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

Evaluation of the grading performance of an ensemble-based microaneurysm detector

Bálint Antal, István Lázár, András Hajdu, Zsolt Török, Adrienne Csutak, Tünde Pető

Abstract— **In this paper, results of a diabetic retinopathy screening experiment are presented which is based solely on the findings of a microaneurysm detector. For this purpose, an ensemble-based algorithm developed by our research group was used; this provided promising results in our earlier experiments. At its best, the 1200 image of the Messidor database is classified by this detector with a sensitivity of 96%, a specificity of 51% and achieved an AUC of 0.87. As anticipated, larger microaneurysm counts are recognized with higher level of certainty. Therefore, this approach might be expected to have good performance in relation to the severity of the disease.**

I. Introduction

Diabetic retinopathy (DR) is one of the most common causes of blindness. In clinical practice there are existing treatments which are very effective in reducing vision loss in case of early detection of the disease. Due to the large number of patients with diabetes, provision of population coverage can be problematic in certain parts of the world. Several efforts have been made to aid the manual screening by an automated image analysis based process. The development of such systems has reached the point where they produce equal results to human experts [1]. One of the key components of DR screening is the localization of microaneurysms (MAs), which are the earliest and at the same time the hallmark lesions of this disease, therefore a prime candidate for automated image analysis. According to the clinical protocol, the presence of a single MA refers to DR.

Our research group has been developing an ensemble-based MA detector. This methods has been proven to be very effective in the Retinopathy Online Challenge [2], where it is currently ranked as first. In clinical grading (both manual and automated) high sensitivity is desired with a fair specificity rate. Since MA detection is a key component in automatic screening, it is critical to have good sensitivity/specificity values for this step.

MAs appear as small, dark, circular objects in retinal images. To develop a successful MA detector, one must have to deal with several difficulties. First, the sensitivity of such algorithms is low. As our previous work showed [3], the first problem can be fixed by setting up an ensemble consisting of ⟨preprocessing methods, candidate extractors⟩ pairs from different approaches. It is also a problem that several false detections appear as MAs can be confused with some features (e.g. pigment, artefacts) and vessel fragments. The solution for the this problem is to establish a voting scheme among these pairs.

In this paper, the results of the MA detector developed by our group is presented. The 1200 images of Messidor database was used for testing purposes. The system was trained on an independent dataset.

The rest of the paper is organized as follows: in section II, a brief description of the ensemble-based microaneurysm detector is provided. Section III contains quantitative results, and the the discussion of lessons learnt in this experiment. Finally, section IV details the conclusions.

II. Our ensemble-based microaneurysm detector

In this section, we present our ensemble-based MA detector, which can be organized into the following steps. All available preprocessing methods and candidate extractors were placed in the ensemble pool (section II-A and II-B). A subset of combinations from the ensemble pool were selected after evaluation on the training set (section II-C). MAs are extracted using selected combinations of test images (section II-D). A spatial voting is performed on the detected MAs (section II-D). The ensemble approach is also shown in Figure 1.

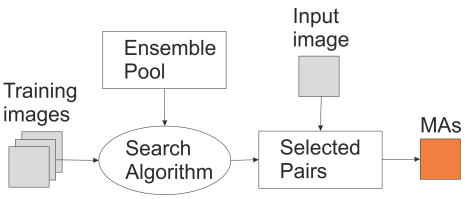

Fig. 1. Flow chart of the ensemble-based framework.

A. Preprocessing methods

In this section, the selected preprocessing methods are presented. These are to be applied before executing MA candidate extraction. These algorithms are collected from current literature recommendations for medical image processing.

1) Walter-Klein contrast enhancement [4]: This preprocessing method aims to enhance the contrast of fundus images by applying a gray level operator, which stretches the histogram of the image according to a real-valued parameter.

Bálint Antal, István Lázár and András Hajdu are with University of Debrecen, Faculty of Informatics, 4010 Debrecen, POB 12, Hungary. Email: {antal.balint, lazar.istvan, hajdu.andras}@inf.unideb.hu

Zsolt Török and Adrienne Csutak are with the University of Debrecen, Medical and Health Science Centre, 4032 Debrecen, Nagyerdei Krt. 98, Hungary. Email: zsolt.torok@astrid.hu, acsutak@dote.hu

Tünde Pető is with the Department of Research and Development, Moorfields Eye Hospital NHS Foundation Trust, 162 City Road, London, EC1V 2PD, UK. Email: Tunde.Peto@moorfields.nhs.uk

2) Contrast Limited Adaptive Histogram Equalization (CLAHE) [5]: The image is split into disjoint regions, and on each region a local histogram equalization is applied. Then, the boundaries between the regions are eliminated with a bilinear interpolation.

3) Vessel removal and extrapolation [6]: This methods investigates the effect of processing images with the complete vessel system being removed, based on the idea proposed in [6]. It extrapolates the missing parts to fill in the holes caused by the removal using inpainting.

4) Illumination equalization [8]: This preprocessing method aims to reduce the vignetting effect, which causes uneven illumination of retinal images. The difference of the desired average intensity and the actual local average intensity to each pixel intensity.

5) No preprocessing: In the ensemble pool, the results of candidate extractors are also included on the original images without any preprocessing.

B. MA candidate extractors

Candidate extraction is a process aiming to spot objects in the image showing MA-like characteristics. Individual MA detectors follow their own way to extract MA candidates. In this section, a brief overview of the candidate extractors involved in our analysis is provided. These algorithms realize different approaches that were described in details in the introduction.

1) Walter et al. [9]: Candidate extraction is accomplished by grayscale diameter closing. That is, the method aims to find all sufficiently small dark patterns on the green channel. Finally, a double thresholding is applied.

2) Spencer et al. [10]: From the input fundus image, the vascular map is extracted by applying twelve morphological top-hat transformations with twelve rotated linear structuring elements. Then, the vascular map is subtracted from the input image, which is followed by a Gaussian matched filtering. The resulting image is then binarized with a fixed threshold. Since the extracted candidates are not precise representations of the actual lesions, a region growing step is also applied to them.

3) Circular Hough-transformation [11]: Following the idea presented in [11], we established an approach based on the detection of small circular spots in the image. Candidates are obtained by detecting circles on the images using circular Hough transformation. With this technique, a set of approximately circle-shaped objects can be extracted from the image.

4) Lazar et al. [12]: Pixel-wise cross-section profiles with multiple orientations are used to construct a multi-directional height map. This map assigns a set of height values that describe the distinction of the pixel from its surroundings in a particular direction. In a modified multilevel attribute opening step a score map is constructed, from which the MAs are extracted by thresholding.

5) Zhang et al. [13]: In order to extract candidates, this method constructs a maximal correlation response image for the input retinal image. This is accomplished by considering the maximal correlation coefficient with five Gaussian masks with different standard deviations for each pixel. The maximal correlation response image is thresholded with a fixed threshold value to obtain the candidates. Vessel detection and region growing is applied to reduce the number of candidates, and to determine their precise size, respectively.

C. Ensemble selection and the spatial voting scheme

First, a pair from each preprocessing method and candidate extractor is formed by generating the output of the candidate extractor on the training images with the given preprocessing method applied. The combination of such pairs is accomplished by merging their outputs. The number of potential combinations is $2^{25}-1$ as a subset of 25 pairs is selected. The optimal combination of pairs is found by simulated annealing [14], which is an effective search algorithm in large search spaces.

This search algorithm requires an energy function to be optimized, which was defined by applying a voting scheme: the candidates extracted by the $\langle$preprocessing method, candidate extractor$\rangle$ pairs are analyzed based on their spatial density. A real value between 0 and 1 are assigned to each according to it. Each combination is characterised by their average sensitivity at seven predefined false positives / image rates [15]. The combination with the highest energy values is selected and used for the classification of the unknown data.

D. MA detection

After the pairs are selected, the actual MA detection on unknown images can be performed. First, MA candidates using all selected pairs are selected. Then, a confidence value to all candidates are assigned as discussed in section II-C. Finally, the detected MAs are extracted by thresholding. With the adjustment of this threshold value, the number of MA candidates can be controlled. This influences the sensitivity/specificity of the system.

III. Results and discussion

The publicly available Messidor database was used to evaluate the grading performance of our detector. This database consists of 1200 images with 45 field of view (FOV) and different resolutions. For each image, a grading score ranging from R0 to R3 is provided. The grades usually denotes to following conditions [16]: a patient with an R0 grade has no diabetic retinopathy (DR). R1 and R2 are background and pre-proliferative retinopathy, respectively. Finally, R3 is denotes proliferative diabetic retinopathy. This grading is based on certain features included in DR grading, examples of which are appearance of MAs, haemorrhages and neovascularization.

A. Training

As there is no training set provided for the Messidor database, the ROC dataset [2] was used as training set. This is the same that that was also utilised in the Retinopathy Online Challenge successfully. The dataset consists of 50

color retinal images with different FOVs and resolutions. The selected ⟨Preprocessing method, candidate extractor⟩ pairs for the optimal ensemble on the data can be seen in Table I.

B. Testing

In our experiment, we classified the retinal images whether they contains signs of DR (R1, R2, R3) or not (R0). The MA detector classifies an image as diseased if it at least one MA is detected (see Fig. 2 for some examples), and healthy otherwise. It is important to note that it is not essential to detect all MAs for grading, but it is essential to detect MAs where DR is present. We measured the performance of the detector at different thresholds. We provided the sensitivity, specificity and the accuracy for several thresholds in Table II. The ROC curve of the detector can be seen in Fig. 3. The area under this curve (AUC) is 0.87, which is similar to the 0.86 reported by Abramoff et al. in [1] on a different dataset.

It is also interesting to see how the different classes recognized at different threshold. According to the protocol used for the grading of the Messidor dataset, an image at R1 level contains at least 1 MA, while this number is 5 for R2 and 15 for R3, respectively. Thus, it is expected that an MA detector recognizes a more severe case easier. This claim is fulfilled by our approach, as it can be seen in Table III that the severity of DR affects the performance of the detector. At every threshold, where the sensitivity is less than 1.0, the more severe case recognized with a higher percentage.

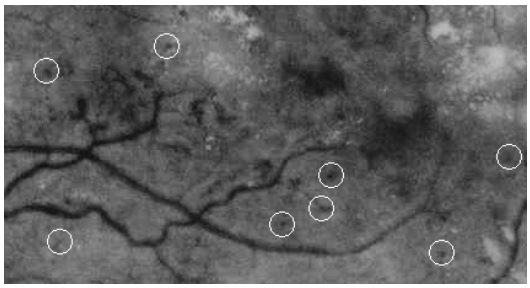

Fig. 2. Detected microaneurysms on a sample image.

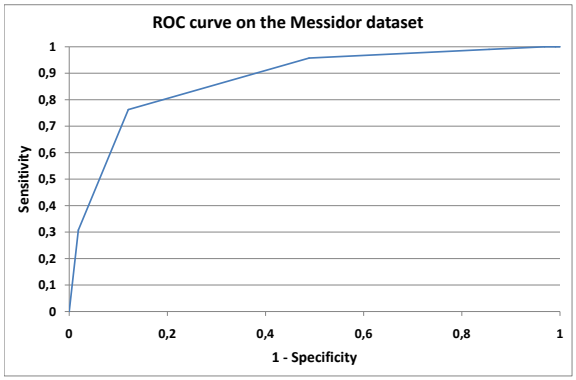

Fig. 3. ROC curve of the detector on the Messidor dataset.

C. Discussion

The most important question is how to select the appropriate threshold for our detector. In Table II, we can see that the most accurate result is achieved with threshold level 0.9. However, most of this accuracy originates from the high specificity level, which is for the sake of efficiency, and sensitivity is more important for a screening system [1]. Thus, we selected the results at the 0.8 threshold value as the more appropriate for the clinical needs, where 96% sensitivity and 51% specificity is achieved. That is, we recognize almost all of the cases where DR is present, and also the half of the healthy ones. For a comparison, the study of Abramoff et al. [1] reported a sensitivity of 90% at a specificity of 54,7%.

It is also important to compare our method to the performance of human graders. In [17], the following sensitivity / specificity values are reported for three human graders: 73% / 89%, 62% / 84%, 85% / 89% with a moderate inter-rater agreement (an average κ-statistic of 0.55). With the formula presented in [1], it can be calculated, that three graders with the grading performance presented above will miss approximately 260 images from the Messidor dataset. Our algorithm missed 293 images, mostly from the healthy ones.

The presented results are promising, however, it can be noted that in our experiment, the number of the healthy and diseased patients are almost equal. This mirrors the real screening situation closely, but not entirely, as in population screening, the proportion of the non-DR ones are about 60%. However, due to its high sensitivity, the detector is expected to perform well also for cases where different proportions are present (e.g. in [1], where only 10% of the images contained no signs of DR).

IV. Conclusions

In this paper, we presented the results of an evaluation of a microaneurysm detector for diabetic retinopathy screening developed by our research group. This detector previously showed its efficiency at an online challenge, where it is ranked as first. We measured the grading performance of this detector using the 1200 images of the Messidor database. We have achieved a 96% sensitivity and 51% specificity with an overall 0.87 AUC, which is similar to the previously reported results on different databases. The results are also compared to the expected performance of human graders, which is similar to ours.

The results presented in this paper are promising. Our MA detector can be further improved by adding more preprocessing methods and MA candidate extractors for the ensemble pool. Furthermore, in an automatic screening system, more components should be taken into consideration besides MA detection. Namely, the system should contains quality assessment to filter out low quality images. Moreover, we can incorporate exudate detection and a module dedicated to the recognition of severely ill retinas. Adding these components is expected to boost the sensitivity / specificity values achieved by only MA detection.

TABLE I

⟨PREPROCESSING METHOD, CANDIDATE EXTRACTOR⟩ PAIRS SELECTED AS MEMBERS OF THE ENSEMBLE BASED ON THE ROC DATASET.

Preprocessing \ Candidate extractor	Walter	Spencer	Hough	Lazar	Zhang
Walter-Klein					•
CLAHE	•			•	
Vessel removal and extrapolation				•	•
Illumination equalization				•	
No preprocessing	•			•	•

TABLE II

SENSITIVITY, SPECIFICITY AND ACCURACY OF THE DETECTOR ON THE MESSIDOR DATASET AT DIFFERENT THRESHOLD LEVELS.

Measure \ Threshold	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Sensitivity	1	1	1	0.99	0.96	0.76	0.31
Specificity	0	0.01	0.03	0.14	0.51	0.88	0.98
Accuracy	0.53	0.54	0.55	0.59	0.75	0.82	0.62

TABLE III

SENSITIVITY OF THE DETECTOR BY THE SEVERITY OF DR ON THE MESSIDOR DATASET.

Class \ Threshold	0.4	0.5	0.6	0.7	0.8	0.9	1.0
R0 (normal fundus image)	0.00	0.01	0.03	0.14	0.51	0.88	0.98
R1 (background retinopathy)	1.00	1.00	1.00	0.97	0.92	0.60	0.18
R2 (pre-proliferative retinopathy)	1.00	1.00	1.00	1.00	0.96	0.72	0.29
R3 (proliferative retinopathy)	1.00	1.00	1.00	1.00	0.98	0.92	0.42

ACKNOWLEDGEMENTS

This work was supported in part by the János Bolyai grant of the Hungarian Academy of Sciences, and by the TECH08-2 project DRSCREEN - Developing a computer based image processing system for diabetic retinopathy screening of the National Office for Research and Technology of Hungary (contract no.: OM-00194/2008, OM-00195/2008, OM-00196/2008). Tünde Pető is supported by the UK NIHR BMRC for Ophthalmology. The Messidor database is kindly provided by the Messidor program partners (see http://messidor.crihan.fr).

REFERENCES

[1] M. Abramoff, *et al.*, "Automated early detection of diabetic retinopathy," *Ophthalmology*, vol. 117, no. 6, pp. 1147–1154, 2010.

[2] M. Niemeijer, *et al.*, "Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 185–195, 2010.

[3] B.Antal *et al.*, "Improving microaneurysm detection in color fundus images by using an optimal combination of preprocessing methods and candidate extractors," *18th EUSIPCO*, pp. 1224–1228, 2010.

[4] T. Walter *et al.*, "Automatic detection of microaneyrysms in color fundus images of the human retina by means of the bounding box closing," *Lecture Notes in Computer Science*, vol. 2526, pp. 210–220, 2002.

[5] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics gems*, vol. IV, pp. 474–485, 1994.

[6] S. Ravishankar, *et al.*, "Automated feature extraction for early detection of diabetic retinopathy in fundus images," in *Computer Vision and Pattern Recognition*, 2009, pp. 210–217.

[7] A. Criminisi, *et al.*, "Object removal by exemplar-based inpainting," in *Computer Vision and Pattern Recognition*, vol. 2, 2003, pp. II–721 – II–728.

[8] A. Hoover *et al.*, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 951 –958, 2003.

[9] T. Walter, *et al.*, "Automatic detection of microaneurysms in color fundus images," *Medical Image Analysis*, vol. 11, pp. 555–566, 2007.

[10] T. Spencer, *et al.*, "An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus," *Computers and Biomedical Research*, vol. 29, pp. 284–302, May 1996.

[11] S. Abdelazeem, "Microaneurysm detection using vessels removal and circular hough transform," *Proceedings of the Nineteenth National Radio Science Conference*, pp. 421 – 426, 2002.

[12] I. Lazar, *et al.*, "Microaneurysm detection in digital fundus images," University of Debrecen, Hungary, Tech. Rep. 2010/14(387), 2010.

[13] B. Zhang, *et al.*, "Detection of microaneurysms using multi-scale correlation coefficients," *Pattern Recogn.*, vol. 43, no. 6, pp. 2237–2248, 2010.

[14] S. Kirkpatrick, *et al.*, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[15] M. Niemeijer, *et al.*, "On combining computer-aided detection systems," *IEEE Transactions on Medical Imaging*, vol. PP, no. 99, pp. 1 –1, 2010.

[16] UK National Screening Committee, *National Screening Programme for Diabetic Retinopathy*. http://www.retinalscreening.nhs.uk/, 2009.

[17] M. Abramoff, *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, February 2008.