward the image's mean intensity, equal to 110 in this example. Each evolving layer encompasses the previous ones. At each level, the shape features, for example $Elong$, $Circ$ and Sol, are computed on the catchment basins and reported on the graph of Fig. 3(c) together with their 1st order linear fit. The elongation function $Elong(i)$ reveals that candidates S_1 and S_4 evolve in a more elongated manner than candidates S_2 and S_3, with larger values for their linear fit intercepts. Indeed, the catchment basins of these vessel segments extend anisotropically, following the direction of the vessel. By contrast, the circularity functions $Circ(i)$ reveal that candidates S_2 and S_3 evolve toward a more circular shape, with positive slopes, particularly for candidate S_3 which corresponds to a MA. Finally, the solidity function $Sol(i)$ reveals that at intensity level 40 for candidate S_1, the catchment basin appears to merge with the rest of the vasculature and drops in solidity with a negative slope. The same goes for candidate S_4 but at a higher
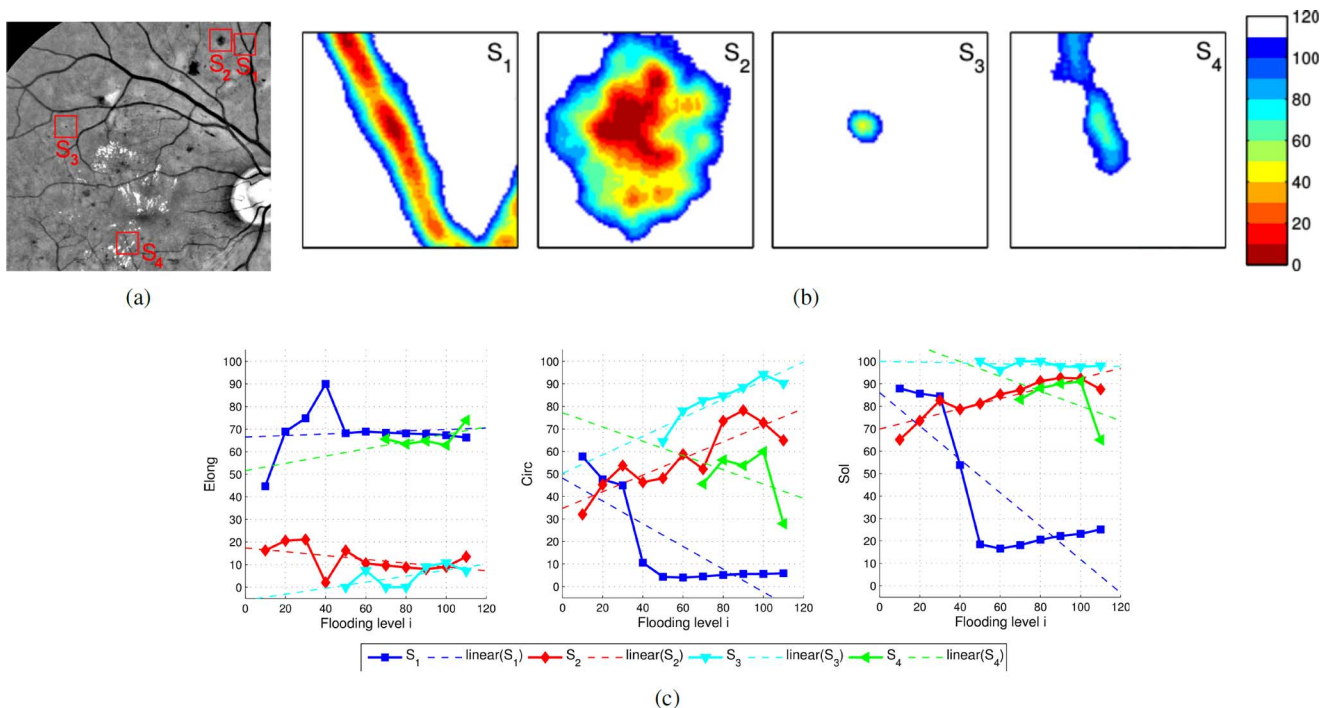
Fig. 3.   DSFs illustrated on 4 candidates extracted from the image in (a). Candidates S_1 and S_4 correspond to vessel segments, S_1 being on a relatively large segment while S_4 is part of a third order segment close to the fovea. Candidate S_2 corresponds to a medium-sized HE, and candidate S_3 to a MA. (b) shows the consecutive catchment basins from $i = 0$ to $i = 110$ according to the colormap given on the right. At each level i and for each candidate, the circularity, elongation and solidity are computed and traced in graphs as functions of i. For ease of visualization, the shape features are reported at every 10 intensity levels. (c) shows the resulting curves $Elong(i), Circ(i)$ and $Sol(i)$ for each candidate and their linear fits (dashed lines).

intensity level (100) since it is part of a much smaller and contrasted vessel than S_1. The solidities of candidates S_2 and S_3, however, either increase (positive slope) for the HE, or stay constant and high (slope close to zero and intercept of almost 100) for the MA, indicating that these structures are isolated from the other catchment basins.

Four more features, computed on the local minima, are added to the features vector: the mean red, green and blue values in I_p and the local contrast [24] in G_p.

F. Classification

To distinguish between lesions and non-lesions, we use a Random Forest (RF) classifier [26]. This powerful approach has been widely used in computer vision over the last few years, due to its numerous advantages. It is convenient for non-linear classification with high-dimensional and noisy data. It is robust against outliers and over-fitting. Moreover, it incorporates an implicit features selection step.

A RF is a combination of T decision trees trained independently using T bootstrap samples drawn with replacement from the training set. Each node is split using the best of a randomly selected subset of m features chosen, according to the decrease in the Gini index [26]. The RF returns, for each candidate, a probability of being a lesion $P(S_j)$, equal to the proportion of trees returning a positive response.

We used the MATLAB interface [27] to the RF implementation in [28]. The classifier relies on two user-defined parameters, T and m, but usually the outcome is not very sensitive

to their values [28]. We thus follow the recommendations provided in [26] and set $T = 200$ trees and $m = [\sqrt{M}]$, where $M = 4 + 6 \times (K + 4)$ is the total number of features.

III. EXPERIMENTAL SETUP

To evaluate the performance of our method on a large variety of images, we used six independent databases, the main characteristics of which are summarized in Table I. This allowed us to evaluate our method with regard to differences in image resolution, FOV, image compression and acquisition system.

Our method was evaluated at two levels. When a delineation of the lesions was provided, our method was evaluated on a per-lesion basis, meaning we analyzed its performance in detecting every single lesion. For DR grading, the per-lesion performance must be as high as possible, because the number of lesions as well as their location and type are crucial to assess DR severity level. When only a diagnosis was provided for each image, our method was evaluated on a per-image basis. This latter evaluation is more interesting from a screening point of view since it evaluates the method's performance in discriminating images with/without signs of DR.

A. Per-Lesion Evaluation

To evaluate the detection of red lesions, we performed a free-response receiver operating characteristic (FROC) analysis. This consists in computing the per-lesion sensitivity and the average number of false positives per image (FPI), for different thresholds on the output probabilities $P(S_j)$. To provide a global detection score, we computed the FROC score as proposed by Niemeijer et al. [29]. It is obtained by averaging

TABLE I
DATABASES USED FOR EVALUATION

Database	Acquisition systems	D (pxl)	FOV	Compression	Reference	Nb. Images	Public
ROCh[29]	multiple	720, 1020, 1340	45°	lossy (JPEG)	Per lesion (MA)	100	yes
Diaretdb1[30]	single	1400	50°	lossless (PNG)	Per lesion (MA, HE)	89	yes
CARA143	multiple	1400, 2200, 3240	45°	lossy (JPEG)	Per lesion (MA, HE)	143	no
Messidor[31]	multiple	870, 1340, 1400	45°	lossless (TIF)	Per image	1200	yes
Erlangen[32]	single	3240	60°	lossy (JPEG)	Per image	45	yes
CARA1006	multiple	1400, 2200, 3240	45°	lossy (JPEG)	Per image	1006	no

D: diameter of the ROI; FOV: field of view.

the sensitivities taken at 7 particular points along the FROC curve. These points correspond to 1/8, 1/4, 1/2, 1, 2, 4 and 8 false positives per image. This evaluation requires a prior manual segmentation of the lesions, to serve as a reference. Thus, we use the following three independent databases:

1) Retinopathy Online Challenge (ROCh) Database [29]: This public database comprises 50 images for training and 50 images for testing. Only the annotations of the training set are publicly provided. The evaluation on the testing dataset is performed through the challenge's website which provides a FROC curve and the corresponding score. It is important to note that only MAs are annotated in this database. Nevertheless, we found it important to evaluate our method on this database, because it has been widely used in the literature, and allows us to compare our method to existing ones [33]. The RF built using the ROCh training images is denoted RF_{roch}.

2) Diaretdb1 Database [30]: This public database comprises 28 training images and 61 testing images. As opposed to ROCh, it provides manual segmentations of both MAs and HEs. The RF built using the training images is denoted RF_{drb1}.

3) CARA143 Database: This private database consists of 143 images collected through the telemedicine platform of Diagnos Inc. (Brossard, QC, Canada). Depending on the client, images are acquired using different retinal cameras (various models of DRS, Zeiss, Topcon and Canon devices) with different resolutions (D between 1400 and 3240 pixels). The images are JPEG compressed at an estimated ratio of 12:1. All the images were analyzed by two human experts. Expert 1 segmented the images twice with an interval of one month between trials in order to assess the intra-expert variability. Expert 1's first segmentation was considered as the reference for validation. A total of 1384 red lesions were annotated, 1071 of which were MAs, 275 were blot HEs and 38 were flame HEs. Using the DR classification provided in Table II, 87 images are classified as R0, 16 as R1, 22 as R2 and 18 as R3. Including 87 images free of DR in the dataset ensured that the prevalence of images with red lesions was similar to the diabetic screening population. The RF built using all 143 images is denoted RF_{cara}. We evaluate our method on this database using 10-fold cross-validation to avoid any dependence on the partitioning of the dataset.

B. Per-Image Evaluation

It takes at least one lesion in an image for the retina to be considered as having signs of DR, while the absence of lesions indicates a healthy retina. Since the output of our red lesion detection method is a set of candidates with a probability associated to each one, we pose that the probability $P(I)$ of image I

TABLE II
RETINOPATHY GRADING IN MESSIDOR DATABASE

Grade	Description	Nb Images
R0	$(\text{N}_{\text{MA}}=0)$ AND $(\text{N}_{\text{HE}}=0)$	546
R1	$(0<\text{N}_{\text{MA}}\leq 5)$ AND $(\text{N}_{\text{HE}}=0)$	153
R2	$(5<\text{N}_{\text{MA}}< 15)$ AND $(0<\text{N}_{\text{HE}}< 5)$ AND $(\text{N}_{\text{NV}}=0)$	247
R3	$(\text{N}_{\text{MA}}\geq 15)$ OR $(\text{N}_{\text{HE}}\geq 5)$ OR $(\text{N}_{\text{NV}}>0)$	254

N_{MA}, N_{HE}, N_{NV}: numbers of MAs, HEs and neovessels, respectively

of having signs of DR is given by the highest probability among its red lesion candidates:

$$P(I) = \max_{j}(P(S_j)). \tag{3}$$

If $P(I) > Th$, Th being the threshold above which a candidate region is considered a lesion, we can consider that the image contains lesions, and thus that it reveals signs of DR. For this validation, we performed a standard receiver operating characteristic (ROC) analysis by varying Th between 0 and 1. We computed the area under the ROC curve (AUC) as a global detection score. We used the three databases presented below, for which a per-image annotation was provided (DR vs. no DR). It should be noted that because no manual lesion segmentations were provided with these databases, we could not train our lesion classifier on these datasets. Instead, we used our previously built classifiers, RF_{roch}, RF_{drb1} and RF_{cara}, to identify the lesions in the new images. Although this might be seen as a disadvantage, we believe that by doing so we were able to assess the performance of our lesion detection under realistic conditions and its robustness in the face of data changes. The same approach was used for validation using the Messidor database in other studies [19], [34], [35].

1) Messidor Database [31]: This public database consists of 1200 fundus images. For each image, a DR grade is provided as well as a risk of macular edema. In this study, we used only the DR grades, which are based on the number of MAs, HEs and neovessels (see Table II).

2) Erlangen Database [32]: This public database consists of 15 images of healthy retinas, 15 images with glaucoma and 15 with DR. We merged the healthy and glaucoma images into a single class (no DR).

3) CARA1006 Database: This private database consists of 1006 images acquired in the context of a large telemedicine screening project. It is independent of CARA143. All the images in CARA1006 were analyzed by two human experts for
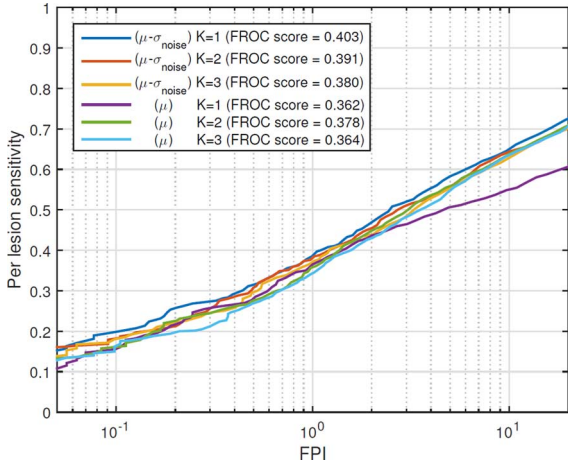
Fig. 4.   Effect of the DSF parameters on lesion detection in the CARA143 database (evaluated by cross validation).

the presence/absence of any signs of DR. The annotations provided by the senior ophthalmologist were considered as the reference for validation. A total of 304 images were labelled with DR, making the prevalence in the dataset comparable to that of actual screening situations where approximately 10% of the images exhibit signs of DR [10].

C. DSF Parameters

In a separate experiment, we evaluated the effects of the two DSF parameters, *i.e.,* the order of the linear fit K and the last flooding level I_{stop}, on the detection performance. We used $1^{st}(y = ax + b)$, $2^{nd}(y = ax^2 + bx + c)$ and $3^{rd}(y = ax^3 + bx^2 + cx + d)$ order linear regressions to model the shape attributes. We also considered two image-specific values for I_{stop} which estimates the retinal background: μ and $\mu - \sigma_{noise}$. We evaluated each pair of parameters by cross validation on CARA143.

IV. RESULTS

A. DSF Parameters

Fig. 4 illustrates the cross-validation FROC curves obtained on CARA143 for different pairs of DSF parameters. The corresponding FROC scores are shown in the legend.

Concerning the order K of the linear fit, there is no significant difference between the results of the three regressions that would justify considering one model over the others. Consequently, because the first order regression parameters (slope and intercept) facilitate the interpretation of the model, we considered in the rest of the study a first order linear regression to compute the DSFs.

Concerning the last flooding level I_{stop}, Fig. 4 shows that, for a first order linear regression, $\mu - \sigma_{noise}$ generates better results than μ (FROC score of 0.403 over 0.362). This finding reveals that $\mu - \sigma_{noise}$ is a better estimate of the retinal background intensity level. Thus, we considered $I_{stop} = \mu - \sigma_{noise}$ for the remainder of the study.

B. Per-Lesion Evaluation

The performance of our method on the ROCh test set is shown in Fig. 5(a), using the three RFs. The best detection accuracy is achieved with RF_{roch}, *i.e.,* when training the RF using images of the same database. According to the evaluation by the challenge's organizers, our method achieved an overall score of 0.420 using RF_{roch}, which ranks it in 4th place among currently published results [15]. With RF_{drb1} and RF_{cara}, the scores are 0.245 and 0.344, respectively.

Fig. 5(b) shows the FROC curves obtained by our method on the Diaretdb1 test set, again using the three RFs for classification. The overall scores obtained with RF_{roch}, RF_{drb1} and RF_{cara} are 0.146, 0.354 and 0.217 respectively.

The performance on CARA143 is illustrated in Fig. 5(c). The scores of RF_{roch} and RF_{drb1} are 0.287 and 0.241 respectively while the cross-validation achieved 0.404.

Fig. 6 shows the performance of the proposed method in detecting each type of red lesions in CARA143, together with the intra- and inter- expert variability (human experts 1 and 2 respectively). Blot HEs are the best detected lesions (FROC-score = 0.493), followed by MAs (FROC-score = 0.385). For the detection of flame HEs, the sensitivity is lower (FROC-score = 0.265). As shown on this graph, most of the lesions that are missed by the detection are close to the vessels (FROC-score = 0.210); they are either on top of or directly linked to visible vessels.

Fig. 7 shows some results of the proposed red lesion detection method on images from the CARA143 database. These results were obtained with a probability thresholding of 0.24, which corresponds to a sensitivity of 0.51 for an average of 3.3 FPI on the cross-validation's FROC curve. The proposed method is able to detect both HEs, such as in Figs. 7(a) and 7(c), and MAs, such as in Fig. 7(b). Most of the missed lesions are directly connected to large vessels such as the two HEs at the bottom of Fig. 7(c). False positive detections, occuring mostly in images without DR, correspond to tiny vessel segments located in between bright reflections of the nerve fiber layer, such as in Fig. 7(d), a frequent pattern observed in retinal images of young adults. Because these local minima are surrounded by very high intensities, even at the last flooding level, the catchment basins are very compact and the DSFs do not allow the classification to discriminate between these vessel segments and real lesions.

C. Per-Image Evaluation

Table III presents the results obtained on Messidor using the three lesion classifiers, RF_{roch}, RF_{drb1} and RF_{cara}, in comparison to the results obtained by Sanchez *et al.* [34] on the same database. RF_{cara} achieves the best performance. The sensitivity, specificity and detection percentages of R1, R2 and R3 are plotted against the probability threshold in Fig. 8(a).

When classifying images without DR against images of all DR stages, the proposed method obtained an AUC of 0.899 and a sensitivity of 93.9% at a specificity of 50%. When classifying non referable images (R0 and R1) against referable images (R2 and R3), our method achieved an AUC of 0.916 and a sensitivity of 96.2% at a specificity of 50%. Fig. 8(a) illustrates the complete ROC curves for these labelling tasks.
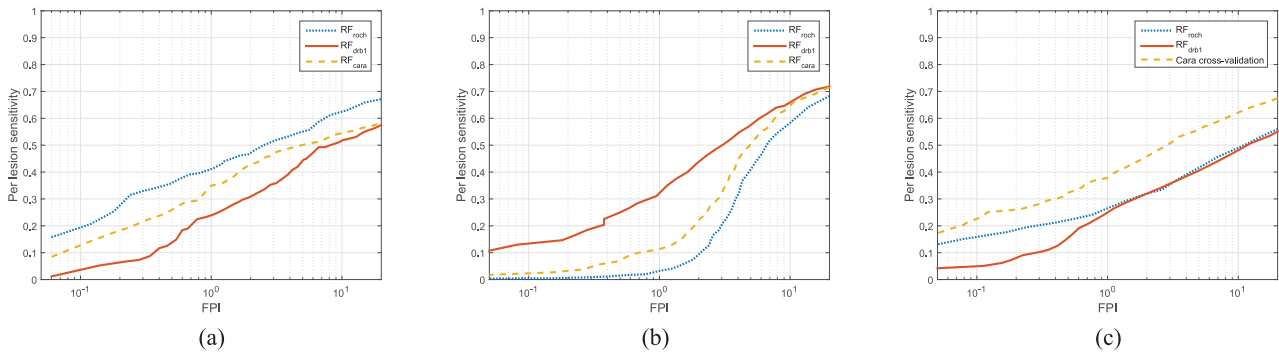
Fig. 5. FROC curves of the proposed method applied on the test sets of the 3 databases, using $\mathrm{RF_{roch}}$, $\mathrm{RF_{drb1}}$ and $\mathrm{RF_{cara}}$. (a) ROCh (b)Diaretdb1 (c) CARA143.

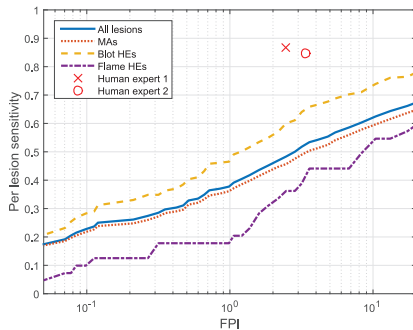Fig. 6. FROC curve per lesion type on CARA143.

On the Erlangen and CARA1006 databases, the best performance was also achieved using $\mathrm{RF_{cara}}$ with an AUC of 0.976. The ROC curves are plotted in Fig. 8(b). For the high resolution Erlangen images, at a set point of 0.74, the method obtained a sensitivity and specificity both of 93.3%. The only image with DR that was missed corresponds to a mild non proliferative DR with only two MAs: the first one is at the edge of the ROI and the second is directly connected to a large vessel. The two false positive images are glaucoma cases.

On the CARA1006 dataset, the method achieved a sensitivity of 96.1% for a specificity of 50%. The AUC is equal to 0.941. Most of the false positive images exhibit bright reflections of the nerve fiber layer, as in Fig. 7(d). Compared to the reference, the other human expert achieved a sensitivity of 92.9% and a specificity of 85.9% (see Fig. 8(b)).

V. DISCUSSION

This extensive validation of the proposed red lesion detection method demonstrates its high potential to be used as part of an automatic telemedicine DR screening system.

A. Performance in Detecting Lesions

On a per-lesion level, the method ranks 4th on the ROCh dataset and thus, has proven to be highly competitive compared to state-of-the-art approaches. The best published method on this dataset [33] combines different individual detection approaches. The second [15] and third place methods are specific to MA detection as they use prior shape knowledge. Our method allows the detection of both MAs and HEs.

On Diaretdb1, which provides reference segmentation of both MAs and HEs, our method, with the classifier trained on the same dataset, achieves better detection than the ensemble-based method in [35]. Unfortunately, Antal *et al.* [35] did not provide a FROC score. However, according to their FROC curve, the sensitivity reaches 0.6 for almost 30 FPI, compared to 6 FPI in our case.

Nevertheless, it is to be noted that for Diaretdb1, the classifier trained using the learning set provided by Diaretdb1 achieves much better results than the other to RF. We can explain this difference by noting that most of the images in Diaretb1 include 3 to 5 dark spots that resemble red lesions but are always located in the exact same position in the images. It is likely that these spots were actually caused by specks of dirt on the lens of the retinal camera. In the learning set of Diaretdb1, these spots are not marked as lesions and thus, the $\mathrm{RF_{drb1}}$ classifier, as opposed to $\mathrm{RF_{roch}}$ and $\mathrm{RF_{cara}}$, "knows" that these candidates are not actual lesions. These false positives can be easily discarded by comparing the right and left fundus images of the same patient and removing detections located at exactly the same position in both images. Unfortunately, the database does not provide the correspondence between images of the left and right eyes so we are not able to discard these false positives.

Even on the highly heterogeneous CARA143 dataset, the proposed method achieves satisfactory results, proving its robustness with respect to differences in image resolution and retinal camera. Compared to our previous paper [18] evaluated on the same CARA143 dataset (FROC score of 0.393), the addition of the adaptive contrast equalization allows the detection of more lesions located in poorly contrasted areas that were otherwise lost at the candidate extraction step. Nevertheless, the proposed method yields a sensitivity lower than that of human experts for the same FPI rate.

The results of the 3 RFs on the 3 databases demonstrate that the best performance is achieved when images of the same dataset are used for training. In fact, the ROCh training set has only MAs labeled as lesions, which explains why $\mathrm{RF_{roch}}$ is less successful on Diaretdb1 and CARA143. Conversely, $\mathrm{RF_{cara}}$, trained with both MAs and HEs marked as lesions, performs better than $\mathrm{RF_{roch}}$ on Diaretdb1 and CARA143 because both types of red lesions are marked as such in the reference sets. This observation underscores the importance of properly labelling the samples in the training set.

The results of our experiments indicate that the DSFs combined with the proposed preprocessing steps are highly discriminant. The main advantage of these shape features is that no pre-
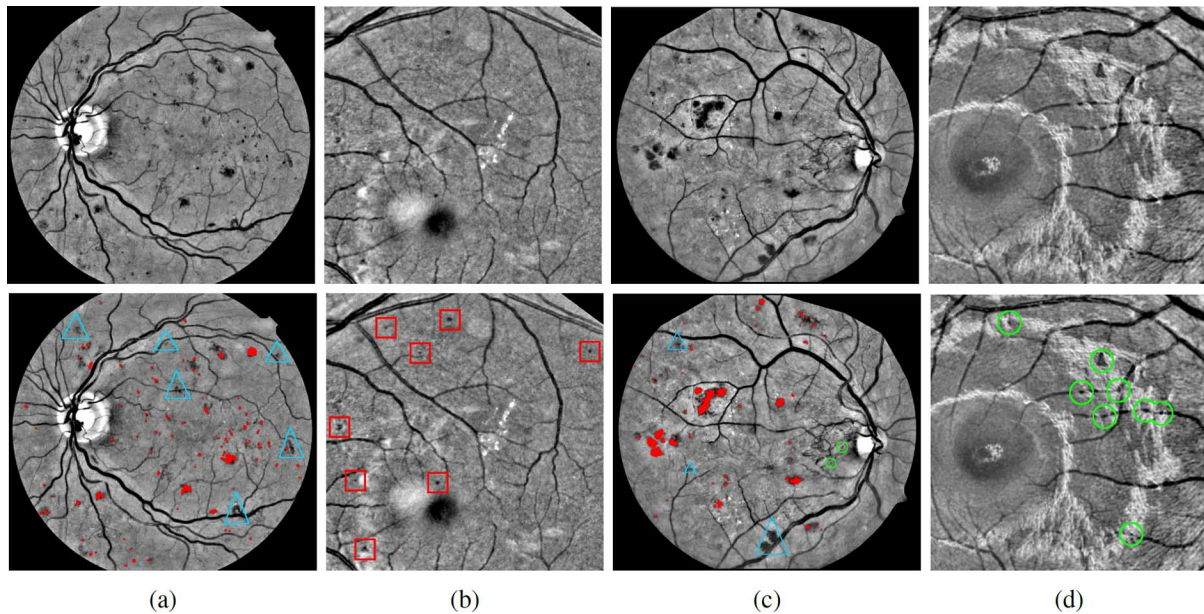
(a)   (b)   (c)   (d)

Fig. 7. Examples of lesion detection. The top row shows the G_p images. The bottom row shows the detected lesions. The probability threshold is set to 0.24, corresponding to 51% per-lesion sensitivity and an average of 3.3 FPI. True positives are filled in red or marked by a red square. False positives are circled in green. Sample false negatives are marked by blue triangles.

TABLE III
ROC RESULTS ON THE MESSIDOR DATABASE

AUC	RF_{roch}	RF_{drb1}	RF_{cara}	Sanchez [34]
R0 vs (R1,R2,R3)	0.859	0.844	**0.899**	0.876
(R0,R1) vs (R2,R3)	0.839	0.889	**0.916**	0.910
R0 vs R1	**0.768**	0.659	0.759	0.721
R0 vs R2	0.856	0.841	**0.907**	0.867
R0 vs R3	0.916	0.958	**0.974**	0.973

cise prior segmentation of the candidates is needed since feature computation is performed at different levels throughout the image flooding. This is supported by the fact that our method adapts well to changes in image resolution. Its performance on the low resolution images of ROCh ($D_{min} = 720$) is comparable to its performance on the high resolution images of CARA143 ($D_{max} = 3240$). We believe that DSFs could help in other detection applications, particularly when the objects to be detected do not show clear boundaries and are difficult to segment precisely.

To compute the DSFs, we used a linear regression to model the shape attribute functions. This model is rather basic and does not provide higher order information. Still, the root mean squared error measures the dispersion around the approximated linear model, and has a large value for functions that are not linear. Moreover, we demonstrated that there is no significant effect of the order of the linear regression on the lesion detection performance. The slope of the first order regression model quantifies the rate at which a catchment basin evolves. This property makes the DSF easier to interpret. Regarding the stop criterion for the flooding, we used an image-specific intensity level equal to $\mu - \sigma_{noise}$ which approximates the minimum value of the retinal background in the presence of noise. The illumination and contrast normalization are crucial for this phase, since

without these preprocessing steps, a global stop criterion would be difficult to select.

The proposed method proves to be very efficient for the detection of blot HEs, even more efficient than for MAs. In fact, blot HEs appear in different shapes and sizes, and no single template can represent the various geometries of these lesions. Template matching methods [12], [14], [15] were designed specifically for MA detection and were evaluated on datasets with only MA annotations, such as ROCh.

Compared to other state-of-the-art approaches [8], [9], [16], [17], our method does not rely on vessel segmentation to remove false candidates prior to classification. In fact, our novel features can distinguish candidates corresponding to segments of large and small vessels from real lesions. However, HEs that clearly appear to be linked to first or second order vessels are still missed. These account for most of the false negatives. Also, the detection of flame HEs, which are more often associated to hypertensive retinopathythan to DR, appears to be almost as challenging. Future work will focus more thoroughly on the detection of these HEs as well as neovessels, which are important features for grading.

The purpose of the per-lesion validation was to evaluate and compare our method to existing red lesion detection algorithms. However, from a clinical point of view, we believe that per-image validation is more appropriate, especially to evaluate our method for screening applications.

B. Performance in Detecting Images With Signs of DR

On the Messidor database, the proposed method, using CARA143 as a training set, outperforms existing DR detection algorithms whether they are based on lesion detection [34], [35] or on global image features [36], [37]. On subsets of the database, image classification using AM-FM texture features [36] achieves an AUC of 0.84 and a content-based image
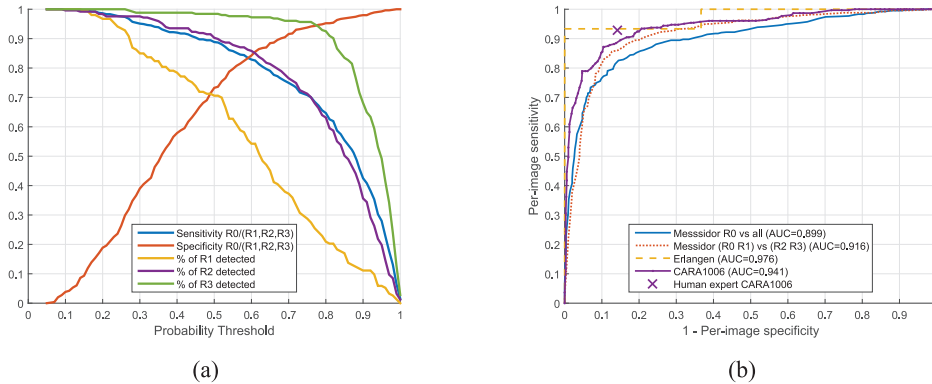
Fig. 8. Performance of RF_{cara} on Messidor (a), Erlangen and CARA1006 databases (b).

retrieval system using a multiple instance learning framework [37] achieves an AUC of 0.881. On the complete dataset, AUCs of 0.875 and 0.876 were reported using respectively an ensemble-based MA detector [35] and using a combination [34] of dark and bright lesion detection systems proposed in [10], [38].

With an AUC of 0.899 on Messidor, the proposed system achieves a performance comparable to that of the two human experts in Sanchez's study on the same dataset ($AUC_A = 0.922$ and $AUC_B = 0.865$) [34], even on the differential DR grading (R0/R1, R0/R2, R0/R3). Our system shows the lowest performance when distinguishing normal images from those with stage 1 DR (R1). However, even for experts, this appears to be the most difficult task [34]. In fact, the most common DR classifications used for clinical diagnosis, such as in [5], suggest that cases with fewer than 5 MAs and no HEs need not be referred.

When classifying Messidor images into non-referable (R0 and R1) and referable (R2 and R3) cases, our system achieves an AUC of 0.916. Previously published results of computer-assisted detection systems for referable DR, using a dataset derived from Messidor, report an AUC of 0.937 [39]. However, this performance is achieved by using the combination of images from both eyes. Unfortunately, we could not perform a direct comparison with [39] because we did not have access to their database (Messidor2).

The results obtained on the Erlangen database indicate that our method achieves good performance even on images of very high resolution. To our knowledge, no other DR detection method has been evaluated on this dataset. However, it is to be noted that the spatial calibration was adjusted to images of 45° FOV, which is not the case for the Erlangen images (FOV of 60°).

On our private database of 1006 highly heterogeneous images, for a specificity equal to that of the human expert (85.9%), the sensitivity of our method reaches 87.7%, which is comparable to that of the human expert (92.9%). Moreover, at a set point of 0.41, the sensitivity is 96.2% for a specificity of 60%. At the same set point on CARA143's FROC curve, the lesion detection achieves a sensitivity of 38% for an average of 1 FPI. This implies that even if most of the lesions are not detected, the detection method performs very well in detecting images with DR, because it takes only one lesion to be detected for an image to be considered as showing signs of DR. However, in a grading scheme, where we want to evaluate DR severity, our lesion detection method might be insufficient in its current implementation. More experiments should be conducted to evaluate the ability of the proposed method to grade images with DR.

Even though the aim of this study was the detection of red lesions, complete DR grading also requires bright lesions such as exudates and cotton wool spots to be considered. In future work, the proposed method will be adapted to finding bright lesions and tested on public databases to evaluate the potential of the proposed DSFs in different detection problems.

C. Computation Time

The overall computation time depends mostly on image resolution. Using an Intel Core i5–2400 CPU at 3.10 GHz, without parallel computing, the execution time of our software written in C++ varies between 4 seconds for the smallest ROCh image ($D = 720$) to nearly 5 minutes for the largest Erlangen image ($D = 3240$). On average, it takes 98 seconds to process an image with D in the range 2000–3000 pixels.

VI. CONCLUSION

A novel red lesion detection method based on a new set of shape features, the DSFs, was presented and evaluated on six different databases. The results demonstrate the strong performance of the proposed method in detecting both MAs and HEs in fundus images of different resolution and quality and from different acquisition systems. The method outperforms many state-of-the-art approaches at both per-lesion and per-image levels. DSFs have proven to be robust features, highly capable of discriminating between lesions and vessel segments. The concept of DSFs could be exploited in other applications, particularly when the objects to be detected do not show clear boundaries and are difficult to segment precisely. Further work focusing on bright lesion and neovessel detection will complete the proposed system and allow automatic DR grading.

REFERENCES

[1] N. Cheung, P. Mitchell, and T. Y. Wong, "Diabetic retinopathy," *Lancet*, vol. 376, no. 9735, pp. 124–36, 2010.

[2] J. Ding and T. Y. Wong, "Current epidemiology of diabetic retinopathy and diabetic macular edema," *Curr. Diabetes Rep.*, vol. 12, no. 4, pp. 346–54, 2012.

[3] J. W. Y. Yau *et al.*, "Global prevalence and major risk factors of diabetic retinopathy," *Diabetes Care*, vol. 35, no. 3, pp. 556–64, 2012.

[4] , D. Gan, Ed., *Diabetes Atlas*, 2nd ed. Brussels: Internatio, 2003.

[5] C. P. Wilkinson *et al.*, "Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales," *Ophthalmology*, vol. 110, no. 9, pp. 1677–82, 2003.

[6] C. Baudoin, B. Lay, and J. Klein, "Automatic detection of microaneurysms in diabetic fluorescein angiographies," *Revue Dépidémiologie et de Santé Publique*, vol. 32, pp. 254–261, 1984.

[7] T. Spencer, R. P. Phillips, P. F. Sharp, and J. V. Forrester, "Automated detection and quantification of microaneurysm in fluorescein angiograms," *Graefes Archives for Clinical and Experimental Ophtalmology*, vol. 230, pp. 36–41, 1992.

[8] M. Cree, J. Olson, K. McHardy, P. Sharp, and J. Forrester, "A fully automated comparative microaneurysm digital detection system," *Eye*, vol. 11, pp. 622–628, 1997.

[9] A. J. Frame *et al.*, "A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms," *Comput. Biol. Med.*, vol. 28, pp. 225–238, 1998.

[10] M. Niemeijer, B. van Ginneken, J. Staal, M. S. A. Suttorp-Schulten, and M. D. Abràmoff, "Automatic detection of red lesions in digital color fundus photographs," *IEEE Trans. Med. Imag.*, vol. 24, no. 5, pp. 584–92, May 2005.

[11] T. Walter *et al.*, "Automatic detection of microaneurysms in color fundus images," *Med. Image Anal.*, vol. 11, no. 6, pp. 555–66, 2007.

[12] G. Quellec *et al.*, "Optimal wavelet transform for the detection of microaneurysms in retina photographs," *IEEE Trans. Med. Imag.*, vol. 27, no. 9, pp. 1230–41, Sep. 2008.

[13] A. Mizutani, C. Muramatsu, Y. Hatanaka, S. Suemori, T. Hara, and H. Fujita, "Automated microaneurysm detection method based on doublering filter in retinal fundus images," in *SPIE Med. Imag. Comput.-Aid. Diagnosis*, 2009, vol. 7260, pp. 72601N–72601N-8.

[14] B. Zhang, X. Wu, J. You, Q. Li, and F. Karray, "Detection of microaneurysms using multi-scale correlation coefficients," *Pattern Recognit.*, vol. 43, no. 6, pp. 2237–2248, 2010.

[15] I. Lazar and A. Hajdu, "Retinal microaneurysm detection through local rotating cross-section profile analysis," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 400–7, Feb. 2013.

[16] C. Sinthanayothin *et al.*, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic Med. A J. Brit. Diabetic Assoc.*, vol. 19, no. 2, pp. 105–12, 2002.

[17] S. Ravishankar, A. Jain, and A. Mittal, "Automated feature extraction for early detection of diabetic retinopathy in fundus images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 210–7.

[18] L. Seoud *et al.*, "Automatic detection of microaneurysms and haemorrhages in fundus images using dynamic shape features," in *Proc. IEEE 11th Int. Symp. Biomed. Imag.*, Beijing, 2014, pp. 101–104.

[19] X. Zhang *et al.*, "Exudate detection in color retinal images for mass screening of diabetic retinopathy," *Med. Image Anal.*, vol. 18, no. 7, pp. 1026–1043, 2014.

[20] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 951–958, Aug. 2003.

[21] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, "Automated microaneurysm detection using local contrast normalization and local vessel detection," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1223–32, Sep. 2006.

[22] A. M. Mendonça, A. Sousa, L. Mendonça, and A. Campilho, "Automatic localization of the optic disc by combining vascular and intensity information," *Comput. Med. Imag. Graph.*, vol. 37, no. 5-6, pp. 409–17, 2013.

[23] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, Apr. 1993.

[24] M. Grimaud, "A new measure of contrast : The dynamics," in *SPIE Image Algebra Morphological Image Process.*, 1992, vol. 1769, pp. 292–305.

[25] N. Beucher and S. Beucher, Hierarchical Queues : General Description and implementation in MAMBA image library ÉCole des Mines de Paris, Tech. rep., 2011.

[26] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[27] Random Forest Implementation for Matlab [Online]. Available: https://code.google.com/p/randomforest-matlab/

[28] A. Liaw and M. Wiener, "Classification and regression by random forest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[29] M. Niemeijer *et al.*, "Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs," *IEEE Trans. Med. Imag.*, vol. 29, no. 1, pp. 185–95, Jan. 2010.

[30] T. Kauppi *et al.*, The DIARETDB1 diabetic retinopathy database and evaluation protocol Tech. Rep, 2007.

[31] E. Decencière *et al.*, "Feedback on a publicly distributed image database: The Messidor database," *Image Anal. Stereology*, vol. 33, pp. 231–234, 2014.

[32] J. Odstrcilik *et al.*, "Retinal vessel segmentation by improved matched filtering: Evaluation on a new high-resolution fundus image database," *IET Image Process.*, vol. 7, no. 4, pp. 373–383, 2013.

[33] E. Trucco *et al.*, "Validating retinal fundus image analysis algorithms: Issues and a proposal," *Invest. Ophthalmol. Vis. Sci.*, vol. 54, no. 5, pp. 3546–3559, 2013.

[34] C. I. Sanchez *et al.*, "Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data," *Invest. Ophthalmol. Vis. Sci.*, vol. 52, no. 7, pp. 4866–71, 2011.

[35] B. Antal and A. Hajdu, "Improving microaneurysm detection using an optimally selected subset of candidate extractors and preprocessing methods," *Pattern Recognit.*, vol. 45, no. 1, pp. 264–270, 2012.

[36] C. Agurto *et al.*, "Multiscale AM-FM methods for diabetic retinopathy lesion detection," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 502–12, Feb. 2010.

[37] G. Quellec *et al.*, "A multiple-instance learning framework for diabetic retinopathy screening," *Med. Image Anal.*, vol. 16, no. 6, pp. 1228–40, 2012.

[38] M. Niemeijer, B. van Ginneken, S. R. Russell, M. S. A. Suttorp-Schulten, and M. D. Abràmoff, "Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis," *Invest. Ophthalmol. Vis. Sci.*, vol. 48, no. 5, pp. 2260–2267, 2007.

[39] M. D. Abràmoff *et al.*, "Automated analysis of retinal images for detection of referable diabetic retinopathy," *JAMA Ophtalmol.*, vol. 131, no. 3, pp. 351–357, 2013.