Classification of Diabetic Retinopathy using Machine Learning

Pilar Pérez Conde, Jorge de la Calleja, Ma. Auxilio Medina, Antonio Benitez
Department of Masters
Universidad Politécnica de Puebla, Puebla, Mexico
{gperez,jdelacalleja,mmedina,abenitez}@uppuebla.edu.mx

***Abstract*—This paper presents a method to classify diabetic retinopathy using fundus images. In our study we categorize the disease into two classes: diabetic retinopathy non-proliferative and diabetic retinopathy proliferative. The method reduces the dimensionality of the images and find features using the statistical method of principal component analysis (PCA). Then, we classify the images using decision trees, the naive Bayes classifier, neural networks, *k*-nearest neighbors and support vector machines. The experimental results show that the naive Bayes classifier obtains the best results with 73.4% of accuracy using a data set of 151 images and testing with different resolutions.**

***Index Terms*—Medical Image Analysis, Machine Learning, Principal Analysis Component, Diabetic Retinopathy.**

I. Introduction

Actually blindness has increased significantly in many countries due to a large augment in diabetic retinopathy, among others causes. According to the World Health Organization (WHO), there are 285 million people worldwide approximately with visually impaired, 39 million are blind and 246 million have low vision; a factor for this increase is the diabetic retinopathy [1]. For example, the presence of diabetes in Mexico has increased in last years, showing amounts between 6.5 and 10 million people with this disease [2], [3], that implies more patients with diabetic retinopathy in its different stages.

One way to identify the presence of diabetic retinopathy is by analyzing the retina using fundus images, that allow experts to find anomalies such as micro-aneurysms, soft exudates, hard exudates and hemorrhages.

Frequently, the classification of the diabetic retinopathy is done manually, however, this is not easy because it requires of skills and experience, and this is also a time-consuming task. Instead, automated methods are more impartial than humans, i.e. they are not subject to the conscious and unconscious prejudices which may affect humans when they are looking medical images.

Several works using machine learning methods have been proposed in order to classify some types of eye diseases. In 2011 *Osareh A. et al* [4] proposed an automatic method to detect exudate regions. They introduced comprising image color normalization, enhancing the contrast between the objects and background, segmenting the color retinal images into homogenous regions using fuzzy c-means clustering, and classifying the regions into exudates and non exudates patches using a neural network. This work achieved 92% sensitivity and 82% specificity. *Niemeijer M. et al* [5] in 2007 proposed an automated system able to detect exudates and cotton-wool spots in color images. Their approach uses *k*-nearest neighbors as the machine learning algorithm. Their system achieved an area under the receiver operating characteristic curve (ROC) of 0.95 and sensitivity/specificity pairs of 0.95/0.88 for the detection of bright lesions of any type, and 0.95/0.86, 0.70/0.93 and 0.77/0.88 for the detection of exudates, cotton-wool spots and drusen, respectively. In 2010, *Silberman N. et al* [6] described an automated system to detect diabetic retinopathy from retinal images. This approach used support vector machines to recognize exudates. They performed an evaluation of optic disc detection. The results obtained were 98.4% for optic disk detection and 87% for exudates detection. In 2011, *Karegowda A. et al* [7] attempted to detect exudates using back propagation neural networks. The significant features were identified from preprocessed images by using two methods: decision trees and genetic algorithms. Their method was able to detect exudates and non-exudates at pixel level. This method showed its best performance with sensitivity of 96.97%, specificity of 100% and classification accuracy of 98.45%. *Kavitha S.* and *Duraiswamy K.* [8] in 2011focused on automatic detection of diabetic retinopathy exudates in color fundus retinal images. A series of experiments on classification of hard and soft exudates were performed using image processing techniques. Exudates were detected with the aid of thresholding color histogram. The overall sensitivity, specificity and accuracy were 89.78%, 99.12% and 99.07%, respectively.

In this work we propose an automated method to classify diabetic retinopathy using fundus images. This method is divided into two stages: the first one is image reduction and feature extraction, and the second one is the classification itself. For the first stage we have used principal component analysis, while decision trees, *k*-nearest neighbors, the naive Bayes classifier, neural networks and support vector machines are used in the second stage. In contrast to the related work, we automatically obtain features that allow us to classify the images into two classes: diabetic retinopathy non-proliferative and diabetic retinopathy proliferative. We also test the method with different image resolutions, and varying the number of

principal components.

The remainder of the paper is organized as follows. In Section 2 we briefly describe the diabetic retinopathy. The proposed method is introduced in Section 3. Experimental results are shown in Section 4. Finally we include in Section 5 some conclusions and present directions of future work.

II. Diabetic retinopathy

Diabetes mellitus is a metabolic disorder which results from a defect in insulin synthesis and secretion or from a resistance of the receptors on target tissues for this hormone. The most significant complication of diabetes mellitus involving the eye and which develops in 85% of all diabetics patients eventually is the retinopathy [9].

Diabetic retinopathy is an abnormality involving the small blood vessels that targets the central region of the eye, for example the macula. In fact, the diabetic retinopathy is a progressive disease and this is the principal factor that causes blindness.

According to the Early Treatment Diabetic Retinopathy Study (ETDRS), the diabetic retinopathy is divided as follows [10]:

- Diabetic retinopathy non-proliferative (DRNP), which is subdivided into mild, moderate, severe and very severe.
- Diabetic retinopathy proliferative (DRP), which is subdivided into early, high-risk and advanced.

III. The automated method to classify diabetic retinopathy

We propose an automated method to classify fundus images of diabetic retinopathy into two types: DRNP and DRP. This method is divided into two stages: 1) image reduction and feature extraction using principal component analysis; and 2) image-based classification using machine learning algorithms, particularly decision trees, *k*-nearest neighbors, the naive Bayes classifier, neural networks and support vector machines. Figure 1 shows the stages of our method. The next subsections describe in detail each one of them.

A. Image reduction and feature extraction

Image reduction and feature extraction is based on the idea of M. Turk and A. Pentland, of creating eigenfaces [11], but in our case we would be creating eigenfundus images. Therefore, we need to extract significant information of an image, encode it as efficiently as possible, and compare one encoded image with a data set using a model encoder based on similarly.

We have used principal component analysis in order to find the principal components of the distribution of fundus images (features) or the eigenvectors of the covariance matrix of the set of fundus images, treating an image as a point (or

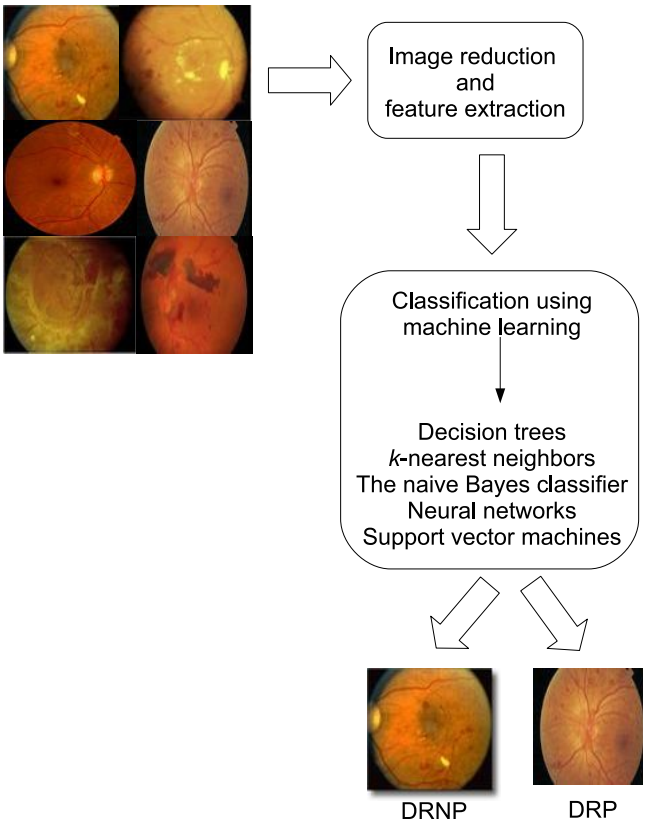


Fig. 1. Stages of the proposed method.

vector) in a very high dimensional space. The eigenvectors are ordered, each one accounting for a different amount of variation among the fundus images.

The basis of the principal component analysis is to find the vectors that best account for the distribution of fundus images within the entire image space. This analysis is as follows:

Given a training data set of fundus images $I_1, I_2, ..., I_M$, where every image I_i will be represented as a vector Γ_i, M is the training data set. Then the average fundus image vector Ψ will be compute as follows:

$$\Psi = \frac{1}{M}\sum_{i=1}^{M}(\Gamma_i) \tag{1}$$

where each fundus image differs from the average by the vector, i.e. it subtracts the mean fundus image:

$$\Phi_i = \Gamma_i - \Psi$$

This set of very large vectors is then subject to principal components analysis, which seeks a set of M ortonormal vectors u_n and the eigenvalues (scalars) λ_k respectively which best describe the distribution of the data. The eigenvectors u_k and eigenvalues λ_k are obtained of the covariance matrix C. The λ_k are the elements in the diagonal of the matrix C.

$$C = \frac{1}{M}\sum_{n=1}^{M}(\Phi_n\Phi_n^T = AA^T) \tag{2}$$

where the matrix $A = [\Phi_1 \Phi_2 ... \Phi_M]$.

The *k-th* vector u_k is chosen such that

$$\lambda_k = \frac{1}{M} \sum_{n=1}^{M} (u_k^T \Phi_n)^2 \quad (3)$$

is a maximum, subject to

$$u_l^T u_k = \delta_{lk} = \begin{cases} 1, & if \quad l = k \\ 0, & otherwise \end{cases} \quad (4)$$

B. Image-based classification

In this second stage we perform the classification task using several machine learning methods. Each algorithm takes the projection of the images onto the principal components as input parameter, and the classification of each one will be the output of this stage.

The classification is done according to the learning processes of each machine learning algorithm, which are described in the next paragraphs.

1) Decision Trees (DT): Decision trees is an algorithm for approximation of discrete-values target functions. This method is robust to noisy and capable of learning disjunctive expressions [12].

DT classify examples by sorting them down the tree from the root to some leaf node, which provides the classification of the example. The classification process starts at the root node of the tree, testing the attribute specified by this node, then moving down the tree branches corresponding to the value of the attribute in the given example. This process is repeated for the subtree rooted at the new node. For details of this algorithm we recommend [12].

2) k-nearest neighbors (k-nn): This algorithm belongs to the instance-based methods, which store all the training examples and delay the classification task until a new instance must be classified [12].

The classification task is as follows: first it finds the set of k nearest neighbors to n where n is a new example to classify. Then the algorithm take the plurality vote of the neighbors (which is the major vote in the case of binary classification). To avoid ties, k always must be chosen as an odd number [13].

k-nn is defined in terms of standard Euclidean distance where $a_r(x)$ is the value of the *i-th* attribute for the instance x and n is the total number of input examples. Therefore, the distance between x_i and x_j is defined as $d(x_i, x_j)$, hence the equation is as follows:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{n} (a_r(x_i) - a_r(x_j))^2} \quad (5)$$

3) Naive Bayes (NB): Naive Bayes is based on the assumption that the quantities of interest are obtained by probability distributions and that optimal decisions can be made by reasoning about these probabilities together with observed data [12].

To determine the best hypothesis from some space H, given the observed training data D, i.e. we choose the most probable hypothesis given the data D plus any initial knowledge about the prior probabilities of the different hypothesis in H. The Bayes Theorem provides a way to compute the probability of an hypothesis based in their priori probability [12].

$$p(h \mid D) = \frac{P(D \mid h) P(h)}{P(D)} \quad (6)$$

where, $P(h)$ indicates the *a priori* probability of h; $P(D)$ is a probability of D without knowledge of h; $P(D|h)$ is a probability of D given h. $P(h \mid D)$ is *a posterior* probability of h.

4) Neural Networks (NNs): Neural Networks are based in the biological processes of human neurons [12]. A neural network is composed by nodes interconnected by directed links. The connection from unit i to unit j helps to propagate the activation a_i from i to j, and each link has a numeric weight $w_{i,j}$ associated with it, which determines the direction and intensity of the connection. Each unit has an artificial input $a_0 = 1$ with an associated weight $w_{0,j}$. Each node j first computes a weighted sum of its inputs in. [13]:

$$in_j = \sum_{i=1}^{n} (w_j a_i) \quad (7)$$

Subsequently it applies an activation function g to this sum to obtain the output:

$$a_j = g(in_j) = \sum_{i=0}^{n} (w_{i,j} a_j) \quad (8)$$

The properties of a neural network are determined by its topology and by the properties of the neurons. There are different topologies such as feedforward (that has connections in only one direction, i.e., this is linearly separable) and the recurrent networks which are not linearly separable and they are composed by several linear functions. A common algorithm to train the networks is base on the gradient descent method, which provides the basis for the back-propagation algorithm [13].

5) Support Vector Machines (SVM): Support vector machines are nonparametric methods which save training examples. They need great potentiality to store all the examples. SVM constructs a maximum separator, a decision boundary with the largest possible distance to example points, they have ability to map the data into a higher-dimensional space using kernels when the data are not linearly separable in the original input space [13].

Thus a maximum separator (hyperplane) is defined as:

$$w \cdot x + b = 0 \tag{9}$$

where x represents some image of the input space; w is a normal vector (weight) and b is the bias term.
The optimal solution to find the parameters that maximizes the margin while correctly classifying all the examples is calculated by the following equation:

$$argmax_\alpha \sum_j \alpha_j - \frac{1}{2} \sum_{j,k} \alpha_j \alpha_k y_j y_k (x_j \cdot x_k) \tag{10}$$

where y is the output f the x input space respectively; $k > 0$ is the fold cross validation. This is subject to the $\alpha_j \geq 0$ and $\sum_j \alpha_j y_j = 0$ constraints.
The weights α_j associated with each data point are zero except for the support vectors, which are the closest points to the separator which keep the separating plane [13]. These support vectors carry on relevant information about the classification task.

When we find a not linearly separable case, we use the kernel functions $K(x_j, x_k)$ to find linear separators in a higher-dimensional feature space. In this case, several kernel exists such as polinomial kernels or gaussian kernels [14].

IV. Experimental results

We performed the experiments using a data set of 151 fundus eye images of a size of 2588×1958 pixels. These images were manually classified by a retinologists into two groups: 109 as DRNP (class 1) and the remaining 42 as DRP (class 2). These images were provided by *El Hospital de Especialidades del IMSS, Puebla, México*.

In order to test our method over different sizes of data sets, we divided the original data set into three subsets: the first one had 42 images (S1), with 21 per class; the second one had 84 images ($S2$), 42 per class; and the third one with the 151 ($S3$), 109 as RDNP and the remainder as DRP. In addition, we tested different resolution sizes for the images, thus images were scaled to a size of 490×490, 979×979 and 1958×1958 pixels, which are labeled as $R1$, $R2$ and $R3$, respectively.
Thus, we obtained the principal components that represented 80%, 85% and 90% of relevant information for the each data set.

The metric used to evaluate the performance of the machine learning methods was accuracy which is described below.

$$Accuracy = \frac{TN + TP}{TN + FP + FN + TP} \tag{11}$$

where TP (True Positive) is the number of correct predictions of a positive example, FP (False Positive) is the number of incorrect predictions of a positive example, TN (True Negative) is the number of correct predictions of a negative instance and FN (False Negative) is the number of correct predictions of a positive instance.

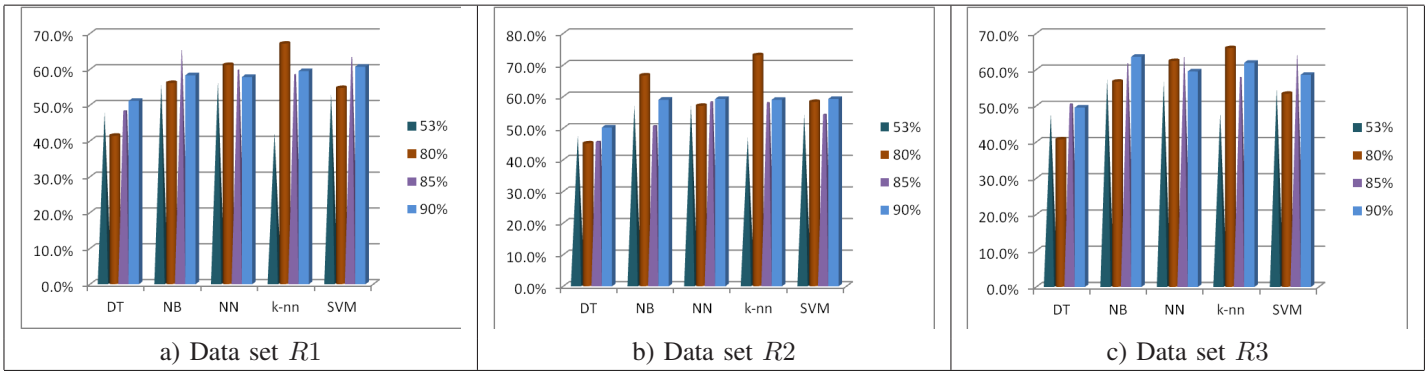
The results that we show later correspond to the average of ten runs. In the following Graphics, we show the accuracy of each data set, and we can observe that the best result are for the data set $S1$ (Figure 2) that was obtained using *k*-nn and $R2$, with 73.1 %; for the data set $S2$ (Figure 3) the best result was obtained using Naive Bayes and $R1$ with 59.9%; and for the data set $S3$ (Figure 4) also Naive Bayes obtained the best result, but using $R3$, with 73.4%.

V. Conclusions

We have presented a method to perform automated classification two types of diabetic retinopathy: DRNP and DRP with fundus images using principal component analysis and machine learning algorithms. The best results were obtained using *k*-nn and Naive Bayes, to obtain them were enough (4 to 5) principal components which represent about 80-85% of the information. Future work will include preprocessing of the images, as well as use a different method to reduce and extract features.

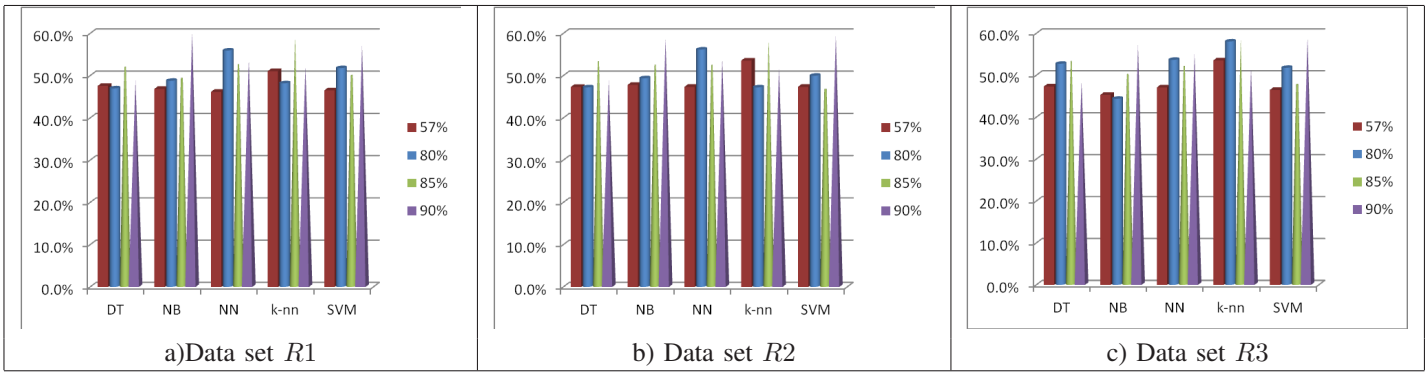
References

[1] WHO, "Diabetes and visual impairment and blindness," 2012. [Online]. Available: http://www.who.int/
[2] INEGI, "Porcentaje de la población con limitación en la actividad según tipo de limitación para cada entidad federativa, 2010." [Online]. Available: http://www.inegi.org.mx
[3] PMFARMA, "Los números de la diabetes en México," Agosto 2007. [Online]. Available: http://www.pmfarma.com.mx/noticias/1359-los-numeros-de-la-diabetes-en-mexico.html
[4] A. Osareh *et al.*, "Automatic recognition of exudative maculopathy using fuzzy c-means clustering and neural networks," in *Medical Image Understanding and Analysis*. BMVA Press, 2001, pp. 49–52.
[5] M. Niemeijer *et al.*, "Automatic detection and differentiation of drusen, exudates, and cotton-wool spots in digiatl color fundus photographs for diabetic retinopathy diagnosis," *Investigative Opthalmology & Visual Science*, vol. 48, no. 5, May 2007.
[6] N. Silberman *et al.*, "Case for automated detection of diabetic retinopathy," in *AAAI Spring Symposium on AI for Development*, 2010.
[7] A. Gowda and e. a. A. Nasiha, "Exudates detection in retinal images using back propagation neural network," *International Journal of Computer Applications*, vol. 25, no. 3, July 2011.
[8] S. Kavitha and K. Duraiswamy, "Automatic detection of hard and soft exudates in fundus images using color histogram thresholding," *European Journal of Science Research*, vol. 48, no. 3, pp. 493–504, 2011.
[9] K. Hem, V. Lalit, and V. Pradeep, *Fluorescein Angiography A users manual*, New Delhi 110 002, India, 2008.
[10] S. Bonafarte, C. García, and M. Davis, *Retinopatía Diabética*, España, S.A., 2006.
[11] M. Turk and A. Pentland, "Eigengaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
[12] T. Mitchell, *Machine Learning*, United States, 1997.
[13] S. Russell and P. Norving, *Artificial Intelligence, a modern approach.*, New Jersey, 2010.
[14] J. Burges Christopher, "A tutorial on support vector machines for recognition," *Data mining and knowledge Discovery*, vol. 2, pp. 121–167, 1998.



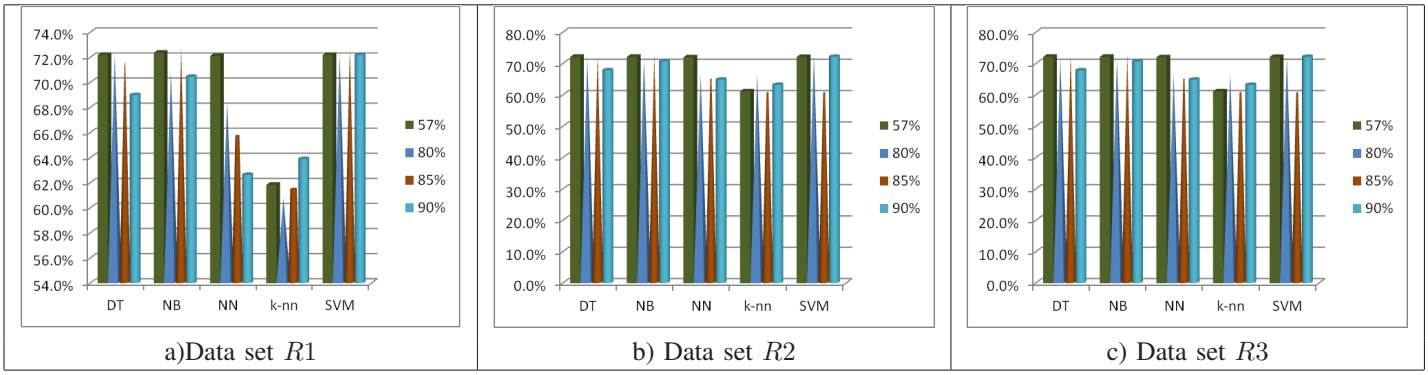
a) Data set $R1$ b) Data set $R2$ c) Data set $R3$

Fig. 2. Accuracy (%) of data set $S1$ (53%, 80%, 85% and 90% represent the relevant information).



a)Data set $R1$ b) Data set $R2$ c) Data set $R3$

Fig. 3. Accuracy (%) of data set $S2$ (57%, 80%, 85% and 90% represent the relevant information).



a)Data set $R1$ b) Data set $R2$ c) Data set $R3$

Fig. 4. Accuracy (%) of data set $S3$ (57%, 80%, 85% and 90% represent the relevant information).