Automatic Identification of Diabetic Retinopathy in Retinal Images Using Ensemble Learning

Flávio Araújo[1], **Romuere Silva**[1], **André Macedo**[1], **Kelson Aires**[1], **Rodrigo Veras**[1]

[1]Computer Department, Federal University of Piauí, Teresina - Brazil

Abstract. *Diabetic Retinopathy (DR) is one of the major complications of diabetes mellitus and can cause blindness. The diagnosis of DR is performed by visual analysis of retinal images being exudates (fat deposits) the mains patterns traced by a specialist doctor. This paper presents a new method for DR detection in color retinal images. The proposed algorithm combines two images classification methodologies using an ensemble learning. The first Methodology extracts the image attributes using Speeded Up Robust Features (SURF) algorithm and determines the presence of DR using a Support Vector Machine (SVM). The second Methodology consists in the classification of an image based on the segmentation of exudates regions. The experimental validation was performed on a public image database, DIARETDB1.*

1. Introduction

The retina is the most internal membrane of the human eye. Digital images of the fundus eye can provide information on pathological changes caused by eye and systemic diseases such as hypertension, arteriosclerosis and diabetes mellitus [Bernardes et al. 2011], [Gutterres et al. 2011].

In many situations, digital image processing techniques can be used to detect changes in the retina. The identification of pathologies of the human eye using these techniques have increased day by day and the main applications intends to identify three diseases: Diabetic Retinopathy (DR), Macular Edema and Glaucoma [Giancardo et al. 2012].

Particularly, DR is the largest cause of vision loss in diabetics [Sopharak et al. 2010]. This disease occurs as a result of vascular changes in the retina, causing swelling of capillaries known as microaneurysms (MA). With the progress of the disease these MAs can rupture, and, eventually, become a source of extravasation of plasma creating regions of fat deposits in the retinal well known as exudates [Giancardo et al. 2012].

In the early stages of DR, the ophthalmologists look for MAs which are small and very difficult to detect, visually. Thus, at more advanced stages, experts search for exudates that usually form clusters that can be distributed throughout the retina, are easily visible and its presence indicates that the patient has DR [Sopharak et al. 2010]. In situations that they occur in the macula region can cause vision loss [Giancardo et al. 2012]. The continuous monitoring of DR is very important for the disease can be diagnosed before onset of symptoms. Furthermore, early treatment may prevent or reduce the vision loss of the patient.

The goal of our work is to propose automatic identification of images with diabetic retinopathy using ensemble learning to combine two Methodologies: The first Methodology extracts the image attributes using Speeded Up Robust Features (SURF) algorithm and determines the presence of DR using a Support Vector Machine (SVM); The second Methodology consists in the classification of an image based on the segmentation of exudates regions. Automatic DR detection is capable of supporting medical decision, improving diagnosis quality and furthermore decreasing the workload of medical professionals. In this work an image that has diabetic retinopathy will be call *pathologic*, and an image without diabetic retinopathy will be call *non-pathological*.

2. Related Works

Several methods for exudate detection are available in the literature, and furthermore there are some problem solving strategies [Sopharak et al. 2010].

[Kavitha and Shenbaga 2005] and [María et al. 2009] used multilevel thresholding to extract the Optic Disc (OD) and the exudates. The former work detected the OD as the converging point of the vessels then classify the other bright regions as exudates. The later made a classification with Multilayer Perceptron (MLP) and Support Vector Machine (SVM).

[Osareh et al. 2002], [Zhang and Chutatape 2005] and [Sopharak et al. 2009] proposed the use of fuzzy k-means algorithm to segment the retina in groups with similar colors. The work of [Osareh et al. 2002] used the RGB color model while [Zhang and Chutatape 2005] used the Luv color space. In both works was used a SVM to separate exudate and non-exudate regions. [Sopharak et al. 2009] used four input attributes for the fuzzy k-means: intensity value, standard deviation of intensity, value of hue and number of edge pixels in a region around the pixel.

Recently [Sopharak et al. 2010] and [Harangi et al. 2012] perform a pixel by pixel classification as belonging or not to a exudate region. The former used a SVM classifier and the last a improved Naive-Bayes classifier. Despite the successful results reported, the use of pixel by pixel classification requires a high computational power for training and classification processes. The approach of [Köse et al. 2012] was based on information that the background images of a healthy retina has regular patterns of color and texture. Therefore, the background image was estimated and other patterns were considered abnormal. The method was able to identify the existence or not of exudates, however detection of the exudates regions is not part of the objectives of this work.

[Rocha et al. 2012] and [Ram and Sivaswamy 2009] were the mains works used as base for the development of our system. [Rocha et al. 2012] constructs a visual word dictionary representing points of interest (PoIs) located within regions marked by specialists that contain lesions associated with DR and classifies the fundus images based on the presence or absence of these PoIs as normal or DR-related pathology. [Ram and Sivaswamy 2009] proposed a method of multi-space clustering to exudate segmentation. The clustering was carried out in two spaces of different attributes and then were combined only for exudate segmentation.

3. Proposed System

The automatic identification of images with diabetic retinopathy proposed in this paper, aims to combine two Methodologies using an ensemble learning. The first Methodology consists in apply the SURF algorithm, it will return a set of points of interest, then visual dictionaries are used in order to generate a single attributes vector to be passed to a SVM that will classify the retinal images; the second Methodology consists in use image processing techniques to segmented the regions of exudates then one MLP is used to classify images of the retina. Figure 1 illustrates the steps of the proposed system.

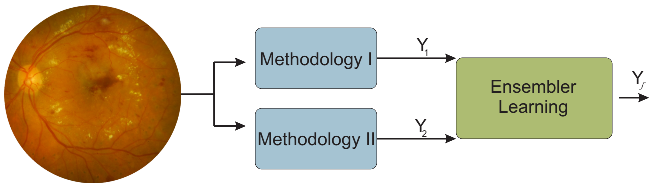


Figure 1. Diagram of the proposed system. Y_1 and Y_2 respectively represent the outputs of the Methodologies I and II, and Y_f is the final output of the system.

3.1. Methodology I

The retinal images classification requires a series of steps. Figure 2 shows a flowchart of how this problem is treated in this Methodology. The attributes extraction is made from an image database, and the visual dictionaries are created from these attributes. These visual dictionaries will generate signatures for each image. After this processing the found values are sent to the classifier which will classify an image as *pathological* or *non-pathological*.

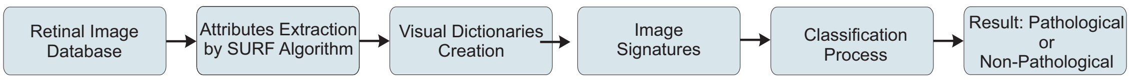


Figure 2. Diagram of the proposed Methodology I.

In this paper, the SURF algorithm [Bay et al. 2006] was used for image attributes extraction. These attributes will generate an attribute vector which will be used by the classifier. There is a problem when we use the SURF algorithm, the attribute vectors generated have a different number of points per image, and this complicates the use of classifiers. The solution of this problems is the use of visual dictionaries [Papa and Rocha 2011].

Visual dictionaries constitute an approach in which each image is treated as a collection of regions. The only important information is the appearance of each region [Agurto et al. 2010]. The visual dictionaries are matrices composed of words (matrix rows) that are made by the most representative image points extracted by the SURF algorithm. These dictionaries are used to capture common properties among regions marked by the SURF algorithm. The dictionary must be large enough to distinguish relevant differences between images but it cannot include irrelevant variations. It is necessary use a clustering algorithm to generate this dictionary. Each value is composed with two types of information, the first half with *pathological* images and the other half with *non-pathological* images. After this each word is compared with each point of others images and thus will form the visual bags (signatures). For the image classification we used the SVM [Haykin 2001].

3.2. Methodology II

The retinal images classification using segmented exudates regions requires a series of steps. Figure 3 shows a flowchart of how this problem is treated in this Methodology. First digital image processing and classification techniques are needed to segment exudate candidates regions, after a new classification is used to decide if an image is *pathological* or *non-pathological*.

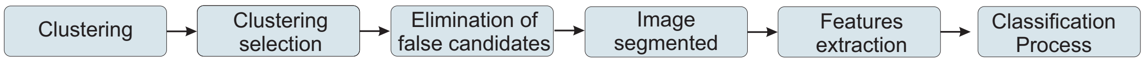


Figure 3. Diagram of the proposed Methodology II.

The first step of the method is the clustering. In this step each retinal image in RGB model was converted to three different color space: Luv, HSV, HSI [Rafael C. Gonzalez 2008]. Using these color spaces were built two feature vector f_1 = (H,S,V,I) and f_2 = (R,G,L,u,v) which were used as input of fuzzy k-means algorithm [Haykin 2001]. The output of this algorithm consists of two images (I_1 and I_2). These images are results of clustering with the feature vector f_1 and f_2.

After obtained the images I_1 and I_2, the clusters that represent the exudates regions were selected. In I_1 the goal was select the clustering corresponding to bright lesions and bright background. Considering that these regions are brightest of the original image it was selected the clustering with the highest intensity value I in HSI space. In I_2 the clustering selected corresponds to the OD and exudates. Because these regions has a yellowish color was selected the clustering of smaller value α for $\alpha = max(R) - max(G)$ in RGB color space. In order to improve the exudates detection are formed two new images I_3 and I_4: the first is formed by all I_2 regions present in I_1 and the second is formed by other I_1 regions not present in I_2. Then, in order to eliminate false candidate regions we followed two steps: the elimination of OD region and the use of classification technical.

We detected the region of the OD, as the focal point of the blood vessels. This technique was used because strategies that use the information that the OD is the region of convergence of the vessels are most successful than techniques based only on the image color properties. Therefore, the blood vessels of the original image (I_o) are segmented by the algorithm proposed by [Zana and Klein. 2001] resulting in a vessel image (I_v). Thereafter, the vessels in I_v are converted in straight lines by application of Hough transform (resulting in I_l) and was performed a search for three square windows of side equal to half radius of the OD (70 pixels) with the largest amount and of straights in I_l. The OD center is chosen as the center of the window that has a higher quantity of white pixels on I_v. This choice was made due to fact that the vessels that converge to OD are greater caliber. The elimination of the OD was performed by removal of the region connected to the center.

Towards to eliminate other false candidate regions were used classification techniques. We tested the classifiers K-Nearest Neighbor (KNN), Multilayer Perceptron (MLP), Radial Basis Function (RBF) and Support Vector Machine (SVM). Details about these classifiers can be found in [Haykin 2001].

After the exudate candidates regions segmentation, we used a MLP for classify the images in *pathological* and *non-pathological*.

3.3. Ensemble Learning

There is not a single classifier that can be considered optimal for all problems [Ponti 2011], therefore, we used an ensemble learning to combine the presented Methodologies. There is no clear guideline to choose a set of learning methods and it is rare when one has a complete knowledge about data distribution and also the about the details of how the classification algorithm behaves. Therefore, in practical pattern classification tasks it is difficult to find a good single classifier [Ponti 2011].

The result combination of the two presented Methodologies is performed using an ensemble learning. In this technical, the responses of two or more predictors (experts) are combined by means of a mechanism that does not involve the input signal [Haykin 2001]. In this paper, we used the Equation 1 for combining the results of two specialists.

$$Y_f = \frac{(Y_1 * A_1) + (Y_2 * A_2)}{A_1 + A_2} \tag{1}$$

where, A_1 e A_2 are respectively the individual accuracy of the Methodologies I and II; Y_1 and Y_2 are respectively the outputs of the Methodologies I and II; Y_f is the combination of Y_1 and Y_2.

4. Results

In order to evaluate algorithm performance, we used sensitivity (S), specificity (SP), positive predictive value (PPV), negative predictive value (NPV) and accuracy (A). All these measures can be calculated based on four values: true positive (TP), the number of images correctly classified as pathological; false positive (FP), the number of non-pathological images wrongly classified as pathological; false negative (FN), the number of pathological images wrongly classified as non-pathological and true negative (TN), the number of images correctly classified as non-pathological [Ram and Sivaswamy 2009].

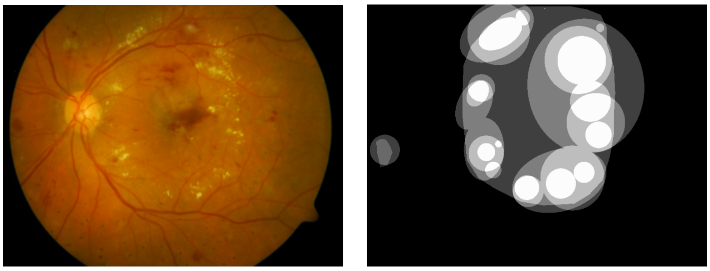
4.1. Image Database

We tested our approach on the publicly available DIARETDB1 color fundus image database [Kauppi et al. 2007]. The DIARETDB1 consists of 89 imagens, all of same size (1500 $\times$ 1152). It is used in detection of exudates works, because it presents the ground-truth spatial coordinates of findings related to four pathologies: hemorrhage, hard exudates, soft exudates and red spots [Giancardo et al. 2012].

On this database, the marking of pathologies was performed by four ophthalmologists, in some images, there was no consensus among all of them. This work considered exudates only regions marked by three of the four ophthalmologists, as suggested by the database authors. Figure 4(a) shows an example of image of DIARETDB1. Figure 4(b) shows the marking of regions of exudates. Lighter regions represent more agreement in the diagnosis. Regions in white represent areas where there was 100% agreement among ophthalmologists.

4.2. Methodology I Results

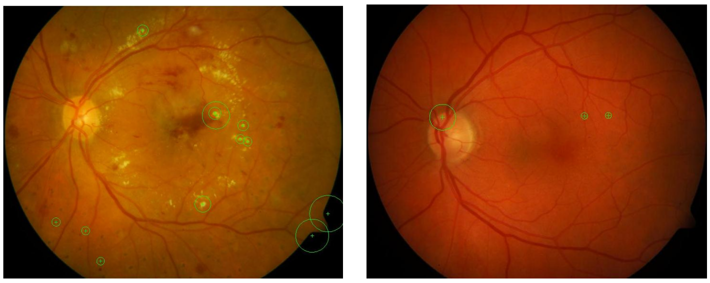
The SURF descriptor was tested with 2, 3 and 4 octaves. Taking into consideration the accuracy rates obtained and the run time to generate results, it was concluded that the



(a) (b)

Figure 4. DIARETDB1 image: a) retina image with various exudates and bleeding, b) ground-truth image of Figure 4(a).

best results was obtained using 4 octaves. To generate the visual dictionaries we use the K-means as clustering algorithm with the K value equal to 128, we test others K values (32,64,256) but this was the valeu with the best result.



(a) (b)

Figure 5. Surf points detected: a) Surf points for the pathological image, b) Surf points for the non-pathological image.

The SVM classifier was performed using linear kernel function. The better results were obtained from these functions in comparing to radial basis functions and tangent sigmoid. The weight to use for the classes was 2000, and the cost parameter was 1.

In order to evaluate classification performance, we used to SVM training 10% of the images from the database. We used 9 retinal images: 4 pathological and 5 non-pathological. To test the classifier, we used all the other images of the database: 35 pathological and 45 non-pathological.

The classifier accuracy was 81.25%. Table 2 shows the values for Sensitivity, Specificity, PPV, NPV and Accuracy.

4.3. Methodology II Results

Towards to initial reference for adjustment of fuzzy k-means, were used parameters presented in [Sopharak et al. 2009]: fuzzy degree 2, number of interactions 200 and maximum error 10^{-6}. For this paper, variations of these parameters were tested aiming find values where there was no difference in the clustering. Thus, the fuzzy degree used was maintained in 2, the maximum number of interactions was fixed in 2000 and maximum error was 10^{-8}. The choice of these values ensured that the randomness of the choice of initial clusters do not interfere in the final result of the cluster. In other works, independent of the choice of initial clusters, the algorithm always converges to the same result.

In order to classified the exudates regions, we use classical literature features [Osareh et al. 2002] and [Sopharak et al. 2010]. These features were divided into two groups: NON-COLOR (6 features) - area, perimeter, circularity, homogeneity and x, y coordinates of the region center; COLOR (18 features) - average and standard deviation of the all component of color model RGB, Luv and HSI. However, tests shown that elimination of some features used initially improved the quality of the classification resulting in final set with 12 features (5 of non-color group and 7 of color group). These 12 features that performed best are: area, perimeter, circularity, average of components (L), (u), (v), (H), (I), (G), standard deviation of the component (G), and x and y coordinates of the region center.

After all steps of the image processing, the images resulting from the algorithm execution contained 6.835 candidate regions. 484 were exudates and 6351 were non-exudates. Aiming to keep the proportion between the number of regions of exudates and non-exudates, we created a set of training with regions present in 25 images: 20 pathologic images (with regions of exudates and non-exudates) and 5 healthy images (with only regions of non-exudates). As the proportion of regions of non-exudates is higher than exudates, in each of the 20 pathologic images, we used for training the percentage of 30% more of non-exudates regions.

The feature vector used for training had 226 exudates and 280 non-exudates candidates. The data used for validation was created with all candidates remaining of the other images, containing a total of 6104 candidates (222 exudates and 5882 non-exudates). The result of this classification is shown in Table 1. The MLP used had one hidden layer with 9 neurons, the learning rate was 0.05 and the mean squared error was 0.01.

Table 1. Classification evaluation results.

	KNN		MLP		RBF		SVM	
	Exu.	N-Exu.	Exu.	N-Exu.	Exu.	N-Exu.	Exu.	N-Exu.
Exu.	21.62%	78.38%	**97.29**%	2.71%	2.70%	97.30%	77.03%	22.97%
N-Exu.	2.50%	97.50%	18.43%	**81.57**%	0.22%	99.78%	16.68%	83.32%

Analyzing the results of Table 1 we observe that the MLP presented the best performance: 97.29% of exudates and 81.57% of non-exudates. Thus, we choose the MLP to classify candidates regions. In a classification performed with candidates of images that had only exudates the success rate was 97.83% and for images that had only non-exudates the success rate was 82.42%.

In order to evaluate the retinal images classification, we extract two attributes from the segmented images: number of regions before the first classification and number of regions removed by classification. We used 45 retinal images for training set: 21 pathological and 24 non-pathological. To test the classifier, we used all the other images of the database: 18 pathological and 26 non-pathological. The MLP used had one hidden layer with 3 neurons, the learning rate was 0.05 and the mean squared error was 0.01.

The classifier accuracy was 84.09%. Table 2 shows the values for Sensitivity, Specificity, PPV, NPV and Accuracy.

4.4. Ensemble Learning

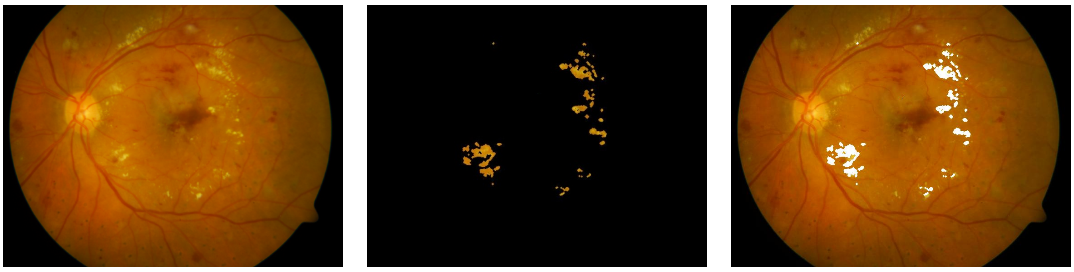
In order to evaluate committee machine performance, we used 50% of the images from the database. It is highlighted that these images were not present in the classifiers training set of the Methodologies I and II. Table 2 shows the results for Methodologies I and II and for the combination these Methodologies using committee machine.

Table 2. Classification Evaluation Results

	S	SP	PPV	NPV	A
Methodology I	80%	83.33%	88.89%	71.43%	81.25%
Methodology II	88%	78.95%	84.62%	83.33%	84.09%
Committee Machine	83.87%	100%	100%	73.68%	88.89%

Analyzing the results of Table 2 we observe that the combination using committee machine increases the individual performance of Methodologies I and II. Our method presented better results than other algorithm in the literature as [Osareh et al. 2002], [Walter et al. 2002], [Sopharak et al. 2009] and [Ram and Sivaswamy 2009].

If after the combination of Methodologies I and II using committee machine the image is classified as pathologic, then the segmented regions using Methodology II are marked as exudates, as shows Figure 6.



(a) (b) (c)

Figure 6. Application of algorithm: a) original image, b) regions segmented as exudates using Methodology II, c) overlaid the result in the original image.

5. Conclusion and Future Work

This paper presented an extended Methodology for diabetic retinopathy identification in retinal images that used ensemble learning for combined two Methodologies. The first Methodology consist in apply the algorithm SURF which will return a number of points of interest, after this the visual dictionaries are used in order to generate a single feature vector to be passed to the classifier that will classify the retina in *pathological* or *non-pathological*; the second Methodology consist in to segment the retinal images to identify exudates regions, then a classifier is used to classify the images in *pathological* or *non-pathological*

Analyzing the obtained results, it is verified that combination using committee machine increased the accuracy of the two Methodologies (improvement of 7.09% comparing as Methodology I and 4.8% comparing as Methodology II).

As future work we propose the study of principal component analysis (PCA) and of new descriptors. The goal of PCA is eliminate to features the least influential in the classification, in order to decrease network complexity and increased its efficiency. The use of new descriptors can increase the efficiency of classification and consequently the result of classification. Another future work proposal is the identification of others retinal pathologies, as: hemorrhage, red spots, macular edema and glaucoma.

References

[Agurto et al. 2010] Agurto, C., Murray, V., Barriga, E., Murillo, S., Pattichis, M., Davis, H., Russel, S., Abràmoff, M., and Soliz, P. (2010). Multiscale am-fm methods for diabetic retinopathy lesion detection. *IEEE Trans. on Medical Imaging*, pages 502–512. 3

[Bay et al. 2006] Bay, H., Tuytelaars, T., and Gool, L. V. (2006). Surf: Speeded up robust features. In *In ECCV*, pages 404–417. 3

[Bernardes et al. 2011] Bernardes, R., Serranho, P., and Lobo, C. (2011). Digital ocular fundus imaging: A review. *Ophthalmologica*, 4(226):161–181. 1

[Giancardo et al. 2012] Giancardo, L., Meriaudeau, F., Karnowski, T. P., Li, Y., Garg, S., Tobin, K. W., and Chaum, E. (2012). Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Medical Image Analysis*, 16(1):216–226. 1, 5

[Gutterres et al. 2011] Gutterres, D. M., Maestri, M., Feller, N., Nedel, L., and Duncan, B. (2011). Análise de imagens de retinografia digital com imagej customizado. *Workshop de Informática Médica*. 1

[Harangi et al. 2012] Harangi, B., Antal, B., and Hajdu, A. (2012). Automatic exudate detection with improved naïve-bayes classifier. In *IEEE International Symposium on Computer-Based Medical Systems*, pages 1–6. 2

[Haykin 2001] Haykin, S. (2001). *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 2 edition. 3, 4, 5

[Kauppi et al. 2007] Kauppi, T., Kalesnykiene, V., Kamarainen, J. K., Lensu, L., Sorri, I., Kalviainen, H., and Pietila, J. (2007). Diaretdb1 diabetic retinopathy database and evaluation protocol. In *Conference on Medical Image Understanding and Analysis*, pages 1–5. 5

[Kavitha and Shenbaga 2005] Kavitha, D. and Shenbaga, S. (2005). A cellular neurofuzzy network for contrast enhancement of fundus images with retinopathies. In *Proceeding ICS'06 Proceedings of the 10th WSEAS international conference on Systems*. 2

[Köse et al. 2012] Köse, C., Sevik, U., Ikibas, C., and Erdöl, H. (2012). Simple methods for segmentation and measurement of diabetic retinopathy lesions in retinal fundus images. In *Computer Methods and Programs in Biomedicine*, pages 274–293. 2

[María et al. 2009] María, G., Clara, I. S., María, I. L., Daniel, A., and Hornero, R. (2009). Neural network based detection of hard exudates in retinal images. *Computer Methods and Programs in Biomedicine*, 93:9–19. 2

[Osareh et al. 2002] Osareh, A., Mirmehdi, M., Thomas, B., and Markham, R. (2002). Comparative exudate classification using support vector machines and neural networks. In *5th International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2, 7, 8

[Papa and Rocha 2011] Papa, J. and Rocha, A. (2011). Image categorization through optimum path forest and visual words. In *Intl. Conference on Image Processing (ICIP)*, pages 3586–3589. 3

[Ponti 2011] Ponti, M. (2011). Combining classifiers: From the creation of ensembles to the decision fusion. In *(SIBGRAPI-T)*, pages 1–10. 5

[Rafael C. Gonzalez 2008] Rafael C. Gonzalez, R. E. W. (2008). *Digital Image Process*. Prentice Hall, 3 edition. 4

[Ram and Sivaswamy 2009] Ram, K. and Sivaswamy, J. (2009). Multi-space clustering for segmetation of exudates in retinal color photographs. In *Annual Int. Conf. of the IEEE Enginnersing in Medicine and Biology*, pages 1437–1440. 2, 5, 8

[Rocha et al. 2012] Rocha, A., Carvalho, T., Jelinek, H., Goldenstein, S., and Wainer, J. (2012). Points of interest and visual dictionaries for automatic retinal lesion detection. *IEEE Transactions on Biomedical Engineering*, 59(8):2244–2253. 2

[Sopharak et al. 2010] Sopharak, A., Dailey, N. M., Uyyanonvara, B., Barman, S., Williamson, T., New, T. N., and Moe, A. Y. (2010). Machine learning approach to automatic exudate detection in retinal images from diabetic patients. *Journal of Modern Optics*, (57):124–135. 1, 2, 7

[Sopharak et al. 2009] Sopharak, A., Uyyanonvara, B., and Barman, S. (2009). Automatic exudates detection from non-dilated diabetic retinopathy retinal images using c-means clustering. *Sensors*, pages 2148–2161. 2, 6, 8

[Walter et al. 2002] Walter, T., Klevin, J., Massin, P., and Erginay, A. (2002). A contribution of image processing to the diagnosis of diabetic retinopathy-detection of exudates in color fundus images of the human retina. *IEEE Transactions on Medical Imaging*, 21(10):1236 – 1243. 8

[Zana and Klein. 2001] Zana, F. and Klein., J. C. (2001). Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE Transactions on Image Processing*, 10(7):1010 – 1019. 4

[Zhang and Chutatape 2005] Zhang, X. and Chutatape, O. (2005). Top-down and bottom-up strategies in lesion detection of background diabetic retinopathy. In *CVPR*. 2