Assessment of four neural network based classifiers to automatically detect red lesions in retinal images

María García [a,*], María I. López [b], Daniel Álvarez [a], Roberto Hornero [a]

[a] *Grupo de Ingeniería Biomédica, E.T.S. Ingenieros de Telecomunicación, University of Valladolid, Paseo Belén 15, 47011 Valladolid, Spain*
[b] *Instituto de Oftalmobiología Aplicada (IOBA), University of Valladolid, Valladolid, Spain*

ARTICLE INFO

ABSTRACT

Diabetic retinopathy (DR) is an important cause of visual impairment in industrialised countries. Automatic detection of DR early markers can contribute to the diagnosis and screening of the disease. The aim of this study was to automatically detect one of such early signs: red lesions (RLs), like haemorrhages and microaneurysms. To achieve this goal, we extracted a set of colour and shape features from image regions and performed feature selection using logistic regression. Four neural network (NN) based classifiers were subsequently used to obtain the final segmentation of RLs: multilayer perceptron (MLP), radial basis function (RBF), support vector machine (SVM) and a combination of these three NNs using a majority voting (MV) schema. Our database was composed of 115 images. It was divided into a training set of 50 images (with RLs) and a test set of 65 images (40 with RLs and 25 without RLs). Attending to performance and complexity criteria, the best results were obtained for RBF. Using a lesion-based criterion, a mean sensitivity of 86.01% and a mean positive predictive value of 51.99% were obtained. With an image-based criterion, a mean sensitivity of 100%, mean specificity of 56.00% and mean accuracy of 83.08% were achieved.

© 2010 IPEM. Published by Elsevier Ltd. All rights reserved.

1. Introduction

Diabetes mellitus is an endocrine disorder that currently affects more than 170 million people worldwide [1]. Moreover, its incidence is expected to increase to 366 million by the year 2030 [1]. Among the visual complications of diabetes, diabetic retinopathy (DR) is of particular importance because it is a leading cause of visual impairment and new-onset blindness in industrialised countries [1,2].

Treatment with laser photocoagulation can reduce the risk for blindness and moderate vision loss by more than 90% [1,3]. Additionally, some studies show that focal/grid laser treatment can reduce macular oedema in some patients [4] more effectively and with fewer side effects than newer treatments with corticosteroid injections [5]. However, early detection of the disease is essential to prevent sight damage. The problem inherent in DR is that the patient is not aware of the disease until the changes in the retina have progressed to a level that treatment will tend to be less effective [6]. Therefore, early diagnosis through regular screening should be attempted [7]. Different strategies have been employed for screening [2]. In this study, digital colour photographs of the retina were analysed. These images are a sensitive and specific means of detecting the early signs of DR when examined by expert readers [8,9].

With the growing incidence of diabetes, the number of images that need to be examined by ophthalmologists can be prohibitively large [7]. Thus, automatic analysis of retinal images to detect clinical signs of DR could be an important help in the diagnosis and screening of the disease.

DR clinical signs include red lesions (RLs), such as microaneurysms (MAs) and haemorrhages (HEs), and bright lesions like hard exudates (EXs) and cotton-wool spots (CWs) [9]. In previous studies we attempted the automatic identification of EXs in retinal images [10,11]. This paper focuses on RLs detection. They appear as reddish patches in retinal images, very similar in colour to blood vessels and retinal background [12]. MAs are usually small and circular, while HEs have varying sizes and shapes [13]. RLs are among the first unequivocal signs of DR [7]. Therefore, their detection would be an important contribution to the screening tasks and serve as a first step to the complete follow-up of DR.

The importance of RLs detection has been established in previous studies. In some works, fluorescein angiographic images were analysed [14,15], as RLs appear more contrasted with the background [7]. However, the risk associated with the dye injection needed in angiography motivated the development of RLs detection techniques in retinal images [7,12,13,16–21].

* Corresponding author. Tel.: +34 983423983; fax: +34 983423667.
 E-mail address: maria.garcia@tel.uva.es (M. García).

Some of the studies developed for RLs detection in colour fundus images were based on the use of NNs. Gardner et al. [17] used a MLP to detect RLs in subimages of size 20×20 pixels. The feature vectors were constructed using the intensity of the pixels in the subimages. Other authors [18] determined if an image contained RLs using SVM. The intensities of the pixels in regions of size 15×15 pixels formed the feature vectors. In [12], the reddish regions in the image were segmented using the *moat* operator. Then, a NN was used to separate RLs and blood vessels. The features used for classification and the type of NN were not specified.

Studies that do not use NNs to detect RLs can also be found in the literature. Grisan and Ruggeri [16] detected the areas of the image where dark pixels appeared with a high density. Authors stated the need of including a classifier to distinguish among true RLs and noise. In [7] several classifiers (linear discriminant, quadratic discriminant and kNN) were studied. They used 21 features to separate RLs from retinal background. The kNN classifier, as well as a bayesian and a Mahalanobis classifiers, were analysed by Ege et al. [13]. The candidate regions were represented by 31 features to separate RLs and bright lesions. Other authors [19] used a bayesian classifier to segment dark and bright regions in retinal images. Linear discriminant analysis was subsequently applied to determine if a dark region was a RL. In this work, 30 features were selected for the classification of RLs. Walter et al. [20] propose a method for microaneurysm detection with four steps: preprocessing, candidate detection, feature extraction and classification using a scheme based on kernel density estimation with variable bandwidth. In this study, a set of 15 features was extracted from candidate lesions. Quellec et al. [21] also detected microaneurysms, using a method based on wavelet transforms in different subbands. The proposed method was applied to different image modalities, including colour fundus photographs.

In this study, four neural network (NN) based classifiers were studied as part of a RLs detection method. Three of them were individual NNs: multilayer perceptron (MLP), radial basis function (RBF) and support vector machine (SVM). The last classifier was constructed as a combination of the previous three using a majority voting (MV) schema. The aim of this study was to assess the ability of these pattern recognition techniques in RLs detection.

2. Database of retinal images

The database used in this study contained 115 images with variable colour, brightness and quality. The images were provided by the "Instituto de Oftalmobiología Aplicada" (IOBA) of the University of Valladolid, Spain. 90 images were captured using a TopCon TRC-50IX mydriatic digital retinal camera at $50°$ field-of-view (FOV). The remaining 25 images were captured with a TopCon TRC-NW6S non-mydriatic digital retinal camera at a FOV of $45°$. Image resolution was 576×768 pixels in 24 bit JPEG format. The RLs in these images were manually marked by an ophthalmologist.

Images came from a clinical set of diabetic patients who were referred to the ophthalmologist for further examination. The images belonged to patients who suffered from mild to moderate nonproliferative DR or to healthy retinas, according to the ophthalmologist. 90 images contained RLs. In the remaining 25 images, the ophthalmologist did not mark any RLs. The 115 images were randomly divided into a training set and a test set:

- The training set contained 50 images with RLs. From this image set, we extracted the regions that were used to perform feature selection and to train the NNs.
- The test set was composed of the remaining 65 images (40 with RLs and 25 without RLs). It was used to validate the effectiveness of the complete algorithm.

3. Methods

3.1. Image preprocessing

The images that appear in a clinical environment are normally variable. The physical features of patients, such as race and iris colour, result in wide variations in the colour of the fundus [22]. Besides, the image acquisition process can produce intra-image variations of luminosity and contrast. Effects like patient movement, poor focus, bad positioning, reflections or inadequate illumination can cause a significant proportion of images to be of such poor quality as to interfere with analysis [23]. These factors need to be minimised before starting the detection of retinal structures or lesions.

In this study, we applied the method proposed by Foracchia et al. [24] in order to normalise the images and increase the contrast between RLs and retinal background [19,24]. Let I be the observed image. The method calculated the original image, $\hat{I}^o$, undistorted by luminosity and contrast variations, as:

$$\hat{I}^o(x, y) = \frac{I(x, y) - \hat{L}(x, y)}{\hat{C}(x, y)}, \tag{1}$$

where $\hat{L}(x, y)$ and $\hat{C}(x, y)$ are estimates of the luminosity and contrast drifts of the image, respectively [24]. They were calculated from the background image, I_b, consisting of the pixels of I that did not belong to lesions or retinal structures. A pixel (x,y) was considered to belong to the background if its intensity was close to the mean intensity in a neighbourhood N centred in (x,y). The similarity measure was the Mahalanobis distance, d_M. Mathematically, this is expressed as [24]:

$$(x, y) \in I_b \quad \text{if} \quad d_M = \left| \frac{I(x, y) - \hat{\mu}_N}{\hat{\sigma}_N} \right| < t \tag{2}$$

where $\hat{\mu}_N$ is the estimated mean value inside N, $\hat{\sigma}_N$ is the estimated standard deviation inside N and t is a threshold value. The estimation of $\hat{L}(x, y)$ and $\hat{C}(x, y)$ required a similar computation of the mean and standard deviation of the intensity values inside a neighbourhood N centred in (x,y). However, for these calculations, only the pixels belonging to I_b were considered.

In order to reduce the computational burden when computing background and drifts, the images were partitioned into square blocks. In this work, a value of $s = 100$ pixels was experimentally chosen, based on a visual inspection of the results for the images in the training set. Besides, we chose $t = 1$. This value ensures that an adequate percentage of pixels in each block were considered as background pixels [24]. The process assumed that at least 50% of the pixels in each block belonged to the background. For the blocks where this condition was not met, we applied an averaging technique [11]. In order to avoid border effects in the image processing, we also created a circular mask for each image that limited the FOV area. The radius of the circle was determined analysing the intensity profile along one image diagonal in the red component. The intensity transitions are more abrupt in the FOV boundary, which permits an estimation of the diagonal of the FOV. Then, a Canny edge detector was applied to extract FOV border pixels. The circular Hough transform with fixed radius was finally applied to the set of border pixels in order to determine the circle centre [25].

Observed image normalisation yielded a colour image with components $\hat{R}^o \hat{G}^o \hat{B}^o$. An example of observed image is shown in Fig. 1(a and b). After normalization, we obtained Fig. 1(c).

3.2. Candidate segmentation

The aim of this stage was to coarsely separate the reddish regions in the image, which were candidate RLs. Segmenta-
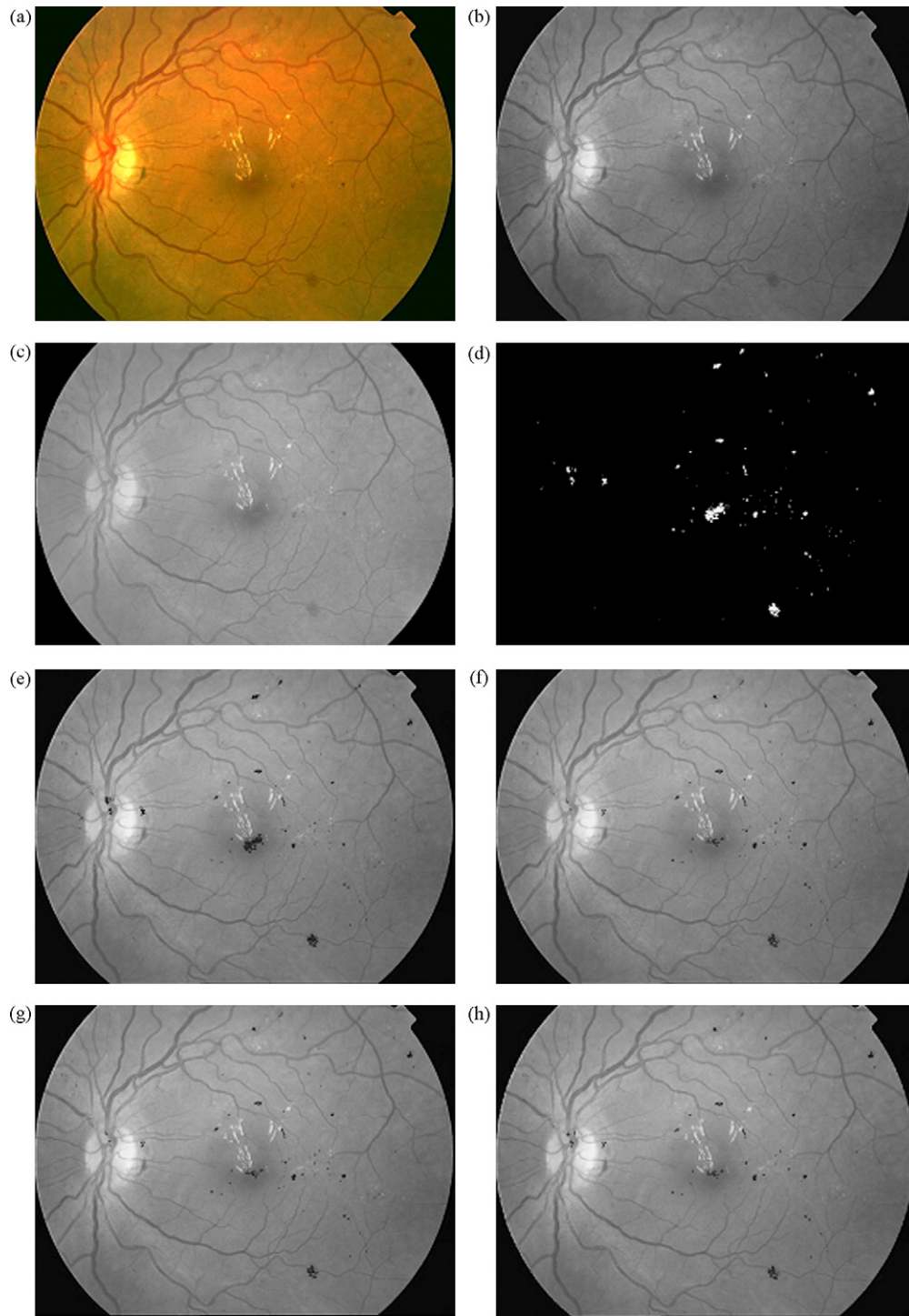
Fig. 1. (a) Original colour image. (b) Greyscale version of the original retinal image. (c) Greyscale version of the image after contrast and luminosity normalization. (d) Candidate RLs regions obtained after segmentation. (e) Result obtained after classification with MLP. RLs are depicted in black and superimposed on a greyscale version of the original image. (f) Result obtained after classification with RBF. RLs are depicted in black and superimposed on a greyscale version of the original image. (g) Result obtained after classification with SVM. RLs are depicted in black and superimposed on a greyscale version of the original image. (h) Result obtained after classification with MV. RLs are depicted in black and superimposed on a greyscale version of the original image. (For interpretation of the references to color in the figure caption, the reader is referred to the web version of the article.)

tion was accomplished using the brightness properties of RLs in $\hat{G}^o$, where they show the highest contrast with the background [7,26]. Although the pixels belonging to RLs show variable intensities, they show lower intensities in $\hat{G}^o$ than their surrounding region.

The segmentation method in this stage is based on the work by Grisan and Ruggeri [16]. For each pixel $p = \hat{G}^o(x, y)$, located at a position (x,y) in the image, a neighbourhood I_p, centred at (x,y) and of size $M \times N$, was defined. A pixel was considered as part of a candidate RL if its intensity was lower than a fraction t_G of the mean intensity in I_p, $\mu(I_p)$. Mathematically, a pixel is considered as part of a RL if [16]:

$$p < t_G \cdot \mu(I_p), \tag{3}$$

The neighbourhood size must be small enough to capture the local properties of RLs, so residual variations of luminosity and contrast do not affect their detection. Besides, it must be large enough to cover extensive HEs. In this study, we used two neighbourhood sizes: 20×20 and 50×50 pixels. Our experiments showed that these values ensured an adequate detection of RLs of different sizes. The choice of the neighbourhood sizes is not critical, but should be adequately chosen to cover large lesions as well as to allow the detection of small RLs. As a consequence, the optimal values would probably be different for different image sizes. We also varied t_G between 0.6 and 1 in steps of 0.01. We experimentally determined that $t_G = 0.98$ ensured that most RLs were detected while some noisy regions were eliminated. The two resulting images were combined using the morphological operation of reconstruction by dilation [27]. The blood vessels were also masked out using the method described in [28] to reduce the computational cost of the classification stage. This method is based on the combination of differential filters with morphological operators. The blood vessels with different widths and at different orientations were detected by means of a three stage process. The first one was a preprocessing phase. In this stage, the background of the image was normalised and a set of directional filters in the four main orientations of $0°$, $45°$, $90°$ and $135°$ was applied to enhance the thin vessels. In the second stage of the algorithm, the vessel centrelines were detected. For this task, a set of candidate points was extracted after the application of four directional Difference of Offset Gaussians filters. The candidate points were subsequently connected to form segments by a region growing process based on statistical measures. Each candidate segment was then confirmed or rejected based on intensity and length measures. In the third phase, the final segmentation of blood vessels was obtained. First, vessels with different widths were enhanced by applying a top-hat transform with structuring elements of different sizes. Then, binary maps of vessels at different scales were created by morphological reconstruction. Finally, these images were combined with the centrelines extracted in the second stage. A region growing procedure was applied successively to the four scales to obtain the final segmentation of vessels [28].

The result of the candidate segmentation of Fig. 1(c) is shown in Fig. 1(d).

3.3. Feature extraction

For the classification stage, it was necessary to represent each region by a feature vector that could be used as input to the NNs. We included colour and shape features, which represented the characteristics that help ophthalmologists to visually distinguish RLs in retinal images. In this way, prior knowledge was incorporated into the classification task. The feature set contained the following 29 features [7,11,18,29,30]:

- Feature 1: Region size (a). Let Ω be the set of pixels belonging to the region. The region size is defined as the number of pixels inside the region.

$$a = \sum_{j \in \Omega} 1 \qquad (4)$$

- Features 2–4: Mean $\hat{R}^o \hat{G}^o \hat{B}^o$ values inside the region ($\mu_{\hat{R}^o}, \mu_{\hat{G}^o}, \mu_{\hat{B}^o}$). These features are calculated as follows:

$$\mu_{\hat{R}^o} = \frac{\sum_{j \in \Omega} \hat{R}^o(j)}{a}$$
$$\mu_{\hat{G}^o} = \frac{\sum_{j \in \Omega} \hat{G}^o(j)}{a} \qquad (5)$$
$$\mu_{\hat{B}^o} = \frac{\sum_{j \in \Omega} \hat{B}^o(j)}{a}$$

- Features 5–7: Standard deviation of the $\hat{R}^o \hat{G}^o \hat{B}^o$ values inside the region ($\sigma_{\hat{R}^o}, \sigma_{\hat{G}^o}, \sigma_{\hat{B}^o}$). They are calculated as:

$$\sigma_{\hat{R}^o} = \frac{\sqrt{\sum_{j \in \Omega} [\hat{R}^o(j) - \mu_{\hat{R}^o}]^2}}{a}$$
$$\sigma_{\hat{G}^o} = \frac{\sqrt{\sum_{j \in \Omega} [\hat{G}^o(j) - \mu_{\hat{G}^o}]^2}}{a} \qquad (6)$$
$$\sigma_{\hat{B}^o} = \frac{\sqrt{\sum_{j \in \Omega} [\hat{B}^o(j) - \mu_{\hat{B}^o}]^2}}{a}$$

- Features 8–10: Mean $\hat{R}^o \hat{G}^o \hat{B}^o$ values around the region ($\mu_{\hat{R}^o}^a, \mu_{\hat{G}^o}^a, \mu_{\hat{B}^o}^a$). They are obtained averaging the intensity of the pixels belonging to a rectangular region, Γ, centred on the object and with a width of 5 pixels. The pixels of Ω are excluded from this region.

$$\mu_{\hat{R}^o}^a = \frac{\sum_{j \in \Gamma} \hat{R}^o(j)}{\sum_{j \in \Gamma} 1}$$
$$\mu_{\hat{G}^o}^a = \frac{\sum_{j \in \Gamma} \hat{G}^o(j)}{\sum_{j \in \Gamma} 1} \qquad (7)$$
$$\mu_{\hat{B}^o}^a = \frac{\sum_{j \in \Gamma} \hat{B}^o(j)}{\sum_{j \in \Gamma} 1}$$

- Features 11–13: Standard deviation of the $\hat{R}^o \hat{G}^o \hat{B}^o$ values around the region ($\sigma_{\hat{R}^o}^a, \sigma_{\hat{G}^o}^a, \sigma_{\hat{B}^o}^a$). Defined as:

$$\sigma_{\hat{R}^o}^a = \frac{\sqrt{\sum_{j \in \Gamma} [\hat{R}^o(j) - \mu_{\hat{R}^o}^a]^2}}{\sum_{j \in \Gamma} 1}$$
$$\sigma_{\hat{G}^o}^a = \frac{\sqrt{\sum_{j \in \Gamma} [\hat{G}^o(j) - \mu_{\hat{G}^o}^a]^2}}{\sum_{j \in \Gamma} 1} \qquad (8)$$
$$\sigma_{\hat{B}^o}^a = \frac{\sqrt{\sum_{j \in \Gamma} [\hat{B}^o(j) - \mu_{\hat{B}^o}^a]^2}}{\sum_{j \in \Gamma} 1}$$

- Features 14–16: $\hat{R}^o \hat{G}^o \hat{B}^o$ values of the region centre ($\hat{R}_C^o, \hat{G}_C^o, \hat{B}_C^o$). Computed as the intensity values of the pixel that corresponds to the centre of mass of the region in the three colour channels.
- Feature 17: Region compactness (C). After calculating the perimeter of the region, p, using the inner boundary tracing algorithm described in [31], C is calculated as:

$$C = \frac{p^2}{a} \qquad (9)$$

- Feature 18: Region edge strength (ES). Defined as:

$$ES = \frac{\sum_{j \in B} I_{prewitt}(j)}{\sum_{j \in B} 1} \qquad (10)$$

where B is the set of boundary pixels identified by the tracing algorithm used previously to calculate the perimeter and $I_{prewitt}$ is the result of applying the Prewitt operator to the green channel of the image.
- Features 19–21: Homogeneity of the region, measured as the Shannon's entropy of the $\hat{R}^o \hat{G}^o \hat{B}^o$ values inside the region ($H_{\hat{R}^o}, H_{\hat{G}^o}, H_{\hat{B}^o}$):

$$H = -\sum_{i=0}^{L-1} P(b_i) \ln[P(b_i)] \qquad (11)$$

In Eq. (11), L is the number of grey levels inside the region and $P(b_i)$ is defined as [32]:

$$P(b_i) = \frac{N(b_i)}{a} \tag{12}$$

where $N(b_i)$ is the number of pixels in the region with intensity equal to b_i.

- Features 22–24: Colour difference of the $\hat{R}^o\hat{G}^o\hat{B}^o$ values ($C^d_{\hat{R}^o}, C^d_{\hat{G}^o}, C^d_{\hat{B}^o}$). These features are computed as:

$$
\begin{aligned}
C^d_{\hat{R}^o} &= \frac{\mu_{\hat{R}^o}}{\mu^a_{\hat{R}^o}} \\
C^d_{\hat{G}^o} &= \frac{\mu_{\hat{G}^o}}{\mu^a_{\hat{G}^o}} \\
C^d_{\hat{B}^o} &= \frac{\mu_{\hat{B}^o}}{\mu^a_{\hat{B}^o}}
\end{aligned}
\tag{13}
$$

- Feature 25: Region circularity (Ci). Defined as:

$$Ci = \frac{p^2}{4 \cdot \pi \cdot a} \tag{14}$$

- Feature 26: Region length (l). Length (in pixels) of the major axis of the ellipse with the same normalized second central moments as the region that is being analysed.
- Feature 27: Region width (w). Length (in pixels) of the minor axis of the ellipse with the same normalized second central moments as the region that is being analysed.
- Feature 28: Aspect ratio of the region (r), measured as the relationship between region length and width:

$$r = \frac{l}{w} \tag{15}$$

- Feature 29: Region eccentricity (e). Defined as the eccentricity of the ellipse that has the same second central moments as the region.

$$e = \sqrt{1 - \frac{w^2}{l^2}} \tag{16}$$

3.4. Feature selection

Classification errors tend to increase with the number of features. Besides, the structure of a NN with many input nodes is more difficult to interpret [33]. These problems can be avoided by selecting the subset of features that is most useful for a certain task [34].

We focused on two statistical model-building strategies, discriminat analysis (DA) and logistic regression (LR). These methods have been widely used for feature selection and data analysis. Unlike LR, DA requires the independent variables to be multivariate normal and homocedastic [35]. In order to decide whether to apply DA or LR, we analysed the normality and homocedasticity of our data under SPSS 14.0®. The following statistical tests were used:

- Kolmogorov–Smirnov test with Lilliefors significance correction and Shapiro–Wilk test for normality.
- Levene's test for homocedasticity.

The results of these tests showed that the normal distribution and homocedasticity could not be verified for all the variables under study. Therefore, LR is more adequate in our case. In this work, LR was used to describe the relationship between a dichotomous dependent variable (the class of a candidate region, RL or non-RL) and 29 independent variables (the extracted features). Model selection was accomplished using a stepwise selection method, where a test for backward elimination followed by a forward selection procedure was performed. The backward elimination procedure was based on the likelihood ratio test. This test performs better than other options like the Wald test. We used the *score* test in the forward selection procedure due to its lower computational cost [34].

In addition, we verified that no significant features were eliminated. For this task, we compared the complete model (including the 29 extracted features) with the reduced model (with only the selected features) using a likelihood ratio test [34]:

$$\lambda = -2 \ln \left[\frac{\text{likelihood of the reduced model}}{\text{likelihood of the complete model}} \right] \tag{17}$$

Let k be the number of variables excluded in the reduced model. Then, λ has a chi-square distribution with k degrees of freedom. This property can be used to obtain a p-value associated to this statistical test. If $p[\chi^2(k) > \lambda] > 0.05$, we can assume that the k variables excluded from the model were not significant for the classification task.

3.5. Classification using NNs

We used several classifiers based on NNs to separate the true RLs from non-RL candidate regions. According to our previous research [10,11] NNs could be useful in the detection of bright lesions associated to DR. In this study, we propose to apply NNs to detect also RLs in retinal images.

In this section, we include a brief description of each classifier and the implementation details for this work. A comprehensive explanation of the NNs used in this work can be found in [36,37].

3.5.1. Multilayer perceptron (MLP)

MLP is an important class of NNs that can represent nonlinear mappings between a set of input variables and a set of output variables. It has been proven [38,39] that a single hidden layer MLP with any bounded nonconstant activation function or any arbitrary bounded activation function which has unequal limits at infinities can form decision regions with arbitrary shapes in multidimensional classification problems. For this reason, we constructed a 3-layer MLP, for which the number of hidden units was experimentally determined. As hidden neuron activation function, we chose the hyperbolic tangent sigmoid function (*tanh*), an antisymmetric function in the interval $(-1, 1)$. Sigmoidal functions satisfy the constraints in [38,39]. In addition, *tanh* improves the learning speed of MLP. In the output layer, we used the logistic sigmoid (*logsig*) activation function, which is also a sigmoidal function defined in $(0, 1)$. This choice permits the interpretation of network outputs as posterior probabilities. If the output is above 0.5 for some input vector, the probability of the corresponding candidate region of being a RL is greater than that of being a non-RL. Therefore, the output threshold was set to 0.5.

Training a NN can be viewed as the minimisation of an error function. The performance of MLP can be improved if suitable error functions and minimisation algorithms are chosen. We used a cross-entropy error function because it simplifies the optimisation process when the output layer activation function is *logsig*. Regarding the minimisation algorithm, we selected the scaled conjugate gradients algorithm, which generally shows faster convergence than gradient descent-based techniques or even conventional conjugate gradient algorithms. In addition, we employed a weight decay regulariser to avoid overfitting and improve generalisation. The regularisation parameter, ν, was experimentally settled.

An example of the classification with MLP is shown in Fig. 1(e).

3.5.2. Radial basis function (RBF)

RBF networks are 3-layer universal approximators. The input layer receives the feature vectors. The neurons in the only hidden

layer are nonlinear and have Green activation functions. They perform a nonlinear transformation from the input space into a high dimensional space, where the patterns are more likely to be linearly separable. The output layer is linear, and supplies the response of the NN.

Native RBFs are constructed with as many hidden neurons as the available training points. This is computationally demanding and may lead to poor generalisation. Generalised RBF networks represent a suboptimal solution that, usually, requires less hidden neurons. In the generalised RBF NN (hereafter, RBF networks), the training involves not only learning the weights, but also the number of radial basis functions and their parameters. It can be regarded as a two-stage process. The first stage involves determining the parameters of the basis functions. In the second stage, the basis functions are kept fixed while the weights connecting the hidden and the output layers are found.

In this study, we chose Gaussian functions as hidden layer activation functions, the most common Green function employed in practice. They are characterised by their centre and *spread* parameter, which represents the width of the Gaussian. For training, we applied the orthogonal least squares algorithm [40]. The optimum *spread* and number of radial basis functions were experimentally determined. The output threshold was set to 0 because the output of RBF is positive for RLs and negative for non-RLs.

An example of classification using RBF can be seen in Fig. 1(f).

3.5.3. Support vector machine (SVM)

SVMs are also 3-layer NNs that can approximate any function. The only hidden layer in the network performs a nonlinear transformation from the input space into a high dimensional space, where a linear decision boundary in the form of optimal hyperplane is constructed. When the classes are not linearly separable, like in our case, the optimal hyperplane is the one that minimises the probability of classification error.

The optimal hyperplane can be determined using the method of Lagrange multipliers. Given the training set, $\left\{(x_i, d_i)\right\}_{i=1}^N$, find the Lagrange multipliers, $\left\{\alpha_i\right\}_{i=1}^N$, that maximise the objective function [36]:

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N \alpha_i\alpha_j d_i d_j K(x_i, x_j), \tag{18}$$

subject to the constraints:

$$\sum_{i=1}^N \alpha_i d_i = 0, \tag{19}$$

$$0 \leq \alpha_i \leq C \quad \text{for} \quad i = 1, \ldots, N, \tag{20}$$

where (x_i, d_i), $i = 1, \ldots, N$ are input vector-desired response pairs. $K(x_i, x_j)$ is called inner-product kernel. The constant C controls the trade-off between the complexity of the NN and number of non-separable points.

In this study, we used a Gaussian inner-product kernel. This is the most common type of kernel, as it is rotationally and translationally invariant. Two parameters needed to be settled: the width of the Gaussian functions (σ) and C. The output given by the implementation of SVM used in this work is the class label of the region represented by the input feature vector [41]. Therefore, it was not necessary to choose an output threshold. An example of classification with SVM is depicted in Fig. 1(g).

3.5.4. Majority voting (MV)

After the configuration and training of the three individual NNs, we studied if the results could be improved by combining the decisions made by these classifiers. We used a static structure, where the responses of several predictors are combined by means of a mechanism that does not involve the input signal. The simplest static combination of experts is a MV scheme. The patterns in the test set were simultaneously presented to the input of the trained MLP, RBF and SVM NNs. Each pattern was classified by MV in the class that at least two of the three individual NNs had chosen. The output of the MV classifier for the example image can be seen in Fig. 1(h).

4. Results

4.1. Performance evaluation

We used the 50 images in our training set to determine the parameters of the algorithm and the 65 images in our test set to assess the performance of the proposed method.

Firstly, the significant features needed to be identified by means of stepwise LR. To assess the adequacy of the selected features, the accuracy obtained with the complete and reduced models was compared and a likelihood ratio test was performed.

The selected features were computed for 1710 RL and 1710 non-RL regions obtained after the segmentation of the 50 training images. These feature vectors formed the training examples. The inputs were normalized (mean = 0, standard deviation = 1) and the data points were randomly presented to the classifiers [36]. Our goal was to find, for each NN, the architecture that led to a minimum prediction risk [42]. We used 10-fold cross-validation to assess the generalization ability of the networks as it is an estimator of the prediction risk [42]. For each parameter combination, we measured the mean sensitivity (SE_{val}), mean specificity (SP_{val}) and mean accuracy (A_{val}) over the validation set. We used A_{val} for model selection as it represents better the trade-off between sensitivity and specificity [10].

After the method parameters were determined, the diagnostic accuracy of the proposed method on the test set was assessed. We compared the results obtained for the four classifiers in terms of a lesion-based criterion (pixel resolution) and an image-based criterion [29]:

- Using a lesion-based criterion, we computed the mean sensitivity (SE_{les}) and positive predictive value (PPV_{les}). SE_{les}, measured the mean number of correctly detected RLs pixels in the test images. For the lesion-based criterion, specificity is not an informative measure. In our images, the number non-RLs pixels is much higher than the number of RLs. Therefore, even if RLs detection is inaccurate, specificity values would be high. PPV_{les}, which accounts for the probability that a detected region is really a RL, was regarded as a more significant measure to evaluate the system [11,29].
- The image-based criterion measured the ability of the algorithm to detect pathological images and separate them from healthy ones on the basis of the presence or absence of RLs. We calculated the mean sensitivity (SE_{im}), mean specificity (SP_{im}) and mean accuracy (A_{im}). To improve SE_{im} without compromising the SP_{im}, we identified those images where less than 30 pixels had been detected as RLs and considered them as belonging to healthy retinas, following the approach in [29].

4.2. Selected features

We computed the 29 features described in Section 3.3 for 1710 RL regions and 1710 non-RL regions extracted after the segmentation of 50 images (training set) from our database. As the normality and homocedasticity hypothesis could not be verified for all the

Table 1
Results of fitting a logistic regression model to the complete and reduced feature sets.

Model	−2 log likelihood	Accuracy (%)
Complete (29 features)	2428.84	84.60
Reduced (16 features)	2447.72	84.50

variables under study, we used SPSS 14.0® to perform LR on these data.

The reduced model included the following 16 features: $\mu_{\hat{G}^o}$, $\mu_{\hat{B}^o}$, $\sigma_{\hat{G}^o}$, $\sigma_{\hat{B}^o}$, $\mu_{\hat{C}^o}^a$, $\sigma_{\hat{C}^o}^a$, $\sigma_{\hat{B}^o}^a$, $\hat{G}_C^o$, ES, $H_{\hat{R}^o}$, $H_{\hat{G}^o}$, $H_{\hat{B}^o}$, $C_{\hat{R}^o}^d$, $C_{\hat{G}^o}^d$, l and r.

Although LR was not used in this work as a classifier, the performance of the complete (29 extracted features) and reduced (16 selected features) LR models can be used to gain insight into the importance of the selected variables. The results of fitting a LR model to the complete and reduced feature sets are summarised in Table 1. According to the data in this table, $\lambda = 18.88$. The number of variables excluded from the model is $k = 13$. This yields a p-value of $p[\chi^2(13) > 18.88] = 0.13$. Since the p-value is greater than a significance level of 0.05, the reduced model did not lose significant information. In addition, Table 1 shows that the classification accuracy of the LR fitted models was similar in both cases. Results in Table 1 were obtained over the 3420 feature vectors described in this section. These were 29-dimensional in the case of the complete model and 16-dimensional for the reduced model.

4.3. Training

The aim of the training phase was to adjust the free parameters of each NN (MLP, RBF and SVM). In this stage, we used the selected 16 features computed for the 1710 RL regions and 1710 non-RL extracted from the training set.

4.3.1. MLP

For MLP, we had to determine the number of hidden neurons and the regularization parameter, ν. We varied the number of hidden neurons from 10 to 60 in steps of 1, while ν was varied from 1 to 10 in steps of 0.1. The maximum values of A_{val} were obtained for $\nu = 4.5$. Besides, we chose a network with 28 hidden units because A_{val} did not significantly improve for larger architectures. At this operating point we obtained $SE_{val} = 84.15\%$, $SP_{val} = 88.48\%$ and $A_{val} = 86.32\%$.

4.3.2. RBF

The optimum *spread* and number of radial basis functions had to be settled. In RBF, the output of a hidden neuron decreases as the distance between the input pattern and the centre of it corresponding radial basis function increases. Therefore, the centres and widths of the radial basis functions determine the receptive field of each neuron [10,11]. To ensure that the whole feature space is adequately covered, we varied both parameters jointly. The number of RBFs ranged from 10 to 150 in steps of 1. The *spread* parameter varied from 1 to 10 in steps of 0.1. The highest A_{val} was obtained for *spread* = 5 and 85 hidden units. At this point we achieved $SE_{val} = 82.92\%$, $SP_{val} = 88.30\%$ and $A_{val} = 85.61\%$.

4.3.3. SVM

We tuned σ and C independently. First, we varied σ between 0.1 and 1 in steps of 0.1, keeping $C = \infty$ (hard margin approach [36]). The values of SE_{val}, SP_{val} and A_{val} tend to be more stable for $\sigma \geq 0.2$. Then, fixing $\sigma = 0.2$, we varied C between 0.5 and 10 in steps of 0.1 (soft margin approach [36]). A_{val} increased for values of C below 1 and decreased for higher values of C. Therefore, we chose $C = 1$. With $\sigma = 0.2$ and $C = 1$, we obtained $SE_{val} = 83.63\%$, $SP_{val} = 86.32\%$ and $A_{val} = 84.97\%$.

Table 2
Performance of MLP, RBF, SVM and MV in the detection of red lesions.

Neural network	Lesion-based criterion		Image-based criterion		
	SE_{les} (%)	PPV_{les} (%)	SE_{im} (%)	SP_{im} (%)	A_{im} (%)
MLP	65.77	75.58	100	52.00	81.54
RBF	86.01	51.99	100	56.00	83.08
SVM	85.38	52.03	100	52.00	81.54
MV	82.74	53.29	100	56.00	83.08

SE_{les}: mean lesion-based sensitivity; PPV_{les}: mean lesion-based positive predictive value; SE_{im}: mean image-based sensitivity; SP_{im}: mean image-based specificity; A_{im}: mean image-based accuracy.

4.4. Testing

Each classifier (MLP, RBF, SVM and MV) was introduced in the classification stage and results on the test set were calculated. The diagnostic accuracy obtained with each classifier was assessed in terms of two different criteria: lesion-based criterion and image-based criterion [29]. The results are summarised in Table 2.

5. Discussion

In this study, we present an automatic method based on NNs to detect RLs in retinal images. The proposed method was tested on 40 images from DR patients that contained RLs and 25 images without RLs. The test set included images captured with different fundus cameras and showed variable colour, brightness and quality. No bad quality images were eliminated. Besides, the test set also contained images with different lesions associated with DR, such as EXs and CWs. Therefore, we evaluated the suitability of the method for the variability of images that could appear in the clinical practice.

In the first step, the images underwent a preprocessing step. However, the image characteristics are so closely related to the physical features of the patient and to the acquisition process that, even after normalisation, the images were heterogeneous. This is especially relevant in the detection of RLs, because the contrast between the retinal background and RLs is smaller than in the case of bright lesions.

Then, candidate RLs were segmented and a set of 29 features was studied. The 16 features with maximum discriminatory power were selected according to a stepwise LR analysis of our data. The selected features were used to train and test four NN based classifiers: MLP, RBF, SVM and MV. The utility of MLP and SVM in the detection of RLs in retinal images has been studied in previous works [17,18]. However, to the best of our knowledge, RBF and MV have not been previously used with this purpose. In addition, the use of a cross-entropy error function and regularisation for MLP has not been investigated in this context.

The performance of the different classifiers was assessed on a test set of 65 images. Table 2 shows that there were minimum differences among the performance of the four classifiers. However, for the RBF and MV classifier we obtained the better image-based results and a good trade-off between the lesion-based statistics. These NNs have not been previously studied for RL detection. The disadvantage of MV over RBF was its computational cost, as it required the training of three NNs. Therefore, attending to performance and complexity criteria, the RBF NN was the preferred classifier.

The image-based sensitivities obtained in this work show that we detected all images with RLs. Consequently, all patients for which DR was detected based on the presence of RLs in their images, would be derived to the ophthalmologist for further examination and treatment. These results can be considered satisfactory according to [43], where the authors state that $SE_{im} \geq 60\%$ maximises

Table 3
Performance of different methods for the detection of red lesions in retinal images.

Reference	Number of test images	Lesion-based results		Image-based results		
		SE_{les} (%)	PPV_{les} (%)	SE_{im} (%)	SP_{im} (%)	A_{im} (%)
Gardner et al. [17]	301[a]	–	–	73.80	–	–
Zhang and Chutatape [18]	30[a]	–	–	90.6	–	–
Sinthanayothin et al. [12]	30	77.50	–	–	–	–
Grisan and Ruggeri [16]	6	–	–	94.00	–	–
Niemeijer et al. [7]	50	31.00	–	100	87.00	–
Ege et al. [13]	134[a]	–	–	83.00	–	–
Grisan and Ruggeri [19]	200[b]	–	–	71.00	99.00	–
Walter et al. [20]	94	88.47[c]	–	97.00[c]	–	–
Quellec et al. [21]	120	89.62[c]	89.50[c]	–	–	–
Proposed method using RBF	115	86.01	51.99	100	56.00	83.08

SE_{les}: mean lesion-based sensitivity; PPV_{les}: mean lesion-based positive predictive value; SE_{im}: mean image-based sensitivity; SP_{im}: mean image-based specificity; A_{im}: mean image-based accuracy.

[a] Number of images used to obtain the training and test sets.
[b] Images digitized from film slides.
[c] Results on the detection of microaneusysms.

the cost-effectiveness in screening for DR. Newer evaluations, like [44], require a minimum SE_{im} of 80% and SP_{im} of 95%. These sensitivity requirements were met with the four classifiers. However, our specificity values are below these figures. The wrongly classified images usually presented retinal background irregularities that appeared with dark or reddish colour in the images. SP_{im} could be improved by changing the output threshold of the NNs, but SE_{im} values would decrease. For a computer-aided DR screening program it is more important to correctly classify all patients with sight-threatening DR (high sensitivity), even if it means misclassifying some healthy subjects in the DR group [10].

Regarding our lesion-based results, we detected a high percentage of RLs with the four classifiers. However, we missed some RLs. The majority of these lesions were connected to the vasculature. In the segmentation stage, they were considered as part of the vascular network and masked out. The PPV_{les} values in Table 2 also show that we detected some non-RL regions in the test images. False positives were mainly small vessel segments that were not eliminated in the segmentation stage.

The detection of RLs has been addressed in previous studies, as described in Section 1. Some of them analysed the ability of NNs to detect RLs in retinal images [12,17,18]. Studies that do not use NNs to detect RLs can also be found in the literature [7,13,19–21]. The results obtained in these studies, as well as the number of test images employed in each work, are summarised in Table 3.

As shown in Table 3, there are differences in the way results are presented in the diverse works. Most of the studies present only image-based statistics, while only one of them report PPV_{les}. Our results are below some of the statistics in Table 3. However, the differences in the databases used and performance measures makes it impossible to objectively compare our results with those reported in these works.

The proposed algorithm presents some limitations that merit consideration. In spite of the normalisation step, the colour and size of RLs are variable. Besides, the preprocessing method was aimed at normalising the contrast and luminosity in the whole image. As a consequence, the contrast between lesions and background is not necessarily improved for all the lesions in the image. These inconvenients add to the fact that it is more difficult to distinguish dark objects from the background in retinal images because the two classes show an important overlapping [18]. Additionally, the blood vessels and RLs show similar colour properties. As the detection of the vessel network is itself a difficult task [18], generally the thin and low contrasted vessels are not correctly masked out. This results in RLs segmentation errors. Additionally, some problems may arise from the size of the database. Some previous works used a smaller number of test images [12,16,18]. However, the size

of the test set may not be large enough to draw general conclusions about the diagnostic value of our method.

Future works will be aimed at improving our results by refining the vessel segmentation method and studying new features for RL segmentation. It is also our intention to address the distinction among RLs in future studies. For this task, the ROC database [45] would be a basic tool in order to compare our results with previous studies.

It is also important to notice that, although microaneurysms and retinal haemorrhages indicate DR, on their own they are not usually responsible for vision loss in diabetic patients. For this reason, future works will be aimed at detecting other types of lesions associated with DR together with RLs in order to separate healthy subjects from DR patients with a higher accuracy and grade the evolution of the disease. We will also try to improve our results by refining the vessel segmentation method and studying new features for RL segmentation.

Acknowledgements

Conflict of interest statement

There are no conflicts of interest that could inappropriately influence this research work.

References

[1] World Health Organisation. Prevention of blindness from diabetes mellitus: report of a WHO consultation in Geneva, Switzerland. Switzerland: WHO Library Cataloguing-in-Publication Data; 2005.
[2] Ong GL, Ripley LG, Newsom RS, Cooper M, Casswell AG. Screening for sight-threatening diabetic retinopathy: comparison of fundus photography with automated color contrast threshold test. Am J Ophthalmol 2004;17:445–52.
[3] Fong DS, Aiello L, Gardner TW, King GL, Blankenship G, Cavallerano JD, et al. Diabetic retinopathy. Diabetes Care 2003;26:226–9.
[4] Diabetic Retinopathy Clinical Research Network. The course of response to focal/grid photocoagulation for diabetic macular edema. Retina 2009;29:1436–43.
[5] Diabetic Retinopathy Clinical Research Network. A randomized trial comparing intravitreal triamcinolone acetonide and focal/grid photocoagulation for diabetic macular edema. Ophthalmology 2008;115, 1447–59, 1459.e1–10.
[6] Taylor CR, Lawrence MM, Salunga AM, Hepworth JT, Crutcher TD, O'Day DM, et al. Improving diabetic retinopathy screening ratios using telemedicine-

based digital retinal imaging technology. Diabetes Care 2007;30:574–8.

[7] Niemeijer M, van Ginneken B, Staal J, Suttorp-Schulten MSA, Abràmoff M. Automatic detection of red lesions in digital color fundus photographs. IEEE Trans Med Imaging 2005;24:584–92.

[8] Lin DY, Blumenkranz MS, Brothers SJ, Grosvenor DM. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. Am J Ophthalmol 2002;134:204–13.

[9] Niemeijer M, van Ginneken B, Russell SR, Suttorp-Schulten MSA, Abràmoff M. Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. Invest Ophthalmol Vis Sci 2007;48:2260–7.

[10] García M, Sánchez CI, López MI, Abásolo D, Hornero R. Neural network based detection of hard exudates in retinal images. Comput Methods Programs Biomed 2009;93:9–19.

[11] García M, Sánchez CI, Poza J, López MI, Hornero R. Detection of hard exudates in retinal images using a radial basis function classifier. Ann Biomed Eng 2009;37:1448–63.

[12] Sinthanayothin C, Boyce JF, Williamson TH, Cook HL, Mensah E, Lal S, et al. Automated detection of diabetic retinopathy on digital fundus images. Diabetic Med 2002;19:105–12.

[13] Ege BM, Hejlesen OK, Larsen OV, Møller K, Jennings B, Kerr D, et al. Screening for diabetic retinopathy using computer based image analysis and statistical classification. Comput Methods Programs Biomed 2000;62:165–75.

[14] Spencer T, Olson JA, McHardy KC, Sharp PF, Forrester JV. An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. Comput Biomed Res 1996;29:284–302.

[15] Frame AJ, Undrill PE, Cree MJ, Olson JA, McHardy KC, Sharp PF, et al. A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms. Comput Biol Med 1998;28:225–38.

[16] Grisan E, Ruggeri A. Segmentation of candidate dark lesions in fundus images based on local thresholding and pixel density. In: Proceedings of the 29th annual international conference of IEEE engineering in medicine and biology society. 2007.

[17] Gardner GG, Keating D, Williamson TH, Elliot AT. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. Br J Ophthalmol 1996;80:940–4.

[18] Zhang X, Chutatape O. Top-down and bottom-up strategies in lesion detection of background diabetic retinopathy. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition. 2005.

[19] Grisan E, Ruggeri A. A hierarchical Bayesian classification for non-vascular lesions detection in fundus images. In: Proceedings of the 3rd European medical and biological engineering conference. 2005.

[20] Walter T, Massin P, Erginay A, Ordonez R, Jeulin C, Klein J-C. Automatic detection of microaneurysms in color fundus images. Med Image Anal 2007;11:555–66.

[21] Quellec G, Lamard M, Josselin PM, Cazugel G, Cochener B, Roux C. Optimal wavelet transform for the detection of microaneurysms in retina photographs. IEEE Trans Med Imaging 2008;27:1230–41.

[22] Osareh A, Shadgar B, Markham R. A computational-intelligence-based approach for detection of exudates in diabetic retinopathy images. IEEE Trans Inf Technol Biomed 2009;13:535–45.

[23] Teng T, Lefley M, Claremont D. Progress towards automated diabetic ocular screening: a review of image analysis and intelligent systems for diabetic retinopathy. Med Biol Eng Comput 2002;40:2–13.

[24] Foracchia M, Grisan E, Ruggeri A. Luminosity and contrast normalization in retinal images. Med Image Anal 2005;9:179–90.

[25] Sánchez CI. Retinal image analysis by mixture model based clustering and discriminant analysis for automatic detection of hard exudates and haemorrhages. A tool for diabetic retinopathy screening. Ph.D. thesis. Valladolid; 2008.

[26] Hoover A, Kouznetsova V, Goldbaum M. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans Med Imaging 2000;19:203–10.

[27] Soille P. Morphological image analysis: principles and applications. New York: Springer-Verlag; 1999.

[28] Mendoça AM, Campilho A. Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. IEEE Trans Med Imaging 2006;25:1200–13.

[29] Osareh A. Automated identification of diabetic retinal exudates and the optic disc. Ph.D. thesis. Bristol; 2004.

[30] Sleigh JW, Steyn-Ross DA, Steyn-Ross ML, Grant C, Ludbrook G. Cortical entropy changes with general anaesthesia: theory and experiment. Physiol Meas 2004;25:921–34.

[31] Sonka M, Hlavac V, Boyle R. Image processing, analysis and machine vision. London: International Thomson Computer Press; 1996.

[32] Loew MH. Feature extraction. In: Sonka M, Fitzpatrick JM, editors. Handbook of medical imaging. Bellingham: SPIE Press; 2000. p. 273–341.

[33] Nadler M, Smith E. Pattern recognition engineering. New York: John Wiley; 1993.

[34] Hosmer DW, Lemeshow S. Applied logistic regression. New York: John Wiley & Sons; 1989.

[35] Jobson JD. Applied multivariate data analysis: categorical and multivariate methods. New York: Springer; 1991.

[36] Haykin S. Neural networks: a comprehensive foundation. Upper Saddle River, NJ: Prentice-Hall International; 1999.

[37] Bishop CM. Neural networks for pattern recognition. New York: Oxford University Press Inc.; 1995.

[38] Hornik K. Approximation capabilities of multilayer feedforward networks. Neural Netw 1991;4:251–7.

[39] Huang G-B, Chen Y-Q, Babri HA. Classification ability of single hidden layer feedforward neural networks. IEEE Trans Neural Netw 2000;11:799–801.

[40] Chen S, Cowan CFN, Grant PM. Orthogonal least squares learning algorithm for radial basis function networks. IEEE Trans Neural Netw 1991;2:302–9.

[41] Intelligent Data Analysis Group. SVM toolbox for Matlab; 2002, http://ida.first.fraunhofer.de/~anton/software.html.

[42] Moody J. Prediction risk and architecture selection for neural networks. In: Cherkassky V, Friedman JH, Wechsler H, editors. From statistics to neural networks: theory and pattern recognition applications. Springer-Verlag; 1994.

[43] Javitt JC, Canner JK, Frank RG, Steinwachs DM, Sommer A. Detecting and treating retinopathy in patients with type I diabetes mellitus. A health policy model. Ophthalmology 1990;97:483–94.

[44] Working Party of the British Diabetic Association. Retinal photography screening for diabetic eye disease. A British Diabetic Association Report, London; 1997.

[45] Niemeijer M, van Ginneken B, Cree MJ, Mizutani A, Quellec G, Sánchez CI, et al. Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. IEEE Trans Med Imaging 2010;29:185–95.