A MULTIPLE CLASSIFIER SYSTEM FOR THE AUTOMATIC LOCALIZATION OF ANOMALIES IN RETINAS OF DIABETIC PATIENTS

M. MORGANTI, A. SPERDUTI, A. STARITA
Dipartimento di Informatica, Università di Pisa,
Corso Italia 40, I-56125 Pisa, Italy.
e-mail: `perso@di.unipi.it`

In this paper we study the performance of several committee techniques for the detection of microaneurysms and hard exudates in images of retinas of diabetic patients. New algorithms for the selection of classifiers to be inserted in the committee are proposed and tested with very good results. Moreover, we show how classifiers can be combined sequentially to further improve the performance of the system. Comparative results are reported.

1 Introduction

Diabetic patients show anomalies of the retina (microaneurysms, cotton wools, hard exudates, etc.) related to the state of gravity of the pathology, up to generate blindness. The automatic detection and localization of small and medium size anomalies in retinas of diabetic patients can be of great help for early diagnosis of the damages to the retina and to evaluate the evolution of the illness with respect to the therapy. This is especially true considering the usual high number of photos gathered daily by a Diabetic out-patients Department.

In this paper we report the experimental results obtained by applying committee of classifiers within the DRA (Diabetic Retina Analyzer) System [1] (see Fig. 1). We have experimented with both well known combination algorithms and new combination algorithms devised by us. Comparative results, showing the superiority of our new algorithms are reported.

2 The Detection and Localization Module

One of the aim of the DRA System, besides the management of all the data concerning a patient, is to automatically locate anomalies which are clinically important, and to recognize them on the basis of their different shapes and colors (microaneurysm present themselves as little dark red spots, hard exudates are white spots with neat contours, etc.). The current version of the system can detect and localize microaneurysms and hard exudates. This is obtained by selecting, through image processing techniques, zones of the retina which with high probability contain anomalies. The selected zones are then

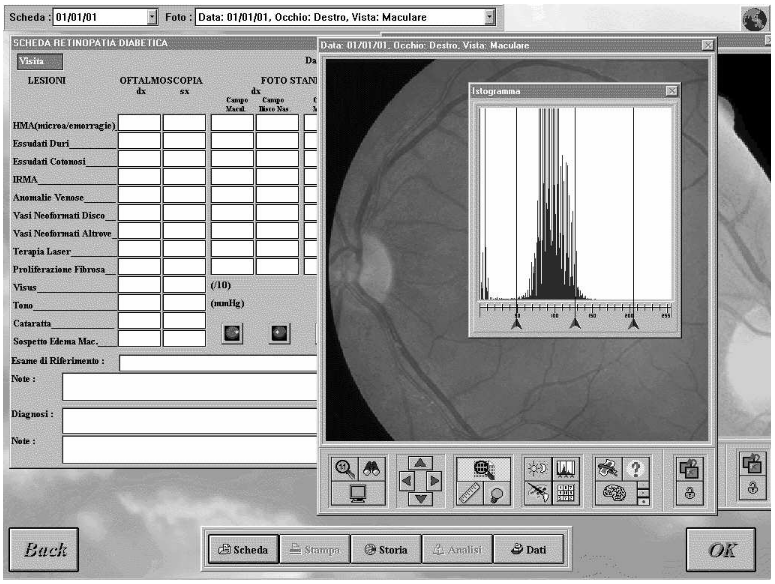

Figure 1: A particular of the graphical interface of the DRA system.

classified as true or false anomalies by specialized neural networks. Specifically, the 512×768 pixels color image of the retina is pre-processed through a specially developed filter which on the basis of the brightness histogram, the green[a] component histogram, and on some statistics computed on a set of about 1,000 different images, implements an automatic stretch filter. The result is a normalized and enhanced version of the image which remarkably improves the quality of the image. The resolution of the image is then changed according to the size of the anomalies which the physician is looking for. Images at higher resolutions allows the system to discover small size anomalies, and so on. In this way, the same localization (*Search*) and detection (*Neural Validation*) functions can be used when looking for anomalies of different sizes. The *Search* function works on the green component of the image and it exploits the observation that microaneurysms and hard exudates are local minima and maxima of the brightness, respectively. Usually, the *Search* function returns about 50-200 candidate points for each type of anomaly within 1-3 seconds on a 100MHz Pentium based Personal Computer. However, while for hard exudates only about 4% of the candidate points are actually false re-

[a] We have selected the green component of the image since it is actually conveying all the relevant information about the anomalies.

sponses, this percentage greatly increases for microaneurysms. This is mainly due to the presence of veins which share several common signal features with microaneurysms. The best way to discriminate between this two class of patterns is to use an adaptive technique. This is because the problem can only be defined through a set of positive and negative examples. On the basis of this consideration, the natural candidates to perform the task are feed-forward backpropagation networks. The task of the *Neural Validation* function is to remove from the list of candidates the ones which are not anomalies. This is obtained by specialized neural networks, one for each type of anomaly, trained on positive and negative examples of images (5×5) returned by the *Search* function.

3 Using Committee Algorithms

The problem faced by the *Neural Validation* module is not easy, and it is difficult to obtain good results by a single neural network. A way to improve the performance is to use committee of classifiers. For this reason, we have experimented with several committee algorithms. Specifically, we have used the following combination techniques: Mean, Bayesian, OPM [2], Voting.

To improve the performance of the system, we have also devised new algorithms for the selection and combination of classifiers. Specifically, we show how to select the classifiers to be inserted in the final system using a greedy approach (*Greedy Selection* and *Pseudo-exhaustive Selection*), and how the sequential application of classifiers can improve the overall performance of the system (*Sequential Cooperation*).

3.1 New Selection Algorithms

It is well known that the error of a committee of classifiers depends both on the mean of the error of the classifiers and on the ambiguity between them [3]. Thus, in general, the insertion of a new classifier in the committee does not necessarily lead to an improvement of the performance. On the contrary, it is possible that the insertion of a new classifier (neural network) in the committee decreases the overall performance. To cope with this problem, we have developed two algorithms for the selection, from a pool of candidates, of the "best" classifiers to be inserted into the committee. For both algorithms, the Validation set is used for early stopping during training, while an additional independent set, called *Combination set*, is used to decide which candidate should be inserted into the committee.

Greedy Selection

The first algorithm is based on a greedy approach: the classifiers are sorted by increasing value of Mean Square Error (MSE) on the validation set; then the MSE of the committee obtained by inserting all classifiers starting from the first one to the i-th is computed on the combination set; finally, only the classifiers which caused a decrease in the MSE of the committees are inserted into the final committee:

Greedy Selection Algorithm
input: Classifiers $C_1,\ldots,C_N$, Validation Set, Combination Set;
output: Committee;

begin
 for i=1 **to** N **do** Error_Val[i] $\leftarrow$ MSE(C_i,Validation Set);
 Sort C_i according to increasing values of Error_Val;
 for i=1 **to** N **do begin**
  Committee $\leftarrow$ Committee $\cup$ C_i;
  Error_Com[i] $\leftarrow$ MSE(Committee,Combination Set);
  end
 Committee $\leftarrow$ C_1;
 for i=2 **to** N **do**
  if(Error_Com[i]<Error_Com[i-1]) **then** Committee $\leftarrow$ Committee $\cup$ C_i;
end

Pseudo-exhaustive Selection

In this case the classifier with smallest MSE on the validation set is inserted into the committee and the MSE on the combination set is used as reference error for the committee. Then, in turn, all committees obtained by adding one of the classifiers are evaluated and the committee with smallest MSE on the combination set is selected. The algorithm stops when no committee generated in this way is better than the current one. Note that the generation of new committees is done with replacement, i.e., a classifier C_i which already is in the current committee can be selected again, leading to a duplication (which can also be multiple) in the new committee. This duplication must be avoided when the combination strategy is Bayesian or OPM, since the combination weight for each single classifier is automatically generated.

Of course, this algorithm is computationally much more expensive than the Greedy algorithm. Thus it is not practical to use it when the number of classifiers is huge. Finally, note that the algorithm is not really exhaustive, since no "backtracking" is allowed when a stop point is reached.

3.2 *Sequential Cooperation*

Usually, when combining several classifiers, the combination is performed in parallel. However, under specific conditions, an improvement on the performance can be obtained by sequentially combining the output of the classifiers.

Let consider a classification task with two classes ω_1 and ω_2. Moreover, let suppose we have available three classifiers, $C_i(\cdot)$, $i = 1, 2, 3$. These three classifiers can be combined in the following way

$$C_{seq}(\boldsymbol{x}) = \begin{cases} C_2(\boldsymbol{x}) & \text{if } C_1(\boldsymbol{x}) = \omega_1 \\ C_3(\boldsymbol{x}) & \text{if } C_1(\boldsymbol{x}) = \omega_2 \end{cases} \tag{1}$$

The resulting classifier $C_{seq}(\cdot)$ will show an improved performance over the single classifiers $C_i(\cdot)$ under specific conditions on the classifiers's sensitivity. Remembering that the sensitivity of a classifier C_i with respect to class ω_j is defined as $S_i(\omega_j) = P(C_i(\boldsymbol{x}) = \omega_j | \boldsymbol{x} \in \omega_j)$, it is possible to derive necessary conditions for C_{seq} to be superior in performance to each single C_i.

To understand this, let study the sensitivity of $C_{seq}(\cdot)$ with respect to class ω_1. It can be expressed as $S_{seq}(\omega_1) = P(C_{seq}(\boldsymbol{x}) = \omega_1 | \boldsymbol{x} \in \omega_1)$, which, because of the fact that $C_{seq}(\boldsymbol{x}) = \omega_1$ *if and only if* $C_1(\boldsymbol{x}) = C_2(\boldsymbol{x}) = \omega_1$ *or* $C_1(\boldsymbol{x}) \neq C_3(\boldsymbol{x}) = \omega_1$, can be written as

$$\begin{aligned} S_{seq}(\omega_1) &= P(C_1(\boldsymbol{x}) = \omega_1,\ C_2(\boldsymbol{x}) = \omega_1 | \boldsymbol{x} \in \omega_1) + \\ &\quad P(C_1(\boldsymbol{x}) \neq \omega_1,\ C_3(\boldsymbol{x}) = \omega_1 | \boldsymbol{x} \in \omega_1). \end{aligned} \tag{2}$$

Applying Bayes rule, the above formula can be rewritten as

$$\begin{aligned} S_{seq}(\omega_1) &= P(C_2(\boldsymbol{x}) = \omega_1 | C_1(\boldsymbol{x}) = \omega_1, \boldsymbol{x} \in \omega_1) P(C_1(\boldsymbol{x}) = \omega_1 | \boldsymbol{x} \in \omega_1) + \\ &\quad P(C_3(\boldsymbol{x}) = \omega_1 | C_1(\boldsymbol{x}) \neq \omega_1, \boldsymbol{x} \in \omega_1) P(C_1(\boldsymbol{x}) \neq \omega_1 | \boldsymbol{x} \in \omega_1). \end{aligned} \tag{3}$$

If we suppose that classifier C_1 is independent with respect to C_2 and C_3, then we have

$$\begin{aligned} S_{seq}(\omega_1) &= P(C_2(\boldsymbol{x}) = \omega_1 | \boldsymbol{x} \in \omega_1) P(C_1(\boldsymbol{x}) = \omega_1 | \boldsymbol{x} \in \omega_1) + \\ &\quad P(C_3(\boldsymbol{x}) = \omega_1 | \boldsymbol{x} \in \omega_1) P(C_1(\boldsymbol{x}) \neq \omega_1 | \boldsymbol{x} \in \omega_1). \end{aligned} \tag{4}$$

Thus, by definition of sensitivity of a classifier, we obtain

$$S_{seq}(\omega_1) = S_1(\omega_1) S_2(\omega_1) + (1 - S_1(\omega_1)) S_3(\omega_1), \tag{5}$$

i.e., the sensitivity of C_{seq} with respect to ω_1 is between the sensitivity of C_2 and C_3 with respect to ω_1, respectively. A similar result can be obtained for ω_2, i.e.,

$$S_{seq}(\omega_2) = S_1(\omega_2) S_3(\omega_2) + (1 - S_1(\omega_2)) S_2(\omega_2). \tag{6}$$

Finally, the performance of C_{seq} can be expressed using $S_{seq}(\omega_1)$ and $S_{seq}(\omega_2)$

$$Performance(C_{seq}) = S_{seq}(\omega_1)P(\boldsymbol{x} \in \omega_1) + S_{seq}(\omega_2)P(\boldsymbol{x} \in \omega_2). \qquad (7)$$

In order for C_{seq} to be superior with respect to each single classifiers C_i, we must have

$$Performance(C_{seq}) > Performance(C_i) \ i = 1, 2, 3. \qquad (8)$$

By using equations (7) and (5-6), it turns out that the following conditions have to be satisfied:

Superiority with respect to C_1 (9)

$$\begin{aligned} S_1(\omega_1)[S_2(\omega_1) - S_3(\omega_1) - 1] \quad &> \quad \frac{P(\omega_1) - 1}{P(\omega_1)}(S_1(\omega_2)[S_3(\omega_2) - S_2(\omega_2) - 1] \\ &\quad + S_2(\omega_2)) - S_3(\omega_1) \end{aligned}$$

Superiority with respect to C_2 (10)

$$S_1(\omega_2)[S_3(\omega_2) - S_2(\omega_2)] \quad > \quad \frac{P(\omega_1)}{1 - P(\omega_1)}(S_1(\omega_1) - 1)[S_3(\omega_1) - S_2(\omega_1)]$$

Superiority with respect to C_3 (11)

$$S_1(\omega_1)[S_2(\omega_1) - S_3(\omega_1)] \quad > \quad \frac{1 - P(\omega_1)}{P(\omega_1)}(S_1(\omega_2) - 1)[S_2(\omega_2) - S_3(\omega_2)]$$

It is not difficult to see that, under the conditions of reasonably balanced a priori probabilities for the two classes ω_1 and ω_2 and given a reasonably good classifier C_1, conditions (10) and (11) are satisfied by the joint events *i)* $S_2(\omega_1) > S_3(\omega_1)$ and *ii)* $S_3(\omega_2) > S_2(\omega_2)$. On the contrary, it is more difficult to understand condition (9), which relates the sensitivity of the three classifiers in a more complex fashion.

4 Experimental Results and Conclusion

In this section we report the experimental results obtained by applying standard committee techniques (Mean, Bayesian, OPM, Voting) without selection of classifiers, with selection by our algorithms (Greedy and Pseudo-exhaustive Selection), and Sequential Cooperation.

The images for microaneurysms (hard exudates), i.e. 5×5 images generated by the *Search* module, are split in four sets: a Training set of 1200 (500) images, a Validation set (for early stopping) of 576 (151) images, a Combination set of 276 images[b], and a Test set of 600 (350) images.

The number of neural networks (classifiers) used in the experiment was 10 for hard exudates and 30 for microaneurysms. They differed both in number of hidden units and in the learning parameters used during learning. For hard exudates the best network was able to obtain an error rate of 3.2. By combining the 10 networks together we obtained the following results for the error rate: Mean, 2.9; Bayesian, 2.6; OPM, 2.0; Voting, 2.0. Because of the good results, no experiment with the selection algorithms and the Sequential Cooperation algorithm was performed. Results for microaneurysms are shown in Table (1). The table reports also results obtained by using invariants. The basic idea is to transform the input using a transformation for whom we want the classification to be invariant (we used clockwise rotations of 90, 180, and 270 degrees, horizontal and vertical reflexive projection), and then to evaluate the output of the networks for these transformed images. The obtained outputs are then used within the committee, as if they were outputs of additional new classifiers[c]. It can be noted that invariants increase the performance of the committee. Results obtained for the selection algorithms and the Sequential Cooperation algorithms are reported in Table (2). It is evident that the performance increases independently of the combination strategy, thus validating the quality of the proposed selection algorithms. The best result is obtained by the Sequential Cooperation algorithm.

We can conclude that the use of committees in the automatic detection of microaneurysms and hard exudates leads to a significant improvement of the performance of the system. Moreover, we have shown that it is possible to further improve the performance by using invariants, new selection algorithms, and sequential cooperation among single classifiers.

Finally, it should be reported that the use of neural networks for the classification of diabetic retinopathy is also explored in [4] where a neural network with an input window of 20×20 pixels is moved across the image. The main difference with respect to our work is that the task in [4] is to automatically classify the images, while the aim of our system is the localization of small anomalies to support early diagnosis of retina pathologies. It must be noted that our task is more difficult, since images which present both small and large size anomalies can easily be classified on the basis of the large size anomalies only. Nevertheless, our system is considerably faster since the neural networks both are smaller (20 versus 200 hidden nodes) and need to be applied only

[b] The selection algorithms have been experimented only over microaneurysms.

[c] Note that the response time of a committee using invariants is slowed down of a factor 5, i.e., the number of transformations, with respect to the same committee without invariants.

Table 1: Microaneurysms: experimental results for different combination algorithms, with (I = Invariants) or without invariants. NN is the best neural network over 30 different networks.

	NN	NN_I	Mean	Mean_I	Bayes	Bayes_I
% Error	12.7	11.4	11.0	10.2	11.0	9.4
	OPM	OPM_I	Voting	Voting_I		
% Error	10.6	9.9	10.3	9.0		

Table 2: Microaneurysms: experimental results, using invariants, for different combination and selection (G = Greedy, E = Exhaustive) algorithms. Results for the Sequential Cooperation algorithm are reported as well (Seq).

	Mean_G	Mean_E	Bayes_G	Bayes_E	OPM_G	OPM_E
MSE	0.0335	0.0334	0.0342	0.0341	0.0342	0.0327
% Error	8.9	9.0	9.0	9.0	9.1	8.3
	Voting_G	Voting_E	Seq			
MSE	-	-	0.0371			
% Error	9.1	8.9	7.8			

on a small fraction of the pixels of the images. Concerning the classification performance, it is not possible to make a direct comparison because of the different formulation of the classification task and to the different sets of images used for the training of the networks.

References

1. E. Casi, A. Ceravola, R. Cionini, A. Sperduti, and A. Starita. Diabetic Retina Analyzer, *Proceedings of the 18th Annual International Conference of IEEE Eng. in Med. and Biol. Soc.* Vol 4, pp. 282-283, Amsterdam, 1996.
2. K. Tumer and J. Ghosh. Error Correlation and Error Reduction in Ensemble Classifiers. *Connection Science*, Vol. 8, n. 3&4, pp. 385-404, 1996.
3. A. Krogh and P. Sollich. Statistical Mechanics of Ensemble Learning. *Physical Review E*, Vol. 55, pp. 811-825, 1997.
4. G. G. Gardner and D. Keating and T. H. Williamson and A. T. Elliott. Detection of Diabetic Retinopathy Using Neural Network Analysis of fundus Images. In *Proc. 2nd Int. Conf. on Neural Networks and Expert Systems in Medicine and Healthcare*, Plymouth, UK, 1996.