STUDIA INFORMATICA 2004
Volume 25 Number 3 (59)

Katarzyna STĄPOR
Politechnika Śląska, Instytut Informatyki
Adrian BRUECKNER
Uniwersytet Śląski, Instytut Matematyki

CLASSIFICATION OF FUNDUS EYE IMAGES USING SUPPORT VECTOR MACHINES FOR SUPPORTING GLAUCOMA DIAGNOSIS

Summary. In this paper the new method for automatic classification of fundus eye images into normal and glaucomatous ones is proposed. The cup region is automatically segmented from fundus eye images taken from classical fundus camera. The proposed method makes use of support vector machines classifier with Gaussian kernel. The mean sensitivity is 85 %, while specificity 90%.

Keywords: SVM classification, MLP classification, glaucoma.

KLASYFIKACJA OBRAZÓW DNA OKA ZA POMOCĄ MASZYN WEKTORÓW PODPIERAJĄCYCH DLA WSPOMAGANIA DIAGNOZOWANIA JASKRY

Streszczenie. W artykule przedstawiono nowa metodę automatycznej klasyfikacji cyfrowych obrazów dna oka na normalne i jaskrowe. Obszar wnęki naczyniowej zostaje automatycznie wysegmentowany na obrazie dna oka pozyskanego z klasycznej funduskamery. Zaproponowana metoda klasyfikacji wykorzystuje maszyny wektorów podpierających z jądrem Gaussowskim. Średnia czułość metody wynosi 85%, a specyficzność 90%.

Słowa kluczowe: klasyfikacja SVM, klasyfikacja MLP, jaskra.

1. Introduction

Glaucoma is a group of ocular diseases characterized by the proceeding optic nerve neuropathy which leads to the rising diminution in vision field, ending with blindness. The correct optic disk (i.e. the exit of the optic nerve from the eye known as "blind spot") structure contains: neuroretinal rim of pink color and centrally placed yellowish cup [7] (Fig. 1). The cup is the area within the optic disc where no nerve fibers and blood vessels are present and in 3D image appears as an excavation. Glaucomatous changes in retina appearance embrace various changes in neuroretinal rim and cup, as the result of nerve fibers damages.

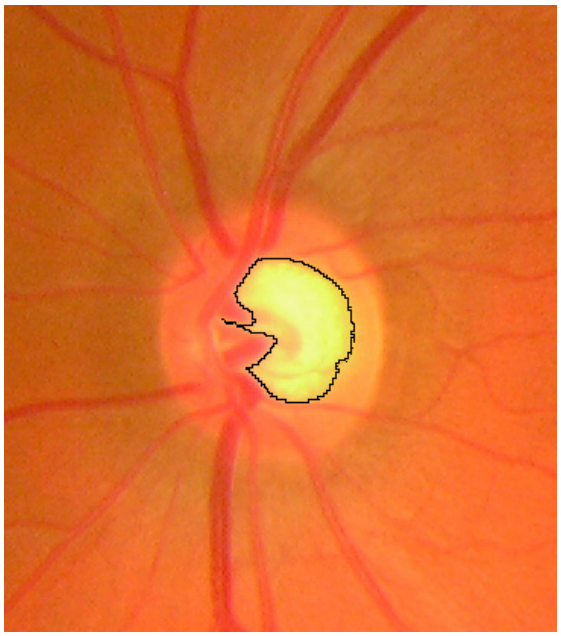

Fig. 1. The segmented cup contour superimposed on the input image
Rys. 1. Kontur wysegmentowanej wnęki nałożony na obraz wejściowy

Optic disc structures evaluation is one of the most important examinations in glaucoma progress monitoring and diagnosis. Searching for glaucoma damages during routine examination is not an easy task and gives uncertain results even with the experienced ophthalmologist [7]. The existing methods of qualitative analysis are very subjective, while quantitative methods of optic disc morphology evaluation (cup to disc ratio, neuroretinal rim area) do not result in full diagnosis. The new methods of analysis based on scanning-laser-tomography are expensive and accessible only in specialized ophthalmic centers.

That is why we have developed a more objective and cheaper method that enables for automatic classification of digital **fundus eye images** (**fei**) taken from classical fundus-camera into normal and glaucomatous ones. This method relies on the fact that shape of the

cup and its numerical characteristics correlate with glaucoma progress. The new method is composed of the following three main stages:

1. Segmentation of the cup region based on fuzzy clustering (described in [10]),
2. Selection of the cup features using genetic algorithms (described in [11]),
3. Classification of fundus eye images using **Support Vector Machine** (**SVM**) classifier.

This article describes the last stage – the new method for classification of fei into normal and glaucomatous ones based on SVM classifier.

A successful fei classification into normal and glaucomatous ones based on the suitable shape descriptors can boost the performance of applications supporting glaucoma diagnosing. Despite its importance, it has received no attention in the literature.

The Authors would like to thank prof. George Michelson from Department of Ophthalmology, Friedrich-Alexander-University Erlangen-Nuremberg for the possibility to access a database with fundus eye images, as well as for medical consultation.

2. Selection of the cup features using genetic algorithms

Due to nerve fibres damages during glaucoma progress, different changes in a shape of the cup are observed. Proper shape feature selection can provide a better classification accuracy due to finite sample size effects. In our approach [11], genetic algorithms were used to select the most significant features from the whole set of the computed 29 geometric features characterizing the shape of the cup region.

A given feature subset is represented as a binary string (a chromosome) of length *n*, with a zero or one in position *i,* denoting the absence or presence of feature *i* in the set (*n* is the total number of available features). Each chromosome is evaluated based on the proposed in [11] fitness function:

$$Fitness = 10^4 \cdot accuracy + 0.4 \cdot zeros$$

where *accuracy* is the accuracy rate that the given subset of features achieves (i.e. the performance of a classifier on a given subset of features), *zeros* is the number of zeros in the chromosome. As a classifier, we used the SVM with Gaussian radial basis kernel in which parameters $\gamma = 2.5$ and *C=100* [9]. The following 3 dimensional feature vector has been selected from a set of 29 features by genetic algorithm [11]:

$$(\phi_2, I_3, R_F)$$

where:

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2$$

is Hu invariant moment, where $\eta_{20}, \eta_{02}, \eta_{11}$ are normalized central moments. Normalized central moment of order (p+q) of a function *f (i,j)* is defined as [5]:

$$\mu_{pq} = \frac{m_{pq}}{(m_{00})^{\alpha}} \quad \alpha = \frac{p+q}{2} + 1$$

where m_{pq} is a spatial central moment of order (p+q) defined as:

$$m_{pq} = \sum_{i=1}^{m}\sum_{j=1}^{n}(i-I)^{p}(j-J)^{q} f(i,j) \quad I = \frac{m_{10}}{m_{00}},\ J = \frac{m_{01}}{m_{00}}$$

$$I_3 = \mu_{20}(\mu_{21}\mu_{03} - \mu_{12}^2) - \mu_{11}(\mu_{30}\mu_{03} - \mu_{21}\mu_{12}) + \mu_{02}(\mu_{30}\mu_{12} - \mu_{21}^2)$$

is compound, invariant moment,

$$R_F = \frac{L_h}{L_V}$$

is Feret coefficient where:
L_h – maximal diameter in horizontal direction,
L_V – maximal diameter in vertical direction.

3. Classification of fundus eye images using SVM classifier

3.1. SVM classifier

The **Support Vector Machine** (**SVM**) [6,9,12] has become an effective tool for pattern recognition because of its high generalization performance. The basic idea of an SVM classifier is illustrated in Fig. 2. which shows the simplest case in which the data vectors (marked by 'X' and 'O') can be separated by a hyperplane. Among all the separating hyperplanes that may exist, SVM classifier seeks for that one which produces the largest **separation margin** (the distance between the hyperplane and the closest data point).

If we have a training set $S = \{(x_i, y_i), 1 \le i \le N\}$ composed of examples $x_i \in R^n$, each belonging to a class labeled by $y_i \in \{1, -1\}$, our object is to find a hyperplane that divides *S*, leaving all the points with the same label on the same side of the hyperplane and also maximize the separation margin. This means that we must find a pair *(w,b)* such that (after some rescaling):

$$y_i (w \cdot x_i + b) \ge 1 \quad i = 1,..., N, w \in R^n, b \in R$$

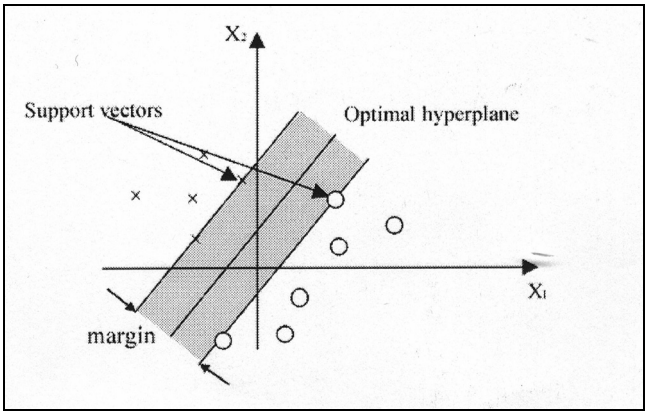

Fig. 2. SVM classification with a hyperplane that maximizes the separating margin between the two classes
Rys. 2. Klasyfikacja SVM za pomocą hiperpłaszczyzny maksymalizującej margines międzyklasowy

The goal of the SVM is to find the **optimal separating hyperplane** (**OSH**) – i.e which will maximize the margin. The problem of finding the OSH is equivalent to the maximization of the function [6,9]:

$$W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j x_i x_j \tag{2}$$

$\alpha_i \geq 0, \quad i = 1,...,N$ Lagrange multipliers [4]

$\sum_{i=1}^{N} y_i \alpha_i = 0$ a constraint

Let $\alpha = (\alpha_1,...,\alpha_N)$ be the solution of (2). The OSH *(w,b)* has the following expansion:

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

while *b* can be obtained from α and from the Kuhn-Tucker conditions [4]:

$$\alpha_i [y_i (w \cdot x_i + b) - 1] = 0 \quad i = 1,...,N$$

If the training examples (x_i,y_i) correspond with nonzero coefficients α_i, then we call them **support vectors**. These are the elements of the training set that lie either exactly on, or inside the decision boundaries of the classifier. They consist of those training examples that are most difficult to classify. The SVM classifier uses these „borderline" examples to define its decision boundary between the two classes. This in philosophy is quite different from a classifier that is based on minimizing learning error alone.

Finally, the decision function of classifying a new data point x can be written as follows:

$$f(x) = \text{sgn}(\sum_{i=1}^{N} \alpha_i y_i x_i \cdot x + b)$$

In the more general case in which the data points are not linearly separable in the **input space**, a non-linear transformation $\Phi(x)$ is used to map the data vector x into a high-dimensional space (called **feature space**) prior to applying the linear maximum-margin classifier. To avoid the potential pitfall of over-fitting in this higher-dimensional space, an SVM uses a **kernel function**:

$$K(x_i, x_j) = \Phi(x_i)\Phi(x_j)$$

in which the nonlinear mapping is implicitly embedded. A function qualifies as a kernel provided that it satisfies Mercer's conditions [9]. We can rewrite equation (2) as follows:

$$W(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

subject to the constraints:

1. $\sum_{i=1}^{N} y_i \alpha_i = 0$
2. $0 \le \alpha_i \le C$

where ***C*** is a **regularization parameter** which controls the tradeoff between complexity of the classifier and the number of nonseparable points. Smaller values of C will maximize the margin while larger C will cause the *OSH* to minimize the number of misclassifying points. Finally, we can achieve a new decision function, as follows:

$$f(x) = \text{sgn}[\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b]$$

α_i, b are constants, all determined through numerical optimization during training.

Several common kernel functions are used to map data into high-dimensional feature space, for example **linear kernel**:

$$K(x, z) = x \cdot z ,$$

polynomial kernel:

$$K(x, z) = (\gamma \cdot x \cdot z + coef)^d$$

where $\gamma, coef$ are constants and d – degree,

Gaussian radial basis kernel:

$$K(x, z) = \exp(-\gamma \cdot | x - z |^2)$$

where γ is a constant.

3.2. Design of SVM classifier for glaucomatous changes detection

Fig. 3 shows the training examples in the feature space (ϕ_2, I_3, R_F). In our experiment, a nonlinear SVM with a Gaussian radial basis kernel was used. During the training phase the regularization parameter C and constant γ need to be determined for the Gaussian SVM classifier. For this purpose we adopt the widely used statistical k-fold cross validation method ($k=10$ in our case) which consist of the following steps: 1) divide randomly all the training examples into k equal-sized subsets, 2) use all but one subset to train the SVM, 3) use the holdout subset to measure classification error, 4) repeat steps 2 and 3 for each subset, 5) average the results to get an estimate of the generalization error of the SVM classifier.

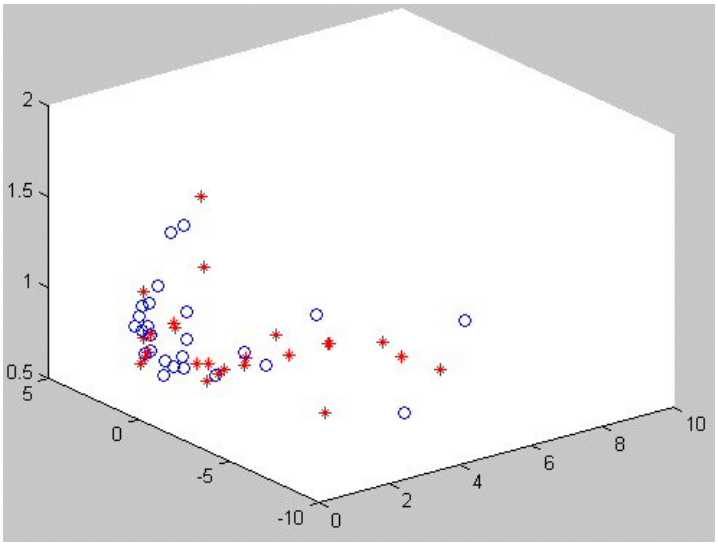

Fig. 3. Training examples in the feature space
Rys. 3. Próbki uczące w przestrzeni cech

The generalization error was measured as a mean classifier accuracy defined as a ratio:

$$(TP+TN)/(TP+TN+FP+FN),$$

where TP – true-positive, FN – false-negative, TN – true-negative, FP – false-positive.

The SVM was trained using this procedure for various parameter settings. In the end, the model with the smallest generalization error was adopted, in our case C = 100 and $\gamma = 2.5$ were used. The number of resulting support vectors for this case was about 12% of the total number of training samples. In Fig. 4 we show a plot of the estimated generalization error rate for the trained SVM classifier with the Gaussian kernel. A generalization level as low as *5%* was achieved under various parameter settings. These results demonstrate that the performance of the SVM classifier is rather robust over the choice of model parameters.

The developed method has been applied into 50 fei of patients with glaucoma and 50 fei of normal patients which where previously examined by conventional methods by an

ophthalmologist. On the acquired from Canon CF-60Uvi fundus-camera images, the cup contour is automatically detected by the segmentation method described in [10]. The exampled fundus eye image with the extracted cup contour is shown in Fig.1.

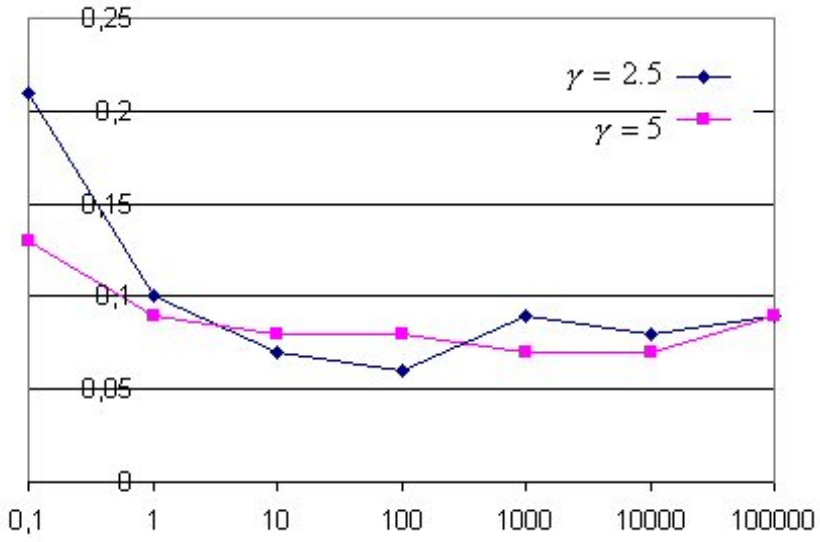

Fig. 4. Plot of generalization errors versus the regularization parameter C, achieved by the SVM classifier using the Gaussian kernel with $\gamma = 2.5,5$

Rys. 4. Wykres błędu generalizacji jako funkcji parametru regularyzacyjnego dla klasyfikatora SVM z jądrem Gaussowskim o parametrze $\gamma = 2.5,5$

Next, for the set of *100* detected cup region images, the suboptimal feature vectors (ϕ_2, I_3, R_F) are calculated (as described in [11]). This set was randomly divided into 2 parts: the training and the testing one. The training set is used to train the SVM classifier as described in the previous section. Classifier performance is then tested on test set using the Gaussian kernel with $\gamma = 2.5$ and C=100. The following results were obtained: the mean sensitivity: *85%* and the mean specificity: *90%*. (*Sensitivity = TP/(TP+FP*), and is the percent of correctly classified glaucomatous/positive cases, *specificity = TN/(TN+FN*) is the percent of correctly classified normal/negative cases).

To compare the performance of our proposed method based on SVM classifier, we used MLP neural network classifier with standard backpropagation learning rule [3]. We used 3-5-2 network architecture (3 inputs, 5 neurons in a hidden layer, 2 output neurons) with learning rate equal to 1. We obtained the following results (on the same feature vector as for SVM classifier): mean sensitivity 46% and mean specificity 78%.

The detection performance was also evaluated quantitatively using receiver operating characteristic (ROC) curves [8]. A ROC curve plots the correct positive detection rate (i.e. TP fraction) versus correct negative detection rate (i.e. FP fraction) for the continuum of the decision thresholds. This permits the system designer to assess the performance of the

recognition system at various operating points (thresholds in the decision rule). The two classifiers were evaluated using the same set of 100 fei. The results are summarized using ROC curves in Fig. 5. As can be seen, the SVM classifier offers the best result in the operating range. These results show that SVM classifier outperforms MLP classifier.

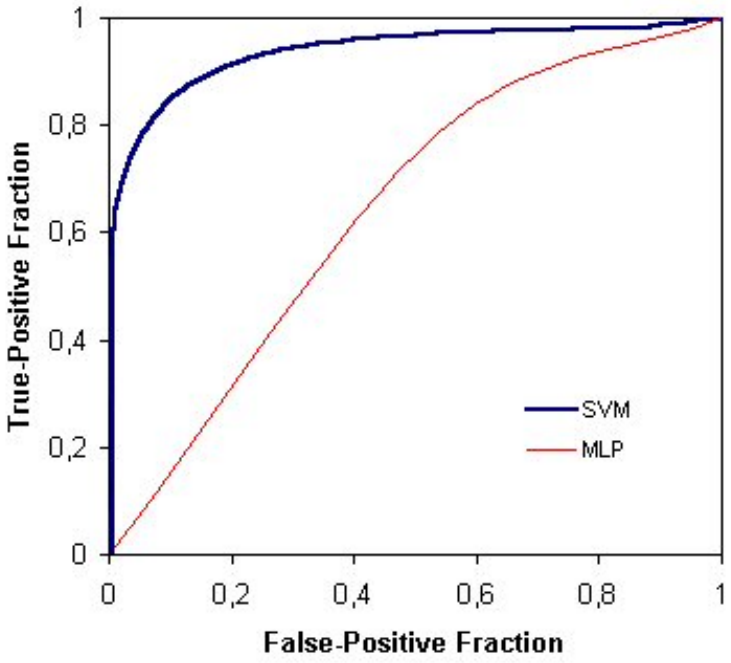

Fig. 5. Diagram of ROC curves for the SVM and the MLP in the classification of normal and glaucomatous patients

Rys. 5. Wykres krzywej ROC dla klasyfikatorów SVM i MLP w zadaniu klasyfikacji normalnych i jaskrowych obrazów dna oka

4. Conclusions

In this work we demonstrated the new method for the detection of glaucomatous changes from fei taken from classical fundus camera in opthalmology. As far as we know, no automatic method for the classification of fei acquired from fundus-cameras into normal and glaucomatous has been reported yet. Our method proves that shape of the cup and its numerical characteristics correlate with progress of glaucoma. The obtained classification results are encouraging. It is expected that the new method, after clinical tests would support glaucoma diagnosis based on digital fei obtained from fundus-camera.

REFERENCES

1. Arabas J.: Lectures on genetic algorithms.WNT, Warsaw 2001.
2. Bezdek J. C.: Pattern recognition with fuzzy objective function algorithms.Plenum Press, New York 1982.
3. Bishop C. M.: Neural networks for pattern recognition. Clarendon Press, Oxford 1995.
4. Fletcher R.: Practical methods of optimization. Wiley-Interscience, Chichester 1987.
5. Gonzalez R.C., Woods R.E.: Digital image processing. Prentice-Hall, 2002.
6. Guyon I. et al: Structural risk minimization for character recognition. Advances in Neural Information Processing Systems, Nr 4, 1992, 471-479.
7. Kanski J. et al. Glaucoma: a color manual of diagnosis and treatment. Butterworth-Heinemann, 1996.
8. Metz M.: Basic principles of ROC analysis. Seminars in Nuclear Medicine, vol. III, nr 4, 1978, 283-298.
9. Mueller K. R. et al.: An introduction to kernel-based learning algorithms, IEEE Trans. Neural Networks, v.12, Nr 2, 2001, 181-201.
10. Stąpor K, Brueckner A.: Segmentation of fundus eye images using fuzzy clustering for supporting glaucoma diagnosing. Studia Informatica, submitted for publication.
11. Stąpor K. et al.: Genetic feature subset selection for classification of eye-cup region in fundus eye images. Studia Informatica, v. 24, Nr 4, 2003, 332-344.
12. Vapnik V.: The nature of statistical learning theory. Springer Verlag, New York 1995.

Recenzent: Dr inż. Krzysztof Fujarewicz

Wpłynęło do Redakcji 31 marca 2004 r.

Omówienie

W artykule przedstawiono nowa metodę automatycznej klasyfikacji cyfrowych obrazów dna oka na normalne i jaskrowe. Opracowana metoda wykorzystuje fakt, że kształt wnęki naczyniowej koreluje z postępem jaskry. Obszar wnęki naczyniowej zostaje automatycznie wysegmentowany na obrazie dna oka pozyskanego z klasycznej funduskamery za pomocą metody opisanej w [10]. Przykładowy kontur wnęki uzwysegmentowany automatycznie z cyfrowego obrazu dna oka pokazano na rys. 1. Dla wysegmentowanych obszarów wnęk naczyniowych obliczone zostają suboptymalne wektory cech wyznaczone za pomocą

algorytmów genetycznych. Zaproponowana metoda klasyfikacji wykorzystuje maszyny wektorów podpierających z jądrem Gaussowskim. Doświadczalnie ustalono optymalne parametry klasyfikatora: $\gamma = 2.5$ i C=100. Dokonano porównania z wielowarstwowym perceptronem jako klasyfikatorem. Uzyskane wyniki pokazały wyższość klasyfikatora SVM. Uzyskano następujące wyniki: średnia czułość *85%*, specyficzność *90%*.

Adresses

Katarzyna STĄPOR: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16, 44-100 Gliwice, Polska, delta@ivp.iinf.polsl.gliwice.pl .
Adrian BRUECKNER, Uniwersytet Śląski, Instytut Matematyki, ul.Bankowa 14, 40-007 Katowice, Polska.