

Comparative Exudate Classification using Support Vector Machines and Neural Networks

Alireza Osareh¹, Majid Mirmehdi¹, Barry Thomas¹ and Richard Markham²

¹Department of Computer Science, University of Bristol, Bristol, BS8 1UB, U.K.
{osareh, majid, barry}@cs.bris.ac.uk

²Bristol Eye Hospital, Bristol, BS1 2LX, U.K.
markham@gifford.co.uk

Abstract After segmenting candidate exudates regions in colour retinal images we present and compare two methods for their classification. The Neural Network based approach performs marginally better than the Support Vector Machine based approach, but we show that the latter are more flexible given criteria such as control of sensitivity and specificity rates. We present classification results for different learning algorithms for the Neural Net and use both hard and soft margins for the Support Vector Machines. We also present ROC curves to examine the trade-off between the sensitivity and specificity of the classifiers.

1 Introduction

Intraretinal fatty (hard) exudates (EXs) are a visible sign of diabetic retinopathy and also a marker for the presence of co-existent retinal oedema. If present in the macular area, oedema and exudates are a major cause of visual loss. Automated early detection of the presence of EXs can assist ophthalmologists prevent the spread of the disease more efficiently. We are working towards an automatic computer assisted system for classification of Diabetic Retinopathy (DR). Identifying the proportion of the colour retinal image that contains exudates is one of our key objectives. In this paper, we briefly present our retinal image segmentation process using Fuzzy C-Means clustering (FCM) to segment candidate EX regions [1], and then concentrate on a comparative analysis of their classification. We apply various configurations of neural networks (NN) and compare their classification performance to Support Vector Machines (SVMs) using hard and soft margins. A NN based on the backpropagation learning method classified the segmented EXs with an overall diagnostic accuracy of 93.4% with 93.0% sensitivity and 94.1% specificity in terms of lesion classification. Similarly, our SVM classifier achieved an overall classification of 90.4% with 83.3% sensitivity and 95.5% specificity. The most common parameter estimation algorithms used to estimate the parameters of a NN are based on the Empirical Risk Minimization (ERM) principle, which can achieve minimum risk on the training set. This is one of the reasons why NNs can get stuck in local saddle points and therefore, they are susceptible to many training problems including overfitting and convergence. In contrast, SVMs follow the Structural Risk Minimization (SRM) principle that results in a classifier with the least expected risk on the test set and hence good generalisation. Also, unlike NNs, SVMs always converge to the same solution for a given data set, regardless of initial conditions. However, in this application, while NNs performed slightly more accurately for EX classification, SVMs are more suitable for sensitivity and specificity control and do not suffer from overfitting

problems and hence generalise better. Overall, our results show the performance of these two different classifiers are close and comparable.

In [2], Wang et al. applied a Bayesian statistical classifier based on colour features to differentiate yellowish lesions (including EXs) from the dark objects. They achieved a global 100% sensitivity and 70% specificity measured on whether EXs were present anywhere in the image. They did not measure lesion-based performance which represents how accurate the system can distinguish EXs among the others lesions. Sinthanayothin [3] identified EXs in graylevel images based on a recursive region growing technique. The sensitivity and specificity reported was 88.5% and 99.7%, however, these measurements were based on 10x10 windows. Gardner et al. [4] used a NN to identify the EX lesions in greylevel images. The authors reported a sensitivity of 93.1%. Again this was the result of classifying whole 20x20 regions rather than a pixel-level classification. One novelty of our proposed method here is that we locate EXs at pixel resolution in colour images and evaluate the performance of the system applying both lesion-based and image-based criteria.

Section 2 briefly outlines our automatic method for identification of the EX pathologies in colour retinal images. Section 3 reviews the features used in the classification stage to distinguish EX candidates from other segmented regions. In Section 4, SVMs and NN classifiers will be compared in how they perform in classifying the EX regions. The paper is concluded in Section 5.

2 Fuzzy C-Means Segmentation

In this study, we used 142 colour retinal images obtained from a non-mydratic retinal camera with a 45° field of view. The image resolution was 760x570 at 24bit RGB. We obtained candidate EX regions by performing FCM segmentation directly on our colour retinal images. This involved two important pre-processing steps. Typically, there is wide variation in the colour of fundus from different patients and is strongly correlated to the person's race and iris colour. In the first step, we performed a normalization of our colour images (see Figure 1(c)). In the next pre-processing step, local contrast enhancement was performed to distribute the values of the pixels around the local mean to facilitate later segmentation (see Figure 1(d)).

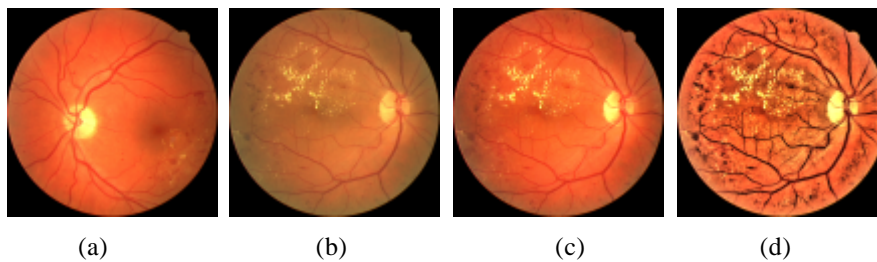


Fig. 1. Colour normalisation and local contrast enhancement: (a) reference image, (b) typical retinal image (including EXs), (c) colour normalised version, (d) after contrast enhancement.

Hard segmentation methods take crisp decisions about regions. However, the regions in an image are not always crisply defined. Fuzzy approaches provide a mechanism to represent and manipulate uncertainty and ambiguity and allow pixels to belong to multiple classes with varying degrees of membership. We segment our retinal images using a two-stage colour segmentation algorithm based on Gaussian-smoothed

histogram analysis and Fuzzy C-Means clustering [5] comprising a coarse and a fine segmentation step. At the coarse stage, an initial classification is performed by interval analysis of the zero-crossings of the histogram second derivative at multiple scales in each colour band. This results in the number of classes (K) and the centre for each cluster. In the fine stage, FCM assigns any remaining unclassified pixels to the closest class based on the minimization of the objective function:

$$J_m(P, V) = \sum_{i=1}^c \sum_{k=1}^n (\mathbf{m}_{ik})^m \|x_k - v_i\|^2 \quad (1)$$

where P is a fuzzy partition of the data ($x_k, k=1, \dots, n$) and V is a vector of cluster centres ($v_i, i=1, \dots, c$). Also, \mathbf{m}_{ik} represent the membership value of x_k to cluster i . These memberships must be between 0 and 1, and \mathbf{m}_{ik} must sum to 1 for all i . The parameter m is a weight that determines the degree to which partial members of a cluster affect the clustering result. The fuzzy partitioning is carried out through an iterative optimisation in order to find both prototypes v_i and membership functions \mathbf{m}_{ik} to minimise J_m . Here, $m = 2$ and the algorithm was iterated until the Euclidean distance between two successive membership values reached 0.5 where FCM could distinguish three different clusters. Figure 2(a) shows the colour segmentation of Figure 1(d). Figure 2(b) shows the candidate EX regions overlaid on the original image.

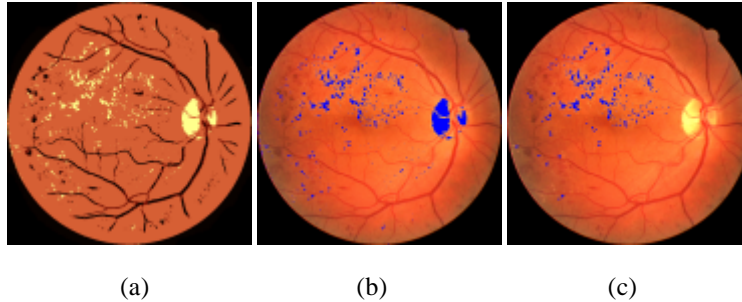


Fig. 2. Colour image segmentation: (a) FCM segmented image, (b) candidate EX regions overlaid on the original image, and (c) final classification.

The full segmentation approach was straightforward to implement, fast and, had fixed parameters, but most importantly it allowed us to segment colour images. To assess the accuracy of the proposed segmentation technique, an expert clinician marked the EX regions in all 75 of our retinal images. Accurate, manual, pixel-level registration of small pathologies like EXs is very difficult due to the wide variability in their colour and size. The FCM-based technique could segment 97% of all the EXs based on the groundtruth. Only extremely faint EXs were not identified. It is worth noting that false positive EX candidates can arise due to general retinal reflections.

3 Feature Extraction

To classify the segmented regions into EX or non-EX classes we must represent them using relevant and significant features. Clinically, ophthalmologists use colour to

differentiate various pathological conditions. Similarly coloured objects like cotton-wool spots and EXs are differentiated with further features such as size, edge strength, shape and texture. The feature set should be selected such that the between-class discrimination is maximised while the within-class discrimination is minimised. Indeed, in order to avoid the curse of dimensionality it is desirable for the feature set to be as small as possible. We selected 18 features after a comparative study of the discriminative attributes of a much larger set. We experimented with a number of colour spaces including RGB, HSI, *Lab* and *Luv* and found that colour spaces which separate luminance and chrominance are more suitable. The extracted feature set comprised *mean Luv* and *standard deviation* of the *Luv* values inside a candidate region, *mean Luv* and *standard deviation* of the *Luv* values around the region, *Luv* values of *region centroid*, *region size*, *region compactness*, and *region edge strength*. To evaluate the usefulness of the selected features, the within-class matrix (S_w) and between-class scatter matrix (S_b) were computed. The value $J = trace(S_b/S_w)$ was determined based on the sequential forward selection search strategy and used as a measure of feature-set efficiency [1]. As expected, features that provide colour information seem to contribute significantly more than the other features. The class separability can naturally improve by including additional features, but at the expense of extra features and classifier complexity.

4 Classification

4.1 Support Vector Machine Classification

SVMs have been successfully applied to a wide range of pattern recognition problems and the reader is referred to [6] for details. Here, we investigate them for classifying the segmented EX lesions. SVMs are based on the SRM principle, in contrast to the ERM principle in NNs, to minimize the error on the training data. SRM minimizes a bound on the test error thus allowing SVMs to generalise better than NNs. For a separable classification task, the idea is to map the training data into a higher-dimensional feature space using a *kernel* function where a separating hyperplane (w, b), with w the weight vector and b the bias, can be found which maximises the *margin* or distance from the closest data points. The optimum separating hyperplane can be represented based on kernel function:

$$f(x) = \text{sign} \left(\sum_{i=1}^n \mathbf{a}_i y_i K(x, x_i) + b \right) \quad (2)$$

where n is the number of training examples, y_i is the label value of example i , K represents the kernel, and \mathbf{a}_i coefficients must be found in a way to maximise a particular Lagrangian representation. Subject to the constraints $\mathbf{a}_i \geq 0$ and $\sum \mathbf{a}_i y_i = 0$, there is a Lagrange multiplier \mathbf{a}_i for each training point and only those training examples that lie close to the decision boundary have nonzero \mathbf{a}_i . These examples are called the *support vectors*. However, in real-world problems data are noisy and in general there will be no linear separation in the feature space. The hyperplane margins can be made more relaxed by penalising the training points the system misclassifies. Hence, the optimum hyperplane equation can be define as

$$y_i (w \cdot x_i + b) \geq 1 - \mathbf{x}_i, \quad \mathbf{x}_i \geq 0 \quad (3)$$

and the following equation is minimized in order to obtain the optimum hyperplane

$$\|w\|^2 + C \sum_{i=1}^n \xi_i \quad (4)$$

where ξ introduces a positive slack variable that measure the amount of violation from the constraints. The penalty C is a regularisation parameter that controls the trade-off between maximizing the margin and minimizing the training error. This approach is called *soft margins* [7]. Therefore, specifying a SVM requires two parameters: the kernel function and the regularisation parameter C . For training the SVM classifier, the Kernel-Adatron technique using a Gaussian kernel was used [6]. Segmentation of our 75 colour images (comprising 25 normal and 50 abnormal images) resulted in 3860 segmented regions consisting of 2366 EXs and 1494 non-EXs. These regions were labelled by a consultant ophthalmologist to create a fully marked groundtruth dataset. To obtain the optimal values for the Gaussian kernel (s) and C we experimented with different SVM classifiers using a range of values. Ten-fold cross-validation was applied to find the best classifier based on validation error. The performance of the selected SVMs was quantified based on its *sensitivity*, *specificity* and the *overall accuracy*. In the first experiment, with no restrictions on the Lagrange multipliers (hard margin), we achieved an overall accuracy of 88.6% with 86.2% sensitivity and 90.1% specificity for $s=0.3$. Figure 3(a) illustrates the generalisation performance of this classifier against varying values of s as well as the number of support vectors in each case. This result represents a good performance over positive and negative cases. To illustrate the effect of the soft margins, we trained the SVMs with s fixed at 0.3 and for a wide range of C values, which was applied as an upper bound to a_i (Figure 3(b)). The best overall accuracy, using the soft margin technique (hereafter referred to as SVM*), increased to 90.4% at $C=1.5$. However, in many medical diagnosis tasks, the overall accuracy is not the most appropriate measure since the balance between false positives and false negatives is very important. Moreover, sometimes the data set within a class is limited and so is statistically under-represented with respect to other classes. Consequently, controlling the performance of a system on a particular class of the data is very important. To do that we applied different misclassification cost C_+ and C_- (giving asymmetric soft margins) for each of the two classes (EXs and non-EXs) to adjust the cost of false positives vs. false negatives. This modifies (4) to the following optimisation problem [8]:

$$\|w\|^2 + C_+ \sum_{i:y_i=1} \xi_i + C_- \sum_{i:y_i=-1} \xi_i \quad (5)$$

subject to (3). Figure 3(c) illustrates the effect of a wide range of upper bounds C_+ on the a_i of the positive class (i.e. EXs) while there is no restriction on the a_i of the negative class which means ($C_- = \infty$). For example, as C_+ decreases, the number of false negatives is increased but at the expense of a decrease in the number of false positives. Therefore, specificity is increased while sensitivity is reduced. In this application, the maximum overall accuracy obtained at $C_+=8.0$ is found to be 89.8% with sensitivity decreased to 85.0% and specificity increased to 93.1% compared against the hard-margin results. The opposite effect can be achieved by considering an upper bound C_- on the a_i of the negative class. In this case, maximum overall accuracy achieved at $C_-=8.5$ is 85.2% with sensitivity increased to 95.1% and specificity decreased to 78.5%. Figure 2(c) shows a typical EX-classified image.

4.2 Neural Network Classification

To assess the performance of the SVMs we also classified our segmented EXs using neural networks [9] with a different number of algorithms and architectures. Again a 10-fold cross-validation technique was used for estimating the generalisation error of all classifiers. We experimented with two different learning methods, standard Back-Propagation (BP) and Scaled Conjugate Gradient (SCG) descent. We investigated a single hidden layer with a range of 2 to 35 hidden units to find an optimum. The network with the smallest validation error was selected as the best classifier and then the selected architecture was tested against an unseen test set. In this way a NN classifier using BP learning performed best in terms of the overall generalisation performance. Table 1 summarises the NN and SVM results. These are the best results from a selection of configurations used for training the classifiers.

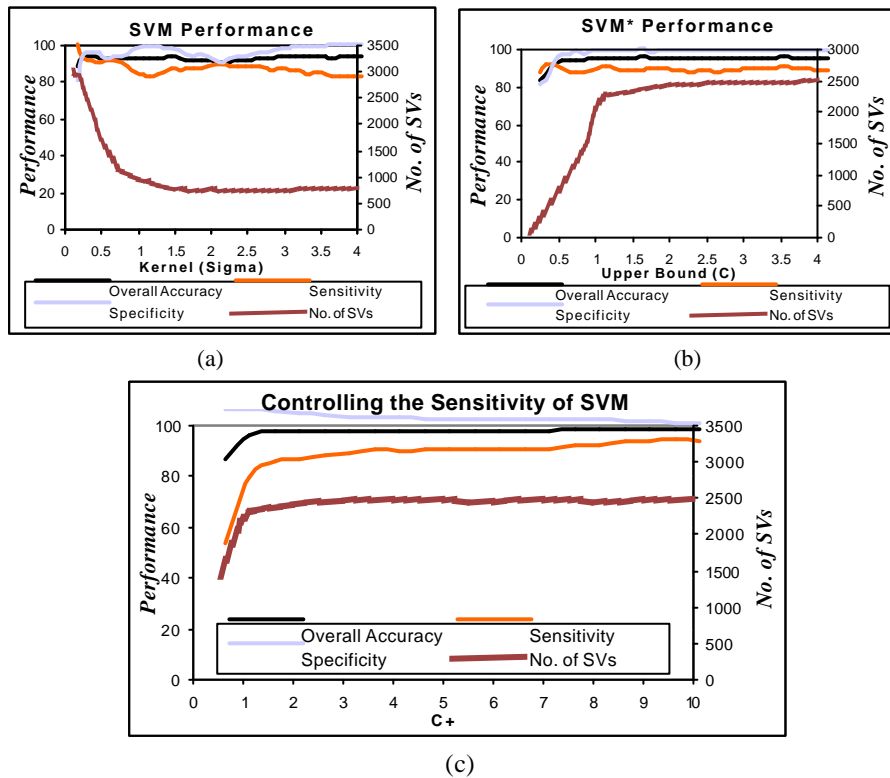


Fig. 3. Generalisation performance of the SVM classifier against (a) different σ values, (b) different C values, (c) different C_+ values.

Although the diagnostic accuracy of the NN classifiers is slightly better than the SVMs, the classifier performances are very close and there is a good balance between sensitivity and specificity in all the cases. However, in most medical applications the overall accuracy is not a sufficient measure to choose the optimal configuration.

Table 1. Performances of different classifiers for lesion-based classification (as %s).

Classifier	Threshold	Accuracy	Sensitivity	Specificity
SVM $s=0.3$	(T=0.0)	88.6	86.2	90.1
SVM* $s=0.3, C=1.5$	(T=0.0)	90.4	83.3	95.5
SVM* $s=0.3, C_+=8.0$	(T=0.0)	89.8	85.0	93.1
SVM* $s=0.3, C=8.5$	(T=0.0)	85.2	95.1	78.5
NN-BP (15 hidden)	(T=0.50)	93.4	93.0	94.1
NN-SCG (15 hidden)	(T=0.45)	92.8	97.9	85.2

In order to assess and analyse the behaviour of the classifiers throughout a whole range of the output threshold values, ROC [10] curves shown in Figure 4 have been produced (with true-positives plotted against the false-positives describing the trade-off between sensitivity and specificity). The bigger the area under the ROC curve, the higher the probability of making a correct decision. The BP and SCG classifiers show a higher performance with areas 0.966 and 0.962 respectively. The SVM (without soft margins) and SVM* (soft margins with $C=1.5$) show slightly lower performance over the entire ROC space with areas 0.907 and 0.924.

So far we have discussed pixel-by-pixel based lesion classification. We can also use our trained classifiers to evaluate the effectiveness of our proposed approach by assessing the image-based accuracy of the system. A population of 67 *different* retinal images were considered (40 abnormal and 27 normal). Each retinal image was evaluated using the NN-BP and SVM* ($C=1.5$) classifiers separately and a final decision was made to show whether the image has some evidence of Diabetic Retinopathy. As Table 2 illustrates the NN-BP classifier could identify affected retinas with 95.0% sensitivity while it recognised 88.9% of the normal images, i.e. the specificity. In the SVM* case the diagnostic accuracy of abnormal images was 87.5% sensitivity and 92.0% specificity for the normal cases. However, as in the lesion-based case, the SVM sensitivity and specificity rates can be easily manipulated by varying the value of C_+ and C to obtain different results.

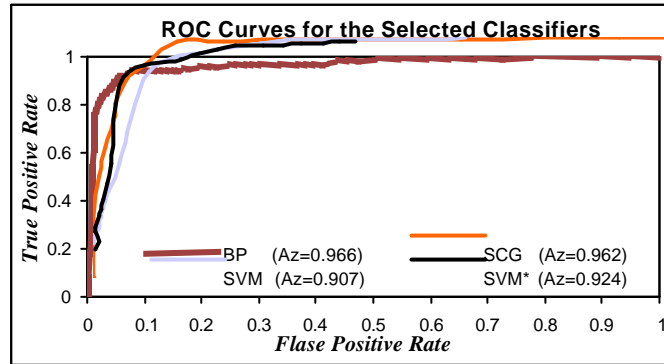


Fig. 4. ROC curves for the classifiers in Table 1 (A_z refers to the area under ROC curve).

EXs usually appear in groups and therefore missing some very faint EXs is not very important. However, when there are only a few new and very faint EXs in the retina the identification task will be more difficult. When we manually analysed the

system's decisions on normal images we found that in most cases where a normal image had been wrongly identified as abnormal only very few false positives had been detected (of the order of 2 or 3 individual false lesions). In such cases we can use proximity information to make a final decision, since EXs usually appear in dense groups rather than randomly scattered across the image.

Table 2. System Performance for assessing the evidence of DR.

Classifier	Image Type	No. of Patients	Detected as Abnormal	Detected as Normal	X=Sensitivity Y=Specificity
BP-NN	<i>Abnormal</i>	40	38	2	X=95.0%
	<i>Normal</i>	27	3	24	Y=88.9%
SVM*	<i>Abnormal</i>	40	35	5	X=87.5%
	<i>Normal</i>	27	2	25	Y=92.0%

5 Conclusion

In this study we investigated SVM and NN classifiers to obtain good class separability between EX and non-EX classes. The results by the two classification approaches are very similar, however, we believe that SVMs are a more practical solution to our application: they have a significant advantage compared to NNs as they can achieve a trade-off between false positives and false negatives using asymmetric soft margins, they always converge to the same solution for a given data set regardless of initial conditions, and finally, they remove the danger of overfitting.

Acknowledgements

The first author is on a scholarship funded by the Iranian Ministry of Science, Research and Technology. The authors also thank the UK National Eye Research Centre for their support.

References

- Osareh, A., Mirmehdi, M., Thomas, B.T., Markham, R.: Classification and Localisation of Diabetic Related Eye-Disease. Proc 7th European Conf. on Computer Vision (2002) 502-516
- Wang, H., Hsu, W., Goh, K., Lee, M.: An Effective Approach to Detect Lesions in Colour Retinal Images. IEEE Conf. on Computer Vision and Pattern Recognition (2000) 181-187
- Sinthanayothin, C.: Image Analysis for Automatic Diagnosis of Diabetic Retinopathy. PhD Thesis, King's College London (1999)
- Gardner, G., Keating, D., Williamson, T., Elliott, A.: Automatic Detection of Diabetic Retinopathy Using an ANN: A Screening Tool. BJO **80** (1996) 940-944
- Lim, Y., Lee, S.: On the Colour Image Segmentation Algorithm Based on the Thresholding and the Fuzzy C-Means Techniques. Pattern Recognition **23**. 9 (1990) 935-952
- Cristianini, N., Taylor, J.S.: An Introduction to Support Vector Machines. Cambridge University Press (2000)
- Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery **2**. 2 (1998) 121-167
- Veropoulos, K., Campbell, C., Cristianini, N.: Controlling the Sensitivity of Support Vector Machines. Proc Int. Joint Conf. on Artificial Intelligence (1999) 55-60
- Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press. (1992)
- Centor, R.M.: Signal Detectability: The use of ROC Curves and their Analysis. Medical Decision Making, (1991) 102-106