



Validation of Microaneurysm-based Diabetic Retinopathy Screening across Retina Fundus Datasets

Luca Giancardo, Fabrice Meriaudeau, Thomas Karnowski, Kenneth Tobin,
Edward Chaum

► To cite this version:

Luca Giancardo, Fabrice Meriaudeau, Thomas Karnowski, Kenneth Tobin, Edward Chaum. Validation of Microaneurysm-based Diabetic Retinopathy Screening across Retina Fundus Datasets. CBMS2013, the 26th International Symposium on Computer Based medical System, Jun 2013, Porto, Portugal. pp.125-130, 10.1109/CBMS.2013.6627776 . hal-00831565

HAL Id: hal-00831565

<https://hal.science/hal-00831565>

Submitted on 17 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Validation of Microaneurysm-based Diabetic Retinopathy Screening across Retina Fundus Datasets

L. Giancardo, *Member, IEEE*, F. Meriaudeau, *Member, IEEE*, T. P. Karnowski, *Member, IEEE*,
K. W. Tobin Jr, *Fellow Member, IEEE* and E. Chaum, *Member, IEEE*

Abstract—In recent years, automated retina image analysis (ARIA) algorithms have received increasing interest by the medical imaging analysis community. Particular attention has been given to techniques able to automate the pre-screening of Diabetic Retinopathy (DR) using inexpensive retina fundus cameras. With the growing number of diabetics worldwide, these techniques have the potential benefits of broad-based, inexpensive screening. The contribution of this paper is twofold: first, we propose a straightforward pipeline from microaneurysm (an early sign of DR) detection to automatic classification of DR without employing any additional features; then, we quantify the generalisation ability of the MA detection method by employing synthetic examples and, more importantly, we experiment with two public datasets which consist of more than 1,350 images graded as normal or showing signs of DR. With cross-datasets tests, we obtained results better or comparable to other recent methods. Since our experiments are performed only on publicly available datasets, our results are directly comparable with those of other research groups.

I. INTRODUCTION

According to estimates of the World Diabetes Foundation, 439 million people will have diabetes mellitus worldwide by 2030. Diabetic Retinopathy (DR) is the leading cause of new cases of blindness among diabetic patients. In the United states alone, 25.8 million people are affected by diabetes mellitus, and this number is projected to grow in the future[1]. Thus effective and inexpensive methods to screen retina fundus images are needed as this large number of patients threatens to overwhelm conventional screening approaches. Since high-quality mydriatic and non-mydriatic fundus cameras are becoming the norm in many clinics, what is needed is a proven highly sensitive automatic method to quickly pre-screen patients before they even reach the ophthalmologist. This will considerably reduce the time and cost of a visit, thus increasing the likelihood of an early detection of DR in diabetic patients and reduce blindness, since timely treatment with laser therapy can reduce the development of severe vision loss on 50% to 60% of cases [1].

These studies were supported in part by grants from Oak Ridge National Laboratory, the National Eye Institute, (EY017065), by an unrestricted UTHSC Departmental grant from Research to Prevent Blindness (RPB), New York, NY, by The Plough Foundation, Memphis, TN and by the Regional Burgundy Council, France.

L. Giancardo is with the Istituto Italiano di Tecnologia (e-mail: luca.giancardo@iit.it).

F. Meriaudeau is with the University of Burgundy, Le Creusot, 71200 France.

T.P. Karnowski and K.W. Tobin, Jr are with the Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA.

E. Chaum is an RPB Senior Scientist. E. Chaum is with the University of Tennessee Health Science Center, Memphis TN 38163 USA.

One of the earliest manifestation of DR are microaneurysms (MAs). As shown in Fig. 1, they appear as small circular dots of the same colour as the retina vasculature (i.e. blood vessels). In fact, they are small, swollen capillaries which can leak blood and fluid into the retina, leading to vision threatening conditions due to other complications (e.g. exudates, macular edema and hemorrhages). As such, automatic MAs analysis detection by itself can be considered the key to trigger the very first alarm for DR detection.

Other authors [2, 3] have presented automated retina image analysis (ARIA) techniques for automatic DR detection. In both cases the fundus images are processed by various modules such as: quality verification, vessel segmentation, red lesion detection and bright lesion detection. Then, the output of the modules is combined and automatically classified with a machine learning technique. An alternative approach is taken by Agurto et al. [4] with a technique entirely based on multiscale amplitude-modulation-frequency-modulation (AM-FM) features that indirectly capture the retina condition by characterising the retina texture as a whole.

In this paper, we present a streamlined ARIA algorithm for automatic DR detection which only requires an initial MA detection step. MA detection has been the topic of many successful approaches which were recently evaluated on a common dataset through the ROC challenge [5]. Our group developed one of such algorithms which works at a very low false positive rate thanks to an approach based on the Radon transform [6]. We test various machine learning techniques to attempt to classify the retina as normal or DR based on the distribution of the MA detection output. These tests take into consideration that fundus images have a great visual variability (especially across different clinical settings) and the generalisation ability of the algorithms need to be evaluated.

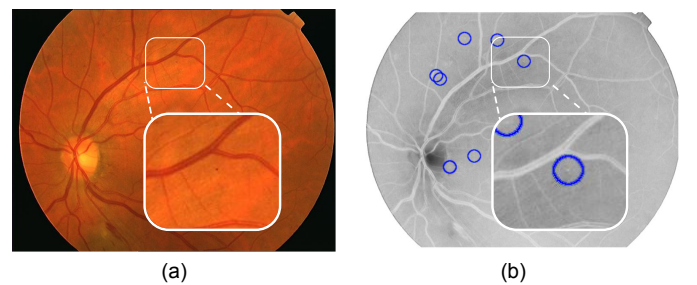


Fig. 1. (a) Example of a retina fundus image with a MA magnified; (b) Inverted green channel of (a) with MAs automatically detected.

This is possible thanks to the cross-dataset approach of the various experiments.

In Section III we present an overview of the MA detection method and we describe our DR classification approach; Section II discusses the characteristics of the datasets employed; Section IV-B presents the results; finally, Section V concludes with a discussion of the results obtained.

II. MATERIALS

The design, implementation and testing of ARIA algorithms requires high quality annotated data. In the literature, there are some publicly available annotated datasets of retinal images which have different goals, characteristics, and levels of completeness [7]. In this paper, we employed three different datasets all containing mydriatic and non-mydriatic macula centred colour fundus images mainly at 45 degrees of field of view: ROC challenge [5], HEI-MED [8] and Messidor [9].

The ROC challenge dataset is composed of 100 images, half of which has been annotated by 4 experts. The only structures annotated are the small red lesions (MAs and round haemorrhages). This dataset has been employed for training the MAs detection algorithm.

The HEI-MED and Messidor dataset are composed respectively of 169 and 1200 images acquired in different centres and from patients of heterogeneous ethnic backgrounds. They provide different types of image-level and lesion-level annotations, however, we will be employing the DR diagnosis for our DR analysis experiments. In the case of HEI-MED, a patient is considered normal if no lesions were found and diagnosed with DR if at least a MA was found by the expert. Messidor, on the other hand, has a direct annotation of the DR condition found together with a severity score.

III. MICROANEURYSMS DETECTION

A. Microaneurysms Detection - Method

In previous papers [10, 6], we have described two methods for the MAs detection (and localisation) based on some useful properties of the Radon transform [11]. It was found that Gaussian-like structures such as MAs have a cliff-like appearance in the Radon space of a local window. This appearance is maintained in highly noisy conditions and in windows containing other structures with high pixel values (outliers). The latter aspect is particularly important because MAs tend to appear very close to the retina vasculature and, in the local window, the vasculature is generally much more visible than the MAs by themselves.

While our first approach [10] relied on a fully rule-based technique, we later discovered that the technique could be considerably improved by compressing the Radon space into a new set of compact features which were later classified through Principal Component Analysis (PCA) and a non-linear Support Vector Machine (SVM) [6]. This approach can be quickly trained through an on-line machine learning strategy, it does not require vessel segmentation and it keeps the image pre-processing at a minimum. A summary of the method follows, for a complete description refer to [6]:

1) *Candidates selection*: The aim of this initial phase is to reduce the computational burden of the Radon analysis. Essentially, it discards all the areas that cannot contain any foreground structure (vessels, lesions, etc.) due to their low intensity. This is accomplished through a fast background subtraction operation, immediately followed by a colour normalisation such that the resulting image can be described by a given Gaussian distribution. The areas selected as candidates are the ones above a very conservative scalar value employed as threshold derived from 5 images from the ROC challenge dataset.

2) *Radon-based features*: The Radon transform is calculated on a local sliding window across the green channel of the original (resized) image. The window is centred on the pixel with the highest local value. Since the local window has a square shape (non-isometric support), each projection of the Radon transform will accumulate an uneven number of pixels, which leads to coefficients biased towards certain projection rays. Therefore each ray is normalised by the number of pixels it crosses, thus obtaining R^i .

Each R^i is represented as the compact feature vector F^i which is computed as follows:

$$R_\mu^i(x) = \frac{1}{\phi} \sum_{n=1}^{\phi} R^i(n, x) \quad 0 \leq x < \rho \quad (1)$$

$$R_\sigma^i(x) = \sqrt{\frac{1}{\phi-1} \sum_{n=1}^{\phi} [R^i(n, x) - R_\mu^i(x)]^2} \quad 0 \leq x < \rho \quad (2)$$

$$F^i = \begin{pmatrix} \frac{d}{dx} R_\mu^i \\ R_\sigma^i \end{pmatrix} \quad (3)$$

where i is the index of a window, R^i is the window in the normalized Radon space having on the horizontal axis the different angles of projections and on the vertical one the number of projections for each angle. ϕ is the number of projection angles and ρ is the number of projection rays.

3) *MAs classification*: After normalisation, F^i is projected to a space of lower dimensionality via Principal Component Analysis (PCA) [12]. The number of dimensions are selected such that the 95% of the original data variance (of the training set) is maintained. The feature vector with lower dimensionality is then classified by a SVM with a radial basis kernel and probabilistic output [13].

The probability of being a MA (P_{ma}) is calculated as a combination of the SVM probability and the average value at the centre of the original window of the equalised image. The probabilities are combined following the unnormalised Bayes rule.

B. Microaneurysms Detection - Generalisation Ability

We evaluate the performance of the algorithm on a synthetic model of a MA with variable dimensions, proportions and noise. The classifier is trained on only 5 representative images from the ROC dataset. These tests quantify the generalisation ability of the algorithm on different MAs. The model f is based on a two dimensional normal distribution N with a maximum height of α and with a support window of 17×17 pixels. The choice of the model was dictated by the general

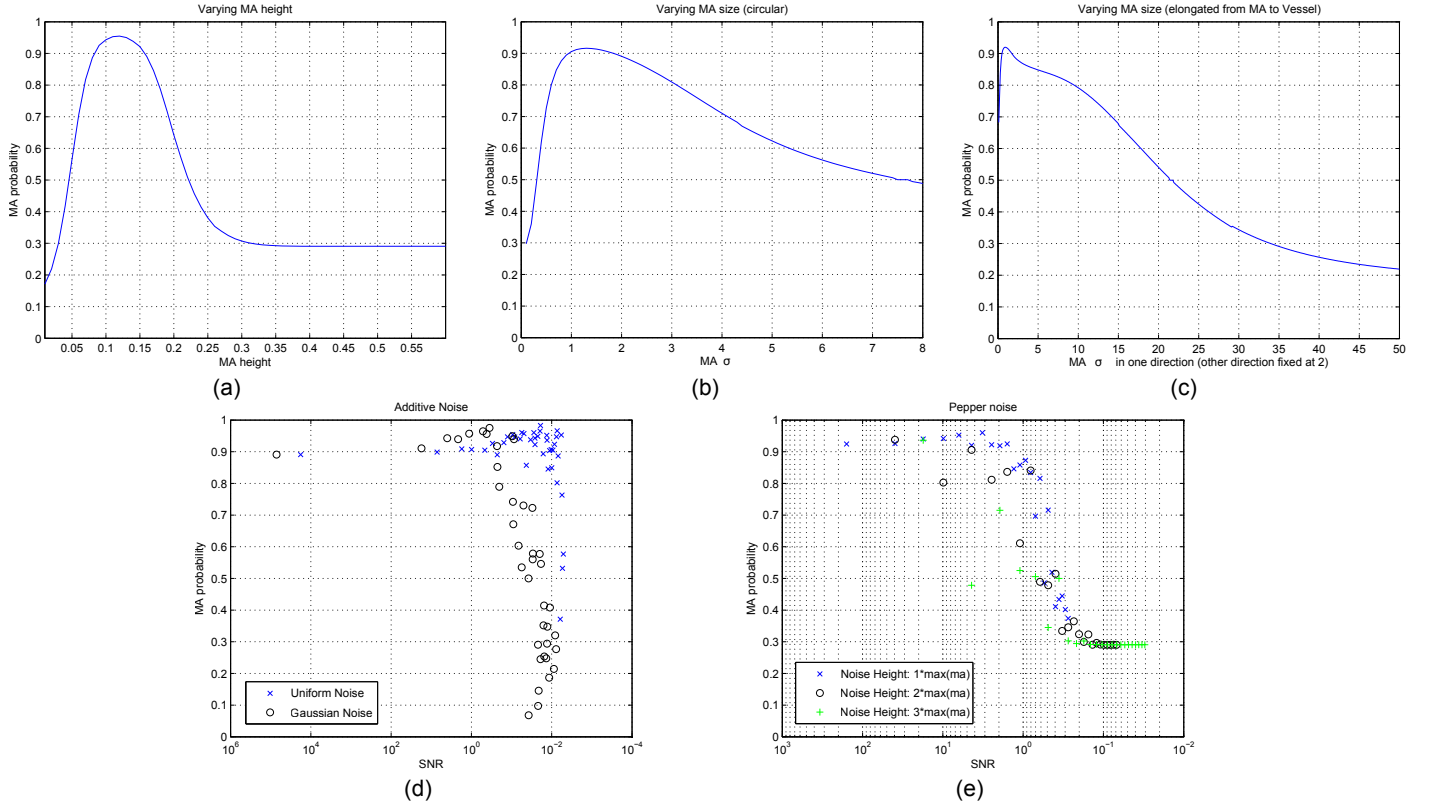


Fig. 2. Tests on Synthetic Microaneurysms. (a) MA detection in function of MA height; (b) MA detection in function of size; (c) MA detection in function of not-proportional size changes (from circular MA to elongated vessel); (d) MA detection in function of two types of additive noise; (e) MA detection in function of pepper noise with different heights.

consent among researchers that MAs mostly appear as 2-D Gaussians [14].

$$N(x, \mu, \Sigma) = \frac{1}{2\pi|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (4a)$$

$$f(x, \mu, \Sigma, \alpha) = \frac{N(x, \mu, \Sigma)}{N(\mu, \mu, \Sigma)} \alpha \quad (4b)$$

where x is a two dimensional vector representing a coordinate in the window; μ is a two dimensional vector always containing the coordinate at the centre of the window $[9 \ 9]^T$ and Σ is the 2×2 covariance matrix.

Fig. 2(a) shows the detection performance as a function of α with $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. The results are themselves normally distributed with a positive detection (i.e. MA prob. ≥ 0.5) with $0.045 \leq \alpha \leq 0.225$. Fig. 2(b) shows how the algorithm can detect circular objects with a great range of sizes. The size variation is simulated as a function of Σ with $\Sigma = \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix}$ and a fixed height ($\alpha = 0.16$). We obtained a positive detection for $0.3 \leq \sigma \leq 7.8$. Fig. 2(c) shows how the classifier can detect MAs that are not perfectly round, and when it stops doing so. We have simulated a round MA that becomes more and more elongated, up until it reaches a “vessel-like” appearance. This

is achieved by using $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & \sigma \end{bmatrix}$, where σ goes from 0.1 to 50.

Fig. 2(d,e) shows various experiments for the detection of MAs in a noisy environment. In all cases the following parameters are used: $\alpha = 0.16$ and $\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$. On the x-axis, the signal to noise ratio (SNR) measures the how much the original signal has been affected by noise:

$$SNR = \frac{(A_f)^2}{(A_{fn} - A_f)^2} \quad (5)$$

where A_f is the sum of all pixels of the MA model, and A_{fn} is the sum of all pixels of the model affected by noise.

IV. DR SCREENING

A. DR Screening - Method

Our approach to DR diagnosis is entirely based on the output of the MA detector previously described. This is possible because the MAs are the very first manifestation of DR, hence if they are successfully detected it is possible to diagnose DR even in its mild form.

The first DR classification strategy implemented is the enumeration of MAs with an estimated probability (P_{ma}) higher than 0.5. An image is deemed to have DR if more than a th

number of MA are identified. By varying th , it is possible to perform a ROC analysis, which will be used as a baseline for the other tests. We note that this simple counting strategy does not make full use of MA probability distribution. Therefore, we estimate this distribution by calculating the histogram of P_{ma} for all the MAs identified in an image with $P_{ma} > 0.5$. The histogram bins are used as a feature vector and then classified with various classification strategies. We chose classifiers that covered the three different classification families described by Jain et al. [15]: *probabilistic* (Naive Bayes), *geometric* (SVMs) and *tree-based* (Random Forest). As a baseline, we report the results of one of the simplest classification method available, nearest neighbour.

Classifier	Feature Set	2 bins	3 bins	4 bins	5 bins
Nearest Neighbour		0.687	0.708	0.683	0.668
Naive Bayes (Gaussian)		0.809	0.819	0.824	0.818
Naive Bayes (Parzen Win.)		0.823	0.829	0.833	0.831
SVM (linear kernel)		0.825	0.834	0.833	0.823
SVM (radial kernel)		0.825	0.831	0.833	0.830
Random Forests		0.806	0.781	0.792	0.758
The results are expressed as AUC for the 3-fold tests on HEI-MED.					

TABLE I
DR CLASSIFIER/FEATURE SELECTION TESTS WITH HEI-MED AS TESTING TARGET.

In Table I, a variable number of histogram bins and classifiers belonging to different families are evaluated on the HEI-MED dataset with a 3-fold evaluation. We note that AUCs do not show significant changes, this is probably due to the low dimensionality and compactness of the samples. The classifier/feature set employed in the testing phase have been highlighted in bold. During this phase it was noted that some FP MAs were identified on the optic nerve (ON) area where the presence of MAs is physiologically impossible. By employing the automatic ON localisation technique described by Tobin et al. [16], we identify the ON area and remove any MAs present. After the ON removal, the average performance improvement was estimated at ~ 0.01 of AUC.

B. DR Screening - Results

Fig. 3 shows the baseline results obtained by counting the MAs on each image. In both datasets the results are surprisingly good for such a simple classification technique. However, it should be remembered that the MA detector introduced is particularly competitive at a low false positive rate which is a desirable characteristic of many DR pre-screening systems.

Fig. 4 shows the tests performed with the best performing classifier/feature sets. In Fig. 4(a), the classifiers are trained on the Messidor and tested on HEI-MED. The best performance are obtained by the Naive Bayes classifier with Parzen window to estimate the samples distribution and a feature

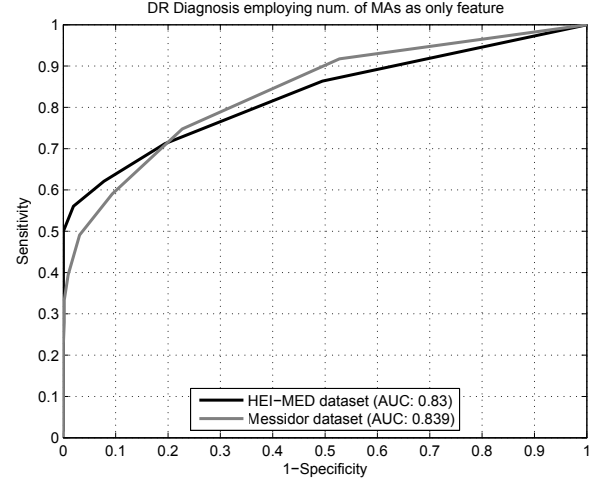


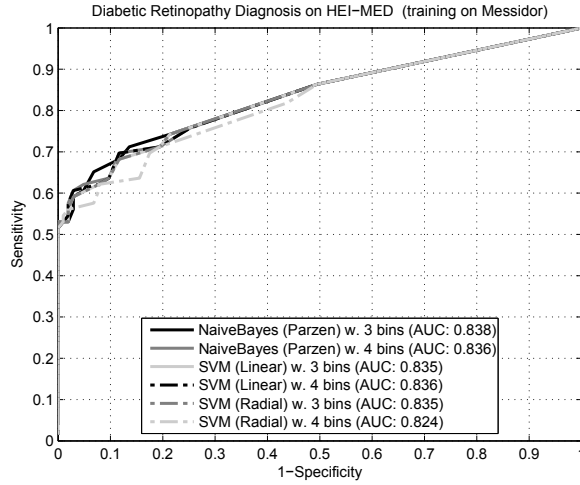
Fig. 3. ROC curve for the DR diagnosis by counting the number of MAs only.

vector with 3 bins. The improvements in comparison with the baseline classification are not substantial, a 0.008 points of AUC. Fig. 4(b) shows the tests on the 1200 images of the Messidor dataset (with HEI-MED as training set). The best performing feature set/classifier are the same as the previous test. However, the improvement in comparison to the baseline is more pronounced: 0.13 AUC. This is even more significant considering the size of this dataset. In Fig. 5, the classifier trained on the full HEI-MED dataset is tested on a subset of Messidor images, e.g. the healthy and the ones showing a high risk of DR. This experiment evaluates the performance of the classifier for the cases that are very urgent. The AUC obtained is 0.95, a performance considerably higher than the one obtained when the mild DR cases are considered.

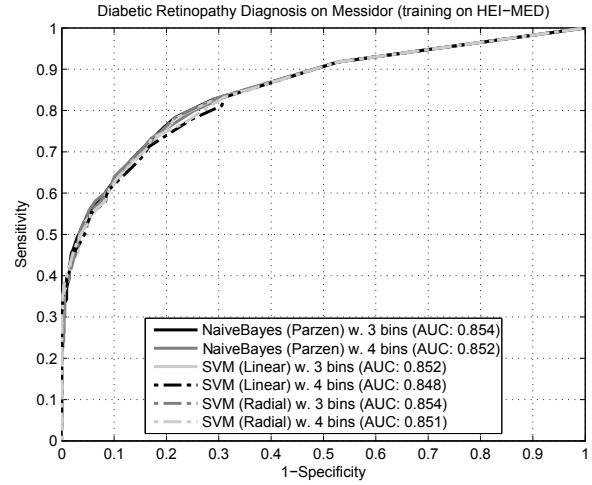
Agurto et al. [4] were the first group that published the results of a DR screening algorithm employing the Messidor dataset, but they only used 400 of the images in the set (the ones labelled as *Lariboisière* in the original metadata). In their tests they obtained a AUC of 0.84, which is already lower in comparison to our results on the complete set (AUC of 0.854). For a fairer comparison, we tested our best performing classifier/feature set on the Lariboisière subset training it on the HEI-MED dataset. The result obtained improved even further, as an AUC of 0.879 was achieved.

More recently, Sánchez et al. [3] perform a thorough evaluation of their ARIA system on the full Messidor dataset and compared it with the diagnosis of other two experts. The authors obtained a sensitivity of 0.92 at 0.5 specificity, which is only slightly higher than ours, i.e. 0.91 sensitivity at 0.5 specificity. However, our results are comparable with the performance of the two human experts, i.e. a sensitivity of 0.91 and 0.94 at 0.5 specificity.

Being based almost entirely on the MAs segmentation, the time required by this classification technique is negligible (< 1 second) once the MAs are segmented.



(a)



(b)

Fig. 4. ROC curves for the DR diagnosis.

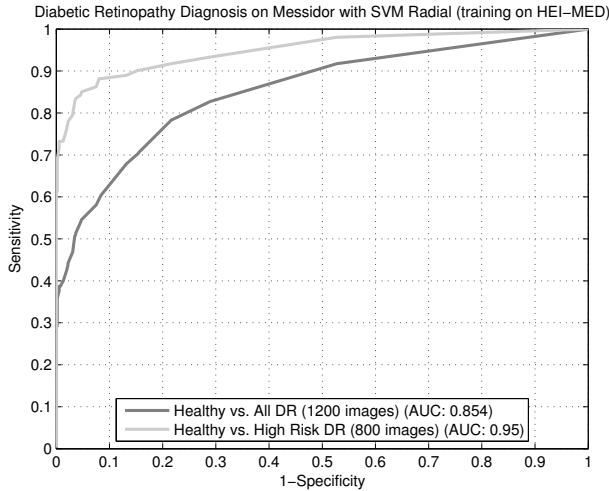


Fig. 5. ROC curve comparing the DR diagnosis for various level of DR.

V. CONCLUSIONS AND DISCUSSION

In this paper, we have presented a streamlined ARIA algorithm for automatic DR detection which only requires an initial MA detection step. The MAs detector, which was already evaluated on the ROC dataset [6], has been quantitatively assessed via a synthetic model of MAs. Then, the DR detection algorithm was tested on two independent datasets obtaining results better or in line with methods tested on the same datasets.

However, our approach has the advantage of requiring features uniquely derived from a single MAs detector, without employing other lesion detectors or image-wide features with obvious computational advantages. In fact, the total time required to generate a complete screening diagnosis (from

an unseen image) is ~ 12 seconds per image on a 1.6 GHz machine with 4 GB of RAM in a Matlab implementation. This aspect leaves to the method untapped potential which can be easily harnessed via additional features.

In order to effectively test the feasibility of these techniques in a healthcare environment, tests with datasets much larger than the ones used need to be performed and compared with the performance of retina experts. Nevertheless, the use of three independent public datasets make a strong point about the effectiveness of the algorithms.

REFERENCES

- [1] Centers for Disease Control and Prevention, "National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the united states," U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, Tech. Rep., 2011.
- [2] M. Niemeijer, M. D. Abramoff, and B. van Ginneken, "Information fusion for diabetic retinopathy cad in digital color fundus photographs," *IEEE Transactions on Medical Imaging*, vol. 28, no. 5, pp. 775–785, 2009.
- [3] C. I. Sanchez, M. Niemeijer, A. V. Dumitrescu, M. S. A. Suttrop-Schulten, M. D. Abramoff, and B. van Ginneken, "Evaluation of a computer-aided diagnosis system for diabetic retinopathy screening on public data," *Invest Ophthalmol Vis Sci*, vol. 52, no. 7, pp. 4866–4871, Jun 2011.
- [4] C. Agurto, V. Murray, E. Barriga, S. Murillo, M. Pattichis, H. Davis, S. Russell, M. Abramoff, and P. Soliz, "Multiscale am-fm methods for diabetic retinopathy lesion detection," *IEEE Transaction on Medical Imaging*, vol. 29, no. 2, pp. 502–512, Feb 2010.
- [5] M. Niemeijer, B. van Ginneken, M. J. Cree, A. Mizutani, G. Quellec, C. I. Sanchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, X. Wu, G. Cazuguel, J. You, A. Mayo, Q. Li, Y. Hatanaka, B. Cochener, C. Roux, F. Karray, M. Garcia, H. Fujita, and M. D. Abramoff, "Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 185–195, Jan. 2010.
- [6] L. Giancardo, F. Meriaudeau, T. Karnowski, Y. Li, K. Tobin, and E. Chaum, "Microaneurysm detection with radon transform-based classification on retina images," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 5939–5942.

- [7] E. Trucco, A. Ruggeri, T. Karnowski, L. Giancardo, E. Chaum, J. Hub-schman, B. al Diri, C. Cheung, D. Wong, M. Abramoff, G. Lim, D. Kant Kumar, B. P. M., N. M. Bressler, H. F. Jelinek, F. Meriaudeau, G. Quellec, T. J. MacGillivray, and B. Dhillon, "Validating retinal fundus image analysis algorithms: Issues and a proposal," *Investigative Ophthalmology & Visual Science, 2013 (in press)*, 2013.
- [8] L. Giancardo, F. Meriaudeau, T. P. Karnowski, Y. Li, S. Garg, K. W. Tobin, Jr, and E. Chaum, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets." *Medical Image Analysis*, vol. 16, no. 1, pp. 216–226, Jan 2012.
- [9] Messidor. (2010) Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology. Online [http://messidor.crihan.fr]. [Online]. Available: http://messidor.crihan.fr
- [10] L. Giancardo, F. Meriaudeau, T. Karnowski, K. Tobin Jr, Y. Li, and E. Chaum, "Microaneurysms detection with the radon cliff operator in retinal fundus images," in *Proceedings of SPIE*, vol. 7623, 2010, p. 76230U.
- [11] J. Resnick, "The radon transforms and some of its applications," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 1, pp. 338–339, 1985.
- [12] C. Bishop, *Pattern Recognition and Machine Learning*, 2nd ed. Springer, 2007.
- [13] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *Advances in Large Margin Classifiers*. MIT Press, 1999.
- [14] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal wavelet transform for the detection of microa-neurysms in retina photographs," *IEEE Transaction on Medical Imaging*, vol. 27, no. 9, pp. 1230–1241, 2008.
- [15] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.
- [16] K. W. Tobin, E. Chaum, V. P. Govindasamy, and T. P. Karnowski, "Detection of anatomic structures in human retinal imagery." *IEEE Transactions on Medical Imaging*, vol. 26, no. 12, pp. 1729–1739, Dec 2007.