Diabetic Retinopathy Image Database(DRiDB): A new database for diabetic retinopathy screening programs research

Pavle Prentašić*, Sven Lončarić*, Zoran Vatavuk†, Goran Benčić†, Marko Subašić*, Tomislav Petković*, Lana Dujmović†, Maja Malenica-Ravlić†, Nikolina Budimlija†, Rašeljka Tadić†
*University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
†Clinical Hospital Center "Sestre Milosrdnice"
Vinogradska cesta 29, 10000 Zagreb, Croatia

***Abstract*—Diabetic retinopathy is one of the leading disabling chronic diseases, and one of the leading causes of preventable blindness in the world. Early diagnosis of diabetic retinopathy enables timely treatment and in order to achieve it a major effort will have to be invested into screening programs and especially into automated screening programs. For automated screening programs to work robustly a representative fundus image database is required. In this paper we give an overview of currently available databases and present a new diabetic retinopathy database. Our database is to our knowledge the first and only database which has diabetic retinopathy pathologies and major fundus structures annotated for every image from the database which makes it perfect for design and evaluation of currently available and new image processing algorithms for early detection of diabetic retinopathy using color fundus images.**

I. Introduction

Diabetic retinopathy (DR) is one of the leading disabling chronic diseases, and one of the leading causes of preventable blindness in the world [1]. It was found to be the fourth most frequently managed chronic disease in general practice in 2009, and the projections go as high as the second most frequent disease by the year 2030 [1]. The global burden of diabetic patients is expected to rise from 171 million in 2000 to 366 million in 2030 [1]. In Europe more than 52.8 million people are diagnosed with diabetes with the number expected to rise to 64 million by 2030. In Croatia about 300 thousand people are estimated to have diabetes and of those only 190 thousand are registered. Early diagnosis of diabetic retinopathy enables timely treatment that can ease the burden of the disease on the patients and their families by maintaining a sufficient quality of vision and preventing severe vision loss and blindness [2]. In addition to the obvious medical benefits, significant positive economical effects are achieved by maintaining patient's workability and self-sustainability.

In order to achieve early diagnosis of diabetic retinopathy a major effort will have to be invested into screening programs. Screening is important as up to one third of people with diabetes may have progressive DR changes without symptoms of reduced vision [3], thus allowing the disease to progress and making treatment difficult. Systematic screening programs for diabetic eye disease have been developed in many countries [4], [5], [6]. In the UK, the NHS Diabetic Screening Program offers annual fundus photography for all patients with diabetes over the age of 12, regardless of their socio-economic status [6].

In current screening programs only color fundus photography is used, and the data are sent to a grading center for reading where expert human readers estimate the disease severity. The main disadvantage is the necessity for qualified experts to grade the images, e.g. in the NHS Diabetes Screening Program one patient's images can be graded by up to four different experts. This standard is impossible to achieve in countries with a shortage of qualified medical personnel.

Fundus imaging has an important role in diabetic retinopathy detection and monitoring because eye fundus is sensitive to vascular diseases and we can consider fundus imaging as a candidate for non-invasive screening. The success of this type of screening approach depends on accurate fundus image capture, and especially on accurate and robust image processing and analysis algorithms for abnormalities detection. Many algorithms have been proposed for fundus image analysis using different methods and approaches but it is sometimes difficult to measure the accuracy and reliability of the proposed algorithms because no commonly accepted and representative fundus image database exists in the public domain.

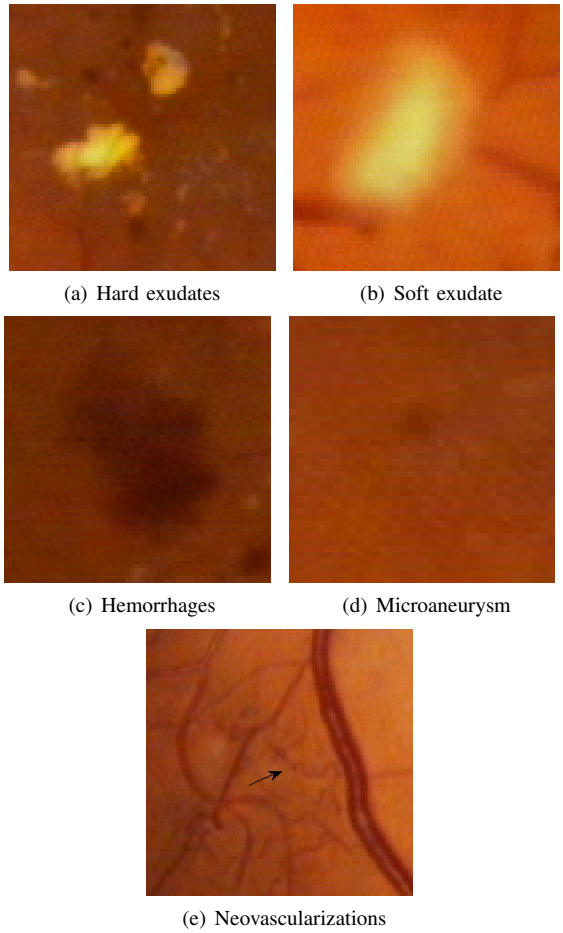
The main contribution of this work is to present a new, complete and publicly available, diabetic retinopathy database, DRiDB which contains the ground truth data from several ophthalmological experts. In Section II typical symptoms of diabetic retinopathy are explained. In Section III current state of the art diabetic retinopathy databases are presented. In Section IV the new database is presented, and finally in Section V we give a brief conclusion with short discussion about expected future work.

II. Diabetic retinopathy

Diabetes is a well known disease and may cause abnormalities in the retina (diabetic retinopathy), kidneys (diabetic nephropathy), nervous system (diabetic neuropathy) and is known to be a major risk for cardiovascular diseases. Diabetic retinopathy is a microvascular complication caused by diabetes which can lead to blindness. In early stages of diabetic retinopathy typically there are no visible signs but the number

and severity of abnormalities increase during the time. Diabetic retinopathy typically starts with small changes in retinal capillaries. The first detectable abnormalities are microaneurysms which represent local enlargements of the retinal capillaries. The ruptured microaneurysms can cause hemorrhages. After a period of time, hard exudates may appear. The hard exudates are lipid formations leaking from weakened blood vessels. As the retinopathy advances, the blood vessels may become obstructed which causes microinfarcts in the retina. These microinfarcts are called soft exudates. Extensive lack of oxygen caused by microinfarcts causes the development of new fragile vessels. This phenomenon is called neovascularization which is a serious eyesight threatening state and may cause sudden loss in visual acuity or even permanent blindness. Examples of microaneurysms, hemorrhages, hard exudates, soft exudates and neovascularization are visible in Fig. 1.

After diagnosis of diabetic retinopathy, regular monitoring is needed due to progressive nature of the disease. Sadly, broad screening cannot be performed due to the fact that fundus image examination requires medical experts. For the screening, automated image processing methods must be developed and to develop automated image processing methods high quality databases for algorithm evaluation are required.



(a) Hard exudates (b) Soft exudate

(c) Hemorrhages (d) Microaneurysm

(e) Neovascularizations

Fig. 1: Abnormal findings in the eye fundus images caused by diabetic retinopathy

III. Publicly available retinal image databases

An overview of all publicly available retinal image databases known to us is given in this section.

A. DRIVE database

The DRIVE (Digital Retinal Images for Vessel Extraction) is a publicly available database, consisting of a total of 40 color fundus photographs [7]. The photographs were obtained from a diabetic retinopathy screening program in the Netherlands. The screening population consisted of 400 subjects between 25 and 90 years of age. Each image has been JPEG compressed, which is common practice in screening programs. Of the 40 images in the database, 7 contain pathology, namely exudates, hemorrhages and pigment epithelium changes. The images were acquired using a Canon CR5 non-mydriatic 3-CCD camera with a 45° field of view (FOV). Each image was captured using 8 bits per color plane at 768×584 pixels. The FOV of each image was circular with a diameter of approximately 540 pixels. The set of 40 images was divided into a test and training set both containing 20 images. Three observers, the first and second author and a computer science student manually segmented a number of images. All observers were trained by an experienced ophthalmologist (the last author). The first observer segmented 14 images of the training set while the second observer segmented the other 6 images. The test set was segmented twice resulting in a set X and Y. Set X was segmented by both the first and second observer (13 and 7 images, respectively) while set Y was completely segmented by the third observer. The performance of the vessel segmentation algorithms was measured on the test set. In set X the observers marked 577,649 pixels as vessel and 3,960,494 as background (12.7% vessel). In set Y 556,532 pixels wew marked as vessel and 3,981,611 as background (12.3% vessel). This database does not contain annotated pathologies and other fundus structures like optic disc and macula.

B. STARE database

The STARE database contains 20 images for blood vessel segmentation; ten of these contain pathology [8]. The slides were captured by a Topcon TRV-50 fundus camera at 35° field of view. Each slide was digitized to produce a 605×700 pixel image, 24 bits per pixel (standard RGB). Two observers manually segmented all the images. On average, the first person labeled 32,200 pixels in each image as vessel, while the second person labeled 46,100 pixels in each image as vessel. A subsequent review indicated that the first person took a more conservative view of the boundaries of vessels and in the identification of small vessels than the second person. Performance was computed with the segmentation of the first observer as the ground truth.

C. ARIA online

This database was created in 2006, in a research collaboration between St. Paul's Eye Unit, Royal Liverpool University Hospital Trust, Liverpool, UK and the Department of Ophthalmology, Clinical Sciences, University of Liverpool, Liverpool, UK [9]. The database consists of three groups; the first group has 92 images with age-related macular degeneration, the second group has 59 images with diabetes and the control

group consists of 61 images. The trace of blood vessels, the optic disc and fovea location was marked by two image analysis experts as the reference standard. The images were captured at a resolution of 768×576 pixels in RGB color with 8-bits per color plane with a Zeiss FF450+ fundus camera at a 50° FOV and stored as uncompressed TIFF files.

D. ImageRet

The ImageRet database was made publicly available in 2008 and is subdivided into two sub-databases, DIARETDB0 and DIARETDB1 [10]. DIARETDB0 contains 130 retinal images of which 20 are normal and 110 contain various signs of diabetic retinopathy. DIARETDB1 contains 89 images out of which 5 images represent healthy retinas while the other 84 have some diabetic retinopathy signs. The images were acquired with a 50° FOV using a fundus camera at a size of 1500×1152 pixels in PNG format. The images were annotated by four experts for the presence of microaneurysms, hemorrhages, and hard and soft exudates. Annotated images from four experts were combined to produce a single ground truth image. There are no manually segmented vessel images in this database.

E. Messidor

The Messidor-project database, with 1200 retinal images, is the largest database currently available on the internet and is provided by the Messidor program partners [11]. The images were acquired by 3 ophthalmologic departments using a color video 3CCD camera on a Topcon TRC NW6 non-mydriatic camera with a 45° FOV. The images were captured using 8 bits per color plane at 1440×960, 2240×1488, or 2304×1536 pixels. 800 images were acquired with pupil dilation (one drop of Tropicamide at 0.5%) and 400 without dilation. The reference standard provided contains the grading for diabetic retinopathy and the risk of macular edema in each image. This database does not contain any other annotations and is used to facilitate studies on computer-assisted diagnoses of diabetic retinopathy.

F. Review

The Retinal Vessel Image set for Estimation of Widths (REVIEW) was made available online in 2008 by the Department of Computing and Informatics at the University of Lincoln, Lincoln, UK [12]. The dataset contains 16 mydriatic images with 193 annotated vessel segments consisting of 5066 profile points manually marked by three independent experts. The images were chosen to assess the accuracy and precision of the vessel width measurement algorithms in the presence of pathology and central light reflex. The 16 images are subdivided into four sets, the high resolution image set (HRIS, 8 images), the vascular disease image set (VDIS, 4 images), the central light reflex image set (CLRIS, 2 images) and the kickpoint image set (KPIS, 2 images).

G. ROC microaneurysm set

The Retinopathy Online Challenge microaneurysm dataset is part of a multi-year online competition of microaneurysm detection that was arranged by the University of Iowa in 2009 [13]. 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 withheld by the organizers. The images were captured using a Topcon NW100, a Topcon NW200 or a Canon CR5-45NM non-mydriatic camera at 45° FOV and were JPEG compressed in the camera. There are three different image sizes present in the database; 768×576, 1058×1061 and 1389×1383 pixels.

H. VICAVR

The VICAVR database is a set of retinal images used for the computation of the A/V ratio [14]. The database currently includes 58 images. The images were acquired with a Topcon NW-100 non-mydriatic camera and are optic disc centered with a resolution of 768×584. The database includes the caliber of the vessels measured at different radii from the optic disc as well as the vessel type (artery/vein) labeled by three experts.

I. HEI-MED

The Hamilton Eye Institute Macular Edema Dataset (HEI-MED) (formerly DMED) is a collection of 169 fundus images to train and test image processing algorithms for the detection of exudates and diabetic macular edema [15]. The dataset is composed of 169 JPEG images compressed at highest quality. Each image of the dataset was manually segmented by Dr. Edward Chaum (an expert ophthalmologist from HEI). He identified all the exudation areas and other bright lesions such as cotton wool spots, drusens or clearly visible fluid occurring on the fundus.

J. Comparison of databases

The presented overview of image databases shows that there is no database which contains both annotated pathologies like microaneurysms, hemorrhages, hard exudates, soft exudates, neovascularizations and normal fundus structures like blood vessels, macula and optic disc. Furthermore, some databases include images that were annotated only by a single expert, which introduces the problem of manual annotation bias. The key features of the evaluated databases are shown in Table I.

IV. Diabetic Retinopathy Image Database

The analysis of the publicly available databases represents a motivation for creation of a comprehensive database with the following desired properties:

- all fundus structures and pathologies are annotated
- at least five experts have annotated each patient image
- at least fifty patients included for statistically valid evaluation of image analysis method
- categorization of disease grade for each patient image

Such a database would be very useful to research community for reliable evaluation and objective comparison of medical image processing algorithms. The database should contain high-quality medical images which are representative of the problem and have been verified by the experts.

In this paper we present a new DR image database that has the above mentioned properties and that has been developed

TABLE I: Existing databases comparison

Database	Regular fundus structures	Pathologies	Neovascularizations	Multiple experts	Disease Grading	Number of images
DRIVE	×			×		40
STARE	×			×		20
ARIA	×			×		92+59+61
ImageRet		×		×		130+89
Messidor					×	1,200
Review	×			×		16
ROC		×		×		50
VICAVR	×			×		58
HEI-MED		×				169
Our database	×	×	×	×	×	50

as a part of our research activities. Regular fundus structures are present in each fundus image and are important because this information can be used to improve the accuracy and robustness of image processing algorithms for detection of diabetic retinopathy pathologies. For example, we can consider an algorithm for hemorrhages detection in color fundus images. Typically, hemorrhages are darker than surrounding background but blood vessels are similar and they are darker than surrounding background too. A typical hemorrhage detection algorithm starts with blood vessel suppression because we want to eliminate similar structures but in order to do this we need to have a vessel detection and extraction algorithm. This is a different problem in comparison to our starting problem and we want to build an algorithm which can detect vessels with high accuracy. There are many different algorithms available for this type of problem but we need a good database which contains manually segmented blood vessels like the DRIVE or STARE database to test the accuracy of proposed vessel detection method. Those databases are good for vessel detection algorithms but they do not contain annotations of diabetic retinopathy pathologies like hemorrhages so the testing results obtained on those databases are sometimes not representative because for example the DRIVE database mainly consists of healthy patients with no signs of hemorrhages and performance of vessel detection algorithms in presence of hemorrhages can be lower if we compare it to performance when no hemorrhages are present. This is the main reason why our database contains manually segmented blood vessels because we wanted to build a new database where comparison of different algorithms used in process of diabetic retinopathy pathologies detection can be reliably measured and compared. Segmented blood vessels can be used to compare the hemorrhage detection algorithms by masking out blood vessels if we want to compare different methods regardless of vessels present in the image. The database has 50 color fundus images of which 36 contain signs of the diabetic retinopathy and 14 which do not contain any signs of the diabetic retinopathy according to all experts who participated in the evaluation. An example of image from the database is visible in Fig. 2.

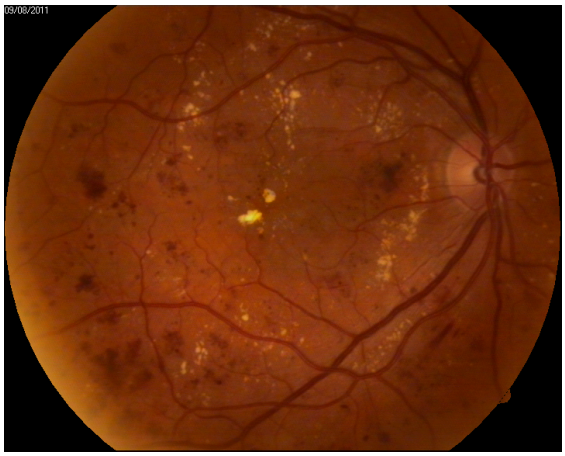


Fig. 2: Image from the new database

The images were taken and selected by medical experts from a university hospital in Zagreb. The distribution of patients does not correspond to any typical population. The diabetic retinopathy signs present vary from almost non existent to cases where new fragile vessels are visible and represent an eye sight threatening state. The images were captured at a resolution of 720×576 pixels in RGB color with 8-bits per color plane with a Zeiss VISUCAM 200 fundus camera at a 45° FOV and stored as uncompressed BMP files. Images were captured with varying flash intensities. The images contain a varying amount of image noise but we can say that images correspond to a good practical situation where the images are comparable and can be used to evaluate the general performance of diagnostic methods.

A set of ground truth images accompanies every color fundus image from the database. For each image from the database five experts independently marked diabetic retinopathy findings. A person with a medical education and specialization in ophthalmology is considered as an expert. A special software was given to the experts to inspect the fundus images and annotate the findings. The experts were asked to annotate the areas related to microaneurysms, hemorrhages, hard and soft exudates. These structures are not present in each image and are important because they can be used not just to measure the performance of image processing algorithms developed for detection of diabetic retinopathy symptoms but they can be used for construction of machine learning based image processing algorithms. An example of hard exudates segmented by one of the experts and superimposed on the original image is shown in Fig. 3.

The experts were asked to mark the blood vessels, optic disc and the macula alongside above mentioned diabetic retinopathy signs. An example of segmented vessel image is visible in Fig. 4.

In the third step the experts performed annotation of neovascularizations. A neovascularization represents a serious eye sight threatening state and may cause sudden loss in visual

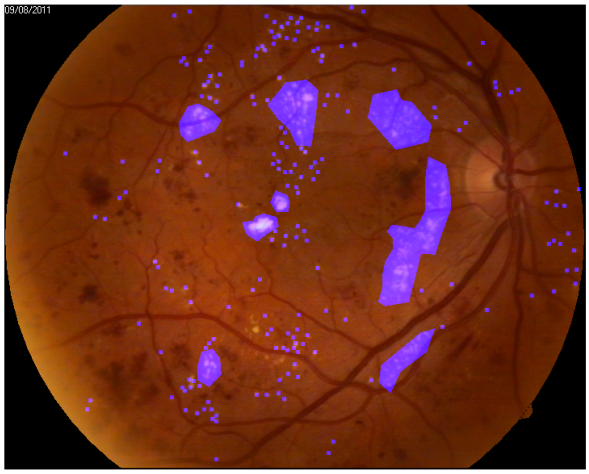


Fig. 3: Original image with exudates superimposed

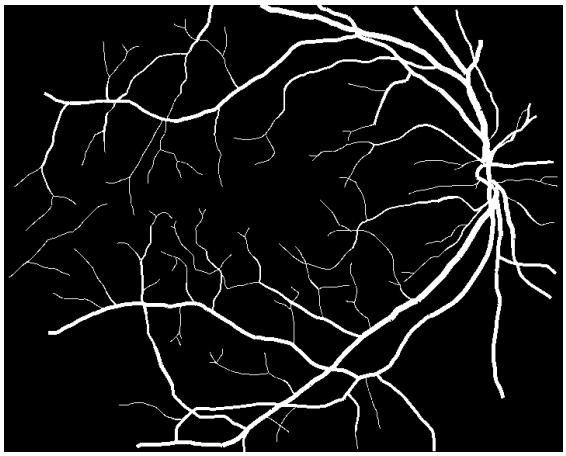


Fig. 4: An example of segmented blood vessels

acuity or even a permanent blindness if not treated accordingly.

Finally, each expert had to provide grading for diabetic retinopathy for each image from the database like in the Messidor database. The experts were instructed to report their confidence for each visual marking. The ground truth confidence levels available are low confidence, medium confidence and high confidence and they represent the certainty of the decision that a marked finding is correct. The experts were taught how to use the image annotation tool but they were not instructed how to annotate their findings to reduce biases introduced by the annotation procedure.

Because blood vessels differ from other structures the image annotation tool can be actually divided into two tools. The first tool is used to mark microaneurysms, hemorrhages, hard exudates, soft exudates, macula and optic disk. Currently, the image annotation tool supports the following graphical directives:

- Centroid
- Polygon region
- Ellipse region

The centroid item is typically used to mark microaneurysms because they represent just a point. Ellipse region is typically used to mark optic disk and macula because they are round. Polygon region is typically used for clusters of exudates and hemorrhages but the expert can use any of mentioned tools for any visual findings as he deems necessary.

Because some of the visual markings are more visible in red-free images the expert can change the annotated image from color to red-free fundus image during the annotation.

The second tool is used for blood vessel segmentation and to mark neovascularizations. It is used for neovascularizations because neovascularizations are actually blood vessels so it was natural to use the blood vessel segmentation tool. The tool is actually a modified version of the Live-Vessel software [16]. Using this software the user opens up an image, clicks the starting seed point of a vessel, and points the mouse to the end of the vessel. The software automatically calculates the best vessel path from the seed point to the mouse position. As the user moves the mouse, the vessel is updated allowing user to control the accuracy of the segmentation with minimal effort. This software is used because it reduces the segmentation time.

The image database has been made available for use by other members of the research community and can be accessed by following the instructions given in [17].

V. Discussion and Future Work

To develop robust image processing methods and use them in medical practice we need high quality databases for software design and evaluation. The method testing must correspond to the strict regulations in the medical treatment and medical research. We proposed a step towards a standardized evaluation of methods for not just detecting visual findings of diabetic retinopathy but for detection of regular fundus image structures. Our database is to our knowledge the first and only database which has both diabetic retinopathy pathologies and major fundus structures annotated. The annotation was done by five experts which is very important because it improves the quality of the segmentation and different segmentation can be used for research on how experts differ among themselves. The size and quality of the database can be improved but DRiDB represents the first complete database that can be used for image processing methods development and evaluation. Future work will include increased number of images in the database and number of experts who will annotate the images.

References

[1] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, "Global prevalence of diabetes: estimates for the year 2000 and projections for 2030," *Diabetes Care*, vol. 27, no. 5, pp. 1047–1053, 2004.

[2] S. J. Lee, C. A. McCarty, H. R. Taylor, and J. E. Keeffe, "Costs of mobile screening for diabetic retinopathy: a practical framework for rural populations," *Aust J Rural Health*, vol. 9, no. 4, pp. 186–192, 2001.

[3] C. A. McCarty, C. W. Lloyd-Smith, S. E. Lee, P. M. Livingston, Y. L. Stanislavsky, and H. R. Taylor, "Use of eye care services by people with diabetes: the Melbourne Visual Impairment Project," *Br J Ophthalmol*, vol. 82, no. 4, pp. 410–414, 1998.

[4] D. A. Askew, L. Crossland, R. S. Ware, S. Begg, P. Cranstoun, P. Mitchell, and C. L. Jackson, "Diabetic retinopathy screening and monitoring of early stage disease in general practice: design and methods," *Contemp Clin Trials*, vol. 33, no. 5, pp. 969–975, 2012.

[5] H. C. Looker, S. O. Nyangoma, D. Cromie, J. A. Olson, G. P. Leese, M. Black, J. Doig, N. Lee, R. S. Lindsay, J. A. McKnight, A. D. Morris, S. Philip, N. Sattar, S. H. Wild, and H. M. Colhoun, "Diabetic retinopathy at diagnosis of type 2 diabetes in Scotland," *Diabetologia*, vol. 55, no. 9, pp. 2335–2342, 2012.

[6] T. Peto and C. Tadros, "Screening for diabetic retinopathy and diabetic macular edema in the United Kingdom," *Curr. Diab. Rep.*, vol. 12, no. 4, pp. 338–345, 2012.

[7] M. Niemeijer, J. J. Staal, M. Ginneken B. v., Loog, and M. D. Abramoff. (2004) DRIVE: digital retinal images for vessel extraction. [Online]. Available: http://www.isi.uu.nl/Research/Databases/DRIVE

[8] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *Medical Imaging, IEEE Transactions on*, vol. 19, no. 3, pp. 203–210, 2000.

[9] (2006) ARIA online, retinal image archive. [Online]. Available: http://www.eyecharity.com/doe/aria_online

[10] T. Kauppi, V. Kalesnykiene, J. K. Kamarainen, L. Lensu, I. Sorri, A. Raninen, R. Voutilainen, H. Uusitalo, H. Kälviäinen, and J. Pietilä, "DIARETDB1 diabetic retinopathy database and evaluation protocol," *Proc Medical Image Understanding and Analysis MIUA*, vol. 1, pp. 3–7, 2007.

[11] (2004) MESSIDOR: Methods for evaluating segmentation and indexing techniques dedicated to retinal ophthalmology. [Online]. Available: http://messidor.crihan.fr/index-en.php

[12] B. Al-Diri, A. Hunter, D. Steel, M. Habib, T. Hudaib, and S. Berry, "REVIEW-a reference data set for retinal vessel profiles," *Engineering in Medicine and Biology Society,30th Annual International Conference of the IEEE*, vol. 2008, pp. 2262–2265, 2008.

[13] M. Niemeijer, "Retinopathy online challenge: Automatic detection of microaneurysms in digital color fundus photographs," *IEEE Trans. Med. Imaging*, vol. 29, p. 185–195, 2009.

[14] (2010) The VICAVR database. [Online]. Available: http://www.varpa.es/vicavr.html

[15] L. Giancardo, F. Meriaudeau, T. P. Karnowski, Y. Li, S. Garg, K. W. Tobin, and E. Chaum, "Exudate-based diabetic macular edema detection in fundus images using publicly available datasets," *Med Image Anal*, vol. 16, no. 1, pp. 216–226, 2012.

[16] M. Poon, G. Hamarneh, and R. Abugharbieh, "Live-vessel: Extending livewire for simultaneous extraction of optimal medial and boundary paths in vascular images," in *Lecture Notes in Computer Science, Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2007, pp. 444–451.

[17] (2013) The Diabetic Retinopathy Image Database. [Online]. Available: http://www.fer.unizg.hr/ipg/resources/image_database