PERGAMON

Computers in Biology and Medicine 28 (1998) 225–238

Computers in Biology and Medicine

A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms

Allan J. Frame[a,*], Peter E. Undrill[a], Michael J. Cree[a], John A. Olson[b], Kenneth C. McHardy[c], Peter F. Sharp[a], John V. Forrester[b]

[a]*Department of Bio-medical Physics and Bio-engineering, University of Aberdeen, Aberdeen, UK*
[b]*Department of Ophthalmology, University of Aberdeen, Aberdeen, UK*
[c]*Diabetic Clinic, Aberdeen Royal Infirmary, Aberdeen, UK*

Received 10 October 1997; accepted 11 May 1998

Abstract

We compared the performance of three computer based classification methods when applied to the problem of detecting microaneurysms on digitised angiographic images of the retina. An automated image processing system segmented 'candidate' objects (microaneurysms or spurious objects), and produced a list of features on each candidate for use by the classifiers. We compared an empirically derived rule based system with two automated methods, linear discriminant analysis and a learning vector quantiser artificial neural network, to classify the objects as microaneurysms or otherwise. ROC analysis shows that the rule based system gave a higher performance than the other methods ($p = 0.92$) although a much greater development time is required.

Keywords: Neural networks; Linear discriminant analysis; Rule based system; Ophthalmology; Computer aided diagnosis

1. Introduction

Diabetic retinopathy (DR) is the ocular manifestation of the systemic disease, diabetes mellitus and is the most common cause of blindness in the UK working population [1]. In current research studies, assessment of DR is made semi-quantitatively by comparing a photograph of the patient's fundus with a standard set of photographs and the stage of DR

* Corresponding author. Tel: +44-1224-681818 x52430; Fax: +44-1224-685645; E-mail: a.frame@biomed.abdn.ac.uk.

0010-4825/98/$19.00
PII: S0010-4825(98)00011-0

graded according to a standard protocol [2, 3]. More accurate quantification of the progress of the disease can be made by identifying and counting some of the lesions on the photograph such as microaneurysms (MA). A positive correlation between the number of MAs and the progression of the disease has been shown [4, 5].

MAs, as seen in normal fundus photographs, are often not readily visualised and may be confused with small dot haemorrhages. Fluorescein angiography increases the visibility of these lesions (a procedure involving the intravenous injection of a solution of sodium fluorescein dye). MAs are small saccular bulges in the walls of the retinal capillary vessels; they therefore fill with dye during the angiography and are classically described as appearing as distinct hyperfluorescent round objects in the angiographic image. In practice, however, the appearance of MAs can deviate from the classical description. In particular, some may appear in association with larger vessels, or they may be a part of a conglomeration of more than one MA. Also, small MAs can look similar to other retinal pathologies or capillary crossings. Therefore the identification of the MAs for counting is not trivial. A representative fluorescein angiogram image of a diabetic patient is shown in Fig. 3a.

Protocols have been developed for manually counting MAs [6], but they are time-consuming and are subject to observer error. They are, therefore, not used in current clinical practice. Automated computer techniques of digital image processing offer a fast, objective and repeatable method of counting MAs, and this approach is currently being investigated in Aberdeen.

Our approach to the detection and quantification of features of interest in an image is based upon the classical model of computer vision. The process involves a number of fundamental steps: acquisition of the image, pre-rocessing, segmentation and classification [7]. Quantification is a trivial step at the end of the process.

We have developed a fully automated image processing system that processes a single frame from a sequence of angiographic photographs. The acquisition and pre-rocessing stages of the system are fully described by Spencer [8], and the segmentation has been updated by Cree [9, 10]. The process is described briefly below.

The output of the segmentation stage is a set of candidate objects that bear resemblance to MAs. A number of shape and grey-scale based features are measured on each candidate and it is these data features that are used to classify each candidate into one of two classes, namely MAs and spurious objects. It is only the MA class that is of interest for quantifying retinopathy.

This paper describes our investigation into the classification stage of the task. We have applied three classification methods and compared their efficacy using receiver operator characteristic (ROC) analysis [11]. The classifiers tested were:

1. An empirically-derived quantitative and logical rule-base (RBS).
2. Linear discriminant analysis (LDA).
3. A Learning vector quantiser (LVQ) artificial neural network (ANN).

The first method (RBS) is a manually derived classifier and consequently requires extensive effort. Furthermore, it can be difficult for a human observer to locate clustering relationships in high-dimensional data. The automated methods (LDA and LVQ) derive classifiers almost

instantaneously with little or no manual intervention, and can handle multi-dimensional data. However, they are not so adept as a human observer at exploiting nonlinear or subtle relationships in the data features. In clinical applications it is important to be able to develop the best classifiers possible, both for accurate quantification and also to avoid errors that may have clinical implications. Therefore the extra time expenditure in manually deriving an RBS may be justified if the RBS can outperform any automated classifiers. In ophthalmic diagnosis all methods have been applied separately [12–14] but no studies have compared several methods on the same data.

In Section 2, a brief description of the image processing strategy up to the point of segmentation of candidates for classification is given. Section 3 outlines the methodology used for training and testing the classifiers. We find it useful to describe the methodology before introducing the three classification methods in Section 4. The performance results of the classifiers and ensuing discussion follow in Sections 5 and 6.

2. Image processing and candidate segmentation

2.1. Acquisition

Before analysis, a photographic angiogram covering a field of 35° of the patient's retina is taken with a conventional fundus imaging camera and captured on 35 mm film. A frame of the early venous stage of the angiogram sequence is chosen, back-illuminated and digitised using a Kodak Megaplus 1.4 CCD camera at 1024×1024 pixel image size and 8-bit precision. The resultant pixel resolution achieved is approximately 8 μm. A flood image is also captured when no film is present in the digitising system, and this is used to correct the digitised image for any nonuniform illumination present in the digitising system.

2.2. Image pre-processing

Once digitised, the system first selects a 512×512 pixel region, centred on the fovea (area of central vision) from the image for analysis. This region is then pre-rocessed to correct for the effect of illumination variations across the image, which are caused by nonuniform illumination of the retina during photography, varying reflectivity of the retina and varying amounts of background fluorescence due to the capillary network in the retina. A technique of shade correction is used to remove illumination variations in the image by smoothing the image with a large scale median filter and subtracting the filtered image from the original. This has the effect of leaving only features of a smaller scale than the filter such as the vessels and the MA in the image [8]. The jth pixel of the original image is referred to below as f_j, likewise the smoothed image as b_j and the shade-corrected image as s_j. The shade-corrected image is then normalised to have the same maximum and minimum grey levels as the original image.

2.3. Segmentation and feature extraction

Once the image has been pre-rocessed, the MA are segmented out using a combination of morphological matched-filtering and region-growing, and the processes are described in detail by Spencer and Cree [8–10]. The image is first transformed using the bilinear top-hat transform, which gives a high degree of discrimination between circular and linear features. A matched-filter with a 2D Gaussian function as the model of a 'classical' MA is applied to the transformed image to enhance any MA-like objects which are present and the resultant image is thresholded to identify their location. The threshold level was deliberately set to ensure that as many MAs as possible would be detected, although this does mean that many spurious objects bearing a similarity to MAs are also segmented. While the spurious objects can be removed at a later stage (classification), any MA not segmented at this stage cannot be reinstated.

A region-growing algorithm is then applied at each candidate position in the shade-corrected image, to 'fill out' and delineate the underlying candidate shape. This is necessary since the morphological filtering applied during segmentation does not preserve the candidate shape. Each segmented candidate is an 8-connected object and is stored in a database for further feature extraction.

Thirteen features are automatically measured and recorded on each segmented candidate:

$$a = \sum_{j \in \Omega} 1 \quad \text{(area)} \tag{1}$$

$$p \quad \text{(perimeter)} \tag{2}$$

$$r = l/b \quad \text{(aspect ratio)} \tag{3}$$

$$c = p^2/4\pi a \quad \text{(circularity)} \tag{4}$$

$$i_i = \sum_{j \in \Omega} f_j \quad \text{(intensity 1)} \tag{5}$$

$$i = \sum_{j \in \Omega} s_j \quad \text{(intensity 2)} \tag{6}$$

$$m_i = i_i/a \quad \text{(mean intensity 1)} \tag{7}$$

$$m = i/a \quad \text{(mean intensity 2)} \tag{8}$$

$$I = \frac{1}{\sigma}(i_i - \bar{x}) \quad \text{(normalised intensity 1)} \tag{9}$$

$$I = \frac{1}{\sigma} i \quad \text{(normalised intensity 2)} \tag{10}$$

$$M_i = \frac{1}{\sigma}(m_i - \bar{x}) \quad \text{(normalised mean intensity 1)} \tag{11}$$

$$M = \frac{1}{\sigma} m \quad \text{(normalised mean intensity 2)} \tag{12}$$

$$t \quad \text{(matched filter seed value)} \tag{13}$$

where Ω is the set of pixels in the candidate, l is the length of the longest-axis of the candidate, b is the breadth of the projection of the candidate onto the axis perpendicular to the longest axis, and f_j and s_j are as defined in the section above. $\bar{x}$ and σ are the mean and standard deviation of the smoothed background image b_j.

The first four shape-based features are derived from a binary representation of each candidate object. The perimeter p is calculated as if the bounding pixels of the candidate were represented by an 8-code chain-code. Each horizontal/vertical link contributes 1 unit and each diagonal link $\sqrt{2}$ units to the perimeter, although this is likely to suffer from some digitisation error, particularly for small candidates. The circularity c gives a measure of the roundness and smoothness of the candidate's boundary. The more complex the boundary, the higher the circularity. The aspect ratio is the longest length of the object divided by the perpendicular breadth. It is particularly useful for rejecting any pieces of vessel inadvertently segmented. The next eight measurements are based on the grey-scale intensity of the candidate taken from the original and shade-corrected images. Various combinations of the integrated, mean and normalised intensities are used. Although they are not fully independent, the various combinations were found useful in constructing the RBS classifier. Finally, a single parameter, t, relating to the intensity value of the matched filter that triggered the candidate is recorded.

It is expected that a true MA has an area between approximately 8 to 170 pixels, a low circularity, and a unity aspect ratio. The intensity of the MA is expected to be high.

3. Training and testing methodology

When considering automated nonparametric classification techniques, each method must be trained to make the correct classification. The normal protocol for doing this is to prepare a large number of observations whose classification is known a priori. In our application, a series of observations were prepared in which an ophthalmologist classified each candidate object as an MA or otherwise. This feature measurement stage was carried out on a training set of 1659 candidates gathered from 68 retinal fluorescein angiographic images. Of the 1659 candidates, 400 were labelled as MA by the ophthalmologist. Each classifier was trained on this data set, which included the candidate classification given by the human expert.

A separate dataset of a further 1067 candidate objects (of which 297 were classified as MA by experts) from another 20 images was prepared for use as an independent testing and validation set. The trained classifier was then used to classify the candidates of the training set.

The expert's opinion of the testing set was not supplied to the classifier, but only used to measure the performance of the classifier as described below.

3.1. ROC methodology

To measure the effectiveness of each classifier and compare the performance of each against the others, receiver operator characteristic (ROC) methodology was used [11]. ROC curves graphically display the relationship between sensitivity and specificity of a classifier as the parameters of that classifier change. Sensitivity and specificity are measures of classification accuracy which emphasise equally the importance of classifying correctly the spurious objects and the MAs.

Let M be the total number of true MAs in the test set and O the total number of objects. For the purpose of our comparison, we consider the set of candidates ($M + O$) to be a finite set, whereby the number of candidates is limited by the output of the image segmentation process and is dependent on the parameters set by that program. We consider the parameters set as being optimum for ensuring that as many MAs as possible are segmented while keeping the number of spurious objects at a manageably low level. In a test run, a classifier labels N_m candidates of the test set as MA of which T_m are correctly labelled (i.e. they are true MA). Likewise, N_o candidates are labelled as spurious objects of which T_o are truly spurious objects. That means that $F_m = N_m - T_m$ candidates labelled as MA are falsely classified, and $F_o = N_o - T_o$ candidates are falsely labelled as spurious objects. The sensitivity S and the specificity P are given by

$$S = \frac{N_m}{M} \quad (\times 100\%) \tag{14}$$

$$P = \frac{N_o}{O} \quad (\times 100\%). \tag{15}$$

An ideal classifier would give 100% sensitivity and 100% specificity, when tested against the gold standard. However this is unachievable in practice because there is an overlap between the feature distributions of the two candidate classes. An index related to the specificity is the false positive fraction (FPF) given by

$$\mathrm{FPF} = 1 - P \tag{16}$$

and the graphical results are plotted in terms of this index rather than the specificity.

A dilemma often arises if one uses (sensitivity, specificity) pairs to compare the performance of two classifiers, since it is not necessarily obvious how to choose between a system that has higher sensitivity but lower specificity than another. This can be resolved by realising that many classifiers can be biased from their optimal operating point, thereby tracing out a path in (sensitivity, specificity) space. (The classifier's achievable specificity for a chosen sensitivity can be read off the graph.) The better classifier is the one whose curve passes closest to 100% sensitivity and 100% specificity (or 0% FPF).

3.2. Analysis of the ROC curves

Fitted ROC curves [15] are derived from the experimentally derived curves or curve segments, using the method of Chakraborty [16]. This allows a comparison of the underlying classifier performance, and the application of statistical techniques to provide an objective comparison of the classification methods. Ideally, the true positive fraction (TPF) and false positive fraction (FPF) are governed by the parametric equations

$$\mathrm{TPF}(z) = 1 - \mathrm{NORMDIST}(z, \mu_s, \sigma_s, \mathrm{TRUE})$$

$$\mathrm{FPF}(z) = 1 - \mathrm{NORMDIST}(z, \mu_n, \sigma_n, \mathrm{TRUE}),$$

where (μ_s, $\mu_n = 0$) and (σ_s, $\sigma_n = 1$) are the means and standard deviations of the signal and noise distributions respectively and the z values span values typically ranging from −4 to 6. NORMDIST(z, μ, σ, TRUE) is a spreadsheet function which determines the area under a normal distribution, with mean μ and standard deviation σ at ordinate z. We utilise a nonlinear curve fitting procedure based on the Levenberg–Marquardt method [17] to derive optimum values of μ_s and σ_s. The area under each curve, A_z, can be computed by numerical integration and the probable errors in the TPF and FPF (determined according to Metz [18]) allow σ_A to be evaluated.

4. Classification methods

The three different classification methods examined represent widely different approaches: a statistical classifier (LDA) a neural classifier (LVQ) and an expert system (RBS). They are described below, starting with the RBS.

4.1. Rule-based analysis

The RBS consisted of a number of quantitative and logical rules which were empirically derived directly from the data available for training. This was achieved by comparing pairs of features from the feature vector (as 2D scatter graphs) and searching for functions of the two features that provide good discrimination where separate distributions exists. An example is shown in Fig. 1, comparing circularity c and the greyscale intensity i of the candidate object. The linear discriminant is superimposed. The equation of this function is

$$1835c - i \geq 1119. \tag{17}$$

If Eq. (17) is satisfied, then the candidate is classified as a spurious object, although it should be noted that this will result in the misclassification of a few MAs. Candidates failing Eq. (17) may or may not be MA and further rules are applied to establish their class. All of the rules were mutually exclusive, therefore no conflict-resolution strategies were required.

A number of such rules were incorporated directly into the image analysis and quantification program. They were established by a number of subjective constraint criteria, leading to an

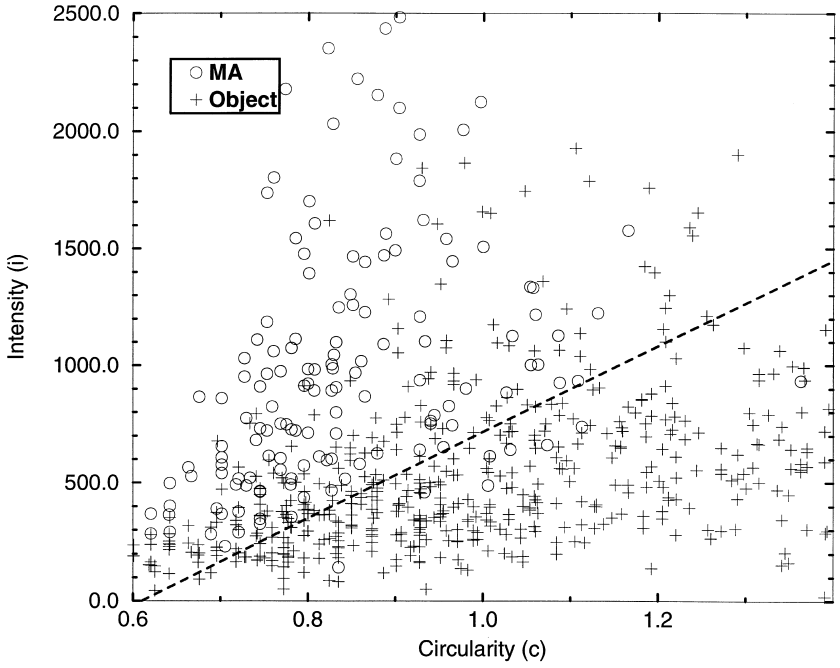

Fig. 1. Circularity vs gray-scale intensity for the training set, with the rule of Eq. (17) overlaid.

effective classifier. To generate the full ROC curve, constant terms, such as those in Eq. (17), were varied to bias the classifier towards choosing more MA or towards spurious objects. This is a multidimensional optimisation problem, and to make the problem manageable it was decided to analyse the behaviour of the classifier at a number of discrete operating points giving the ROC curve in Fig. 2.

About 1 man-month of effort was needed to construct the RBS, therefore, to justify the extra development time for the RBS it should have a better classification performance than the automated systems.

4.2. Linear discriminant analysis

LDA is a well-understood statistical classifier that calculates a linear combination of features which provides a single rule to classify the candidates as MAs or spurious objects. LDA makes a number of assumptions about the underlying distribution of the data, namely that each class is normally distributed and they all have equal covariance matrices. In practice, the method is very robust to departures from these assumptions [19].

LDA places a hyperplane in feature space, orthogonal to the line joining the means of the distributions of the two classes. Any candidate which lies in that part of feature space on the side of the hyperplane including the MA distribution mean is labelled as an MA. Any candidates lying on the other side of the hyperplane are, therefore, spurious objects. If the assumptions given above are known to be satisfied, then it can be shown that LDA places the hyperplane to give the optimal discrimination between the two classes [19].

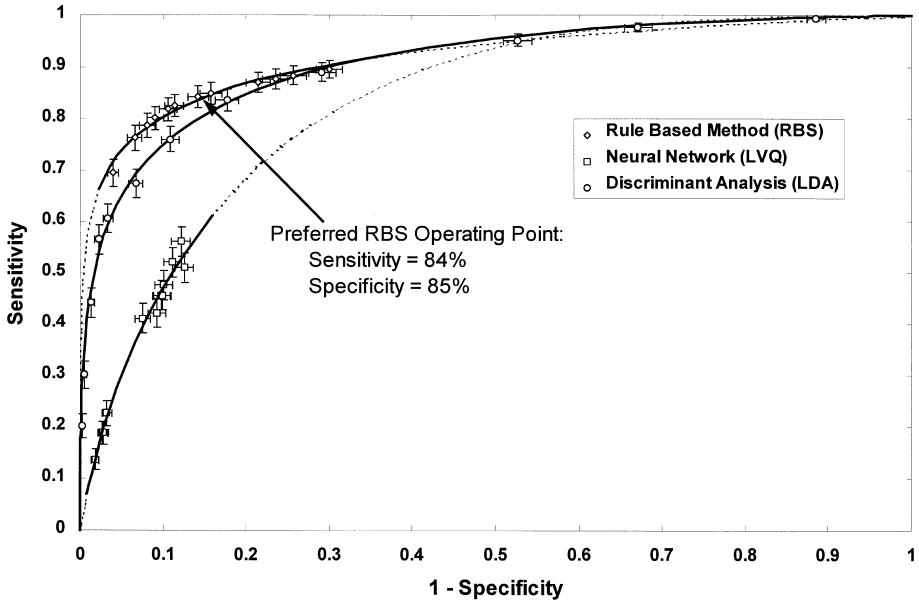

Fig. 2. ROC curves of all classifiers. The curve shown using solid lines over the measurement range and dotted lines outside these limits.

To produce an ROC curve, the constant term in the LDA function was varied, which has the effect of moving the hyperplane between the centre of mass of each of the two classes in feature space, thus varying the number of candidates classified into each class and, therefore, the sensitivity and specificity.

4.3. Learning vector quantisation artificial neural network

LVQ is a type of neural network which makes the classification by creating a number of free parameter vectors (*codebook vectors*) in the feature space, corresponding to different classes. Each class may have a number of such vectors. Each candidate object is classified depending on which codebook vector is closest (Euclidean distance), and the training set is used to set the codebook vectors to the correct classes. This has the advantage of adding a nonlinearity to the classification, where the amount of nonlinearity is proportional to the number of codebook vectors (the linear case occurs when the number of codebook vectors is equal to the number of classes).

The LVQ package used was a public domain ANN simulator package, LVQ-PAK [20]. This was deliberately chosen as the LVQ-PAK implementation was readily available, easy and fast to use, and had been recommended as the most useful and applicable ANN system [21]. The time required to train the network is an important factor in the practical use of such a system. Although the number of codebook vectors can be varied to increase the accuracy of the classification, this also increases the time taken to train the network and, therefore, there is an upper limit to the number of vectors above which training time becomes impractical.

As the number of codebook vectors changes, the sensitivity and specificity also change. However, there does seem to be a limit to the values the sensitivity/specificity pair can take, namely that there is an upper limit to the sensitivity which cannot be breached by increasing the number of codebook vectors. This may be because of normalisation problems in the LVQ algorithm. LVQ is based on the k-nearest neighbour classifier, and this is highly dependent on the magnitudes of the feature vector in each dimension, namely that differences in those magnitudes will bias the classifier. This can be addressed by normalising the data before classification, but this may be suboptimal and limit the performance of the classifier. We are currently investigating normalisation techniques further to try and raise the limit to the sensitivity. This limit occurred when the number of codebook vectors was 400, and in this case the full training and testing process took 5 min on a Sun SPARCstation 5.

5. Results

Each classifier was trained on the training set, and then tested against the test set in the manner described above. The classifiers were biased from their operating points in the manners described above and experimental ROC curves or curve segments obtained. Using the methods outlined in Section 3, analytical ROC curves were fitted to the experimental data and these are plotted on Fig. 2. The curves are shown as a solid line over the range of measurement, and dotted outside these limits. Also shown on Fig. 2 are the error estimates derived using the method of Section 3.

5.1. Measurement of A_z

The area under each optimal ROC curve (A_z) is shown in Table 1 together with the standard deviation. Applying a one-tailed z-test to the area data confirms that the RBS method is statistically superior to the LDA method, but only at the $p = 0.92$ level.

The ROC curve can be deceptive at first glance, although the RBS curve looks to be only slightly higher on the graph than the LDA curve, for a typical specificity of 90% (at least that which would be expected by clinicians) the LDA has only 74% sensitivity compared to the RBS's 82%.

In practice, a specific value of the sensitivity/specificity pair must be chosen for operating the classifier. In our image processing system, the RBS was implemented with values of sensitivity and specificity chosen to be 84% and 85%, respectively. These were chosen in conjunction with

Table 1
A_z values for each classifier, with standard deviations and hypothesis test results

	A_z	σ_A	Hypothesis
RBS	0.9175	$\pm$0.0060	$A_{RBS} > A_{LDA}$: $p \sim 0.92$
LDA	0.9071	$\pm$0.0044	–
LVQ	0.8293	$\pm$0.0074	$p \gg 0.99$

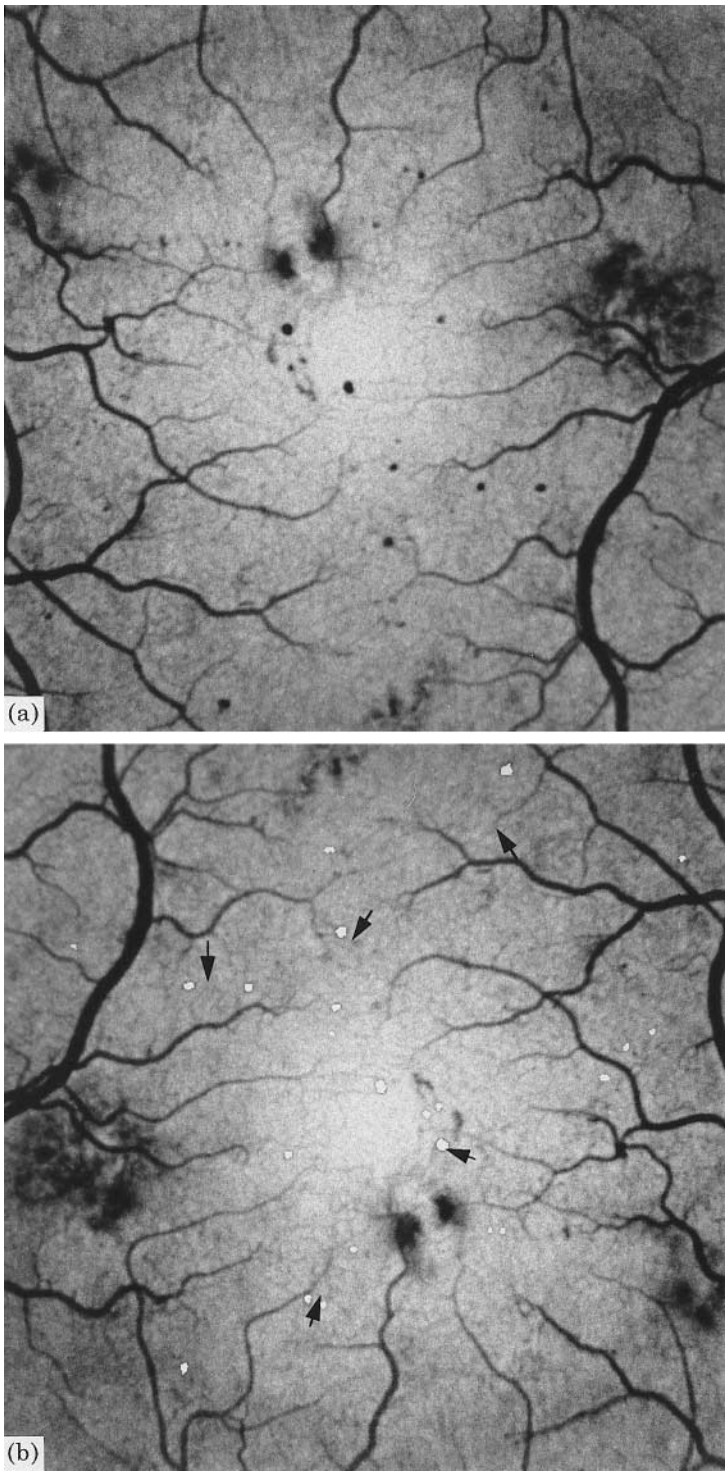

Fig. 3. (a) Fluorescein angiogram of a patient with diabetes. The MA are the small black dots. (b) Same angiogram image, with detected MA highlighted in white (some examples arrowed).

an ophthalmologist as being the values which had the highest specificity that did not compromise the specificity unduly. To demonstrate the classification of the MA directly, Fig. 3b shows the MA visible in Fig. 3a, as detected by the RBS.

6. Conclusion

We have demonstrated the application of manual and automated classification techniques to a difficult clinical diagnostic problem and ROC analysis was used to compare the methods. Both of the automated classification methods (LDA and LVQ) have the advantage of requiring little user intervention and needing only a short training time although extra time was required for ROC analysis. The extra effort required to construct the RBS (1 man-month in this case) was rewarded by better performance but only at the $p = 0.92$ level, and again ROC analysis was slow In circumstances where classification performance is diagnostically critical we therefore conclude that it may be more appropriate to consider manually derived classification techniques.

7. Summary

We compared the performance of three computer based classification methods when applied to the problem of detecting microaneurysms on digitised angiographic images of the retina. Microaneurysms are small bulges in the capillaries of the retina that are important indicators of the progression of the disease Diabetic Retinopathy. Quantitation is not currently implemented in clinical practise due to the large inter- and intra-observer variability. We have developed a fully automated image processing system which segmented candidate objects (microaneurysms or spurious objects) directly from the digitised images. Further processing produces a list of morphological and intensity-based features on each candidate for which can then be used train and test the classifiers. A set of 1659 candidate objects was used to train each classifier, and a further set of 1067 candidates was used as an independent testing set. We compared a rule based system, consisting of a number of quantitative and logical rules, with two automated methods, linear discriminant analysis and a learning vector quantiser artificial neural network. Each system was required to examine all the features for each candidate object and classify it as either a microaneurysm or a spurious object. ROC analysis was used to examine the performance of the trained classifiers when applied to the test set, and it was concluded that the greater development time required for the rule based system was justified as it gave a higher performance than the other methods.

Acknowledgements

Allan Frame is in receipt of a Medical Research Council studentship. The image processing system described is financially supported by Scotia Pharmaceuticals Ltd.

References

[1] I.M. Ghafour, D. Allan, W.S. Foulds, Common causes of blindness and visual handicap in the west of Scotland, Br. J. Ophthalmol. 67 (1983) 209.
[2] Diabetic Retinopathy Study Research Group, A modification of the Airlie House classification of diabetic retinopathy, Investigative Ophthalmology and Visual Science 21 (1981) 107.
[3] Early Treatment of Diabetic Retinopathy Study Group, Grading diabetic retinopathy from stereoscopic color fundus photographs: an extension of the Airlie House classification, Ophthalmology 98 (1991) 786.
[4] R. Klein, S.M. Meuer, S.E. Moss, B.E.K. Klein, The relationship of retinal microaneurysm counts to the four year progression of diabetic retinopathy, Archives of Ophthalmology 107 (1989) 1780.
[5] E.M. Kohner, M. Sleightholm, KROC collaborative study group, does microaneurysm count reflect the severity of early diabetic retinopathy, Ophthalmology 93 (1986) 586.
[6] C. Baudoin, F. Maneschi, G. Quentel, G. Soubrane, T. Hayes, G. Jones, G. Coscas, E.M. Kohner, Quantitative evaluation of fluorescein angiograms: microaneurysm counts, Diabetes 32(Suppl. 2) (1983) 8.
[7] R.C. Gonzalez, R.C. Woods, Digital Image Processing, Addison-Wesley, Reading, MA, 1992.
[8] T. Spencer, J.A. Olson, K.C. McHardy, P.F. Sharp, J.V. Forrester, An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus, Computers and Biomedical Research 29 (1996) 284.
[9] M.J. Cree, J.A. Olson, K.C. McHardy, J.V. Forrester, P.F. Sharp, Automated microaneurysm detection, IEEE International Conference on Image Processing 3 (1996) 699.
[10] M.J. Cree, J.A. Olson, K.C. McHardy, P.F. Sharp, J.V. Forrester, A fully automated comparative microaneurysm digital detection system, Eye, 1998, in press.
[11] C. Metz, ROC methodology in radiologic imaging, Investigative Radiology 21 (1986) 720.
[12] D. Keating, E. Mutlukan, A. Evans, J. McGarvie, B.E. Damato, A back propagation neural network for the classification of visual field data, Physics in Medicine and Biology 38 (1993) 1263.
[13] J. Caprioli, Discrimination between normal and glaucomatous eyes, Investigative Ophthalmology and Visual Science 33 (1992) 153.
[14] H. Hirsbrunner, F. Fankhauser, A. Jenni, A. Funkhouser, Evaluating a perimetric expert system: experience with Octosmart, Graefe's Archive for Clinical and Experimental Ophthalmology 228 (1990) 237.
[15] J.A. Swets, R.M. Pickett, The evaluation of diagnostic systems: methods from signal detection theory, New York, Academic Press, 1982.
[16] D.P. Chakraborty, Observer performance methodology in medical imaging, Tutorial Notes, SPIE Medical Imaging '96, 1996.
[17] W. Press, B. Flannery, S. Teukolsky, W. Vetterling, Numerical recipes in C, Cambridge University Press, 1992.
[18] C.E. Metz, Basic priciples of ROC analysis, Seminars in Nuclear Medicine 8 (1978) 283.
[19] J.M.O. Mitchell, Classical Statistical Methods, in: D. Michie, D.J. Spiegelhalter, C.C. Taylor (Eds.), Machine Learning, Neural and Statistical Classification, Ellis Horwood, London, 1994.
[20] T. Kohonen, J. Kangas, J. Laaksonen, K. Torkkola, LVQ_PAK: A program package for the correct application of learning vector quantization algorithms, Proceedings of the IEEE International Joint Conference on Neural Networks, 1992.
[21] J. Henery, Review of previous empirical comparisons, in: D. Michie, D.J. Spiegelhalter, C.C. Taylor (Eds.), Machine Learning, Neural and Statistical Classification, Ellis Horwood, New York, 1994.

Allan Frame received his BSc from Heriot-Watt University in Edinburgh and his M.Sc. from Aberdeen University. He is currently completing his Ph.D. thesis on "Computerised analysis of the retinal vessels" under the supervision of Peter Undrill, and this paper has arisen from this research. Other aspects to this research included texture analysis of retinal image features and model-based analysis of the vessel structure at different scales. During his Ph.D. research Allan was supported by a grant from the Medical Research Council, and he is now working with EDS Upstream Energy in Aberdeen.

Peter Undrill is Senior Lecturer and head of computer research in the department of biomedical physics and bioengineering. His main research interests are in medical image analysis, particularly in the digital processing of the mammogram and orthopaedic radiographs, mult-modality methods, image registration, image segmentation and three-dimensional modelling.

Dr Michael Cree completed his Ph.D. researching image reconstruction construction techniques for novel gamma camera designs at the University of Canterbury, New Zealand. He then worked on a short contract researching coherent optical measurement techniques with the Machine Vision Team, Industrial Research Ltd., NZ. He then worked for two and a half years with the University of Aberdeen, developing computed image processing and quantification techniques for monitoring the progression of diabetic retinopathy from retinal images. Michael is now with the dept. physics and electronic engineering, University of Waikato, NZ. He is a member of the IEEE.

John Olson received his M.B. and Ch.B. degrees from the University of Aberdeen and is currently completing his thesis on ''Digital imaging of the diabetic retina''. He is senior registrar in medical ophthalmology at Aberdeen Royal Infirmary, and his main research interest is in digital imaging of the diabetic retina.

Kenneth McHardy received his M.B. and Ch.B. degrees from the University of Aberdeen. He also received his MD from Aberdeen on the ''Measurement of acute changes in nutrient oxidation and protein turnover: studies in the role of insulin in man''. He is consultant physician at Aberdeen Royal Infirmary, with special interest in diabetic medicine. He is also senior lecturer in postgraduate medicine at the University of Aberdeen. He has a number of research interests in diabetic medicine, postgraduate medicine and digital imaging of the diabetic retina.

Peter Sharp is Professor of medical physics and head of department of biomedical physics and bioengineering. His research extends across a wide range of medical imaging topics, particularly in digital ophthalmology, radioisotope imaging and the measurement of medical image quality. Current interests include studies of retinal perfusion and the development of the scanning laser ophthalmoscope for three-dimensional studies of the retina.

John Forester is Professor of ophthalmology and head of department of ophthalmology, University of Aberdeen Medical School. Along with Professor Sharp and colleagues he has been responsible for instigating many research initiatives into the application of computers to ophthalmology including micro-aneurysm quantitation and turnover, retinal perfusion and the clinical evaluation of digital processing of retinal images.