Retinopathy Online Challenge: Automatic Detection of Microaneurysms in Digital Color Fundus Photographs

Meindert Niemeijer*, Bram van Ginneken, *Member, IEEE*, Michael J. Cree, *Senior Member, IEEE*, Atsushi Mizutani, Gwénolé Quellec, Clara I. Sánchez, *Member, IEEE*, Bob Zhang, Roberto Hornero, *Member, IEEE*, Mathieu Lamard, Chisako Muramatsu, Xiangqian Wu, Guy Cazuguel, *Member, IEEE*, Jane You, Agustín Mayo, Qin Li, Yuji Hatanaka, Béatrice Cochener, Christian Roux, *Senior Member, IEEE*, Fakhri Karray, María García, *Student Member, IEEE*, Hiroshi Fujita, *Member, IEEE*, and Michael D. Abràmoff, *Member, IEEE*

***Abstract*—The detection of microaneurysms in digital color fundus photographs is a critical first step in automated screening for diabetic retinopathy (DR), a common complication of diabetes. To accomplish this detection numerous methods have been published in the past but none of these was compared with each other on the same data. In this work we present the results of the first international microaneurysm detection competition, organized in the context of the Retinopathy Online Challenge (ROC), a multi-year online competition for various aspects of DR detection. For this competition, we compare the results of five different methods, produced by five different teams of researchers on the same set of data. The evaluation was performed in a uniform manner using an algorithm presented in this work. The set of data used for the competition consisted of 50 training images with available reference standard and 50 test images where the reference standard was witheld by the organizers (M. Niemeijer, B. van Ginneken, and M. D. Abràmoff). The results obtained on the test data was submitted through a website after which standardized evaluation software was used to determine the performance of each of the methods. A human expert detected microaneurysms in the test set to allow comparison with the performance of the automatic methods. The overall results show that microaneurysm detection is a challenging task for both the automatic methods as well as the human expert. There is room for improvement as the best performing system does not reach the performance of the human expert. The data associated with the ROC microaneurysm detection competition will remain publicly available and the website will continue accepting submissions.**

***Index Terms*—Computer aided detection, computer aided diagnosis, diabetic retinopathy, fundus photographs, retina, Retinopathy Online Challenge (ROC) competition.**

Manuscript received August 04, 2009; revised September 29, 2009; accepted September 30, 2009. First published October 09, 2009; current version published January 04, 2010. This work was supported in part by the National Eye Institute (R01 EY017066), in part by the Netherlands Organization for Scientific Research (NWO), and in part by Research to Prevent Blindness. *Asterisk indicates corresponding author.*

*M. Niemeijer was with the Department of Electrical and Computer Engineering and the Department of Ophthalmology and Visual Sciences, University of Iowa, Iowa City, IA 52242 USA. He is now with the Image Sciences Institute, Utrecht, 3584 CX Utrecht, Netherlands. (e-mail: meindert@isi.uu.nl).

B. van Ginneken is with the Image Sciences Institute, 3584 CX Utrecht, The Netherlands.

H. Fujita, C. Muramatsu, and A. Mizutani are with the Department of Intelligent Image Information, Graduate School of Medicine, Gifu University, Gifu 501-1193, Japan.

Y. Hatanaka is with the Department of Electronic Systems Engineering, the University of Shiga Prefecture, Hikone 522-8533 , Japan.

M. García and R. Hornero are with the Biomedical Engineering Group (GIB), University of Valladolid, 47002 Valladolid, Spain.

C. Sánchez was with the Biomedical Engineering Group (GIB), University of Valladolid, 47002 Valladolid, Spain. She is now with the Image Sciences Institute, 3584 CX Utrecht, The Netherlands.

A. Mayo is with the Department of Statistics and Operative Investigation, University of Valladolid, 47002 Valladolid, Spain.

G. Quellec was with the Institut Telecom, 83818 Brest, France and with Inserm U650, 29609 Brest, France. He is now with the Department of Ophthalmology and Visual Sciences and the Department of Biomedical Engineering, University of Iowa, Iowa City, IA 52242 USA.

G. Cazuguel and Christian Roux are with Institut Telecom, 83818 Brest, France, and with Inserm U650, 29609 Brest, France.

M. Lamard and B. Cochener are with the University of Bretagne Occidentale, 29238 Brest, France, and with Inserm, U650, 29609 Brest, France.

M. Cree is with the Department of Engineering, University of Waikato, Hamilton 3240, New Zealand.

B. Zhang and F.Karray are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, N2L 3G1 ON Canada.

X. Wu is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China.

J. You and Q. Li are with the Department of Computing, Hong Kong Polytechnic University, Hong Kong.

M. Abràmoff is with the University of Iowa Hospitals and Clinics, Department of Ophthalmology and Visual Sciences, Iowa City, IA 52242 USA and the University of Iowa, Department of Electrical and Computer Engineering, Iowa City, IA 52242 USA.

Digital Object Identifier 10.1109/TMI.2009.2033909

I. Introduction

DIABETIC RETINOPATHY (DR) is a frequent microvascular complication of diabetes and the most common cause of blindness and vision loss in the working population of the western world [1]. It has been shown that early detection of DR helps prevent blindness and visual loss [2]. Most screening programs use nonmydriatic digital color fundus cameras to acquire color photographs of the back of the eye, the retina. These photographs are then examined for the presence of lesions indicative of DR (including microaneurysms, hemorrhages, exudates and cottonwool spots).

Development of systems to automate DR screening have received a lot of attention from the research community. Recently two large studies of automated screening systems have appeared in the literature [3], [4]. These systems are designed to perform triage in an automated fashion. The software stores those exams that, after analysis, appear to have no visible signs of the presence of diabetic retinopathy at some level of severity. The other exams, that possibly contain diabetic retinopathy related lesions

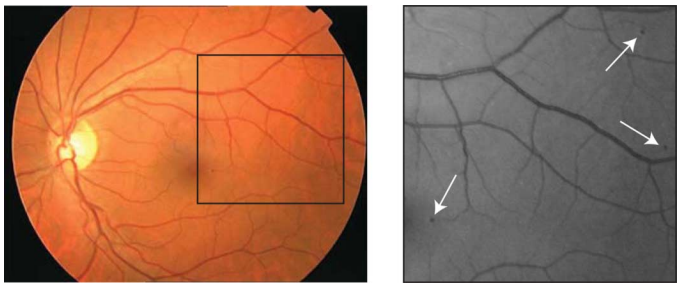


Fig. 1. Digital color fundus photograph containing microaneurysms. This image is one of the images from the test set of the **ROC** microaneurysm detection competition. To the right an enlarged part of the green plane of the image is shown with microaneurysms indicated.

or that have an image quality that prevents automated analysis are evaluated by a human expert. In this manner these automated systems can reduce the workload associated with large scale screening.

One of the most important steps in the automated screening of DR is the detection of microaneurysms. Microaneurysms are small outpouchings in capillary vessels. The capillary vessels are normally not visible in color fundus photographs but due to the local increase in size, microaneurysms appear as small dots between the visible retinal vasculature (see Fig. 1). Microaneurysms are amongst the first clinical signs of the presence of diabetic retinopathy. However, it is important to note that, while a critical component of any DR screening system, detection of microaneurysms is not equivalent to detection of DR. In some cases other types of lesions may appear first and it has been shown that also detecting these lesions increases the performance of an automated DR screening system [5]. In the past a number of different methods for the automated detection of microaneurysms have been proposed (see Section II). However, these methods were tested on different datasets using various evaluation measures that made direct comparison of performance impossible.

To drive the development of ever better image analysis methods, groups have established publicly available, annotated image databases in various fields. Examples for retinal images are the STARE [6], [7], DRIVE [8], [9] and MESSIDOR [10] databases. Even though evaluations on these types of jointly used datasets are already a large improvement over evaluation on separate datasets, there are many metrics that can be used to evaluate system performance. Even while using the same evaluation measures, implementation details of these metrics may have an influence on the final result. The drive for better image analysis methods was pushed even further by making available publicly accessible annotated datasets in the context of online, standardized evaluation, leading to an "asynchronous competition." The Middlebury Stereo Vision website [11] was the first of these. In an asynchronous competition all results are evaluated using the same evaluation software, but groups can submit results continuously. This kind of joint evaluations on a common dataset have the potential to steer future research by showing the failures of certain techniques and guide the practical application of techniques in clinical practice.

The first international synchronous competition in the retinal image analysis field, the Retina Online Challenge (**ROC**) microaneurysm detection competition, is organized by three of the authors (M. Niemeijer, B. van Ginneken, and M. D. Abràmoff). The goal of this work is to present the reference database for automated microaneurysm detection in color fundus photographs for diabetic retinopathy screening used in the **ROC**. Additionally, the results of a variety of different detection methods on the reference database are presented and each of the methods is briefly discussed. The algorithm used for the evaluation is described. A detailed analysis of the produced results giving a good overview of the state-of-the-art is provided.

The paper is structured as follows. A discussion of previous work is given in Section II. Section III describes the competition and the reference database. The different methods used by the participants in the **ROC** are discussed in Section IV. The validation methodology used to evaluate the different algorithms is presented in Section V. The experiments and results are shown in Section VI and the paper ends with a discussion and conclusion in Section VII.

II. Previous Work

The oldest published, peer reviewed journal paper on automatic microaneurysm detection known to us is the 1984 paper by Baudoin *et al.* [12] which describes a mathematical morphology based detection approach for microaneurysms in fluorescein angiograms. By using the morphological top-hat transformation with a linear structuring element at different orientations, connected, elongated structures (i.e., the vessels) were distinguished from unconnected circular objects (i.e., the microaneurysms). A number of additional papers presenting enhanced algorithms based on this technique were published. Spencer, Cree, Frame, and coworkers [13]–[16] added a shade-correction preprocessing step and a matched filtering postprocessing step to the basic top-hat transform based detection technique. After detection and segmentation of candidate microaneurysms, various shape and intensity based features were extracted and a classifier was used to separate the real microaneurysms from spurious responses. The advantage of using fluorescein angiography images of the fundus is that the contrast between the microaneurysms and the retinal background is larger than in digital color photographs. However, the intravenous use of fluorescein is relatively complicated and associated with a mortality of 1:222 000 [17] and this makes the application of fluorescein angiograms for large-scale screening purposes impractical.

A modified version of the top-hat based algorithm has been applied to high-resolution, redfree fundus photographs by Hipwell *et al.* [18]. Fleming *et al.* [19] improved on this method by locally normalizing the contrast around candidate lesions and eliminating candidates detected on vessels through a local vessel segmentation step. Walter *et al.* [20] described an alternative mathematical morphology based detection method. The microaneurysm detection step was preceded by an image normalization step that compensates for the presence of bright lesions, a common occurrence in diabetic retinopathy. The detection method was based on diameter closing and addresses a shortcoming of the top-hat based detection method by eliminating candidates located on tortuous vessels. After application of an automatic thresholding scheme, features were extracted from the candidate objects and they were classified using a classifier,

TABLE I
DIFFERENT TYPES OF IMAGES IN THE **ROC** DATASET

	Resolution ($height \times width$ in pixels)	Coverage of the retina	Number in training set	Number in test set
Type I	768×576	45°	22	22
Type II	1058×1061	45°	3	6
Type III	1389×1383	45°	25	22

that was first trained with example candidates from a training set.

Niemeijer *et al.* [21] presented a hybrid scheme that used both the top-hat based method as well as a supervised pixel classification based method to detect the microaneurysm candidates in color fundus photographs. The pixel classification method was trained using example pixels from both the vasculature and "red lesions" (i.e., microaneurysms and hemorrhages). After training, the detector would detect all retinal vessel and red lesion pixels in an image. After eliminating all connected, elongated structures the remaining objects were considered candidate microaneurysms. This method allowed for the detection of larger "red lesions" (i.e., hemorrhages) in addition to the microaneurysms using the same system. A large set of additional features, including color, was added to those described in [15], [16]. A classifier, trained with example candidates from a training set, distinguished between real and spurious candidate lesions.

A number of other approaches, not based on mathematical morphology, for the detection of red lesions in color fundus photographs have also been described. Sinthanayothin *et al.* [22] applied a recursive region growing procedure that segmented both the vessels and red lesions in a fundus image. To detect the vessels and vessel segments in this result a neural network was used. Any remaining objects after removal of the detected vasculature were identified as microaneurysms. In a recent work, Quellec *et al.* [23] described a supervised microaneurysm detection method based on template matching in wavelet-subbands. In the training phase the optimal adapted wavelet transform for detecting microaneurysms was found using the lifting scheme framework. Template matching in the wavelet domain was used to detect likely locations of microaneurysms. By applying a threshold on the matching result the microaneurysms are found.

None of the above methods was compared on the same dataset. The most common way of reporting algorithm performance was free-response receiver operating characteristic (FROC) analysis [24] on a per lesion basis [14], [15], [18]–[21]. In FROC analysis the sensitivity of the system is plotted against the average number of false positive detections per image. Additionally, a number of papers reported the result of an receiver operating characteristic (ROC) analysis [25] on a per image basis [18], [19], [21]. In ROC analysis, the true positive rate is plotted against the false positive rate.

III. RETINAL ONLINE CHALLENGE: MICROANEURYSM DETECTION

We established the Retinal Online Challenge (**ROC**). The goal of **ROC** is to organize several competitions, focussed on various important challenges in automated detection of retinal disease. The first competition was focused on microaneurysm detection because microaneurysm detection is a critical challenge for automated diabetic retinopathy screening. This first **ROC** was structured as follows: the **ROC** was announced in February 2008 and the image data was made available online[1] at the same time. The training data included a multireader reference standard and was meant for training and preliminary testing of the participant's methods. Of the test data only the images were made available; the reference standard was kept private. The main reason for structuring the competition like this is to prevent the use of the reference standard of the test set in the training process and to keep the evaluation fair.

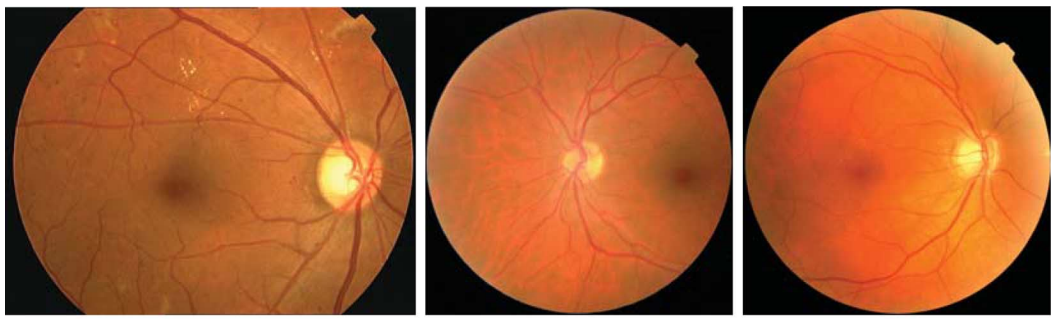
A. Reference Image Database

A set of 100 digital color fundus photographs was selected by the first author (M. Niemeijer) from a large dataset (150 000 images) as acquired in a diabetic retinopathy screening program [26]. The inclusion criteria were that the screening program ophthalmologists had marked the image as containing microaneurysms and had not marked it as ungradable. Since multiple screening sites, with different camera types, are involved in the screening program, the images in the **ROC** set are relatively heterogeneous. Three different types of images with different resolutions are present in the dataset (type I to III, see Table I). The images were captured using either a Topcon NW 100, a Topcon NW 200, or a Canon CR5-45NM and this resulted in two differently shaped FOVs. An example of all types of images is shown in Fig. 2. All images were in JPEG format and compression was set in the camera. The substantial black background around the FOV present in the original type II and III images was cut off using software that can perform certain image processing operations on JPEG images without recompressing them. This complete set was randomly split into a training and a test set each containing 50 images.

B. Reference Standard

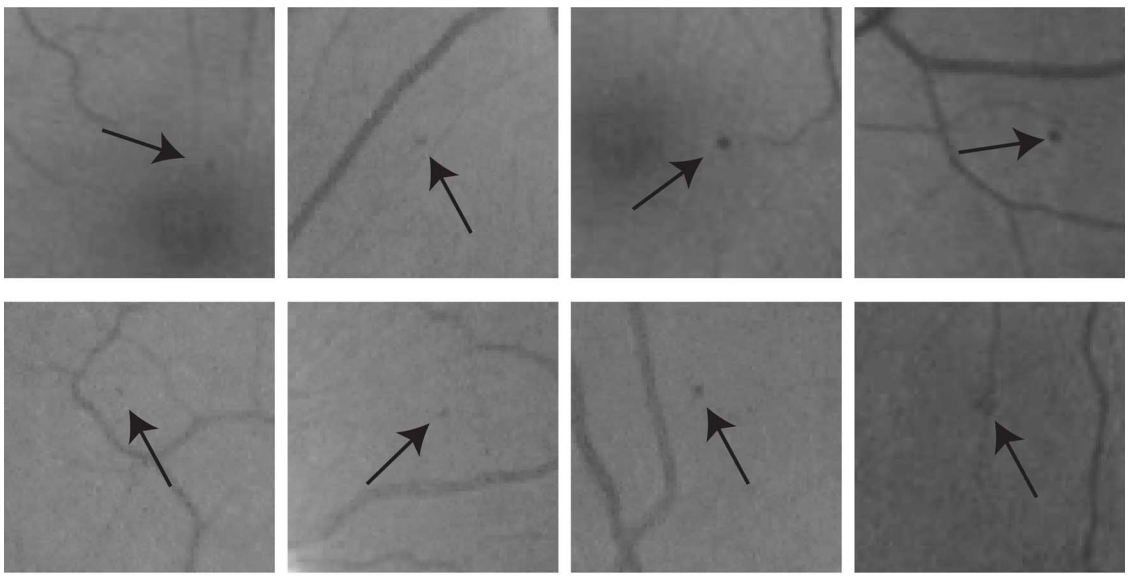
Four retinal experts, all from the Department of Ophthalmology at the University of Iowa, were asked to annotate all microaneurysms and all irrelevant lesions in all 100 images in the test and training set. The class of irrelevant lesions was specifically added to address the problem that there may be objects in the image that are not microaneurysms but that may be picked up as such by an automated program. As these objects are similar in appearance to microaneurysms these detections are not real false positives (examples of these types of objects include hemorrhages and pigment spots). Irrelevant objects are not included in the reference standard and detections on these objects in the test set were not counted as a false positive. The TruthseekerJ program (available on the **ROC** website) for online annotation

[1]http://roc.healthcare.uiowa.edu



(a) (b) (c)

Fig. 2. Representative example images of the three types of images (see Table I) present in the training and test set of the ROC microaneurysm detection competition. (a) Type I. (b) Type II. (c) Type III. Note that the images (b) and (c) shown here were preprocessed by removing a large part of the black image background around the FOV. All images have been scaled to equal height for display.



(a) (b) (c) (d)

Fig. 3. Examples of the different categories of microaneurysms as indicated by the expert. Both images in a column are in the same category. (a) Subtle. (b) Regular. (c) Obvious. (d) Close to the vasculature.

was used to indicate the center location of the microaneurysm or irrelevant lesion in each image. The annotations were exported as a file in XML format that contained the center locations for all microaneurysms and irrelevant lesions in each image in the set.

1) Training Set: For the training set, a logical OR was used to combine the lesion locations annotated by the four experts—thus ensuring that the reference dataset was highly sensitive to lesions, as it required only one retinal expert to detect a lesion for it to be marked as a "microaneurysm."

2) Test Set: To compare with a human observer, the way in which the reference standard for the test set was determined was changed from the way this was done for the training set. One of the four retinal experts was randomly selected and his annotations were not used in determining the reference standard for the test set. This expert's performance is shown in the results as the human expert. The annotations of the remaining three retinal experts were combined into the final reference standard. All lesions for which at least two experts assigned the label "microaneurysm" were assigned this label in the final reference standard for the test set. All other lesions marked by only one expert were assigned the label "irrelevant." In total 1614 objects were marked in the test set, of which 343 objects were assigned the label "microaneurysm." In the final set there were 10 images that did not contain any microaneurysms; they did contain "irrelevant" lesions.

3) Subdivision Into Lesion Categories: Microaneurysms are heterogeneous in their appearance (local contrast and color) and their location with respect to the normal anatomy on the retina. These factors can influence the ability of automated detection systems to find them. We have asked an expert to assign each reference-standard identified microaneurysm in the test set to one of three classes (i.e., subtle, regular, obvious) based on their local contrast and/or visibility in the image [see Fig. 3(a)–(c)]. Also, we asked the expert to indicate for each lesion whether it was located "close" to the vasculature. "Close" was defined as less than or equal to the diameter of the lesion away from the vessel [Fig. 3(d)]. Many of the tested

methods (see Section IV) rely on a vessel segmentation step to eliminate false positive lesion detections on the vasculature. From the literature [21] we know this may lead to difficulties in the detection of lesions located close to the vasculature. Of the 343 objects labeled as "microaneurysm" in the reference standard 92 were labeled "subtle," 192 were labeled "regular," 59 were labeled "obvious" and 37 of the 343 objects were close to the vasculature.

IV. Methods

The **ROC** dataset was downloaded by 26 groups and of these, 6 groups sent in their results in time to be included in this work. Though the organizers (M. Niemeijer, B. van Ginneken, and M. D. Abràmoff) have also published an algorithm for "red lesion" detection [21], because of their obvious conflict of interest they did not participate in this comparison. One group decided they did not want their results to appear in this paper after they were announced. Below are short descriptions of each of the 5 remaining methods. For each of the methods references to more detailed information regarding the methods and techniques used have been provided.

A. Method 1: Valladolid [27]

In this method candidate detection is performed on the green plane of the color image. The image is first resized so that the field of view has a certain width and the image is normalized by subtracting an estimate of the image background. The estimate is determined by median filtering the image using a large kernel. On the normalized image intensities, the candidate detection step is performed using an unsupervised mixture model based clustering method. The authors assume all pixels in the image are part of one of three classes: class 1 (background elements), class 2 (foreground elements, such as vessels, optic disk and lesions), and class 3 (outliers). A three class Gaussian mixture model is fit to the image intensities and a group of microaneurysm candidates are segmented by thresholding the fitted model. Vessel segmentation is performed to remove those detected candidates that lie on the vasculature. Using logistic regression, a likelihood for each of the remaining microaneurysm candidates is generated based on their color, shape and texture characteristics. The training data provided by the **ROC** was used to train the system.

B. Method 2: Waikato [28]

The Waikato Microaneurysm Detector was designed to be useful in an eye screening context; thus the reliable detection and localisation of microaneurysms in a retinal image is not the primary aim. This function is only useful in that it provides a mechanism to detect diabetic retinopathy on the basis of presence of microaneurysms. The algorithm used is a variant of the microaneurysm detection algorithm developed by Spencer *et al.* [14] and extended/used by Cree *et al.* [15] for the detection of microaneurysms in fluorescein angiographic images. For the Waikato Microaneurysm Detector the algorithm has been modified to work on full-color retinal images. The green plane of the retinal images is used to find all candidate objects. After normalization by subtracting a median filtered version of the image, noise is removed by median filtering using a small kernel. The Waikato Microaneurysm Detector performs a top-hat transform by morphological reconstruction [29] using an elongated structuring element at different orientations which detects the vasculature. After removal of the vasculature and a microaneurysm matched filtering step the candidate positions are found using thresholding. Each of the candidates is segmented using region growing. A number of features based on the color, intensity and shape of the candidates are extracted [30]. A Bayesian classifier is used to assign a likelihood to each of the found candidate objects that it is a true microaneurysm. The ROC training data was not used, instead the system was trained using an independent set of images.

C. Method 3: Latim [23]

This method assumes that microaneurysms at a particular scale can be modeled with 2-D, rotation-symmetric generalized Gaussian functions. It then uses template matching in the wavelet domain to find the microaneurysm candidates. The wavelet decomposition of an image produces several subimages of coefficients (called subbands), each of them containing information at a specific scale/frequency and along a specific direction. These subbands contain either more or less relevant information to describe microaneurysms. By ignoring high- and low-frequency subbands, noise and slow image variations are removed from the analysis. To perform template matching in the wavelet domain the wavelet transform of the Gaussian microaneurysm model at several different standard deviations is used. The wavelet transform of the model is restricted in a window which is moved across the image. The coefficients of the wavelet transform of the model are compared to the coefficients of the wavelet transform of the image. Locations in the image where the difference is below a certain threshold are designated as microaneurysm locations. The parameters of the matched model determine the size of the detected microaneurysm. Each detected location is assigned a likelihood based on the local difference in wavelet coefficients. In addition to the method as described in [23] a vessel segmentation step based on wavelet analysis of the image was added to eliminate candidates on the vasculature specifically for the data used in the **ROC**. The training data provided by the **ROC** was used to train the system.

D. Method 4: Ok Medical [31]

To detect the microaneurysm candidates this method uses a multiscale Bayesian correlation filtering [32] approach. In this approach responses from a Gaussian filterbank are used to construct probabilistic models of an object and its surroundings. By matching the filterbank outputs in a new image with the constructed (trained) models a correlation measure is obtained. When the responses of the correlation filtering are larger than a certain threshold, the detected locations are regarded as candidate microaneurysm locations. An adaptive thresholding scheme is applied to segment the vasculature and all candidates on the vasculature are removed. The candidate microaneurysms are segmented using region growing and a large set of features based on shape, grayscale/color pixel intensity, responses of

Gaussian filter-banks and correlation coefficient values as described in [21] are extracted from each candidate. The minimum and maximum values of each feature for all true lesions are placed in a discrimination table. This is used to remove any candidates whose features are below the minimum or greater than the maximum defined in the discrimination table. The remaining candidates after this stage are classified as true red lesions. To do a soft classification the maximum filter response for that particular lesion is used. The training data provided by the **ROC** was used to train the system.

E. Method 5: Fujita Lab [33]

The algorithm starts by preprocessing the image. To reduce differences in intensity and contrast between the images, brightness correction, gamma correction, and contrast enhancement [34] are applied. Further processing is performed exclusively on the green-plane of the color image. Images smaller than the largest size image in the **ROC** dataset are resized using bicubic scaling. Initial microaneurysm detection is performed using a modified double ring filter. The original double ring filter [35] was designed to detect areas in the image in which the average pixel value is lower than the average pixel value in the area surrounding it. Instead, the modified filter detects areas where the average pixel value in the surrounding area is lower by a certain fraction of the number of pixels under the filter in order to reduce false positive detections on small capillaries. After the microaneurysm detection step many false positive candidates on the vasculature remain. To remove these the vasculature is detected using an original double ring filter with different parameter settings. All candidates determined to lie on the vasculature are removed. A region growing procedure is used to segment the detected microaneurysm candidates. A set of 12 features is extracted from each of the segmented candidate objects; these include shape, intensity, color and contrast features. An artificial neural network was trained to distinguish the true lesions from the spurious detections. The training data provided by the **ROC** was used to train the system.

V. Validation

Each of the methods produced an XML file with the final detection results. The format of the XML file was defined by the **ROC** organizers. For each of the findings i in the complete set of findings R three parameters were stored; its location, the likelihood that the detected object was a microaneurysm p_i as well as the index of the image where i was detected x_i with $x \in I$ where I is the complete test set and $x \in [1 \dots 50]$. For each finding in R a match was made with the reference standard to determine the algorithm performance. The reference standard consisted of individual findings j collected in a set T (either irrelevant or true lesion findings, see Section III-B). Each of the individual findings has three parameters; a location, a radius r_j and the index of the image where it is located x_j. The radius is meant to reflect the approximate scale of the object as indicated by the experts that made the reference standard. To match the algorithm results to the reference standard Algorithm 1 was used.

Algorithm 1 ROC evaluation algorithm

Input: R and T.

Output: a paired set of sensitivities and false positive rates.

Select all unique values of p_i and store them in set L. Sort set L in ascending order.

for $k = 1$ to $k = |L|$ **do**

  Select all findings i in R for which $p_i \geq L_k$, where L_k is the k-th element of L, and store them in set F.

  for $l = 1$ to $l = |F|$ **do**

    Determine the reference standard finding $j \in T$ at distance d, closest to F_l in the same image.

    if $d < r_j$ **then**

      if $j =$ true finding **then**

        $TP = TP + 1$

        remove j from T.

      end if

    else

      $FP = FP + 1$

    end if

  end for

  Calculate and output $\text{sensitivity} = (TP/|T_{\text{true}}|)$ where T_{true} is the set of true microaneurysms in the reference standard.

  Calculate and output $\text{false positive rate} = (FP/|I|)$.

  Restore T to include all original reference standard findings.

end for

The output of the **ROC** evaluation algorithm is then used to generate an FROC curve [24] that plots the sensitivity against the average number of false positive detection per image. To facilitate comparison between the different methods the FROC curve is summarized in a single number. This number is based on the achieved sensitivity at a set of particular false positive per image rates. We have selected the following points: $(1/8), (1/4), (1/2), 1, 2, 4$, and 8. To obtain the final score the sensitivities at these specific points are averaged.

VI. Experiments and Results

All the results received from the participants were analyzed using the validation algorithm as outlined in Section V. This algorithm produced a set of FROC curves for each of the different methods. The different curves represent the performance of the algorithm for various groups of microaneurysms (i.e., all, subtle, regular, obvious, and close to the vasculature). Fig. 4 shows the

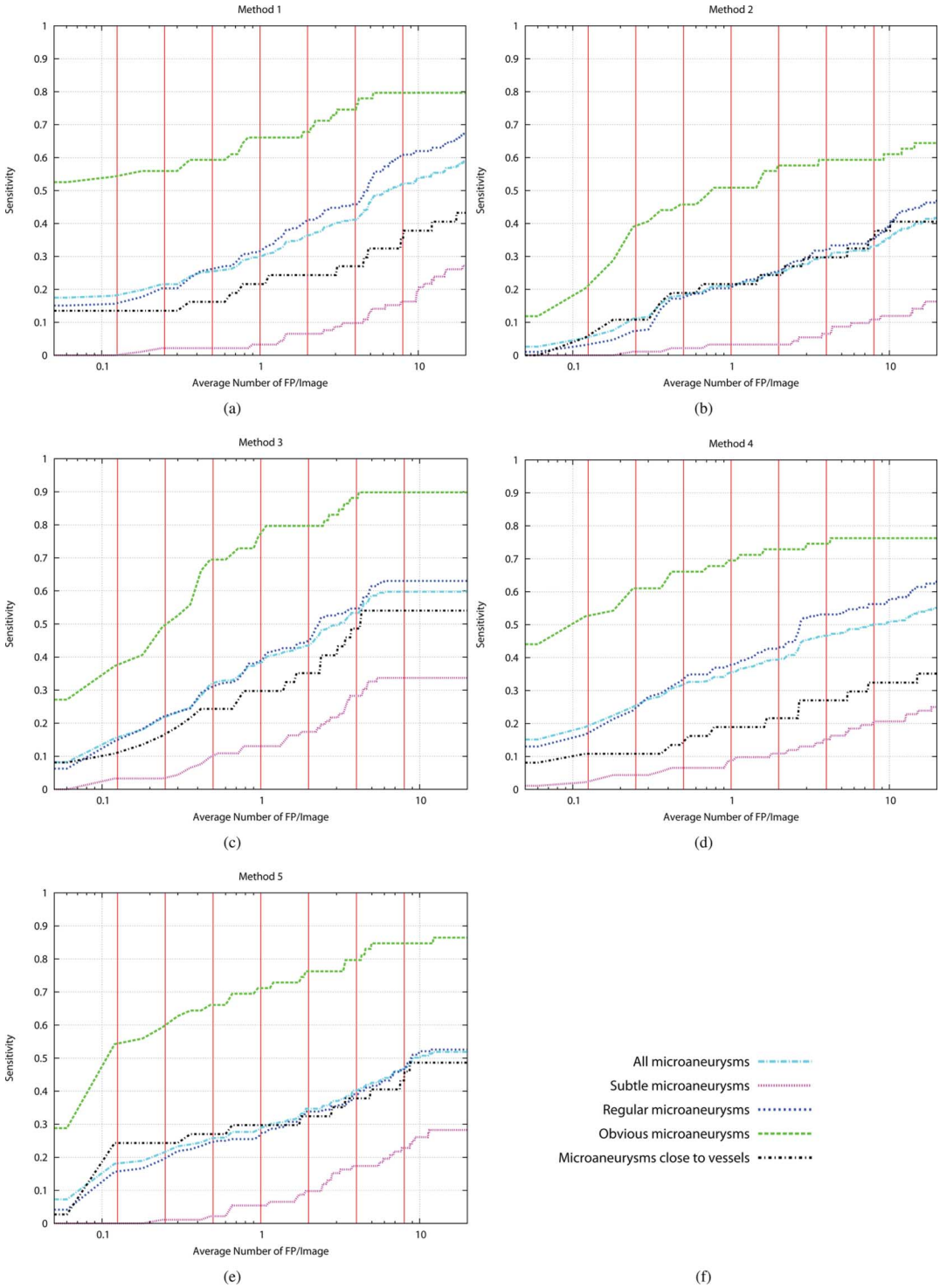


Fig. 4. (a)–(e) The FROC curves for each of the individual methods on the test data. These curves plot the sensitivity versus the average false positive rate for the different categories of microaneurysms. Please note the x-axis scale is logarithmic. The vertical red lines in the image indicate the false positive rates where the sensitivity of the methods are measured to determine the final score (See Table II). (f) Legend.

FROC curves of each of the methods as well as the measurement points, indicated by the vertical red lines, that the final system score is based on. To facilitate comparison of the performance of the different methods on the various types of lesions Fig. 5 shows the scores of all systems on each of the different types of microaneurysms. In Fig. 5 the results of the fourth retinal expert are also included. As the expert just provided a binary opinion his performance is given as a single point on the FROC curve. The final scores of the various methods as shown in Table II are based on all microaneurysms found in the test set.

When we assume that the false positive rate of the fourth expert (1.08 FP/image) provides an indication of the "clinically acceptable" false positive rate we can obtain greater insight into the performance of the systems when set for use in a clinical en-

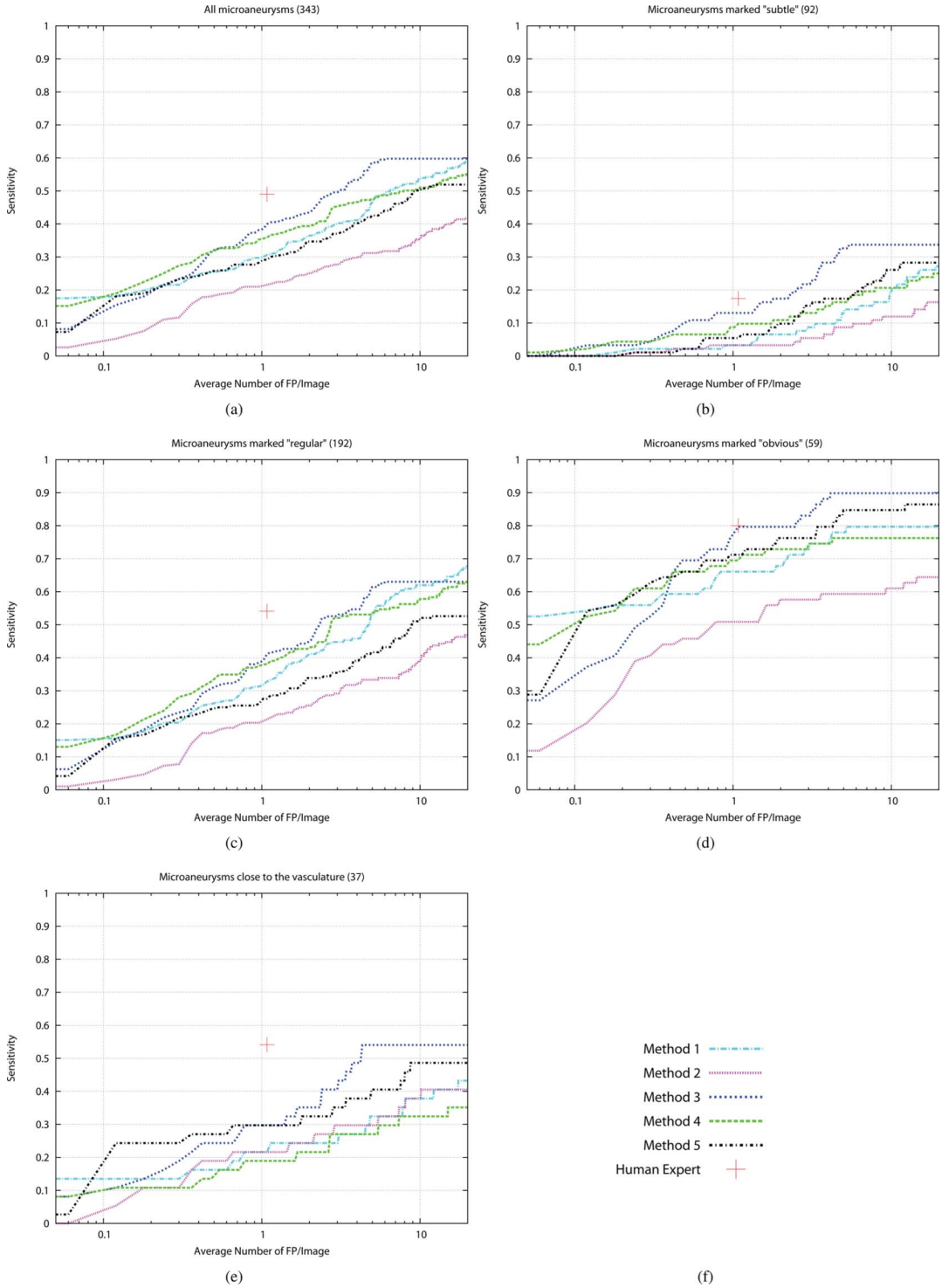


Fig. 5. (a)–(e) The FROC curves for each of the individual categories of microaneurysms. These curves plot the sensitivity versus the average false positive rate for the different methods. Please note the x-axis scale is logarithmic. (f) Legend.

vironment. Table III shows the sensitivities of the systems at this false positive rate for all different categories of microaneurysms, it also includes the rank of each method for each category.

VII. Discussion and Conclusion

A new publicly available database for microaneurysm detection in digital retinal color photographs has been presented as well as the results of five different microaneurysm detection methods on this database. The competition website where the database is available for download will remain open for new submissions and the goal of the **ROC** microaneurysm detection competition image database is to be a reference database for future work on automatic microaneurysm detection.

When examining Fig. 4 two different performance behaviors for the five methods can be distinguished; Method 1 and 4 seem to have substantial performance differences between the

TABLE II
SENSITIVITIES OF THE DIFFERENT METHODS AT THE VARIOUS MEASUREMENT POINTS. ALL MICROANEURYSMS FROM THE TEST SET ARE INCLUDED

Average nr. of false positives per image:	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	1	2	4	8	Final score
Method 1	0.19	0.22	0.25	0.30	0.36	0.41	0.52	0.32
Method 2	0.06	0.11	0.18	0.21	0.25	0.30	0.33	0.21
Method 3	0.17	0.23	0.32	0.38	0.43	0.53	0.60	0.38
Method 4	0.20	0.27	0.31	0.36	0.39	0.47	0.50	0.36
Method 5	0.18	0.22	0.26	0.29	0.35	0.40	0.47	0.31

TABLE III
SENSITIVITIES OF THE DIFFERENT METHODS AT THE AVERAGE FALSE POSITIVE RATE OF THE HUMAN EXPERT (1.08 FP/IMAGE) FOR THE VARIOUS CATEGORIES OF MICROANEURYSMS. THE RANKING OF THE METHODS FOR A PARTICULAR CATEGORY IS GIVEN BETWEEN BRACKETS BEHIND THE SENSITIVITIES

Lesion category:	All	Subtle	Regular	Obvious	Close to vessel
Method 1	0.31 (3)	0.03 (4)	0.33 (3)	0.66 (4)	0.22 (2)
Method 2	0.21 (5)	0.03 (4)	0.21 (5)	0.51 (5)	0.22 (2)
Method 3	0.40 (1)	0.13 (1)	0.41 (1)	0.80 (1)	0.30 (1)
Method 4	0.36 (2)	0.10 (2)	0.38 (2)	0.69 (3)	0.19 (3)
Method 5	0.29 (4)	0.05 (3)	0.28 (4)	0.71 (2)	0.30 (1)

different categories of microaneurysms while Method 2, 3, and 5 have similar performance for "All," "Regular" and "Close to vessels" microaneurysms with a large difference only between the "Subtle" and "Obvious" categories. Both Method 1 and 4 have a substantially lower performance on the microaneurysms that are close to vessels. The fact that a lesion is close to a vessel is independent from whether that lesion is in one of the other categories (i.e., subtle, regular or obvious). Therefore, if a method has no trouble detecting lesions close to the vasculature, one would expect a very similar performance to the complete set of microaneurysms. This is the case for Methods 2, 5, and to a lesser extent 3, as shown in Fig. 4(b), (e), and (c), respectively. In case of Methods 1 and 3 this is most likely due to an oversegmentation of the vasculature leading to the removal of microaneurysms close to the vasculature.

From Fig. 5 the results of the various methods show that only for the microaneurysms marked "subtle" and "obvious" does the best performing computer method approach the human expert. For every other category of microaneurysms as well as for all microaneurysms the human expert is well ahead of the computer methods. This shows there is still room for improvement in the automatic system performance. The performance of the human observer also indicates the difficulty of the task. For the lesions labeled as "subtle" the observer achieved less than 0.20 sensitivity. This also may indicate that this particular expert may have been less sensitive than some of the other experts involved in making the reference standard. This is also indicated by the number of false positive detections the expert generated which was also small.

The method with the highest final score, Method 3, outperforms the other methods in all categories especially when looking at the area under the FROC curve between 0 and 8 FP per image. For the "regular" and "overall" sets of microaneurysms the difference is limited but this particular method does a good job detecting "subtle" microaneurysms as well as those close to the vasculature. When examining the results at a "clinically acceptable" false positive rate (see Table III) the ranking of the different methods for all microaneurysms is exactly the same as the one given by the final score. For the other categories of microaneurysms Method 3 is also ranked number one although it is a shared first place for those microaneurysms that are close to the vasculature.

We received feedback from several of the participants that the data used for the **ROC** was "difficult" due to the presence of noise, compression artifacts and the general image quality. The images were selected randomly out of a vast dataset of images of patients with diabetes, and we have not specifically selected these images (see Section III-A): this is the quality of data typical in the screening project that provided the images used in this study. The images have not been recompressed or altered except in the way described in Section III-A after reception from the screening sites. Another issue that made analysis more difficult was the fact that the data was heterogeneous with different resolutions and cameras being used. The data was also acquired without dilation of the pupil which also leads to greater variations in image quality, but is again a consequence of the real world type of data as collected in the screening program. It is important to point out that the results obtained by the different methods on the **ROC** data are likely not indicative of the performance the method may achieve on different, more homogeneous data (e.g., from a single camera using the same resolution).

Although the final goal of these algorithms is early detection of DR, this competition was only about detecting a specific subset of all lesions that can occur in DR, namely microaneurysms. Detecting microaneurysms is only a single, though essential, step in early detection of DR. DR detection systems are evaluated on the overall presence of DR in the patient, not the number of microaneurysms in a single image. Depending on the level of DR that an automated system is required to detect (for example "referrable retinopathy," or "more than minimal retinopathy" [4]), researchers have used different thresholds on the number of microaneurysms in an image. Systems with differing performance on microaneurysm detection may have better (or worse) performance on DR detection.

We have conducted an additional experiment to examine the capability of the various systems to detect images containing signs of DR by assuming the presence of DR is indicated solely by the presence of microaneurysms. In our experiment the maximum likelihood lesion in the image (that is not irrelevant) indicates likelihood of the whole image to contain signs of DR. Since the testing set contained just 10 images with only irrelevant objects (i.e., without microaneurysms) this experiment is biased, additionally the use of the maximum rule for lesion likelihood combination is likely not optimal. However,

TABLE IV
RESULTS OF AN ADDITIONAL EXPERIMENT WHERE A "PER IMAGE" EVALUATION WAS DONE TO SEE THE EFFECT ON THE RANKING OF THE METHODS. NOTE THAT THE AREA UNDER THE ROC CURVE (A_z) VALUE OF THE HUMAN OBSERVER IS BASED ON ONLY A SINGLE POINT ON THE ROC CURVE

	A_z
Method 1	0.80
Method 2	0.72
Method 3	0.87
Method 4	0.89
Method 5	0.88
Human Expert	0.96

this bias does not affect the ability of the experiment to determine if the ranking of the methods changes when applying a per image evaluation. From the results of this additional experiment, given in Table IV, it is clear that there was an influence. Method 3, the best performing detection method in the competition is now ranked third, although the differences between the three top ranking methods are very small. The change in ranking may be caused by the fact that the per image evaluation using the maximum rule shifts the emphasis from "lesion sensitivity" to "accurate lesion likelihood assignment." Even a single false positive with a high likelihood in one of the ten negative images will have a large impact on the final result given the biased set. Another interesting result of this experiment is that the Human Expert observer achieves an area under the ROC curve of 0.96 despite not detecting the majority of the subtle microaneurysms. This indicates that while finding individual microaneurysms is important in detecting patients with DR, it may not be necessary to detect every individual lesion in all images to screen for DR.

As the previous paragraph's results showed, the choice of evaluation criterion can influence the ranking of the methods. We have chosen to use FROC analysis for the **ROC** competition even though this limits the possibilities for statistical analysis. FROC analysis also lacks an area under the curve summary variable that ROC analysis does have. Our motivation to choose FROC over other evaluation measures was that it is generally regarded as the evaluation measure most closely resembling clinical practice in diagnostic tasks that involve lesion localization [36]. The averaging of the sensitivities of the methods at a fixed set of false positive rates provides, in our view, an adequate summary variable. Proposed FROC alternatives such as AFROC [37] that compensate for some of the earlier mentioned weaknesses of FROC analysis make additional assumptions which may not hold for the data used in this study [36].

As we have stated in the introduction, none of the previous work on microaneurysm detection was compared on the same data. The **ROC** microaneurysm detection competition is unique in the sense it makes a dataset and evaluation methodology available for future work on the detection of microaneurysms. The results from the participants and the data will remain available on the competition website that will remain open and accept additional submissions. We hope that the availability of this data will stimulate additional groups to pursue research in this field. The results of the five teams that participated in the **ROC** provide a great starting point for future developments.

ACKNOWLEDGMENT

The work of R. Hornero and M. García was supported in part by "Ministerio de Ciencia e Innovación" under project TEC2008-02241. H. Fujita, C. Muramatsu, A. Mizutani, and Y. Hatanaka would like to acknowledge the contributions of T. Hara and S. Suemori to their part of this work. B. Zhang, X. Wu, J. You, Q. Li, and F. Karray of the OK Medical team would like to thank the Hong Kong Government General Research Fund (GRF) and The Hong Kong Polytechnic University Research Grant for their support. The Waikato Medical Research Foundation financially supported the development of the Waikato microaneurysm detector.

REFERENCES

[1] D. C. Klonoff and D. M. Schwartz, "An economic analysis of interventions for diabetes," *Diabetes Care*, vol. 23, no. 3, pp. 390–404, 2000.
[2] G. H. Bresnick, D. B. Mukamel, J. C. Dickinson, and D. R. Cole, "A screening approach to the surveillance of patients with diabetes for the presence of vision-threatening retinopathy," *Opthalmology*, vol. 107, no. 1, pp. 19–24, 2000.
[3] S. Philip, A. D. Fleming, K. A. Goatman, S. Fonseca, P. McNamee, G. S. Scotland, G. J. Prescott, P. F. Sharp, and J. A. Olson, "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.
[4] M. D. Abràmoff, M. Niemeijer, M. S. A. Suttorp-Schulten, M. A. Viergever, S. R. Russell, and B. van Ginneken, "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, no. 2, pp. 193–198, 2008.
[5] M. Niemeijer, M. D. Abràmoff, and B. van Ginneken, "Information fusion for diabetic retinopathy CAD in digital color fundus photographs," *IEEE Trans. Med. Imag.*, vol. 28, no. 5, pp. 775–785, May 2009.
[6] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imag.*, vol. 19, no. 3, pp. 203–210, Mar. 2000.
[7] A. Hoover and M. Goldbaum, "Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels," *IEEE Trans. Med. Imag.*, vol. 22, no. 8, pp. 951–958, Aug. 2003.
[8] J. Staal, M. Abràmoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge based vessel segmentation in color image of the retina," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 501–509, Apr. 2004.
[9] M. Niemeijer, J. J. Staal, B. van Ginneken, M. Loog, and M. D. Abràmoff, "Comparative study of retinal vessel segmentation methods on a new publicly available database," in *Proceedings of SPIE: Med. Imag.*, 2004, vol. 5370, pp. 648–656.
[10] Methods to evaluate segmentation and indexing techniques in the field of retinal ophthalmology [Online]. Available: http://messidor.crihan.fr
[11] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int. J. Comput. Vis.*, vol. 47, pp. 7–42, 2002.
[12] C. E. Baudoin, B. J. Lay, and J. C. Klein, "Automatic detection of microaneurysms in diabetic fluorescein angiographies," *Revue D'Épidémiologie et de Sante Publique*, vol. 32, pp. 254–261, 1984.
[13] T. Spencer, R. P. Phillips, P. F. Sharp, and J. V. Forrester, "Automated detection and quantification of microaneurysms in fluorescein angiograms," *Graefe's Arch. Clin. Exp. Ophthalmol.*, vol. 230, pp. 36–41, 1992.
[14] T. Spencer, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester, "An image-processing strategy for the segmentation and quantification in fluorescein angiograms of the ocular fundus," *Comput. Biomed. Res.*, vol. 29, pp. 284–302, 1996.
[15] M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester, "A fully automated comparative microaneurysm digital detection system," *Eye*, vol. 11, pp. 622–628, 1997.
[16] A. J. Frame, P. E. Undrill, M. J. Cree, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester, "A comparison of computer based classification methods applied to the detection of microaneurysms in ophthalmic fluorescein angiograms," *Comput. Biol. Med.*, vol. 28, pp. 225–238, 1998.

[17] L. Yannuzzi, K. Rohrer, and L. Tindel, "Fluorescein angiography complication survey," *Ophthalmology*, vol. 93, pp. 611–617, 1986.

[18] J. H. Hipwell, F. Strachant, J. A. Olson, K. C. McHardy, P. F. Sharp, and J. V. Forrester, "Automated detection of microaneurysms in digital red-free photographs: A diabetic retinopathy screening tool," *Diabetic Med.*, vol. 17, pp. 588–594, 2000.

[19] A. D. Fleming, S. Philip, K. A. Goatman, J. A. Olson, and P. F. Sharp, "Automated microaneurysm detection using local contrast normalization and local vessel detection," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1223–1232, Sep. 2006.

[20] T. Walter, P. Massin, A. Erginay, R. Ordonez, C. Jeulin, and J.-C. Klein, "Automatic detection of microaneurysms in color fundus images," *Med. Image Anal.*, vol. 11, no. 6, pp. 555–566, 2007.

[21] M. Niemeijer, B. van Ginneken, J. Staal, M. S. A. Suttorp-Schulten, and M. D. Abràmoff, "Automatic detection of red lesions in digital color fundus photographs," *IEEE Trans. Med. Imag.*, vol. 24, no. 5, pp. 584–592, May 2005.

[22] C. Sinthanayothin, J. F. Boyce, T. H. Williamson, H. L. Cook, E. Mensah, S. Lal, and D. Usher, "Automated detection of diabetic retinopathy on digital fundus images," *Diabetic Med.*, vol. 19, pp. 105–112, 2002.

[23] G. Quellec, M. Lamard, P. M. Josselin, G. Cazuguel, B. Cochener, and C. Roux, "Optimal wavelet transform for the detection of microaneurysms in retina photographs," *IEEE Trans. Med. Imag.*, vol. 27, no. 9, pp. 1230–1241, Sep. 2008.

[24] P. Bunch, J. Hamilton, G. Sanderson, and A. Simmons, "A free response approach to the measurement and characterization of radiographic-observer performance," *J. Appl. Photogr. Eng0*, vol. 4, pp. 166–172, 1978.

[25] C. E. Metz, "ROC methodology in radiologic imaging," *Investigative Radiology*, vol. 21, no. 9, pp. 720–733, 1986.

[26] M. D. Abràmoff and M. S. A. Suttorp-Schulten, "Web-based screening for diabetic retinopathy in a primary care population: The EyeCheck project," *Telemedicine e-Health*, vol. 11, no. 6, pp. 668–674, 2005.

[27] C. I. Sánchez, R. Hornero, A. Mayo, and M. García, "Mixture model-based clustering and logistic regression for automatic detection of microaneurysms in retinal images," in *SPIE Medical Imaging 2009: Computer-Aided Diagnosis*, N. Karssemeijer and M. L. Giger, Eds., 2009, vol. 7260, p. 72601M.

[28] M. J. Cree, The Waikato Microaneurysm Detector Univ. Waikato, Tech. Rep., 2008 [Online]. Available: http://roc.healthcare.uiowa.edu/results/documentation/waikato.pdf

[29] L. Vincent, "Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms," *IEEE Trans. Image Process.*, vol. 2, no. 2, pp. 176–201, Apr. 1993.

[30] M. J. Cree, E. Gamble, and D. Cornforth, "Colour normalisation ro reduce inter-patient and intra-patient variablility in microaneurysm detection in colour retinal images," in *Proc. WDIC2005 ARPS Workshop Digital Imag. Comput.*, 2005, pp. 163–168.

[31] B. Zhang, X. Wu, J. You, Q. Li, and F. Karray, "Hierarchical detection of red lesions in retinal images by multiscale correlation filtering," in *SPIE Medical Imaging 2009: Computer-Aided Diagnosis*, N. Karssemeijer and M. L. Giger, Eds., 2009, vol. 7260, p. 72601L.

[32] S. B. Isard, J. Sullivan, A. Blake, M. Isard, and J. Maccormick, "Bayesian object localisation in images," *Int. J. Comput. Vis.*, vol. 44, no. 2, pp. 111–135, 2001.

[33] A. Mizutani, C. Muramatsu, Y. Hatanaka, S. Suemori, T. Hara, and H. Fujita, "Automated microaneurysm detection method based on double ring filter in retinal fundus images," in *SPIE Medical Imaging 2009: Computer-Aided Diagnosis*, N. Karssemeijer and M. L. Giger, Eds., 2009, vol. 7260, no. 1, p. 72601N.

[34] Y. Hatanaka, T. Nakagawa, Y. Hayashi, M. Kakogawa, A. Sawada, K. Kawase, T. Hara, and H. Fujita, "Improvement of automatic hemor-rhage detection methods using brightness correction on fundus images," in *SPIE Medical Imaging 2008: Computer-Aided Diagnosis*, M. L. Giger and N. Karssemeijer, Eds., 2008, vol. 6915, p. 69153E.

[35] Y. Hatanaka, T. Nakagawa, Y. Hayashi, A. Aoyama, X. Zhou, T. Hara, H. Fujita, Y. Mizukusa, A. Fujita, and M. Kakogawa, "Automated detection algorithm for arteriolar narrowing on fundus image," in *Proc. 2005 IEEE Eng. Med. Biol. 27th Annu. Conf.*, 2005, pp. 286–289.

[36] P. M. DeLuca, Jr., A. Wambersie, and G. F. Whitmore*, J. ICRU*, vol. 8, no. 1, pp. 31–35, 2008.

[37] D. P. Chakraborty and L. H. Winter, "Free-response methodology: Alternate analysis and a new observer-performance experiment.," *Radiology*, vol. 174, no. 3, pp. 873–881, 1990.