33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Automated Diagnosis of Referable Maculopathy in Diabetic Retinopathy Screening

Andrew Hunter, James A. Lowell, Bob Ryder, Ansu Basu, David Steel

Abstract—**This paper introduces an algorithm for the automated diagnosis of referable maculopathy in retinal images for diabetic retinopathy screening. Referable maculopathy is a potentially sight-threatening condition requiring immediate referral to an ophthalmologist from the screening service, and therefore accurate referral is extremely important. The algorithm uses a pipeline of detection and filtering of "peak points" with strong local contrast, segmentation of candidate lesions, extraction of features and classification by a multilayer perceptron. The optic nerve head and fovea are detected, so that the macula region can be identified and scanned. The algorithm is assessed against a reference standard database drawn from the Birmingham City Hospital (UK) diabetic retinopathy screening programme, against two possible modes of use: independent screening, and pre-filtering to reduce human screener workload.**

I. Introduction

DIABETIC retinopathy is a leading cause of blindness affecting 95% of type 1 diabetics within 15 years of onset [1], of whom 2% may become blind and 10% suffer severe visual impairment. Annual screening is recommended to identify progression allowing early treatment, which may prevent up to 90% of cases of blindness [2]. Grading standards for screening include background and sight-threatening retinopathy, the latter being particularly important to detect reliably. A major category of sight-threatening retinopathy is referable maculopathy, which may be defined as the presence of one or more exudates or three or more HMAs (Haemorrhages or microaneurysms) within the macula region – that is, within an optic disk diameter of the centre of the fovea.

This paper introduces an algorithm for the detection of referable maculopathy, and evaluates it against a reference dataset taken from a diabetic retinopathy screening programme. The algorithm grades the image to ensure it is sufficiently clear to permit lesion detection, detects and measures the optic nerve head (ONH), finds the fovea, then detects and counts exudates and HMAs (hemorrhages and microanerysms). It is evaluated for two modes of use: first, as a standalone system replacing the human grader; second, as a pre-screening stage that identifies and discards clearly non-referable patients, passing on all positives and a proportion of negatives to a human screener.

Manuscript received March 26, 2011.

Andrew Hunter is at the University of Lincoln, UK. (ahunter@lincoln.ac.uk).

James Lowell is working in the image processing software industry.

David Steel is at Sunderland Eye Infirmary, UK, (dhwsteel@hotmail.com).

Bob Ryder and Ansu Basu are at City Hospital, Birmingham, UK.

II. Background

There is a substantial literature on analysis of retinal images relevant to screening for diabetic retinopathy. Space precludes a detailed discussion here; a good survey is presented in [3]. The majority of this literature relates to the segmentation of key features including exudates, HMAs, ONH and fovea; however, there is relatively little that addresses the diagnosis of retinopathy using these features. It is worth noting that studies of segmentation and classification algorithms evaluated on a per lesion basis are difficult to interpret for two reasons: the datasets used critically affect performance as diabetic screening program images often include unclear and difficult to classify images and lesions; and the step from lesion identification to patient diagnosis is non-trivial.

Hipwell [4] and Larson [5] developed systems to diagnose retinopathy by detection of HMAs alone, relying on the prevalence of these to compensate for the lack of explicit detection of exudates or other signs. Usher [6] combined detection of exudates and HMAs. Abrahamoff [7] combined white lesion, HMA and image quality enhancement. Philip [8] evaluated an automated system that combined MA detection and image quality analysis, designed to detect presence of absence of any retinopathy (whether background or referable) against a large dataset of 6722 patients in a retinal screening programme. They report 97.9% sensitivity and 67.4% specificity for detection of referable/observable retinopathy or maculopathy (compared to 99.1% sensitivity and 67.4% by manual screening). Although system performance is far short of manual screening, if used to pre-filter patients the system potentially allows a 60% decrease in manual screening; the cost saving for the Scottish NHS being estimated at £200,000 per annum [9]. More recently, they have investigated a system that combines HMA and exudate detection, showing that detecting both and combining evidence gives better results than MA alone [10]. This system uses a weighted sum of evidence from several sources, including multiple lesion candidates and taking into account their position with respect to the fovea.

Lesion detection algorithms use a range of techniques. First, potential lesion areas must be segmented, typically by contrast enhancement and thresholding [11], Fuzzy C Means [12] region growing [13] or filtering techniques [11]. Some authors assume any segmented region is a lesion-of-interest.

Others, recognizing that retinal images often contain benign distractors (e.g. drusen, light artifacts, visually broken-up tiny blood vessels, choroidal vessels) use a second stage [6,12] where features are extracted and a more sophisticated classifier determines whether a segmented region is actually a lesion. In our experience the latter approach provides higher levels of per-lesion performance and thus ultimately better subserves diagnosis of diabetic retinopathy.

III. Referrable Maculopathy Screening Algorithm

We use two macular-centred images, one from each eye of the patient; images are converted to grey-scale by extracting the green component. We detect and measure the optic nerve head using the algorithm described in [14]. Then, the fovea is located by selecting the maximum correlation point with a 40×40 Gaussian filter with standard deviation $\sigma = 22$ pixels [15]. The image clarity is checked to determine if it is gradeable using the algorithm presented in [16], which yields a quality metric on a scale of 1-5. If the image is gradeable (quality 1-3 corresponds to "machine gradeable"), we locate retinal lesions using the algorithm below.

1. *Candidate points*. We detect *peak points* – pixels that have (for white lesion detection) intensity at least as high as their four-connected neighbors. There is guaranteed to be at least one peak point in any white lesion. This stage typically produces several thousand *candidate* points.

2. *Winnowing*. We rate the "lesionness" of each candidate point by using a robust local contrast measure. We calculate the maximum contrast between the peak point and eight circular regions, C_i radius $r_i \in \{1, 2, 3, 5, 8, 12, 15, 18\}$, each one pixel wide. We represent each circular region using the *lightest* pixel value, and choose the *darkest* such value, so calculating the highest contrast between peak and circular maximum; see equations (1) and (2):

$$I_i = \max_{\mathrm{x}}\{I_x : x \in \mathsf{C}_i\} \tag{1}$$

$$l = I_p - \min_{\mathrm{i}}\{I_i\} \tag{2}$$

Where I_x is the intensity of pixel x, I_p is the intensity of the peak point, and l is the lesionness measure. This process exploits the observation that for a real lesion it should be possible to draw a circle around the lesion that is wholly contained in the background retina, and therefore has strong contrast to the peak from all its pixels. We extract the fifty candidate white lesions with the highest lesionness measures for further analysis. An analogous process is used to identify fifty dark lesion candidates, using "trough" points and reversing the minimum and maximum operators in equations (1) and (2) to calculate the lesionness measure. This algorithm is related to the morphological hit-and-miss procedure for lesion detection [17].

3. *Segmentation*. We attempt to segment the candidate, bearing in mind that real lesions may be small, blurred and/or have low contrast, and that non-lesions may be highly irregular shaped noise patches. We used a specialized variant of region growing that grows to and beyond the candidate boundary, and then exploits the expected shape characteristics to identify the correct segmentation. In brief, the algorithm starts from the peak point and grows the candidate lesion by acquiring (for white lesions) the brightest neighboring pixel. Having first calculated a smoothed gradient magnitude image using (3), we then compute (on each iteration) a gradient contrast measure using (4-6):

$$\mathbf{\Gamma} = \frac{-\|\nabla \mathbf{I}\|}{\max(\|\nabla \mathbf{I}\|)} \tag{3}$$

$$\mathcal{B}_i = \mathcal{R}_i \oplus S - \mathcal{R}_i \tag{4}$$

$$w_i = 1.0 + 0.4^i \tag{5}$$

$$g_i = w_i \left(\frac{\sum_{x \in \mathcal{B}_i} \Gamma_{\mathbf{x}}}{\#\mathcal{B}_i} - \frac{\sum_{x \in \mathcal{R}_i} \Gamma_{\mathbf{x}}}{\#\mathcal{R}_i} \right) \tag{6}$$

where g_i is the gradient contrast measure on iteration i, $\|\nabla \mathbf{I}\|$ is the gradient magnitude image smoothed with a Gaussian filter, $\mathcal{R}_i$ is the candidate region at iteration i, S is a 3×3 structuring element, $\oplus$ is the morphological dilation operator, $\mathcal{B}_i$ is the boundary region at iteration i, w_i is the iteration weight and $\#X$ is the cardinality (number of elements) of set X. The iteration weight w_i compensates for the fact that earlier iterations, which involve few pixels and prominently include the peak point, may demonstrate very high contrast due to noise.

On each iteration we also calculate a compactness measure:

$$c_i = \frac{\#\mathcal{B}_i{}^2}{\#\mathcal{R}_i} \tag{7}$$

We run the region growing algorithm for a maximum of 250 iterations, or until $c_i > 30$. As lesions are reasonably compact, the latter condition invariably occurs when a non-lesion candidate is being segmented, leading to a highly convoluted shape, at which point there is no value in continuing the region growing.

Then, we select the iteration with the highest value of g_i, and this indicates the correct lesion segmentation, $\mathcal{R} = \mathcal{R}_j : j = \arg\max_{\mathrm{i}} g_i$. As shown in [18], this approach segments the lesion with great accuracy, which in turn is critical in ensuring that the features extracted in the next stage are effective. We then define the lesion boundary, $\mathcal{B} = \mathcal{R} \oplus S - \mathcal{R}$ and interior, $\mathcal{R}^- = \mathcal{R} \ominus S$, where $\ominus$ is the morphological erosion operator.

For dark lesions the algorithm works just as above, except that we acquire the darkest neighbor during region growing.

4. *Feature extraction*. We extract features that help to distinguish lesions from non-lesions. These features characterize the shape, colour and texture of the candidate, and the contrast between it and the surrounding retinal background. Our approach is to extract a wide range of features, then to use feature selection techniques to reduce the number used by the classifiers. The full set of features extracted includes the following.

Shape features. We extract from $\mathcal{R}$ the *area* (number of pixels), *perimeter (in pixels)*, *equivalent diameter* (of circle of this area), *major* and *minor axis length*, *eccentricity* (*minor*/*major* axis length), *convex area* (area of convex hull), *solidity* (*area/convex area*) and *compactness*.

Contrast measures. We extract the contrast-of-means between $\mathcal{R}$ and $\mathcal{B}$, and between $\mathcal{R}^-$ and $\mathcal{B}$, separately for each of the Hue, Saturation and Luminosity (Intensity)

channels of the HSI colour representation. We also use the maximum contrast, $g = \max g_i$ from the region growing phase. Lesions typically have strong contrast with the surrounding retinal background. Note that these features are critically dependent on good segmentation.

Color features. We analyze the components of the RGB, HIS and CIE color models separately. For each of these, we determine the *mean*, *standard deviation*, *skew* and *kurtosis* of $\mathcal{R}$.

Texture features. We characterize the texture of $\mathcal{R}$ using co-occurrence matrix features. We discretize to four intensity bins (each containing 25% of values), producing 4×4 co-occurrence matrices; four such matrices are formed for right, down, down-right and down-left directions respectively. We then extract five features from each co-occurrence matrix: the *energy*, *correlation*, *contrast*, *homogeneity* and *entropy*:

$$\mathrm{E} = \sum_{\mathrm{i,j}} \mathrm{p}\,(\mathrm{i},\mathrm{j})^2 \tag{8}$$

$$\rho = \frac{\sum_{\mathrm{i,j}} \mathrm{ij}.\,\mathrm{p}\,(\mathrm{i},\mathrm{j})^2 - \mu_x \mu_y}{\sigma_{\mathrm{x}} \sigma_{\mathrm{y}}} \tag{9}$$

$$\mathrm{C} = \sum_{\mathrm{i,j}} (\mathrm{i}-\mathrm{j})^2 \mathrm{p}\,(\mathrm{i},\mathrm{j}) \tag{10}$$

$$\mathrm{H} = \frac{\sum_{\mathrm{i,j}} \mathrm{p}(\mathrm{i},\mathrm{j})}{1 + (\mathrm{i}-\mathrm{j})^2} \tag{11}$$

$$\mathrm{E} = -\sum_{\mathrm{i,j}} \mathrm{p}\,(\mathrm{i},\mathrm{j}).\log \mathrm{p}(\mathrm{i},\mathrm{j}) \tag{12}$$

where $\mathrm{p}(\mathrm{i},\mathrm{j})$ is the probability of co-occurrence of i^{th} row and j^{th} column, $\mu_x, \mu_y, \sigma_{\mathrm{x}}$ and σ_{y} are the means and standard deviations of the row and column marginal probabilities. We take the mean, standard deviation, skew and kurtosis of the Fourier transform of the green channel in a 16×16 tile centred at the peak point, and the first 10 principal components of the Fourier transforms, which may characterize periodic texture features. We also apply a Prewitt edge detection filter to the 16×16 tile, and take the mean, standard deviation, skew, kurtosis and first 10 principal components of this.

5. *Feature selection*. We select effective features for classification of both white and dark lesions by using a hierarchical feature selection strategy, as described in [20]. The final selected feature sets contain 58 features for white lesions, and 18 features for dark lesions. For brevity we do not list these, but note that a selection of the shape, color, contrast and texture features above are combined.

6. *Classification.* We perform separate classification for dark and white lesions. For each of these, we use ten-fold cross validation to train Multilayer Perceptron neural networks. We used a combination of back propagation (100 iterations) and conjugate gradient descent (500 iterations) to optimize the networks; each network had one hidden layer with 28 and 10 hidden units for the white and dark lesion networks respectively, these settings being determined empirically. The networks are optimized to estimate the posterior class probability in the two-class classification problem (lesion or non-lesion candidate) using the cross-entropy error function. Input features were normalized to range [0,1] using the minimax method. We do not attempt to distinguish lesion sub-type (e.g. exudates versus drusen) as in this application very high sensitivity is required and it is safer to treat all lesions as positives.

7. *Diagnosis*. Having estimated the probability that each candidate is a lesion, we need only to set confidence thresholds for white lesions and dark lesions, and we can then apply a decision rule to diagnose referable maculopathy. In contrast to other authors, we directly follow the standard definition of referable maculopathy: we diagnose on the basis that one or more exudates (white lesions) or three or more dark lesions (HMAs) are identified in the macular region in either eye, using the prior ONH segmentation and fovea detection algorithms to identify the region of interest.

TABLE I
PRE-FILTER PERFORMANCE WITH DIFFERENT QUALITY THRESHOLDS, PER PATIENT

Se.	Sp.	TP	FN	TN	FP	Ung	Scr
97%	80%	48	1	320	78	47	173
100%	95%	27	0	298	15	153	195

Top row: images of quality grade 1-3 (according to [16], equivalent to "human-gradeable) processed. Bottom row: images of quality grade 1-2 processed. Se=Sensitivity, Sp=Specificity, TP=true positive, FN=false negative, TN=true negative, FP=false positive, Ung=ungradeable, Scr=TP+FP+Ung=patients passed on for human grading.

IV. EVALUATION

We evaluated the algorithm using 1000 macula-centered fundal images from 507 randomly sampled patients of the Birmingham City Hospital diabetic screening programme. Fourteen patients have only a single image, the balance (493) having two (one per eye). Images were acquired using a Canon CR6 45MNf fundus camera with a 45 degree field of view, 760×570 resolution. 86 images were classified as ungradeable due to cataract or capture errors, including poor illumination. Of the remaining images, 702 (77%) have no abnormalities, 212 have some retinopathy (23%) of which 61 (12%) have referable maculopathy. The images contain many distractors; for example, 152 contain benign drusen which must be successfully distinguished from exudates.

The patients were classified by two clinicians who identified all lesions within the macular region. Where there were disparities (in 130 images) a third clinician (consultant ophthalmologist) adjudicated. The adjudications were very balanced (54% agreement with the first clinician, 46% with the second). It was observed that these cases largely correspond to medium quality images making lesion classification difficult.

To evaluate the system for automated screening, we first remove all patients with an ungradeable image (on the basis that these would necessarily be referred). We set the neural network decision thresholds to attempt to achieve UK National Screening Committee recommendations of at least 80% sensitivity and 95% specificity per patient. With this decision rule the system achieved 80% sensitivity, 93% specificity on gradeable images. The corresponding performance for individual white lesion candidates was 91%

sensitivity and 91% specificity; for dark lesion candidates 98% sensitivity and 72% specificity.

For pre-filtering, we set the decision thresholds to achieve as close as possible to 100% sensitivity. In this mode the system effectively filters out "clearly healthy" images, passing on any dubious cases for human screening. We achieve (with respect to gradeable images) 97% sensitivity and 80% specificity, with a single false positive image (due to a medium quality image that is defined as acceptable by a clinician, but is sufficiently blurred to prevent our algorithms functioning). In this mode of use 173 patients would be referred for human screening.

We can filter out the single false negative image by changing the gradeability threshold so that only high quality images are analyzed (i.e. using a tighter gradeability threshold than is required by a human grader). On this basis we achieve 100% sensitivity at 95% specificity on the assessed images, although we pass on 153 patients as having at least one image unsuitable for automated assessment (from 170 unsuitable images), as opposed to 47 patients from 86 human-ungradeable images. The very significant gain in specificity is noteworthy, indicating that poor image quality is a key trigger for misclassification of lesions. Using this approach in total 195 patients are referred onto human screening – potentially a reduction of 70% in the workload compared with human screening of all images. This is broadly consistent with the results obtained by Philip [8] for "any disease" referral.

We note however that the evaluation has two limitations. First, the number of cases of referable maculopathy is fairly small (49 patients), so that we cannot be confident the results would carry over at 100% sensitivity to a larger study. Second, the study uses images from a single site and camera, and it is likely that performance would drop using images with different optical characteristics.

V. Conclusion

This paper has presented an algorithm for automated detection of referable maculopathy. The algorithm uses a pipeline of progressively more discriminative stages to find and classify candidate retinal lesions. The accurate segmentation of lesions is critical to the extraction of discriminative features. In contrast to other approaches in the literature we directly model the referable maculopathy screening criteria by counting lesions within the fovea. We have demonstrated a high level of performance for automated detection of referable maculopathy. For future work we will further improve lesion segmentation, feature selection and classification, and evaluate performance across a wider range of datasets taken from multiple retinopathy screening programmes.

Acknowledgment

This work was supported under Diabetes UK grant No. BDA:RD00/0002033.

References

[1] R. Klein, B. Klein *et al.* "The wisconsin epidemiologic study of diabetic retinopathy ii. prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years," *Archives of Ophthalmology*, vol. 102, pp. 520–526, 1984.

[2] P.H. Scanlon, R. Malhotra, G. Thomas, C. Foy, J.N. Kirkpatrick, N. Lewis-Barned, B. Harney, and S.J. Aldington, "The effectiveness of screening for diabetic retinopathy by digital imaging photography and technician ophthalmoscopy," Diabetes UK. *Diabetic Medicine*, vol. 20, pp. 467 474, 2003.

[3] J. Winder, P. Morrow *et al.* "Algorithms for digital image processing in diabetic retinopathy", *Computerized Medical Imaging and Graphics*, vol. 33 no. 8, pp. 608-622, 2009.

[4] J.H. Hipwell, F. Strachan *et al.* "Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool", *Diab Med* vol. 17, pp. 588–94, 2000.

[5] M. Larsen, T. Gondolf *et al.* "Assessment of automated screening for treatment-requiring diabetic retinopathy", *Curr Eye Res* vol. 32, pp. 331-336, 2007.

[6] D. Usher, M. Dumskyj *et al.* "Automated detection of diabetic retinopathy in retinal images: a tool for diabetic retinopathy screening," *Diabet Med*, vol. 21, pp. 84–90, 2004.

[7] M.D. Abramoff, M. Niemeijer *et al.* "Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes", *Diabetes Care* vol. 31, pp. 193-198, 2008.

[8] S. Philip, A.D. Fleming et al. "The efficacy of automated 'disease/no disease grading for diabetic retinopathy in a systematic screening programme", *Br. J. Ophthalmol.* vol. 91 pp.1512–1517, 2007.

[9] G.S. Scotland, P. McNamee *et al.* "Cost-effectiveness of implementing automated grading within the national screening programme for diabetic retinopathy in Scotland", *Br J Ophthalmol* vol. 91 pp. 1518-1523, 2007.

[10] A.D. Fleming, K.A. Goatman *et al.* "The role of haemorrhage and exudate detection in automated grading of diabetic retinopathy", *Br. J. Ophthalmol.* vol. 94, pp. 706-711, 2010.

[11] T. Spencer, J.A. Olson *et al.* "An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescien angiograms of the ocular fundus," *Computers and biomedical research*, vol. 29, pp. 284–302, 1996.

[12] A. Osareh, M. Mirmehdi *et al.* "Automatic recognition of exudative maculopathy using fuzzy c-means clustering and neural network," in *Medical Image Understanding and Analysis*, E. Claridge and J. Bamber, Eds. 2001, pp. 49–52, BMVA Press.

[13] C. Sinthanayothin, J.A. Boyce *et al.* "Automated detection of diabetic retinopathy on digital fundus images," Diabetic Medicine, vol. 19, pp. 105–112, 2002.

[14] J.A. Lowell, A. Hunter, D. Steel, B. Ryder, E. Fletcher. "Optic Nerve Head Segmentation" *IEEE Transactions on Medical Imaging*, 23 (2), 256-264, Feb 2004.

[15] C. Sinthanayothin, J.A. Boyce, H.L. Cook, and T.H. Williamson, "Automated localization of the optic disc, fovea, and retinal blood vessels from digital colour fundus images," *Br J. Ophthalmol*., vol. 83, pp. 902–910, 1999.

[16] A. Hunter, J.A. Lowell *et al.* "An Automated Retinal Image Quality Grading Algorithm," Proc. IEEE EMBC 2011, Boston USA.

[17] G. E. Øien, P. Osnes. "Diabetic retinopathy: Automatic detection of early symptoms from retinal images", *Proc. Norwegian Signal Processing Symposium*, 1995.

[18] E. Massey, A. Hunter *et al.* "A robust lesion boundary segmentation algorithm using level set methods," In: *World Congress on Medical Physics and Biomedical Engineering* , Munich, Germany, 2009.

[19] O. Brinchmann-Hansen, O. Engvold. "Microphotometry of the blood column and light streak on retinal vessels in fundus photographs," *Acta Ophthalmologica, Supplement*, vol. 179, pp. 9–19, 1986.

[20] A. Hunter, J. Lowell *et al.* "Quantification of diabetic retinopathy using neural networks and sensitivity analysis." In: Proc. Artificial Neural Networks in Medicine And Biology, 13-16 May, Goteborg, Sweden, 2000.