DIABETES TECHNOLOGY & THERAPEUTICS
Volume 10, Number 2, 2008

DOI: 10.1089/dia.2007.0239

Automated Detection of Diabetic Retinopathy: Results of a Screening Study

MANAL BOUHAIMED, M.B.Ch.B., Ph.D., F.R.C.S. (Ophth.),[1]
ROBBIE GIBBINS, M.R.C.G.P., MD.,[2] and DAVID OWENS, C.B.E., M.D., F.R.C.P.[2]

ABSTRACT

Background: This study evaluated the operating characteristics of a reading software (Retinalyze® System, Retinalyze A/S, Hørsholm, Denmark) for automated prescreening of digital fundus images for diabetic retinopathy.

Methods: Digital fundus images of patients with diabetes were retrospectively selected from the Bro Taf diabetic retinopathy screening program in Wales, UK in the period of 2002–2004, which has been superseded by the Diabetic Retinopathy Screening Service for Wales. A gold standard reference was defined by classifying each patient as having or not having diabetic retinopathy based on overall visual grading of the digitized images using the Bro Taf reading protocol. Automated grading was applied using automated red or bright lesion detection at varying detection sensitivities and adjusting for image quality. Operating characteristics included sensitivity, specificity, positive predictive values, and negative predictive values (PPV and NPV, respectively).

Results: Automated analysis of four hundred fundus photographs of 192 eyes from 96 patients with diabetes was performed. The automated red lesion detection had a sensitivity of 82%, specificity of 75%, PPV of 41%, and NPV of 95%. Combined automated red and bright lesion detection yielded a sensitivity of 88%, specificity of 52%, PPV of 28%, and NPV of 95%. Performance of the combined red and bright lesion detection at elevated thresholds in images of good quality demonstrated a sensitivity of 93%, specificity of 78%, PPV of 46%, and NPV of 98%.

Conclusions: Prescreening for diabetic retinopathy by automated detection of single fundus lesions seem to be achieved with minimal false negativity and can help to decrease the burden of manual diabetic retinopathy screening.

INTRODUCTION

THE INCIDENCE OF DIABETES is rising rapidly, and the number of patients with vision-threatening diabetic retinopathy (DR) is increasing globally.[1] In order to thwart the progression of this problem, it has been suggested that screening be done using automated digital techniques that would assist current procedures, which depend solely on manual labor and visual inspection.[2–7] This is particularly relevant when fundus photography services are used to remedy logistic problems in providing qualified biomicroscopic fundus examination to all patients with diabetes.[6] In the present study we examined the ability of an automated digital sys-

[1]Departments of Community Medicine and Surgery, Kuwait University, Kuwait.
[2]Diabetes Research Unit, Cardiff University, Wales, United Kingdom.

tem to determine those who will need biomicroscopic fundus examination in a series of untreated (for retinopathy) patients with diabetes having various grades of retinopathy. The study involved a consecutive series of fundus photographs obtained from patients attending the Bro Taf Area Health Authority Diabetic Retinopathy Screening Service. This Area Health Authority is centered in South Wales with a population of 730,000, of which about 20,000 persons have diabetes mellitus. The screening service is a community-based program designed to screen all persons with diabetes over the age of 12 years who are not currently under the care of the Hospital Eye Service.

The specific objective was to test the ability of this automated system to identify those individuals with no DR in this subpopulation of patients with diabetes who had not previously undergone retinal photocoagulation. In this current study, we examined the performance of automated fundus image analysis of red or bright lesion detection, both individually and combined, in patients with diabetes from the screening population of the Bro Taf program. In addition, automated quantification of image quality was described.

SUBJECTS AND METHODS

This study was retrospective and was approved by the local institutional review board following the tenets of the Declaration of Helsinki. A total of 458 digitally acquired retinal fundus photographs (451 photographs of 45° and seven photographs of 30°) were retrospectively retrieved of 100 patients attending the Bro Taf screening program of South Wales. The patients were selected sequentially in day-clusters from the registered patients in two catchment areas. For each eye 45° fundus photographs of the macular and nasal fields were acquired with Topcon (Tokyo, Japan) digital cameras after installation of 1% tropicamide and 2.5% phenylephrine hydrochloride eye drops. If the photographer deemed it necessary additional 30° photographs were acquired for some patients as well as 45° photographs of other fields. This decision was made by the photographer in cases of poor visualization of the fundus based on media opacity, e.g., cataracts or corneal scarring.

A retinopathy grade and an image quality grade of each eye of each patient according to the Bro Taf protocol (Table 1) were obtained from the clinical database of the center. The Bro Taf Protocol graded retinopathy (DR) as follows: no DR (0), background DR (1), minimal preproliferative DR (2), moderate preproliferative DR (3a), severe preproliferative DR (3b), and proliferative DR (4). Grading was done by trained personnel in accordance with a standardized grading protocol. Each image was evaluated by a team of senior clinician, diabetologist, and ophthalmologist. Hence inter-observer variation was not assessed. There was an ongoing quality assurance program that ensured a sensitivity and specificity for detecting DR in excess of 90% for manual screening.

Table 1. The Bro Taf Protocol for DR Screening

Grade	*Characteristics*
0	No DR
1	Minimal NPDR (≤5 Ma >1 DD fovea and/or 1 Hm >1 DD fovea)
2a	Mild NPDR (>5 Ma and ≤2 Ma <1 DD fovea and/or ≥2 Hm >1 DD fovea and/or Ex outside arcades and/or Ex outside arcades and/or ≤5 CWS; questionable IRMA)
2b	Features of 2a and >2 Ma <1 DD fovea and/or Ex within arcades >1 DD fovea
2c	Circinate Ex within arcades >1 DD fovea
3a	Moderate NPDR (>5 CWS and/or multiple Hm)
3b	Severe NPDR (venous irregularities [beading, loops, reduplication] and definite IRMA)
4	Proliferative DR (NVD on disc and/or NVE and/or preretinal Hm and/or fibrous tissue)
5	Advanced diabetic eye disease (vitreous Hm and/or fibrosis with traction and/or recent retinal detachment)

Characteristics are defined from the initiation of the time of study. Ma, microaneurysms; DD, disc diameter; Hm, hemorrhage; Ex, exudates; CWS, cotton-wool spots; NVD, new vessels on disc; NVE, new vessels elsewhere; IRMA, intraretinal microvascular abnormalities.

Data collection

Once the images were produced the process of evaluation began with manual grading as per the Bro Taf protocol (Table 1). Images were then transferred onto named CD-ROMs, and automated grading was carried out using the Retinalyze® System (Retinalyze A/S, Hørsholm, Denmark) in a manner that was completely independent of manual grading. In order to be completely accurate, in terms of matching the manual and automated grading, additional information in paper form verifying all filenames and links between patients and images was also provided. All patients had at least four images—one nasal and one macular of each eye—except for two patients, where one had a macular image missing, while for the other a nasal image was missing. All images were obtained with a field angle 45° except for five who had additional 30° images.

Data entry and statistical analysis

Manual grading using a well-established protocol (Table 1) was considered the gold standard in detecting DR in this study. Manual per-patient grade was obtained as the maximal Bro Taf grade of the two eyes. The Bro Taf grades 0 (no DR) and 1 (minimal nonproliferative DR [NPDR]) were mapped into one category of not-manifest DR. The Bro Taf grades equal to or above 2a (mild NPDR) were mapped into one category of manifest DR.

Patients were classified as having automatically detected DR if the algorithm identified one or more lesions based on the red lesion detection or the bright lesion detection in any of the images from that patient. The accuracy of the detection was thought to be associated with the quality of the image, and as such differentiation based on good and poor image quality was also studied. The clinical relevance of the detected lesions was characterized by sensitivity (proportion of patients with DR as per Bro Taf protocol who have positive automated results) and specificity (proportion of patients who do not have DR who have negative automated results) along with their 95% confidence intervals (CIs).

After the manual grading, one person entered the clinical data and the link between images and patient IDs into a Microsoft (Redmond, WA) Access 2000 database. Another person visually inspected all images and entered a right or left indicator for each image file into the database. All images were visually inspected a second time to assure that all right and left indicators had been entered correctly. All the clinical data were then checked for consistency with the paper files and corrected if necessary.

Software and automated lesion detection

The automated analysis was performed using a modified version of the lesion detection software Retinalyze version 1.0.6.1, which, at the time of the study, was validated for commercial use. This modified version was more sensitive to large red and bright lesions than previous versions. The automated lesion detection was based on advanced mathematical analysis of the gray-level intensity of the images, where the periphery of potential lesions was established from each of a number of seedpoints. The optic nerve head and the vessel tree were automatically identified and were used to exclude lesion candidates. A measure of visibility was assigned to each potential red and bright lesion, and lesions exceeding user-supplied visibility thresholds were automatically detected and displayed by the system, overlaid on the original image. The visibility thresholds of the red and bright lesion detection controlled the sensitivity (and the resulting specificity) of the automated detection of each lesion type.

Table 2. Performance of "Red Lesion" Detection in Identifying Patients with DR

Automated red lesion detection	*Bro Taf grade*		
	Manifest (DR)	*No (DR)*	*Total*
Positive	14	20	34
Negative	3	59	62
Total	17	79	96

The Bro Taf grade for a patient is obtained as the maximal grade of each of the patient's eyes. Patients with grades 0 or 1 considered as having not manifest DR. Patients with grades equal to or above 2a are considered having manifest DR and should automatically be identified by the system. Sensitivity = 82%, specificity = 75%, positive predictive value = 41%, negative predictive value = 95%.

Table 3. Performance of the Combined Red and Bright Lesion Detection in Identifying Patients with DR

Automated red and bright lesion detection	*Bro Taf grade*		
	Manifest (DR)	*No (DR)*	*Total*
Positive	15	39	53
Negative	2	41	43
Total	17	79	96

Sensitivity = 88%, specificity = 52%, positive predictive value = 28%, negative predictive value = 95%.

Automatically detected red lesions indicated the presence of microaneurysms and/or hemorrhages, while bright lesions pointed to hard exudates and/or cotton-wool spots.

Automated image quality detection

The digital image quality was quantified by a contrast quality measure. This quality measure determined whether the image was suited for automated image analysis and may occasionally differ from human visual evaluation of image quality. The quality measure determined the level of visibility of details in a given image. Poor image quality may appear because of technical failures such as insufficient illumination of the retina or because of pathologies such as cataract. A single quality threshold controlled the level at which an image was automatically deemed of unacceptable quality. A prescreening strategy was applied for the automated detection in which all images of patients with a single image below the quality threshold automatically were referred for human grading.

Initial parameter setting

The following parameter settings were used for the images:

1. The visibility threshold of the red lesion detection was 1.4.
2. The visibility threshold of the bright lesion detection was 2.4.
3. The image quality threshold was 0.58.
4. For the images of angle 45° an expected disk diameter of 77 was used.
5. For the images assumed to be of angle 30° an expected disk diameter of 144 was used.

RESULTS

A total of 100 patients (200 eyes) were photographed, and their DR status was assessed by visual grading of images selected. Four patients were excluded from the analysis: two had previous photocoagulation, and two others had images missing. The analyses were thus performed on 96 patients, first with the detection of red lesions and then the combination of red and bright lesions stratified by image quality and threshold setting.

The automated red lesion detection had a sensitivity for detecting eyes with DR of 82% (95% CI 57–96%) and a specificity for detecting eyes without DR of 75% (95% CI 64–84%). Three patients were false-negative; all three patients had Bro Taf grades equal to 2a indicating mild NPDR. A total of 20 false-positive patients (Bro Taf grades 0 and 1) were detected by the red lesion detection (Table 2).

When combining the automated red and bright lesion detection the specificity decreased to 52% (95% CI 40–63%), but the sensitivity increased to 88% (95% CI 64–99%). Hence only two patients were now false-negative. However, 18 more (38 patients in total) were now false-positive. In the two false-negative patients, no red or bright lesions were detected, but they had Bro Taf grades 2a indicating mild NPDR (Table 3).

A further analysis was carried out (using red and bright lesion detection) after exclusion of 15 patients in whom at least one image was below the image-quality threshold. Specificity decreased to 51% (95% CI 38–63%), and sensitiv-

Table 4. Performance of the Combined Red and Bright Lesion Detection in Identifying Patients with DR in Images of Good Automatically Evaluated Quality

Automated red and bright lesion detection with good image quality	*Bro Taf grade*		
	Manifest (DR)	*No (DR)*	*Total*
Positive	13	33	46
Negative	1	34	35
Total	14	67	81

Patients with at least one image of poor quality were excluded from the analysis. Sensitivity = 93%, specificity = 51%, positive predictive value = 28%, negative predictive value = 97%.

Table 5. Distribution of Patients with at Least One Image with Automatically Detected Low Image Quality

	Number of patients				
	Automated red and bright lesion detection, low image quality				
Presence of DR, Bro-Taf grade	Negative		Positive		Total
No/minimal DR					
0	6	7 (58%)	2	5 (42%)	12
1	1		3		
DR					
2a	1		0		
2b	0	1 (33%)	1	2 (67%)	3
3a	0		1		
3b	0		0		
Total	8		7		15

ity increased to 93% (95% CI 66–99%) (Table 4). Now there was only one false-negative that too was of DR grade 2a. It was verified by visual inspection that all the 15 patients were correctly referred for human evaluation of the images; the patients had either dense cataracts or at least one image of poor technical quality. The distribution of the excluded patients with respect to Bro Taf grades is presented in Table 5.

In this study the red and bright visibility thresholds were increased to around 2.0 and 3.1, respectively, without compromising the sensitivity (Table 6). In this situation, the number of false-positive patients further decreased by 18 patients (to 15) (specificity of 78% and 95% CI 66–87%). There was no associated increase in number of false-negative patients (sensitivity of 93%, 95% CI 66–100%). Thus, Table 6 and Table 4 are directly comparable. The difference between the two tables was that Table 4 was based on a visibility threshold setting (used in Tables 2–5) giving good concordance between human visual impression and the automated lesion detection, whereas the threshold setting of Table 6 was optimized for screening, i.e., while maintaining the sensitivity of Table 4 the specificity of the combined red and bright lesion detection was maximized. Switching between Table 4 and Table 6 was solely a matter of choosing between two sets of user-supplied threshold values.

DISCUSSION

The results of this study were obtained using a specific automated image analysis system, which is no longer available because of a commercial decision. This decision had no impact on this work. We documented that digital image analysis can achieve considerable effectiveness in identifying patients without the need for DR

Table 6. Performance of Combined Red and Bright Lesion Detection at the Elevated Thresholds 2.0 (Red) and 3.1 (Bright)

	Bro-Taf grade		
Automated red and bright lesion detection at elevated image thresholds	Manifest (DR)	No (DR)	Total
Positive	13	15	28
Negative	1	52	53
Total	14	67	81

Patients with at least one image of poor quality are excluded from the analysis. Sensitivity = 93%, specificity = 78%, positive predictive value = 46%, negative predictive value = 98%.

screening from a pool of patients with possible retinopathy, particularly if the image is of good quality. Data published by Larsen et al.[7] reported similar considerable effectiveness, with automated red lesion detection demonstrating a specificity of 71% and a sensitivity of 97% in detecting DR. Although the sensitivity in our study falls below this, our work indicates that after automated screening, only between 7% and 18% of retinopathy will be missed.

Missing retinopathy is an important issue. However, as can be seen from Table 2, when classifying patients using only the automated red lesion detection, three false-negative patients appeared—all of whom had mild grade 2a NPDR. One of these was detected by the bright lesions detection, while another patient was detected by the automated quality measure. Our results suggest that the latter group will have very mild retinopathy that can be picked up on a future screen, and as such the potential risk of missing a diagnosis of such levels of retinopathy is not great. This is particularly true because the risk of progression from mild to treatment requiring retinopathy is slow with the mean duration of mild NPDR estimated to be about 4 years.[8]

Although the specificity of automated screening was low, the burden on the clinic of manual screening can clearly be reduced by such an automated system. An automated prescreening system that automatically categorizes the fundus of patients with diabetes into those who should have their photographs visually graded for retinopathy and those who need not have formal screening should be able to accomplish at least two functions. First is the identification of images that are unsuitable for automated lesion detection (i.e., low-quality images caused by refractive media opacity or technical failures in the photographic process). Photographs with such characteristics should be automatically designated to visual inspection. Second is the identification of normal or near-normal fundi that need not be manually graded. Additional enhancement appears to be possible by incorporating good quality images and combining this with automated detection of red and bright fundus lesions. At the enhanced threshold setting of a combined red and bright lesion and detection in a good quality image, 15 of 67 patients (22%) without retinopathy were incorrectly classified as requiring manual grading by the system. This decreases the burden on manual screening by at least two-thirds.

Our findings should be interpreted in the light of the fact that in clinical practice, DR screening concerns individual patients rather than single eyes. However, the decision to refer the patient for ophthalmic examination or treatment is made even if one eye has reached or exceeded a threshold level of retinopathy. Therefore we decided to evaluate patients using the eye most involved rather than individual eyes. Since this study is evaluating the performance of an automated screening system, which is based on single eye evaluations, the validity of this protocol should not be different even with binocular evaluations. Also, the effects of potentially confounding pathologic posterior segment changes such as drusens, chorioretinitis patches, etc., were not investigated in the present study because of the limiting sample size. However, previous work by Larsen et al.[7] suggests that the algorithm for red lesion detection for the same automated system was virtually insensitive to bright lesions of any underlying pathology. It was, however, sensitive to microaneurysms and hemorrhages, regardless of whether these are caused by DR or other types of retinal diseases that produce such microangiopathy.

In conclusion, this study advocates the detailed assessment of retinopathy at or above a minimum retinopathy threshold. Such a task of making a fine distinction between the presence and absence of any retinopathy is rarely addressed outside epidemiologic studies since in clinical practice we deal essentially with treatment thresholds. However, populations that undergo systematic fundus photographic screening have been found to have a favorable outcome on improving visual prognosis. Consequently, the task of automatically assisting the separation of patients with retinopathy from patients without retinopathy is a clinically meaningful one. This was described by other studies,[4,5] where it has been demonstrated that automated image analysis, which detects the presence or absence of retinopathy, can be used as a first-step screening tool.

ACKNOWLEDGMENTS

We acknowledge Dr. Lukman Thalib (Biostatistics) and Dr. Suhail Doi (Endocrinology and Clinical Epidemiology) from the Faculty of Medicine at Kuwait University for their critical comments. We also acknowledge the Retinalyze group (Jannik Godt, Maja Jensen, and Niels Hartvig) for giving us the benefit of assessing their system during this project.

REFERENCES

1. Klein R, Klein BE, Moss SE: How many steps of progression of diabetic retinopathy are meaningful? The Wisconsin epidemiologic study of diabetic retinopathy. Arch Ophthalmol 2001;119:547–553.
2. Jelinek HJ, Cree MJ, Worsley D, Luckie A, Nixon P: An automated microaneurysm detector as a tool for identification of diabetic retinopathy in rural optometric practice. Clin Exp Optom 2006;89:299–305.
3. Fleming AD, Philip S, Goatman KA, Olson JA, Sharp PF: Automated assessment of diabetic retinal image quality based on clarity and field definition. Invest Ophthalmol Vis Sci 2006;47:1120–1125.
4. Hansen AB, Hartvig NV, Jensen MS, Borch-Johnsen K, Lund-Andersen H, Larsen M: Diabetic retinopathy screening using digital non-mydriatic fundus photography and automated image analysis. Acta Ophthalmol Scand 2004;82:666–672.
5. Usher D, Dumskyj M, Himaga M, Williamson TH, Nussey S, Boyce J: Automated detection of diabetic retinopathy in digital retinal images: a tool for diabetic retinopathy screening. Diabet Med 2004;21:84–90.
6. Larsen N, Godt J, Grunkin M, Lund-Andersen H, Larsen M: Automated detection of diabetic retinopathy in a fundus photographic screening population. Invest Ophthalmol Vis Sci 2003;44:767–771.
7. Larsen M, Godt J, Larsen N, Lund-Andersen H, Sjølie AK, Agardh E, Kalm H, Grunkin M, Owens DR: Automated detection of fundus photographic red lesions in diabetic retinopathy. Invest Ophthalmol Vis Sci 2003;44:761–766.
8. Tung TH, Chen SJ, Shih HC, Chou P, Li AF, Shyong MP, Lee FL, Liu JH: Assessing the natural course of diabetic retinopathy: a population-based study in Kinmen, Taiwan. Ophthalmic Epidemiol 2006;13:327–333.

Address reprint requests to:
Manal Bouhaimed, M.B.Ch.B., Ph.D., F.R.C.S. (Ophth.)
Departments of Community Medicine and Surgery
Faculty of Medicine
Kuwait University
P.O. Box 24923 SAFAT
State of Kuwait 13110

E-mail: manal_q8@hsc.edu.kw

manalbouhaimed@gmail.com

This article has been cited by:

1. Anton M. Kolomeyer, Bernard C. Szirth, Khadija S. Shahid, Gina Pelaez, Natasha V. Nayak, Albert S. Khouri. 2013. Software-Assisted Analysis During Ocular Health Screening. *Telemedicine and e-Health* **19**:1, 2-6. [Abstract] [Full Text HTML] [Full Text PDF] [Full Text PDF with Links]
2. Zsolt Torok, Tunde Peto, Eva Csosz, Edit Tukacs, Agnes Molnar, Zsuzsanna Maros-Szabo, Andras Berta, Jozsef Tozser, Andras Hajdu, Valeria Nagy, Balint Domokos, Adrienne Csutak. 2013. Tear fluid proteomics multimarkers for diabetic retinopathy screening. *BMC Ophthalmology* **13**:1, 40. [CrossRef]
3. Arulmozhivarman Pachiyappan, Undurti N Das, Tatavarti VSP Murthy, Rao Tatavarti. 2012. Automated diagnosis of diabetic retinopathy and glaucoma using fundus and OCT images. *Lipids in Health and Disease* **11**:1, 73. [CrossRef]
4. Helen K. Li, Mark Horton, Sven-Erik Bursell, Jerry Cavallerano, Ingrid Zimmer-Galler, Mathew Tennant, Michael Abramoff, Edward Chaum, Debra Cabrera DeBuc, Tom Leonard-Martin, Marc Winchester. 2011. Telehealth Practice Recommendations for Diabetic Retinopathy, Second Edition. *Telemedicine and e-Health* **17**:10, 814-837. [Abstract] [Full Text HTML] [Full Text PDF] [Full Text PDF with Links]
5. Jang Pyo Bae, Kwang Gi Kim, Ho Chul Kang, Chang Bu Jeong, Kyu Hyung Park, Jeong-Min Hwang. 2011. A Study on Hemorrhage Detection Using Hybrid Method in Fundus Images. *Journal of Digital Imaging* **24**:3, 394-404. [CrossRef]
6. B. Dupas, T. Walter, A. Erginay, R. Ordonez, N. Deb-Joardar, P. Gain, J.-C. Klein, P. Massin. 2010. Evaluation of automated fundus photograph analysis algorithms for detecting microaneurysms, haemorrhages and exudates, and of a computer-assisted diagnostic system for grading diabetic retinopathy. *Diabetes & Metabolism* **36**:3, 213-220. [CrossRef]
7. Lars BäcklundFinding a Role for Computer-Aided Early Diagnosis of Diabetic Retinopathy . [CrossRef]
8. Rajiv B. Khandekar, Subramanium Tirumurthy, Saleh Al-Harby, N.S.D. Moorthy, Ikram Amir. 2009. Diabetic Retinopathy and Ocular Co-Morbidities Among Persons with Diabetes at Sumail Hospital of Oman. *Diabetes Technology & Therapeutics* **11**:10, 675-679. [Abstract] [Full Text PDF] [Full Text PDF with Links]
9. Paolo S. Silva, Jerry D. Cavallerano, Lloyd M. Aiello. 2009. Ocular telehealth initiatives in diabetic retinopathy. *Current Diabetes Reports* **9**:4, 265-271. [CrossRef]
10. 2008. Current literature in diabetes. *Diabetes/Metabolism Research and Reviews* **24**:7, i-ix. [CrossRef]