# Automated diagnosis of ophthalmic conditions based on retinal image processing and machine learning techniques

**Christopher Seymour[1], Stelios Pavlidis[1*], Allan Tucker[1], Steve Counsell[1], Stephen Swift[1*]**
[1]School of Information Systems, Computing and Mathematics, Brunel University, UK
[*]Contact e-mail Stelios.Pavlidis@brunel.ac.uk

## Abstract

Glaucoma and diabetic retinopathy are two common medical disorders, associated with eye damage and progressive loss of vision. Given the availability of effective treatment, at the initial stages of disease development, early diagnosis is crucial. Here, we explore the application of various classification methodologies, based on processing of images of the fundus of the eye. We show that high accuracy of automatic classification is achieved using relatively simple techniques, which have the potential to aid the expert and reduce the time and cost of the diagnostic process.

## 1 Introduction

Glaucoma and diabetic retinopathy constitute serious eye disorders and leading causes of blindness, affecting millions of people worldwide [Pascolini and Mariotti, 2011]. It has been established that early treatment of these conditions can be extremely beneficial, hence prompt diagnosis is essential [Tapp et al., 2003, NHS, 2010]. Currently, however the diagnostic process is laborious and time consuming, with patients often required to wait weeks for test results. Diagnosis is largely based on manual inspection of images of the fundus, that is, the interior surface of the eye, including the retinal optic disc and macula [Saine and Tyler, 2006]. The cost of monitoring the population for such conditions is considerably high, due to the need for expensive equipment and medical expertise. According to official data the cost for sight test provision in the UK for the period between 2005 and 2006 was over £197 million [National Statistics, 2005]. Consequentially, the development of automated methodologies for diagnosis, that can reduce cost and time and assist clinical decision making, holds significant potential.

A variety of techniques have been applied in image processing of ophthalmic conditions. An investigation by [Hoover and Goldbaum, 2003] into detecting the optic disc in retinal images reported 89% correct detection, while more recently [Aquino et al., 2010] reported a correct identification rate of 99%. In [Priya and Aruna, 2001] the authors describe the classification of age related macular degeneration by detecting changes in the vascular system and patterns in the retina using probabilistic neural networks, achieving 95% accuracy. Regarding glaucoma and diabetic retinopathy, [Abdel-Ghafar and Morris, 2007] report on an attempt of limited success to automatically distinguish between optic disc images of healthy and unhealthy individuals, with only 65% accuracy. In [Nayak et al., 2009] the authors describe a methodology for automatic classification of images into glaucoma or healthy, based on various morphological features of the optic disc, with accuracy of 90%. A summary of the supervised classification techniques proposed for the detection of glaucoma, through the use of fundus images, can be found in [Acharya et al., 2011]. Nevertheless, automatic classification of retinal images, and especially in glaucoma detection, remains an issue that is open to further investigation. A study by [Greaney et al., 2002] into distinguishing a normal eye from one with glaucoma showed that algorithms proved no better than experienced observers. Hence a combination of machine learning methods and expert opinion is considered to be the optimal choice.

Here, we explore the application of a variety of classification techniques, for the discrimination of fundus images from healthy individuals and patients suffering from either glaucoma or diabetic retinopathy. We show that highly accurate classification can be achieved using a relatively simple approach, based on colour histograms. Histograms for colour image representation have been used extensively, as they constitute a simple way of representing the characteristics of an image and can serve for identifying objects in images [Swain et al., 1991]. For example, [Hijazi et al., 2009] have proposed a histogram based approach for the detection of age related macular degeneration.

In addition the Weighted Kappa metric [Cohen, 1968] is utilised to evaluate the consistency of results produced by different classification methods employed in our study.

The rest of the paper is organized as follows: in Section 2 we provide a description of the analytical approach and the dataset utilised to perform the analysis. Section 3 presents the obtained results. Finally, Section 4 provides some concluding remarks and suggestions for future work.

## 2 Materials and Methods

For the purposes of the analysis discussed here a training set consisting of 45 fundus images, of size 3504 by 2336 pixels, was utilised. The images, obtained from the Gold

Standard Database for Evaluation of Fundus Image Segmentation Algorithms [GSDEFISA, 2012], are split equally over the 3 conditions (15 for each one), that is, healthy, glaucoma and diabetic retinopathy. Figure 1 shows an example of three images of the eye fundus, corresponding to each of the discussed ophthalmic conditions.

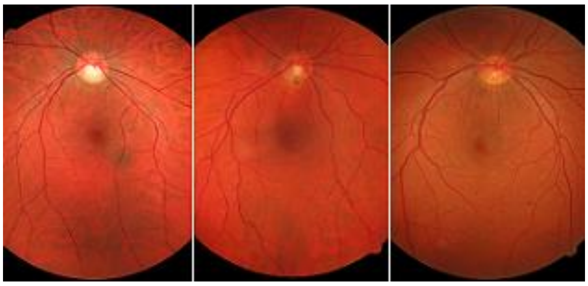

**Fig. 1.** Fundus images corresponding to healthy, glaucoma and diabetic retinopathy, respectively

First, a colour histogram for each image was prepared. Then based on all images in the dataset an average histogram was obtained for each channel and ophthalmic condition. Figure 2 displays the average colour histogram resulting from the images for the red, green and blue channels (A, B, C respectively), for each condition.

Visual inspection of the distributions on Figure 2A suggests that red pixel intensity values between 155 and 185 separate out diabetic retinopathy while values between 185 and 205 separate glaucoma. Consequentially, the average count of pixels and their maximum intensity value, within these ranges, for the red channel, were chosen as candidate features for the subsequent classification (features $X_1$ to $X_4$ respectively). For the green channel (Figure 2B), only the average count ($X_5$) and maximum intensity value ($X_6$) for the range between 15 and 85 were chosen. Notably, the green channel distributions are more similar, which is not surprising given the colour of fundus images. In the same fashion, for the blue channel (Figure 2C) we chose to extract the average count ($X_7$) and maximum of colour values ($X_8$) between 10 and 45.

In this way, we obtained a total of eight features, $X_1$ to $X_8$, in the order described above, that were utilised for the subsequent classification of images into healthy, glaucoma or diabetic retinopathy. All prepossessing steps were implemented in Java.

Classification was applied with the use of the WEKA data mining software [Hall et al., 2009]. Each unique combination of variables, i.e. 255 cases, was tested with a number of different classifiers, using 10-fold cross-validation. This is a well-known approach known as wrapper feature selection, where the search for the subset of most suitable attributes is based on the performance of the classification process itself [Kohavi and John, 1997]. Further evaluation of the quality of the obtained classification results was performed with the Weighted Kappa metric. Following the strategy described above allowed us to identify the combination of features and classifiers producing the most accurate result.

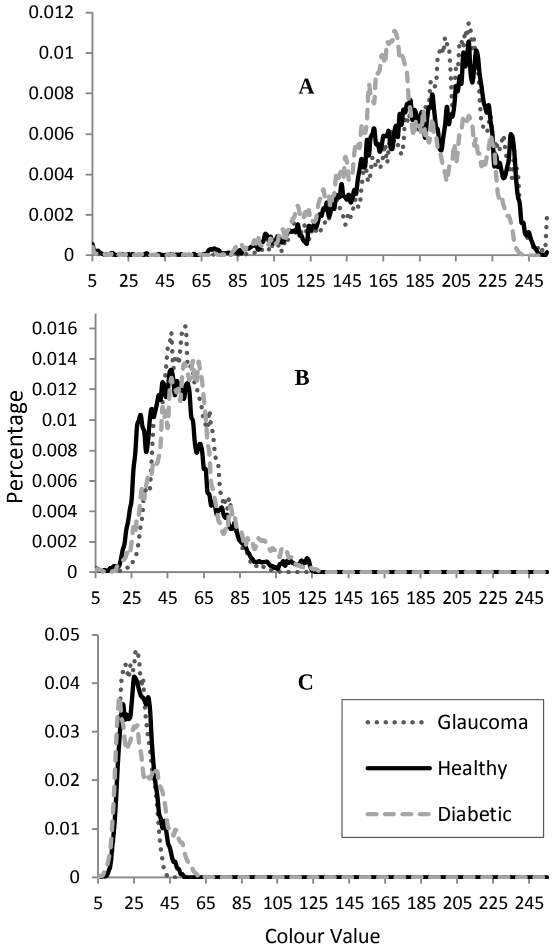


**Fig. 2.** Average histogram of images in the utilised dataset for red (A), green (B) and blue colour channels (C), per condition

## 3 Results

Initially, utilising WEKA and feature selection, a classification accuracy of about 70% was reached by the best performing classifiers. These include best-first decision tree (BFTree), nearest neighbour (IBk), instance based (IB1) and Kstar classifiers, as shown on table 1. Here, zeros indicate features not employed while ones features employed for classification.

**Table 1.**Classification results following feature selection

| **Classifier** | Feature | | | | | | | | Accuracy |
|---|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | |
| BFTree | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 71.11% |
| IBk | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 71.11% |
| IB1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 71.11% |
| KStar | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 69.63% |

As table 1 reveals, only 3 to 4 of the initially selected features are active for optimal accuracy and their combinations differ depending on the chosen classifier. Features $X_7$ and $X_8$ were present in all four of the top results while $X_1$, $X_5$ and $X_6$

**Table 2.** Weka Summary Statistics for all four models

| | 3 Way Model | D. Retinopathy | Glaucoma | Healthy |
|---|---|---|---|---|
| | MultilayerPerceptron | RandomCommittee | IBk | MultilayerPerceptron |
| Correctly Classified | 35 (77.78%) | 40 (88.89%) | 41 (91.11%) | 39 (86.67%) |
| Incorrectly Classified | 10 (22.22 %) | 5 (11.11 %) | 4 (8.89%) | 6 (13.33%) |
| Kappa statistic | 0.67 | 0.75 | 0.79 | 0.71 |
| Mean absolute error | 0.25 | 0.11 | 0.09 | 0.18 |
| Root mean squared error | 0.38 | 0.33 | 0.30 | 0.33 |
| Relative absolute error | 55.67 % | 24.81 % | 19.85% | 39.88% |
| Root relative squared error | 79.79 % | 70.51 % | 63.07% | 71.08% |

were absent. Figure 3 shows the influence of each variable over a 50% accuracy range for all 16,320 experiments (255 feature combinations by 64 alternative classification techniques).

Variables with a very high or very low observation rate exhibited the greatest impact on classification accuracy. An observation rate around 50% for a particular feature signifies that the presence or absence of that feature has little impact on classification accuracy. Feature $X_8$ (maximum intensity value of blue channel between 10 and 45) exhibited the largest positive impact on classification accuracy, with very high observation rate for accuracy over 50%.

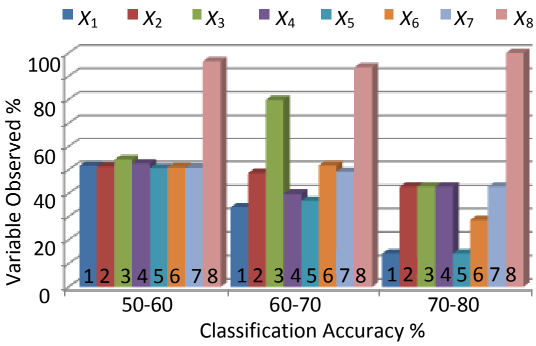


**Fig. 3.** Graph showing the influence of each feature on classification accuracy

Both variables for the green channel led to very poor accuracy rates. In the highest accuracy band variables $X_1$, $X_5$ and $X_6$ are characterised by very low observation rates. In fact the combination of $X_5$ and $X_6$ performed significantly worse than the majority classifier at under 20%.

### 3.1 Binary classification

While the observed accuracy of 71.11% is a significant improvement on the performance of the majority classifier at 33.33%, it is admittedly not particularly high. Hence, we proceeded further in an effort to improve the obtained results. In particular, classification was reduced to a binary problem, treating each condition separately, as present or absent. Images were first split into two classes, condition healthy and other, then split again into glaucoma and other and finally into diabetic retinopathy and other. Classification was applied separately in each case. This methodological approach produced considerably higher accuracy rates, as shown on Table 2. Here we observed an average rise in accuracy level equal to 18.45%, producing an average classification accuracy of 88.89%, using 10-fold cross validation. Table 3 shows the features utilised by each of the top performing binary models. Evidently, the values of the weighted kappa metric, which reveals the agreement between the predicted and actual classes, are quite high, with an average for the binary approaches equal to 0.75.

Combining the three models for each condition served as foundation to build our final classifier. In particular, classification was performed three times, according to each binary model and the final verdict was based on the agreement between the methods. For example, if an image is found not healthy, in the first instance, then corresponding to glaucoma in the second and finally not retinopathy in the third, it is classified as glaucoma. The classifier was tested against the original training set of retinal images, showing very high accuracy of 93.33%, classifying correctly 42 of the 45 images. Two of the images were identified with two conditions, one diagnosed as both healthy and containing signs of glaucoma and the other as healthy and containing signs of diabetic retinopathy. A third image was not identified with any of the 3 classes, that is, in each one of the three classification processes the image was assigned to the category others.

## 4 Conclusions

By exhaustively combining feature selection with a variety of classification techniques we developed a classifier that performs really well on the test set achieving an accuracy of 93.33%. The obtained results demonstrate that a simple approach such as the utilisation of colour histograms can achieve excellent classification accuracy, which is an interesting observation. More complex techniques using image segmentation and machine vision techniques dis-

**Table 3.** Features used in top models for each condition

| Condition Present | Classifier | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | $X_7$ | $X_8$ |
|---|---|---|---|---|---|---|---|---|---|
| Healthy | MultilayerPerceptron | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| Glaucoma | IBk | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| Diabetic Retinopathy | RandomCommittee | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |

cussed in the literature have been shown to produce comparable or even less accurate classifications.

Nevertheless, here we report on a preliminary stage of research and there is clearly considerable room for further investigation. For example, it would be interesting to examine at what level of image resolution this approach produces satisfactory results. Furthermore, the eight variables chosen during feature selection were identified by manually choosing regions within the colour histograms that appear as good candidates to separate the discussed ophthalmic conditions. A deeper investigation into how to choose the optimal regions could be beneficial. In addition, the methodology could also be combined with optic disc detection prior to feature selection, which may have a positive impact on the classification process, given that the area surrounding the disc does not contain any useful information. Finally, application of the discussed methodology to larger datasets would be more revealing of the method's merit and efficiency.

Although expert opinion remains useful and necessary in medical examination of ophthalmic conditions the utilisation of automated methods to assist diagnosis, as suggested in [Greaney et al., 2002], can simplify and speed up the diagnostic process.

## Acknowledgments

This work is based on the final year project of Christopher Seymour, in the School of Information Systems, Computing and Mathematics, Brunel University and funded by the EPSRC grant "Data Integrity and Intelligent Data Analysis Techniques Applied to a Glaucoma Progression Dataset" (EP/H019685/1) in collaboration with Moorfields Eye Hospital.

## References

[Abdel-Ghafar and Morris, 2007] R. A. Abdel-Ghafar and T. Morris. Progress towards automated detection and characterization of the optic disc in glaucoma and diabetic retinopathy. *Medical Informatics and the Internet in Medicine*, 32(1):19–25, 2007.

[Acharya et al., 2011] U. R. Acharya, S. Dua, X. Du, V. S. Sree, and C. K. Chua. Automated diagnosis of glaucoma using texture and higher order spectra features. *IEEE Transactions on Information Technology in Biomedicine*, 15(3)449–455, May 2011.

[Aquino et al., 2010] A Aquino, M.E. Gegundez-Arias and D. Marin. Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. *IEEE Transactions on Medical Imagînging*, 29(11):1860-1869, 2010.

[Cohen, 1968] Jacob Cohen. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4): 213-220, Oct 1968.

[GSDEFISA, 2012] Gold Standard Database for Evaluation of Fundus Image Segmentation Algorithms (2012), last viewed on 24/04/2012, <http://www5.informatik.uni-erlangen.de/research/data/fundus-images/>

[Greaney et al., 2002] Michael J. Greaney, Douglas C. Hoffman, David F. Garway-Heath, Mamdouh Nakla, Anne L. Coleman, and Joseph Caprioli; (2002) "*Comparison of Optic Nerve Imaging Methods to Distinguish Normal Eyes from Those with Glaucoma.*" In: Investigative Ophtalmology and Visual Science 43(1): 140 – 145.

[Hall et al., 2009] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann and Ian H. Witten. The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, 11(1), 2009.

[Hijazi et al., 2009] Mohd Hanafi, Ahmad Hijazi, Francs Coenen and Yalin Zheng. A Histogram Approach for the Screening of Age-Related Macular Degeneration. *In proceedings of the 13th annual conference on Medical Image Understanding and Analysis*, July 2009.

[Hoover and Goldbaum, 2003] A. Hoover and M. Goldbaum. *Locating the optic nerve in retinal image using the fuzzy convergence of the blood vessels.* Medical Imaging, IEEE, 22 (8): 951 – 958, 2003.

[Kohavi and John, 1997] Ron Kohavi and George H. Kohn. Wrappers for feature subset selection. Artificial Intelligence, 97 (1–2):273–324, 1997.

[Nayak et al., 2009] Jagadish Nayak, Rajendra Acharya U, P. Subbanna Bhat, Nakul Shetty and Teik-Cheng Lim. Automated Diagnosis of Glaucoma Using Digital Fundus Images. *J Med Syst*, 33:337–346, 2009.

[NHS, 2010] Glaucoma treatment (2010), last viewed on 24/04/2012<http://www.nhs.uk/Conditions/Glaucoma/Pages/Treatment.aspx>

[National Statistics, 2005] National Statistics (2005). General Ophthalmic Services: Ophthalmic Statistics for England: 1994-95 to 2004-05. NHS, The Information Centre for Health and Social Care.

[Pascolini and Mariotti, 2011] Donatella Pascolini and Silvio Paolo Mariotti. Global estimates of visual impairment: 2010. *British Journal of Ophthalmology*, 2011.

[Priya and Aruna, 2011] R Priya and P Aruna. *Automated diagnosis of Age-related macular degeneration from color retinal fundus images.* In 2011 3rd International Conference on Electronics Computer Technology. IEEE, pp. 227-230, 2011.

[Swain at al., 1991] M.J. Swain, and D.H. Ballard. Color Indexing. *International Journal of Computer Vision*, 7(1), pp 11-32, 1991.

[Saine and Tyler, 2006] Patrick J. Saine and Marshall E. Tyler. *Ophthalmic Photography: Retinal Photography, Angiography, and Electronic Imaging*. 2nd edition, Butterworth-Heinemann, Boston, 2006.

[Tapp et al., 2003] Tapp RJ, Shaw JE and Harper CA. The prevalence of and factors associated with diabetic retinopathy in the Australian population. *Diabetes Care* 26(6):1731-7, june 2003.