Neural network based detection of hard exudates in retinal images

María García[a,*], Clara I. Sánchez[a], María I. López[b], Daniel Abásolo[a], Roberto Hornero[a]

[a] Biomedical Engineering Group, E.T.S. Ingenieros de Telecomunicación, University of Valladolid, Campus Miguel Delibes, Camino del Cementerio s/n, Valladolid, Spain
[b] Instituto de Oftalmobiología Aplicada, University of Valladolid, Valladolid, Spain

ARTICLE INFO

ABSTRACT

Diabetic retinopathy (DR) is an important cause of visual impairment in developed countries. Automatic recognition of DR lesions in fundus images can contribute to the diagnosis of the disease. The aim of this study is to automatically detect one of these lesions, hard exudates (EXs), in order to help ophthalmologists in the diagnosis and follow-up of the disease. We propose an algorithm which includes a neural network (NN) classifier for this task. Three NN classifiers were investigated: multilayer perceptron (MLP), radial basis function (RBF) and support vector machine (SVM). Our database was composed of 117 images with variable colour, brightness, and quality. 50 of them (from DR patients) were used to train the NN classifiers and 67 (40 from DR patients and 27 from healthy retinas) to test the method. Using a lesion-based criterion, we achieved a mean sensitivity (SE_l) of 88.14% and a mean positive predictive value (PPV_l) of 80.72% for MLP. With RBF we obtained $SE_l = 88.49\%$ and $PPV_l = 77.41\%$, while we reached $SE_l = 87.61\%$ and $PPV_l = 83.51\%$ using SVM. With an image-based criterion, a mean sensitivity (SE_i) of 100%, a mean specificity (SP_i) of 92.59% and a mean accuracy (AC_i) of 97.01% were obtained with MLP. Using RBF we achieved $SE_i = 100\%$, $SP_i = 81.48\%$ and $AC_i = 92.54\%$. With SVM the image-based results were $SE_i = 100\%$, $SP_i = 77.78\%$ and $AC_i = 91.04\%$.

1.   Introduction

Diabetic retinopathy (DR) is a visual complication of diabetes, which has become the most common cause of visual impairment among people of working age in industrialized countries [1]. Laser photocoagulation can slow down the progression of DR if detected in its early stages. However, this is not an easy task because DR patients perceive no symptoms until visual loss develops. This happens in the later stages of the disease, when treatment is less effective. To ensure that treatment is received on time, diabetic patients need to undergo a yearly eye fundus examination [2]. Including digital photographs of the retina in the screening protocol is a sensitive and specific means for detecting early clinical signs of retinopathy in at-risk populations [3]. As demonstrated in [4], manual and automated analysis of digital retinal images reach satisfactory results in the detection of DR and diabetic macular oedema. Moreover, their inclusion in the exam protocol showed a high performance in the detection of patients who need further ophthalmological assessment [4].

The growing incidence of diabetes increases the number of images that need to be reviewed by physicians. In addition, the high cost of examinations and the lack of specialists prevent many patients from receiving effective treatment. Computer

aided detection of retinal lesions associated with DR offers many potential benefits. In a screening setting, it allows the examination of a large number of images in less time and more objectively than current observer-driven techniques [5]. In a clinical setting, it can be an important diagnostic aid and can reduce the workload of trained graders and, therefore, reduce costs [5].

DR signs include red lesions, such as microaneurysms (MAs), intraretinal microvascular abnormalities (IRMAs) and haemorrhages (HEs), and bright lesions like hard exudates (EXs) and soft exudates or cotton–wool spots (CWs). Other bright lesions that can appear in fundus images are drusen, which are associated with age related macular degeneration (AMD) [6]. Sometimes they appear similar to exudates and may confound automatic systems. The small laser scars caused by panretinal photocoagulation also appear as bright areas in retinal images belonging to DR patients that have received this treatment. Therefore, their presence is an additional difficulty for bright lesion detection. Retinal exudates can represent the only visible sign of DR in some patients [6]. Moreover, it is important to distinguish among lesion types as they have different diagnostic importance and management implications [6]. This paper focuses on EXs detection and their differentiation from other bright areas in our images, specifically CWs and the optic disk (OD). As EXs are among common early clinical signs of DR [7], their detection would be an important contribution to the screening tasks and could serve as a first step towards a complete monitoring and grading of the disease. EXs are lipid and lipoprotein deposits, white, yellowish or waxy, that appear as compact patches with well-defined borders in retinal images.

Many attempts to detect these lesions can be found in the literature. Some works [8,9] used the high luminosity of EXs to separate them from the background by thresholding. In [8] a global threshold was used for the detection of EXs after image enhancement. The wide variability among images made the method dependent on user intervention to select the adequate threshold for each case. A similar approach was used in [9], where a combination of global and local thresholds was applied to segment EXs. This method, however, needed a previous definition of a region of interest by an operator. Edge detectors, combined with other techniques, were applied in [10–14]. A two-stage method was applied in [10]. First, the borders of candidate EXs and HEs were detected. Then, colour, shape and location properties were used to separate both types of lesions. Other authors [11,12] also used edge detection in combination with different techniques to detect bright lesions. In [12] authors also created a model of the retina to automatically identify the presence of lesions in the macular region. Edge and brightness information was combined in [13] to identify EXs and the severity of DR. In [14] the borders of EXs were extracted using a mathematical morphology based method. However, these studies [8–14] did not explicitly address the differentiation among lesion types.

Some works [15,16] focused on the study of the ability of Bayesian classifiers to detect retinal lesions. The intensity of a set of pixels was used to train the classifier in [15]. On the contrary, the training set in [16] was created by extracting several features from image regions. A statistical classifier was also used in a previous work by our group, where Fisher's linear discriminant analysis was part of a method to segment EXs [17]. The classification rule was automatically adapted to each image. The distinction between CWs and EXs was faced in [16,17], but not in [15].

A method for detecting EXs and CWs, and separate them from drusen was proposed in [6]. A *k*NN classifier was used in a first stage to detect candidate bright lesions. Linear discriminant analysis was subsequently used to differentiate among lesion types. Neural network (NN) based classifiers have also been used [18–20]. A multilayer perceptron (MLP) was used in [18] to determine the presence of bright lesions in image regions of size 20×20 pixels. In [19] candidate bright regions were segmented using Fuzzy *C*-means clustering. A support vector machine (SVM) was used afterwards to determine if a segmented region is an EX or another type of bright region. A similar approach was used in [20], where MLP and SVM NNs were analysed and compared.

In this study, we examined and compared the ability of three different types of NNs to detect EXs in retinal images: MLPs, radial basis function (RBF) networks and SVMs. We developed an automatic method that was suitable for the great variability of images that ophthalmologists could find in their daily practice. Therefore, it could be used as a clinical aid in the diagnosis and monitoring of DR.

2.   Image database

A total of 117 images with variable colour, brightness, and quality were used in this study. All the images were provided by the Instituto de Oftalmobiología Aplicada (IOBA) of the University of Valladolid, Spain. 31 of these images were captured using a TopCon TRC-NW6S non-mydriatic retinal camera at a field-of-view (FOV) of 45°. The remaining 86 images were captured with a TopCon TRC-50IX mydriatic retinal camera at 50° FOV. Due to the wider FOV, the images taken with the TopCon TRC-50IX cover more retinal area and the image details appear bigger. On the other hand, images captured with the TopCon TRC-NW6S showed, generally, better quality and contrast. Image resolution was 576×768 pixels in 24 bit JPEG format. An experienced ophthalmologist manually marked all EXs in the images. These annotations have been considered as the reference standard to compare our results with.

Images came from a clinical set of 106 diabetic patients who were referred to the ophthalmologist for further examination. Therefore, our database contained two images (right and left eye) from 11 patients and the remaining 95 images belonged to different subjects. 90 of these images belonged to patients who suffered from mild to moderate non-proliferative DR, according to an expert. In the remaining 27 images the ophthalmologist did not mark any EXs. The images from DR patients also contained CWs (in 19 images) and HEs (in 87 images). However, drusen were not present in any of the images.

The 117 images were divided into a training set and a test set in a pseudo-random manner:

- The training set contained 50 images from DR patients, with the only restriction that two images from the same patient could not be included in this set. In this way, the ability of the

training set to represent the variability of the population is greater. 1940 segmented regions were extracted from these images after the segmentation stage of our method. They were labelled as EXs or non-EXs according to the annotations of the ophthalmologist. In this way, we created a fully labelled groundtruth dataset to train the NNs with these 50 images.

- The remaining 67 images (27 from healthy retinas and 40 from DR patients) form the test set. It has been used to validate the effectiveness of the complete algorithm by comparing our results with the expert-labelled images.

3. Methods

The proposed method can be divided in four steps, which are explained in detail in the next subsections: luminosity and contrast normalization, segmentation, feature extraction and classification using MLP, RBF or SVM.

3.1. Luminosity and contrast normalization

The great variability within and between retinal images in a clinical environment makes it hard to distinguish retinal features and lesions in some areas. A preprocessing step is necessary to normalize the images and to increase the contrast between EXs and the background.

We applied the method proposed in [21], based on the work developed in [22], to the green channel of the images. Being I the observed image, the method estimates the original undistorted image, $\hat{I}^o$, as [21]:

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

where $\hat{L}(x, y)$ and $\hat{C}(x, y)$ are estimates of the luminosity and contrast drifts of the image, respectively. They are obtained from the background image, I_b, consisting of the pixels of I that do not belong to lesions or retinal structures [21]. The process of computing I_b, however, does not rely on a previous detection of these lesions and retinal structures. Instead, a pixel (x,y) is considered to belong to the background if its intensity is close to the mean intensity in a neighbourhood N centred in (x,y). The similarity measure is the Mahalanobis distance, d_M, between the intensity of pixel (x,y) and the mean intensity value inside N. Mathematically, this is expressed as [21]:

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

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. In this study, we chose $t = 1$. Using this value, we considered an adequate percentage of pixels in each square as background pixels [21].

The subsequent estimation of $\hat{L}(x, y)$ and $\hat{C}(x, y)$ requires a similar computation of the mean and standard deviation of the intensity values inside a neighbourhood N centred in (x,y). However, in this case, only the pixels belonging to I_b are taken into account for these calculations.

In order to reduce the computational burden when computing background and drifts, the images were partitioned into square blocks. The process assumed that at least 50% of the pixels in each square were background pixels. This may be untrue where large retinal structures or lesions appear. To avoid their influence, squares with mean luminosity higher than given threshold were not considered for drift estimation [21].

To make the algorithm more robust, we also discarded squares with mean luminosity above the threshold when estimating I_b. Squares with less than 50% of background pixels were not considered for $\hat{L}(x, y)$ and $\hat{C}(x, y)$ computation either. Estimates for the removed squares were calculated averaging the values obtained for the surrounding squares.

An original observed image is shown in Fig. 1(a). After the preprocessing step, we obtained the image of Fig. 1(b).

3.2. Segmentation

In this stage we tried to find the candidate EX regions that had to be classified by the NNs in the next step. As EXs appear as bright regions in the preprocessed image, a first approach to detect candidate lesions was the extraction of the bright pixels in this image. However, even after the image normalization process, residual variability in the luminosity or in the pigmentation of the retinal background makes the utilization of an approach based on a global thresholding very difficult [23]. Hence, we also made use of the local properties of EXs and combined global and adaptive histogram thresholding methods.

The global and local histograms of the preprocessed image showed one maximum, corresponding to the background. The tail on the right of this maximum corresponded to the bright regions we wanted to segment. We set a threshold at the grey level of this tail for which the histograms decreased to the 10% of the maximum. To calculate local histograms, the image was partitioned into square blocks of side 200 pixels (zero-padding when necessary). The locally and globally segmented images were combined using the AND operator.

The papillary region was also masked out to reduce the computational burden of the classifiers using the method described in [17]. The OD is a bright region in the image and the main vessels emanate almost vertically from it. These properties were used to detect this retinal structure in two stages. First the candidate OD centres are selected using a combination of mathematical morphology (alternative sequential filters) and regional maxima detection. In a second stage, the vertical blood vessels in the preprocessed image are detected using a two-dimensional matched filter with Gaussian cross-profile in the vertical direction [24]. Then, the Hough transform is used to detect vertical lines in the neighbourhood of each candidate point. The candidate with more pixels belonging to vertical lines passing through it is considered the OD centre [17]. The boundary of the OD is approximated as a circle. This approach is considered accurate enough to remove the OD as a candidate EX region [20].

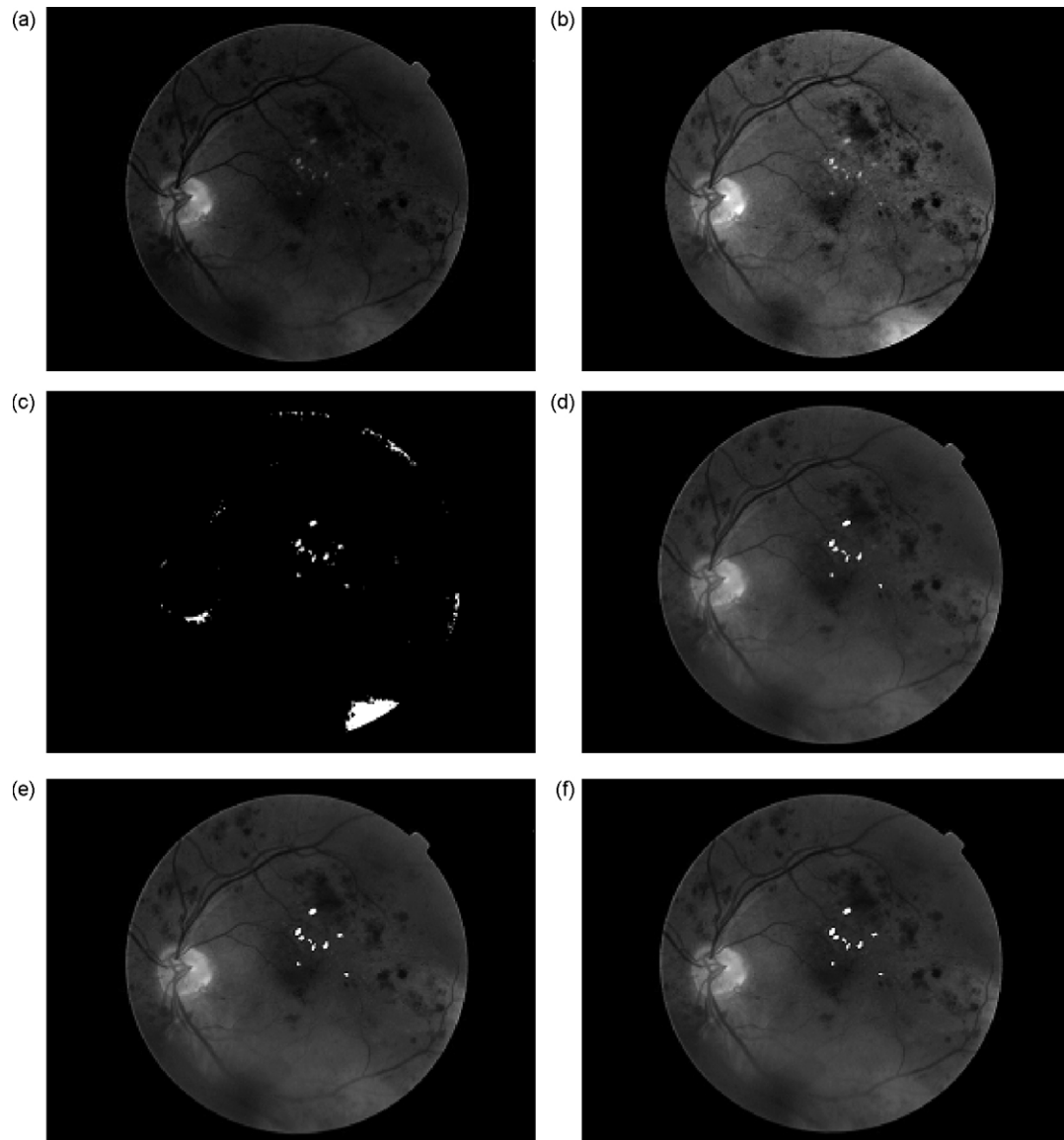Fig. 1(c) is the result of segmenting the image shown in Fig. 1(b).

Fig. 1 – (a) Green channel of the original retinal image. (b) Green channel of the same image after contrast and luminosity normalization. (c) Candidate EX regions obtained after segmentation. (d) Result obtained after classification with MLP. Detected EXs are marked in white and superimposed on a greyscale version of the original image. (e) Result obtained after classification with RBF. Detected EXs are marked in white and superimposed on a greyscale version of the original image. (f) Result obtained after classification with SVM. Detected EXs are marked in white and superimposed on a greyscale version of the original image.

3.3. Feature extraction

In order to classify candidate regions as EX or non-EX, we extracted a set of features from each region and used them as inputs to the NNs. To select an adequate set of features, we focused on those characteristics that ophthalmologists use to visually distinguish EXs from the retinal background and other retinal lesions or structures. We also tried to keep an adequate dimensionality of the feature set, as misclassification probability tends to increase with the number of features and the structure of the classifier is more difficult to interpret [25].

We selected 18 visually distinctive features of EXs, which include colour and shape features [20]:

- Mean RGB values inside the region (1–3), obtained as an average of the RGB values of the pixels that belong to the region.
- Standard deviation of the RGB values inside the region (4–6), obtained as the standard deviation of the RGB values of the pixels that belong to the region.
- Mean RGB values outside the region (7–9), obtained as an average of the RGB values of the pixels that belong to a rectangular area around the region (5 pixels distance).
- Standard deviation of the RGB values outside the region (10–12), obtained as the standard deviation of the RGB values of the pixels that belong to a rectangular area around the region (5 pixels distance).

- RGB values of the region centroid (13–15), measured as the intensity values in the three colour components of the centre of the region.
- Region size (16), measured as the number of pixels inside the region.
- Region compactness (17), measured as the ratio between the square of the perimeter and the area of the region (both measured in number of pixels). To obtain the perimeter, a boundary tracing algorithm needs to be used [26].
- Region edge strength (18), measured as an average of the edge values in the perimeter of the region. The edge values were obtained after the application of a Prewitt operator [26].

3.4. Classification

Three different neural networks were investigated. For the training and testing of MLP, we used the NETLAB software developed within the Neural Computing Research Group of Aston University, UK [27]. RBF was developed using the neural network toolbox included in Matlab®. SVM was implemented using the SVM toolbox for Matlab® developed by the Intelligent Data Analysis Group of Fraunhofer FIRST [28].

The architecture of the three classifiers contained an input layer with 18 nodes and a single neuron output layer. If the output was greater than a given threshold, the region was considered EX. Otherwise, it was labelled non-EX.

3.4.1. Multilayer perceptron

Multilayer feed-forward networks, like MLPs, are an important class of NNs that can represent non-linear functional mappings between a set of input variables and a set of output variables [29,30]. A MLP with enough units in a single hidden layer can approximate any function, provided the activation function of the neurons satisfies some general constraints [31,32]. From these considerations, we decided to use a MLP with one hidden layer, for which the optimum number of hidden neurons was experimentally determined. As neuron activation function in the hidden layer, we chose the hyperbolic tangent sigmoid function (tan-sigmoid), an antisymmetric function in the interval $(-1, 1)$. Tan-sigmoid satisfies the constraints in [31] and [32]. Moreover, it improves the learning speed of MLP [29]. In the output layer, we used the logistic sigmoid activation function, that also satisfies the aforementioned constraints and whose outputs lie in the range $(0, 1)$. This choice was motivated by the fact of interpreting the outputs of the network as posterior probabilities [30].

The problem of training a NN can be formulated in terms of the minimization of an error function. The choice of a suitable error function and minimization algorithm can improve the performance of MLP. It has been demonstrated [30] that a cross-entropy error function simplifies the optimization process when the logistic sigmoid activation function is used in the output layer. Therefore, we considered this function as an appropriate choice in our study.

Regarding the minimization algorithm, several choices are available for MLP. We selected the scaled conjugate gradients algorithm, for which the error function is guaranteed not to increase during training [30]. Moreover, it generally shows

faster convergence when compared to gradient descent-based techniques or even conventional conjugate gradient algorithms [30].

To avoid overfitting and improve generalization, we employed a weight decay regularizer [30]. The regularization parameter, ν, was experimentally settled.

If the output is above 0.5 for some input vector, the probability of the image region represented by that input vector of being an EX is greater than that of being a non-EX. Therefore, the output threshold was set to 0.5. An example of the classification with MLP is shown in Fig. 1(d).

3.4.2. Radial basis function

RBF is also a universal approximator [29,30]. Its architectural model comprises three layers. The input layer receives the feature vector and has as many neurons as the length of this vector. The only hidden layer in the network performs a non-linear 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 network to the activation pattern [29].

The hidden layer activation functions must satisfy some conditions derived from Tikhonov's regularization theory [33] to be called *radial basis functions*. Among all the possibilities, we chose Gaussian functions because they are commonly employed in practice, as they are rotationally and translationally invariant [29]. They are characterized by their centre and *spread* parameter, which represents the width of the Gaussian.

Native RBFs require as many hidden neurons as the available training points. This is computationally demanding and may lead to poor generalization ability. Therefore, a suboptimal solution must be found by means of *generalized RBF networks* that, usually, require less hidden neurons [29]. However, their training involves learning not only the weights, but also the number of radial basis functions, the position of their centres and their width. To accomplish this task, we applied the *orthogonal least squares* algorithm [34]. 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 EXs and negative for non-EXs.

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

3.4.3. Support vector machine

Like the previous approaches, SVMs can approximate any function [29]. As in RBFs, the input space is mapped into a high dimensional feature space. Then, the hyperplane that maximizes the margin of separation between classes is constructed. The points that lie closest to the decision surface are called *support vectors* and directly affect its location. When the classes are non-separable, the optimal hyperplane is the one that minimizes the probability of classification error [28]. In this case, the problem can be solved using the method of *Lagrange multipliers* [29]. Given the training set, $\{(x_i, d_i)\}_{i=1}^{N}$, find the Lagrange multipliers, $\{\alpha_i\}_{i=1}^{N}$, that maximize the objective function:

$$Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j d_i d_j K(x_i, x_j) \qquad (3)$$

subject to the constraints:

$$\sum_{i=1}^{N} \alpha_i d_i = 0 \tag{4}$$

$$0 \le \alpha_i \le C \quad \text{for } i = 1, \ldots, N \tag{5}$$

(x_i, d_i), $i = 1, \ldots, N$ represent the input vector-desired response pairs and $K(x_i, x_j)$ is called inner-product kernel. C is an upper bound for the Lagrange multipliers and controls the trade-off between machine complexity and number of non-separable points [29].

Depending on the selected inner-product kernel, we can construct different types of machines, including two-layer perceptrons and RBFs [29,35]. However, there are some differences with the classical algorithms:

- In the SVM two-layer perceptron, the number of hidden neurons and the weights are automatically determined by the number of support vectors and their values, respectively. In classical MLP, they have to be experimentally determined [29].
- For the RBF type of SVM, the number of radial basis functions and their centres are also automatically determined by the algorithm. In classical RBF classifiers; they have to be determined in advance, usually by some kind of clustering method [29,35].

We employed a Gaussian kernel (i.e. a RBF type of SVM), the most common type of kernel used in practice [30]. Two parameters need to be optimized: the width of the Gaussian functions (σ) and C. The output given by the implementation of SVM we have used in this work is the class label of the region represented by the input feature vector. Therefore, it was not necessary to choose an output threshold. An example of classification with SVM is depicted in Fig. 1(f).

4. Parameter optimization of the NN classifiers

We trained the classifiers with 970 EX and 970 non-EX regions, manually extracted after the segmentation of the 50 images in the training set. Each of these regions was represented by

its corresponding 18-element feature vector. These feature vectors formed the training examples of the classifiers. The inputs were normalized (mean = 0, standard deviation = 1) and the data points were randomly presented to the networks [29]. Our goal was to find an optimal architecture of the NNs that led to a minimum prediction risk [36]. We used 10-fold cross-validation to asses the generalization ability of the networks as it is an estimator of the prediction risk [36].

For the different networks and network architectures, we measured the mean sensitivity (SE), specificity (SP) and accuracy (AC) obtained for the validation set. We used AC for model selection as it represents better the trade-off between sensitivity and specificity.

4.1. MLP

We had to determine the optimum number of hidden neurons for MLP and the optimum regularization parameter, v. We varied the number of hidden neurons from 10 to 50 in steps of 1. The regularization parameter was varied from 1 to 10 in steps of 0.5. The results for $v = 2$, $v = 2.5$ and $v = 3$ were very similar and higher than for other values of v. We chose $v = 3$ because it increases the generalization ability of the network, as the number of effective parameters of the network is a decreasing function of v [37]. Besides, we chose a network with 31 hidden units because AC did not significantly improve for larger architectures, as shown in Fig. 2. This optimum operating point is marked with the symbol '♦' in Fig. 2. For this configuration, we obtained SE = 91.34%, SP = 92.06% and AC = 91.70% over the validation set. The MLP with this configuration was trained for an average number of 1000 training cycles, obtaining a mean RMS error on the validation set of 0.12.

4.2. RBF

The optimum *spread* and number of radial basis functions had to be settled. The output of any 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 area in the input space to which each neuron responds. To ensure that the whole feature space is adequately covered, we varied both parameters jointly. The number of RBFs ranged from 10 to 120, while the spread parameter varied from 1 to 10 in steps of 0.5.
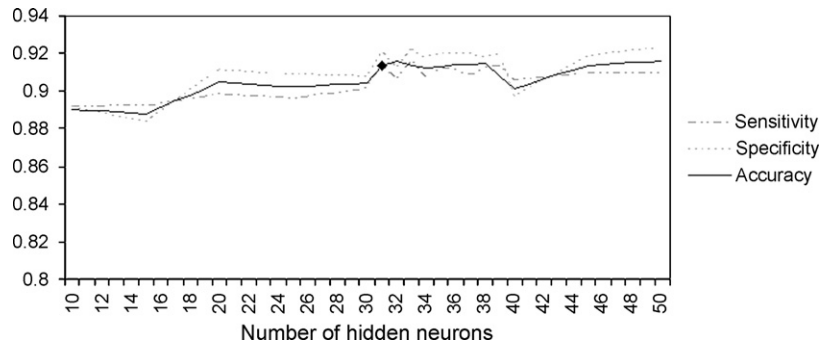

Fig. 2 – **Results of the training stage for the MLP neural network with $v = 3$. The optimum point is marked with the symbol '♦'.**
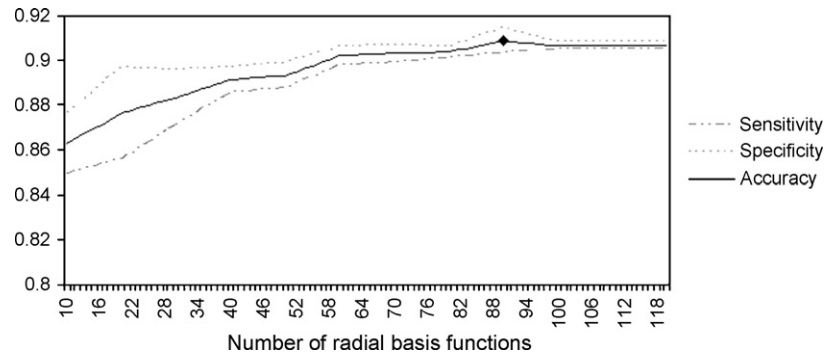
Fig. 3 – Results of the training stage for the RBF neural network with spread fixed to 3.5. The optimum point is marked with the symbol '♦'.

The lowest values of AC were obtained for values of the *spread* parameter below 3.5. The variation of AC with the number of RBFs was very similar for the remaining values of *spread*. Therefore, fixing *spread* = 3.5 the whole input space is adequately covered while, at the same time, each neuron responds to a different region of this input space.

Fig. 3 shows the training results when *spread* = 3.5. We chose a network with 90 RBFs (point marked with the symbol '♦' in Fig. 3) because AC did not significantly improve with more units. At this point, we obtained the following results for the validation set: SE = 90.31%, SP = 91.44% and AC = 90.88%. The implementation of the generalized RBF network used in this work adds a new neuron to the architecture in each cycle. Therefore, for the optimum operating point, 90 training cycles were needed. The mean RMS error on the validation set was 0.24.

4.3. SVM

We had to optimize σ and C. The parameters were independently tuned. First, we varied σ between 0.1 and 1 in steps of 0.1, while C was fixed to ∞ (*hard margin approach* [29]). Similar values of SE, SP and AC were obtained for $0.2 \leq \sigma \leq 0.8$. This range of values of σ, keeping $C = \infty$, was tested on the unseen images. We chose $\sigma = 0.5$ because the results did not significantly vary for larger values. Then, fixing $\sigma = 0.5$, we varied C between 0.5 and 8, in steps of 0.1 (*soft margin approach* [29]). The results obtained in this step are depicted in Fig. 4. AC increases for C below 2.5 and then shows small variations. Therefore, we

chose $C = 2.5$ as the optimum value, as indicated with the symbol '♦' in Fig. 4. With $\sigma = 0.5$ and $C = 2.5$ we obtained SE = 90.21%, SP = 91.96% and AC = 91.08% over the validation set. The implementation needed an average number of 259.40 iterations to train the classifier with the optimum configuration. The RMS error, averaged over the validation set, was 0.36. It is worth noting that this error has been computed using the output of this implementation of SVM, which is equal for all the examples belonging to the same class.

5. Comparison of NN classifiers

Each network was introduced in the classification stage and the complete algorithms were tested on a new set of 67 unseen images (test set).

The performance was assessed in terms of an image-based criterion and a lesion-based criterion (pixel resolution) [20]. The image-based criterion accounted for the ability of the algorithm to separate pathological images from normal ones on the basis of the presence or absence of EXs. With a lesion-based criterion, we examined the number of EXs in the images that were correctly detected. For this lesion-based criterion, specificity is not an informative measure. In a retinal image, the number of true negatives (TNs) is generally much higher than the number of false positives (FPs). As a consequence, even if EXs detection is inaccurate, specificity values would be high. The positive predictive value (PPV_l), which measures the probability that a detected region is really an EX, was regarded
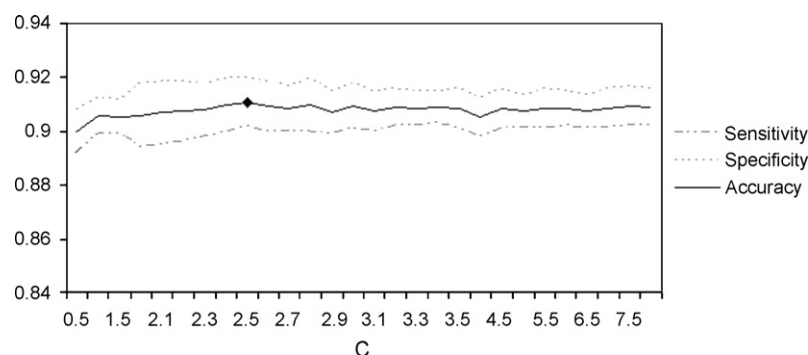

Fig. 4 – Results of the training stage for the SVM neural network with $\sigma = 0.5$. The optimum point is marked with the symbol '♦'.

Table 1 – Performance of MLP, RBF and SVM for EXs detection					
Neural network	Lesion-based criterion		Image-based criterion		
	SE_l (%)	PPV_l (%)	SE_i (%)	SP_i (%)	AC_i (%)
MLP	88.14	80.72	100	92.59	97.01
RBF	88.49	77.41	100	81.48	92.54
SVM	87.61	83.51	100	77.78	91.04

SE_l = mean lesion-based sensitivity, PPV_l = mean lesion-based positive predictive value, SE_i = mean image-based sensitivity, SP_i = mean image-based specificity, AC_i = mean image-based accuracy.

as a more significant criterion to evaluate the performance of the system [20]. We also used the property if EXs of usually appearing in groups in retinal images to improve the image-based specificity (SP_i) of the system without compromising its image-based sensitivity (SE_i). We used an approach similar to that presented in [20]. We identified those images where less than 30 pixels (0.0068% of the total number of pixels in the image) had been detected as EX and considered them as belonging to healthy retinas. This threshold has been experimentally settled by inspection of the results obtained for the images in our database.

The results obtained with the three NNs on the test set are summarized in Table 1.

6.   Discussion and conclusion

In this study, we compared the ability of three neural network based classifiers to distinguish EXs from retinal background in 27 healthy retinas and 40 images of retinas with DR. The proposed method is completely automatic once the classifiers are configured.

Prior to classification, we preprocessed the image in order to normalize the variability in colour, brightness and contrast that images can show in a clinical environment. Subsequently, the candidate EXs regions were detected by means of a segmentation algorithm. The parameters of this algorithm were chosen in order to achieve a high sensitivity (detect a high percentage of true EXs), even if that meant a lower specificity (detecting also some non-EXs regions). The following stages would increase specificity by discriminating between true EXs and FPs. However, the true EXs that were discarded in the segmentation stage could not be recovered. Hence, the selection of the segmentation threshold affects the final classification rates. If it was chosen at the point where the histogram decreased to the 5% of its maximum, PPV_l would increase for the three NNs (89.47% for MLP, 85.27% for RBF and 88.48% for SVM). However, SE_l would decrease (81.88% for MLP, 82.53% for RBF and 82.34% for SVM). Image-based results show a similar behaviour. SP_i would increase (96.29% for the three NNs) while SE_i would decrease (97.50% for MLP, 97.50% for RBF and 95.00% for SVM). On the contrary, if the segmentation threshold was moved to the point where the histogram decreased to the 15% of its maximum, PPV_l would decrease for the three NNs (74.88% for MLP, 67.95% for RBF and 77.04% for SVM). However, SE_l would increase (90.46% for MLP, 88.57% for RBF and 89.18% for SVM). Regarding image-based results, SE_i would equal the results in Table 1 (100% for the three NNs) and SP_i would decrease (77.78% for MLP, 18.51% for RBF and 62.96%

for SVM). For other values of the segmentation threshold, the results obtained followed a similar tendency.

In the classification stage, we explored the ability of MLP, RBF and SVM to segment EXs in retinal images. To the best our knowledge, RBF has not been previously investigated with this purpose. We assessed the performance of RBF on the same database that we used for MLP and SVM and carried out a comparative study. In addition, the use of a cross-entropy error function and regularization for MLP has not been previously investigated in this context.

We achieved similar results with the three NNs. Using a lesion-based criterion, we obtained a mean sensitivity (SE_l) of 88.14% and a mean positive predictive value (PPV_l) of 80.72% for MLP. With RBF we achieved SE_l = 88.49% and PPV_l = 77.41% and with SVM the lesion-based results were SE_l = 87.61% and PPV_l = 83.51%. Regarding the image-based results, with MLP we obtained a mean sensitivity (SE_i) of 100%, a mean specificity (SP_i) of 92.59% and a mean accuracy (AC_i) of 97.01%. We achieved SE_i = 100%, SP_i = 81.48% and AC_i = 92.54% using RBF and SE_i = 100%, SP_i = 77.78% and AC_i = 91.04% using SVM. Additionally, each NN has certain advantages. The complexity of MLP was significantly lower than that of RBF or SVM for a similar degree of accuracy (the optimal RBF needed 90 hidden neurons and the optimal SVM needed 678 support vectors). By contrast, the training complexity in RBF is typically lower than in MLP (due to the presence of nearly flat regions in the error function and the slower convergence that the effects of the different weights can cause) or SVM (due to the complexity involved in the resolution of a quadratic problem) [29]. With respect to SVM, its convergence to an optimum solution is guaranteed [20,29].

These results can be considered satisfactory according to [38], where the authors state that an image-based sensitivity equal or greater than 60% maximizes cost-effectiveness in screening for DR. Newer evaluations, like that of the British Diabetic Association [39], require a minimum image-based sensitivity of 80% and image-based specificity of 95%. These sensitivity requirements were met with the three NNs. However, our specificity values are below these figures. Image-based specificity could be improved by changing the output threshold of the networks. For MLP, setting the output threshold to 0.7, we achieved SE_i = 95% and SP_i = 96.3%. Equal results were obtained with an output threshold of 0.3 for RBF. The output given by the implementation of SVM employed in this work is only the class label. Thus, it is not possible to improve image-based SP_i in the same way. However, lesion-based sensitivity is reduced if these output thresholds are used (SE_l = 85.48% for MLP and SE_l = 75.41% for RBF). It is worth noting that it would be also possible to define two output

Table 2 – Summary of previous studies concerning the detection of EXs in retinal images						
Detection method	Database (number of test images)	Lesion-based results		Image-based results		
		SE_l (%)	PPV_l (%)	SE_i (%)	SP_i (%)	AC_i (%)
Osareh [20]	67	90.0	89.3	95.0	88.9	–
Zhang and Chutatape [19]	30[a]	88.0	84.0	–	–	–
Gardner et al. [18]	179	93.1[b]	–	–	–	–
Walter et al. [14]	15	92.8	92.4	100	86.7	–
Li and Chutatape [12]	35	–	–	100	–	74.0
Wang et al. [15]	154	–	–	100	70.0	–
Niemeijer et al. [6]	300	–	–	95.0	86.0	–
Sánchez et al. [17]	58	88.0	–	100	100	100

SE_l = mean lesion-based sensitivity, PPV_l = mean lesion-based positive predictive value, SE_i = mean image-based sensitivity, SP_i = mean image-based specificity, AC_i = mean image-based accuracy.

[a] Number of images used to obtain the training and test sets.

[b] Obtained for patches of size 20×20 pixels.

thresholds, with the limitation that patients presenting diagnostic values between both cut-off points will not be classified. This approach would lead to a decrease in the number of misclassified cases. However, a proportion of regions would not be classified in any of the classes, especially faint EXs or noisy background regions that appear very similar in retinal images. In addition, one of the advantages of defining a single output threshold for the MLP NN is that it allows for the interpretation of its outputs as posterior probabilities. For these reasons, we used a single threshold in the present study. All regions in our data sets could be classified, although both false positive and false negative cases were present.

Previous studies [18–20] also analysed the ability of NNs to detect EXs in retinal images. Their results and the number of test images in their database are summarized in Table 2. Osareh [20] used 10-fold cross-validation to optimize a MLP and a SVM classifiers. The input vectors to both NNs were formed by the same 18 features that were used in our work. Zhang and Chutatape [19] selected the parameters of a SVM classifier with five-fold cross-validation. A smaller feature set, comprised of five features, was used to describe EXs regions. The 30 images in their database were not divided into training and test sets as in the present work. The training and test sets were composed of bright areas extracted from the complete database of images. Only lesion-based statistics were reported in this study. The results presented in the study by Gardner et al. [18] were obtained for image patches of size 20×20 pixels. The intensities of the pixels inside the patches formed the inputs of this classifier. Therefore, no colour or shape features were derived for each region. The use of cross-validation and image based-results were not reported.

Other studies did not use NNs to detect EXs. The results obtained in some of them and their database specifications are also included in Table 2. Walter et al. reported both lesion-based and image-based results. However, the distinction between EXs and CWs was not addressed in this work. Li and Chutatape [12] only reported image-based statistics. Results about the number of EXs detected in different areas of the retina were also provided for two images in their database.

This permitted the grading of the severity of RD, as the presence of lesions affects vision more seriously if they are located in certain regions of the retina than in others. The results presented by Wang et al. [15] accounted for the ability of their method to detect bright lesions, not only EXs. No lesion-based results were reported in this study either. It is worth noting that these methods [12,14,15] did not require feature extraction. The study by Niemeijer et al. [6] is particularly notable for the effort made in distinguishing among lesion types. To accomplish this task, a large set of 86 features was studied in the final classification stage. Finally, in a previous work by our group [17], the three colour components of each pixel were used to classify it as EXs or non-EXs. The classification rule was automatically adapted to each image and, therefore, a different rule was obtained for every retinal photograph. The use of cross-validation was not mentioned in [6,12,14,15,17].

Although our lesion-based statistics do not reach those reported in some of these works, comparisons should be made with caution. The lack of a public database and a common way to measure performance makes it impossible to objectively contrast our results with those reported in literature. Moreover, the use of 10-fold cross-validation for model selection should be taken into account in these comparisons.

Our image-based results are similar to those reported in many of the cited studies. In addition, the image-based sensitivities in Table 1 show that we detected all images with signs of DR using the three NNs. Consequently, no patient suffering from the disease is prevented from receiving treatment. We also detected some FPs in the images of healthy subjects. These wrongly classified images usually presented yellowish artefacts and retinal background irregularities that appear similar to EXs. These artefacts could be removed by changing the output threshold, as discussed previously. However, 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 is at the expense of misclassifying some healthy subjects in the DR group [5].

We should also point out that we considered manually annotated images by expert clinicians as a reference standard. Although this may reflect better a screening scenario, in a clinical situation slit-lamp biomicroscopy performed by

ophthalmologists should be considered as the reference standard [4,5]. Olson et al. [4] compared slit-lamp biomicroscopy with optometrists, conventional photography (colour slides), manually graded digital images and automatically graded digital images. When detecting DR (based on the detection of MAs and HEs), their automatic method reached $SE_i = 83\%$ and $SP_i = 71\%$ using two-field images and $SE_i = 79\%$ and $SP_i = 80\%$ when only one image was used. Moreover, the automatic method used in [4] achieved a higher sensitivity than optometric examination (with $SE_i = 75\%$ for two field images, no results with one image were reported). However, specificity values did not reach those of other modalities ($SP_i = 89\%$ for manual examination using colour slides, $SP_i = 79\%$ for manual examination of digital images and $SP_i = 82\%$ for optometric examination) [5]. Our results correlate with these findings, where we have obtained $SE_i = 100\%$ for the three NNs but lower SP_i values.

The proposed algorithm presents some limitations that should be mentioned. In spite of the normalization step, the colour and size of EXs vary even within the same eye. This makes it difficult to detect all of them. Failing to detect some faint EXs in an image may not be crucial for the diagnosis of DR because EXs usually appear in groups. However, if only a few dim EXs appeared in the images, our method could show a limited performance. Additionally, some problems may arise from the size of the database. Firstly, although some previous works use a training set of a size similar to ours [20], it may not be enough to represent the great variability of EXs. It would be desirable, upon availability of more images, to train the classifiers on a larger data set and verify if the same optimum points are obtained. Secondly, the test set is not large enough to draw general conclusions about the diagnostic value of our method. This fact would be especially meaningful if the method was applied in a screening program. The rate of DR cases in a screening data set is generally lower than the rate of healthy images [4,7]. In our case, where images come from a clinical database, the number of images from DR patients exceeds the number of images from healthy subjects. It would be necessary to assess the performance of our method on a larger database of images with variable degrees of DR and that includes images with drusen. Finally, it is worth noting that some parameters of the method may need to be tuned when images with different size and quality are studied. One is the size of the square blocks employed for the luminosity and contrast normalization stage. It must be large enough to contain sufficient background pixels in each square but low enough to consider that C and L vary slowly in each square [21]. Therefore, if very different image sizes were to be used, it should be adequately adapted. The segmentation threshold of 10% may also be decreased for higher quality images in order to improve specificity while maintaining also a high sensitivity. A final consideration is that, as mentioned previously, the optimum operating points of the NNs may be different when a new training set is used.

Future works will be aimed at improving our results by studying the discriminatory power of new features [6] for this task. We will also try to detect other types of lesions associated with DR (CWs, HEs and MAs) in order to separate healthy subjects from those who suffer from DR more precisely and grade the evolution of the disease.

Conflict of interest

None.

Acknowledgements

REFERENCES

[1] G.L. Ong, L.G. Ripley, R.S. Newsom, M. Cooper, A.G. Casswel, Screening for sight-threatening diabetic retinopathy: comparison of fundus photography with automated color contrast threshold test, Am. J. Ophthalmol. 137 (2004) 445–452.

[2] D.S. Fong, L. Aiello, T.W. Gardner, G.L. King, G. Blankenship, J.D. Cavallerano, F.L. Ferris, R. Klein, Diabetic retinopathy, Diabetes Care 26 (2003) 226–229.

[3] D.Y. Lin, M.S. Blumenkranz, S.J. Brothers, D.M. Grosvenor, 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. 134 (2002) 204–213.

[4] J.A. Olson, F.M. Strachan, J.H. Hipwell, K.A. Goatman, K.C. McHardy, J.V. Forrester, P.F. Sharp, A comparative evaluation of digital imaging, retinal photography and optometrist examination in screening for diabetic retinopathy, Diabet. Med. 20 (2003) 528–534.

[5] N. Patton, T.M. Aslam, T. MacGillivray, I.J. Deary, B. Dhillon, R.H. Eikelboom, K. Yogesan, I.J. Constable, Retinal image analysis: concepts, applications and potential, Prog. Retin. Eye Res. 25 (2006) 99–127.

[6] M. Niemeijer, B. van Ginneken, S.R. Russell, M.S.A. Suttorp-Schulten, M. 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. 48 (2007) 2260–2267.

[7] R. Klein, B.E. Klein, S.E. Moss, M.D. Davis, D.L. DeMets, The Wisconsin epidemiologic study of diabetic retinopathy VII. Diabetic nonproliferative retinal lesions, Ophthalmology 94 (1987) 1389–1400.

[8] N.P. Ward, S. Tomlinson, C.J. Taylor, Image analysis of fundus photographs. The detection and measurement of exudates associated with diabetic retinopathy, Ophthalmology 96 (1989) 80–86.

[9] R. Phillips, J. Forrester, P. Sharp, Automated detection and quantification of retinal exudates, Graefe's Arch. Clin. Exp. Ophthalmol. 231 (1993) 90–94.

[10] K. Akita, H. Kuga, A computer method of understanding ocular fundus images, Pattern Recogn. 15 (1982) 431–443.

[11] H. Li, O. Chutatape, Fundus image features extraction, in: Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 2000, pp. 3071–3073.

[12] H. Li, O. Chutatape, Automated feature extraction in color retinal images by a model based approach, IEEE Trans. Biomed. Eng. 51 (2004) 246–254.

[13] G. Zahlman, B. Kochner, I. Ugi, D. Schuhmann, B. Lisenfeld, A. Wegner, M. Obermaier, M. Mertz, Hybrid fuzzy image

processing for situation assessment: a knowledge-based system for early detection of diabetic retinopathy, IEEE Eng. Med. Biol. Mag. 19 (2000) 76–83.

[14] T. Walter, J.C. Klein, P. Massin, A. Erginay, A contribution of image processing to the diagnosis of diabetic retinopathy—detection of exudates in color fundus images of the human retina, IEEE. Trans. Med. Imag. 21 (2002) 1236–1243.

[15] H. Wang, W. Hsu, K.G. Goh, M.L. Lee, An effective approach to detect lesions in color retinal images, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head Island, SC, USA, 2000, pp. 181–186.

[16] B.M. Ege, O.K. Hejlesen, O.V. Larsen, K. Møller, B. Jennings, D. Kerr, D.A. Cavan, Screening for diabetic retinopathy using computer based image analysis and statistical classification, Comput. Methods Programs Biomed. 63 (2000) 165–175.

[17] C.I. Sánchez, R. Hornero, M.I. López, M. Aboy, J. Poza, D. Abásolo, A novel automatic image processing algorithm for detection of hard exudates based on retinal image analysis, Med. Eng. Phys. 30 (2008) 350–357.

[18] G.G. Gardner, D. Keating, T.H. Williamson, A.T. Elliot, Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool, Br. J. Ophthalmol. 80 (1996) 940–944.

[19] X. Zhang, O. Chutatape, 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, San Diego, CA, USA, 2005, pp. 181–186.

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

[21] M. Foracchia, E. Grisan, A. Ruggeri, Luminosity and contrast normalization in retinal images, Med. Image Anal. 9 (2005) 179–190.

[22] M.J. Cree, J.A. Olson, K.C. McHardy, P.F. Sharp, J.V. Forrester, The preprocessing of retinal images for the detection of fluorescein leakage, Phys. Med. Biol. 44 (1999) 293–308.

[23] E. Grisan, A. Ruggeri, Segmentation of candidate dark lesions in fundus images based on local thresholding and pixel density, in: Proceedings of the 29th Annual International Conference of the IEEE EMBS, Lyon, France, 2007, pp. 6735–6738.

[24] S. Chaudhuri, S. Chaterjee, N. Katz, M. Nelson, M. Goldbaum, Detection of blood vessels in retinal images using two-dimensional matched filters, IEEE Trans. Med. Imag. 8 (1989) 263–269.

[25] M. Nadler, E. Smith, Pattern Recognition Engineering, John Wiley, New York, 1993.

[26] M. Sonka, V. Hlavac, R. Boyle, Image Processing, Analysis and Machine Vision, International Thomson Computer Press, London, 1996.

[27] I.T. Nabney, NETLAB Algorithms for Pattern Recognition, Springer, London, 2004.

[28] Intelligent Data Analysis Group, SVM Toolbox for Matlab, http://www.kernel-machines.org, 2002.

[29] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice-Hall International, New Jersey, 1999.

[30] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, New York, 2004.

[31] K. Hornik, Approximation capabilities of multilayer feedforward networks, Neural Netw. 4 (1991) 251–257.

[32] G.B. Huang, Y.Q. Chen, H.A. Babri, Classification ability of single hidden layer feedforward neural networks, IEEE Trans. Neural Netw. 11 (2000) 799–801.

[33] A.N. Tikhonov, On solving incorrectly posed problems and method of regularization, Doklady Akad. Nauk 151 (1963) 501–504.

[34] S. Chen, C.F.N. Cowan, P.M. Grant, Orthogonal least squares learning algorithm for radial basis function networks, IEEE Trans. Neural Netw. 2 (1991) 302–309.

[35] B. Schölkopf, K.K. Sung, C.J.C. Burges, F. Girosi, P. Niyogi, T. Poggio, V. Vapnik, Comparing support vector machines with Gaussian kernels to radial basis function classifiers, IEE Trans. Signal Process 45 (1997) 2758–2765.

[36] J. Moody, Prediction risk and architecture selection for neural networks, in: V. Cherkassky, J.H. Friedman, H. Wechsler (Eds.), From Statistics to Neural Networks: Theory and Pattern Recognition Applications, Springer-Verlag, Berlin, 1994.

[37] J.E. Moody, The effective number of parameters: an analysis of generalization and regularization in nonlinear learning systems, in: J.E. Moody, S.J. Hanson, R.P. Lippmann (Eds.), Advances in Neural Information Processing Systems 4, Morgan Kaufmann, San Mateo, 1992, pp. 847–854.

[38] J.C. Javitt, J.K. Canner, R.G. Frank, D.M. Steinwachs, A. Sommer, Detecting and treating retinopathy in patients with type I diabetes mellitus. A health policy model, Ophthalmology 97 (1990) 483–494.

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