Evaluation of a System for Automatic Detection of Diabetic Retinopathy From Color Fundus Photographs in a Large Population of Patients With Diabetes

Michael D. Abràmoff, md, phd[1,2,3]
Meindert Niemeijer, phd[3,4]
Maria S.A. Suttorp-Schulten, md, phd[5]

Max A. Viergever, phd[4]
Stephen R. Russell, md[1,2]
Bram van Ginneken, phd[4]

OBJECTIVE — To evaluate the performance of a system for automated detection of diabetic retinopathy in digital retinal photographs, built from published algorithms, in a large, representative, screening population.

RESEARCH DESIGN AND METHODS — We conducted a retrospective analysis of 10,000 consecutive patient visits, specifically exams (four retinal photographs, two left and two right) from 5,692 unique patients from the EyeCheck diabetic retinopathy screening project imaged with three types of cameras at 10 centers. Inclusion criteria included no previous diagnosis of diabetic retinopathy, no previous visit to ophthalmologist for dilated eye exam, and both eyes photographed. One of three retinal specialists evaluated each exam as unacceptable quality, no referable retinopathy, or referable retinopathy. We then selected exams with sufficient image quality and determined presence or absence of referable retinopathy. Outcome measures included area under the receiver operating characteristic curve (number needed to miss one case [NNM]) and type of false negative.

RESULTS — Total area under the receiver operating characteristic curve was 0.84, and NNM was 80 at a sensitivity of 0.84 and a specificity of 0.64. At this point, 7,689 of 10,000 exams had sufficient image quality, 4,648 of 7,689 (60%) were true negatives, 59 of 7,689 (0.8%) were false negatives, 319 of 7,689 (4%) were true positives, and 2,581 of 7,689 (33%) were false positives. Twenty-seven percent of false negatives contained large hemorrhages and/or neovascularizations.

CONCLUSIONS — Automated detection of diabetic retinopathy using published algorithms cannot yet be recommended for clinical practice. However, performance is such that evaluation on validated, publicly available datasets should be pursued. If algorithms can be improved, such a system may in the future lead to improved prevention of blindness and vision loss in patients with diabetes.

Diabetes Care **31:193–198, 2008**

Diabetic retinopathy blinds ~25,000 patients with diabetes annually in the U.S. alone and is the main cause of blindness in the U.S. and Europe working-age populations (1). Almost 50% of the 18 million patients with diabetes in the U.S do not undergo any form of regular documented dilated eye exam (2). This is in spite of overwhelming scientific evidence that this, if combined with appropriate management, can prevent up to 95% of cases of vision loss and blindness (3–10) and also in spite of guidelines by the American Diabetes Association and the American Academy of Ophthalmology that advise an annual dilated eye exam for most patients with diabetes (11). Digital photography of the retina examined by ophthalmologists or other qualified readers has been shown to have sensitivity and specificity comparable with or better than indirect ophthalmoscopy by an ophthalmologist (12,13) and has been proposed as an approach to make the dilated eye exam available to underserved populations that do not receive regular exams by ophthalmologists. If all of these populations were to be served with digital imaging, the number of retinal images to be evaluated annually is 32 million (~50% of patients with diabetes and at least two photographs per eye) (13,14).

Over the last few years, we and others have developed and published computer algorithms that can aid the ophthalmologist with the evaluation of digital fundus photographs for early forms of diabetic retinopathy in a community (screening) population (15–27). Our studies on small numbers of patients show that such algorithms have the capacity to perform comparably with retinal specialists and outperform most other published algorithms on limited tasks including:

- correct localization of the optic disc in 999 of 1,000 retinal images (28)
- segmentation of retinal vessels at an accuracy of 94.2% in 20 retinal images (29)
- detection of hemorrhages, microaneu-

rysms, and vascular abnormalities with sensitivity of 100% and specificity of 87% in 100 retinal images (30)

- detection and separation of exudates, cotton-wool spots, and drusen with a lesion sensitivity of 95% and specificity of 88% in 300 retinal images (31)
- detection of retinal images with insufficient image quality with an accuracy of 97.4% in 1,000 retinal images (32).

Though these algorithms were targeted to these narrowly focused tasks, they can potentially be combined into a complete system for the detection of diabetic retinopathy in a screening setting, meaning in a population in which the incidence of newly diagnosed diabetic retinopathy is <30% (14,33–35). No studies have examined the performance of a combined system, built from published algorithms, for the detection of diabetic retinopathy on a large group of patients from a true screening population, i.e., with a high proportion, of up to 90%, of normal appearing fundi and a low proportion, of up to 30%, of fundoscopically detectable diabetic retinopathy. Typical screening projects consist of multiple sites and often use multiple types of digital fundus cameras, with camera-operators with varying levels of experience, resulting in retinal images that exhibit considerable variation in size, resolution, and image quality.

The gold standard for the evaluation of diabetic retinopathy is seven-field stereo fundus photography read by trained readers in accordance with the Early Treatment Diabetic Retinopathy Study (ETDRS) standard (36), and ideally the performance of a complete system would be compared with the evaluation of ETDRS stereo slides taken from the same patients on the same day. However, the expense of such a study with seven-field stereo photography by trained photographers and evaluation by ETDRS readers on a sufficiently large number of patients can only be warranted if preliminary studies show that such a research effort may be worthwhile.

The EyeCheck project for online diabetic retinopathy detection in the Netherlands uses so-called "nonmydriatic" fundus cameras. Over 20,000 exams have been performed to date, with two photos per eye, on patients not known to have diabetic retinopathy and read by a single ophthalmologist (of three participating ophthalmologists) according to a strict protocol; prevalence of referable diabetic retinopathy ranges between 5–10% in this population (33).

The present study was designed as a preliminary study to determine how well the performance of a combination of published algorithms for automated detection of diabetic retinopathy holds up in comparison with the clinical evaluation of a single retinal specialist on the same 10,000 retinal exams, consisting of 40,000 retinal images obtained with a variety of different retinal cameras from the EyeCheck project.

RESEARCH DESIGN AND METHODS

In this retrospective study, we selected 10,000 consecutive patient visits from 5,692 unique patients. These visits consisted of four retinal photographs (two left, two right eye), as well as textual information about age, A1C, sex, and presence of risk factors over the period 2003–2005, from the EyeCheck project, which has been described elsewhere (33). Patients are photographed annually in community health centers according to an imaging protocol described below. The study protocol was approved by the institutional review board of the University of Iowa. Because of the retrospective nature of the study and because full anonymity was maintained throughout the study, informed consent was judged not to be necessary by the review board.

Inclusion/exclusion criteria
Patients were included if they had a diagnosis of diabetes according to World Health Organization criteria and were aged 18 years and over. Patients were excluded if they had a diagnosis of diabetic retinopathy, had a previous visit to an ophthalmologist for a dilated retinal eye exam, had only one eye photographed, had less than two photographs per eye, or had no demographic data available. Once a patient is evaluated as having referable retinopathy, he or she cannot be rephotographed. Thus, the exams were always of patients either up until then known not to have referable diabetic retinopathy based on a previous exam or until then not known to have diabetic retinopathy.

Imaging protocol
Patients were photographed with "nonmydriatic" digital retinal cameras by trained technicians, at 10 different sites, using either the Topcon NW 100, the Topcon NW 200 (Topcon, Tokyo, Japan), or the Canon CR5–45NM (Canon, Tokyo, Japan) nonmydriatic cameras. Across sites, four different camera settings were used: *1*) 640 × 480 = 0.3 megapixels (Mp) and 45° field of view (fov), *2*) 768 × 576 = 0.4 Mp and 35° fov, *3*) 1,792 × 1,184 = 2.1 Mp and 35° fov, and *4*) 2,048 × 1,536 = 3.1 Mp and 35° fov; all images were JPEG compressed at the minimal compression setting available. For each exam, four images were acquired, two of each eye, one centered on the fovea and one centered on the optic disc. The exam, date of birth, duration of diabetes, and A1C status were transmitted over the internet to the Web site (www.eyecheck.nl) and evaluated by a single out of three ophthalmologists, based on the International Clinical Diabetic Retinopathy Disease Severity Scale, as either "no referable retinopathy" or "referable retinopathy;" sufficient image quality was also evaluated (37). This evaluation was documented online and was immediately available to the primary care physician and patient at the community health center. Patients were recommended for follow-up for imaging for 1 or 2 years, depending on risk factors not available for this study. For this study, the following were available from each exam: the anonymized retinal images, the age of the patient when the photographs were taken, an anonymized sequential unique site identification number, an anonymized unique patient identification number, and an aggregate human expert evaluation of the four images in each exam for: unacceptable quality; no referable retinopathy or no apparent diabetic retinopathy; or referable retinopathy, at least mild nonproliferative retinopathy.

System of computer algorithms for evaluation of retinal images for diabetic retinopathy
After automatic cropping of the black border, all retinal images were automatically resampled to 640 × 640 pixels. Then, the following sequence of image operations was performed on each image:

- Automatic determination of the probability that the image has a sufficient quality for evaluation, based on the presence of particular image structures with a particular spatial distribution, as described previously (32). After preliminary testing, the threshold for insufficient image quality was set so that ~80% of exams judged by the human expert to have unacceptable quality were rejected, while 20% of exams
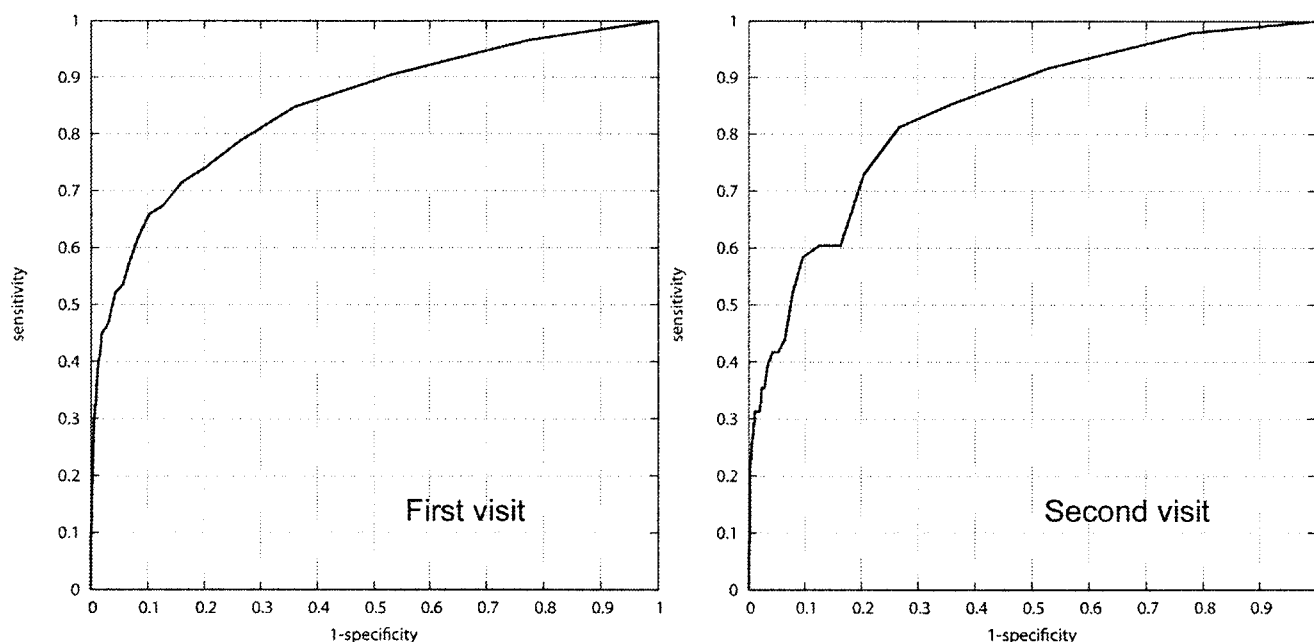
Figure 1—*ROC curve for algorithm versus human expert on 4,739 first visits (A) and 1,701 second visits (B). y-axis, sensitivity of the system at different thresholds; x-axis, 1-specificity. Area under the curve is 0.85 (first visit) and 0.84 (second visit).*

judged by the human expert to have acceptable image quality were rejected by the system as having unacceptable quality.

- Automatic segmentation of the vessels using pixel feature classification as we have described previously, resulting in a vessel probability image, which is necessary to exclude false-positive "red lesions" (29,38).
- Automatic detection and masking of the optic disc using a method that determines where blood vessels converge, as we have described previously (28).
- Automatic detection of red lesions—microaneurysms, hemorrhages, or vascular abnormalities—using feature classification, as we have described previously, resulting in a red lesion probability map (30).
- Automatic detection of "bright lesions"—exudates, cotton-wool spots, and drusen—using pixel feature classification, resulting in a combined probability map for exudates, cotton-wools spots, and drusen (the differentiation between drusen and the other two is not used by the combined system we present here), as described previously (31).

We have used these machine learning algorithms as is, in other words, no retraining of the algorithms was performed for this study. The total runtime of the complete system for a single image on a standard Windows PC was 7 min, and one exam thus took $4 \times 7 = 28$ min. To quickly evaluate the complete set of 10,000 exams, we implemented our system on a cluster of 30 PCs with an automatic job distribution system.

Outcome parameters and data analysis

Several features such as the number and probability of the lesions in all four images of an exam as detected by the red and bright lesion detection algorithms were used to create the per-exam retinopathy probability. Each of the four images, from both eyes, were therefore considered independently. Combined with the output of the image level image quality algorithm, two probabilities were thus created at the exam level, namely the probability that an exam has sufficient quality for evaluation and the probability that that exam shows referable retinopathy. The image quality threshold is a number that determines when an image is ungradable. By varying the thresholds for red and bright lesions, the sensitivity and specificity of the system compared with the clinician's reading can be made to vary. Sensitivity and specificity of the complete system relative to the clinical evaluation are calculated at each threshold setting. These sensitivity/specificity pairs are used to create receiver operator characteristic (ROC) curves, showing the sensitivity and specificity at various thresholds. The area under the ROC curve is regarded as a comprehensive measure of system performance: an area of 1.0 has sensitivity = specificity = 1 and represents perfect detection, while an area of 0.5 is the performance of a system that essentially performs a coin toss on average. From the ROC curve, the "optimal threshold" is the subjective point at which we deemed sensitivity and sensitivity to be optimal for a screening setting.

In addition, we defined the number needed to miss (NNM) [= 1/(false negatives/all negatives)] as the number of normal exams after which, on average, one case of referable retinopathy will be missed for a specific threshold setting. The effect of camera type and administration of pharmacologic dilation, if any, on system performance was evaluated by determining the sensitivity and specificity of the algorithm at the optimal threshold.

RESULTS— Of the 5,692 patients included, 3,107 had one exam, 1,601 had two exams, and 986 had three or more exams over the time period. All included patients had two photographs of each eye taken (see inclusion criteria). The average age of these 5,692 patients at the time of their first exam was 61.8 years (±1 SD 12.81 years); 3,833 (49.85%) were male, and 3,856 (50.15%) were female. Of 10,000 exams, 9,017 (90%) were evaluated by an ophthalmologist as having no apparent retinopathy, 498 (4.98%) as

having at least minimal nonproliferative retinopathy, and 485 (4.85%) as having insufficient image quality on one or more images.

The system detected 2,311 of 10,000 exams as not having sufficient quality, and 7,689 of 10,000 exams were determined by the system to have sufficient image quality for all images in the exam. The 7,689 exams were from 4,739 unique patients, average age 59.9 ± 11.7 years. Of the 7,689 exams with system-determined sufficient image quality on all four images, 7,229 (94%) were evaluated by the human expert as having no referable retinopathy, 378 (4.92%) as having referable retinopathy, and 81 (1.05%) as having insufficient image quality on one or more images. As mentioned in the RESEARCH DESIGN AND METHODS section, the quality threshold for the automatic system was set so that 80% of the 9,515 exams that were thought by humans to have sufficient quality were allowed to pass.

We performed a limited estimate of the reading by single human expert by selecting a random set of 500 from the 7,689 exams. The exams in this set were read independently by three masked retinal specialists. Compared with the original reading, their sensitivity/specificity (referable retinopathy or not) was 0.85/0.89, 0.73/0.89, and 0.62/0.84.

If the human expert was taken as the reference standard or "truth," the area under the ROC curve was 0.84 for all 7,689 sufficient image quality exams, 0.85 for all 4,739 patients on their first visit, and 0.84 for the 1,701 patients with a second visit (Fig. 1). The optimal threshold had a sensitivity of 0.84 and a specificity of 0.64. At this point, 4,648 of 7,689 exams (60%) were correctly identified by the system as having no referable retinopathy, 59 of 7,689 exams (0.8%) did have referable retinopathy but were missed by the system, 319 of 7,689 exams (4%) were correctly identified as having referable retinopathy (Fig. 2*A* and *B*), and 2,581 of 7,689 exams (33%) were false-positives, i.e., the system estimated them as having diabetic retinopathy, while the human expert did not. The number needed to falsely miss one case (NNM) of referable diabetic retinopathy, at this point, was 80, meaning that for every 80 exams, one case of referable retinopathy is missed at this chosen "optimal" threshold. Figure 2*C* shows an example of such a false negative, missed exam.

If the sensitivity of the system was increased, by adjusting the threshold, the
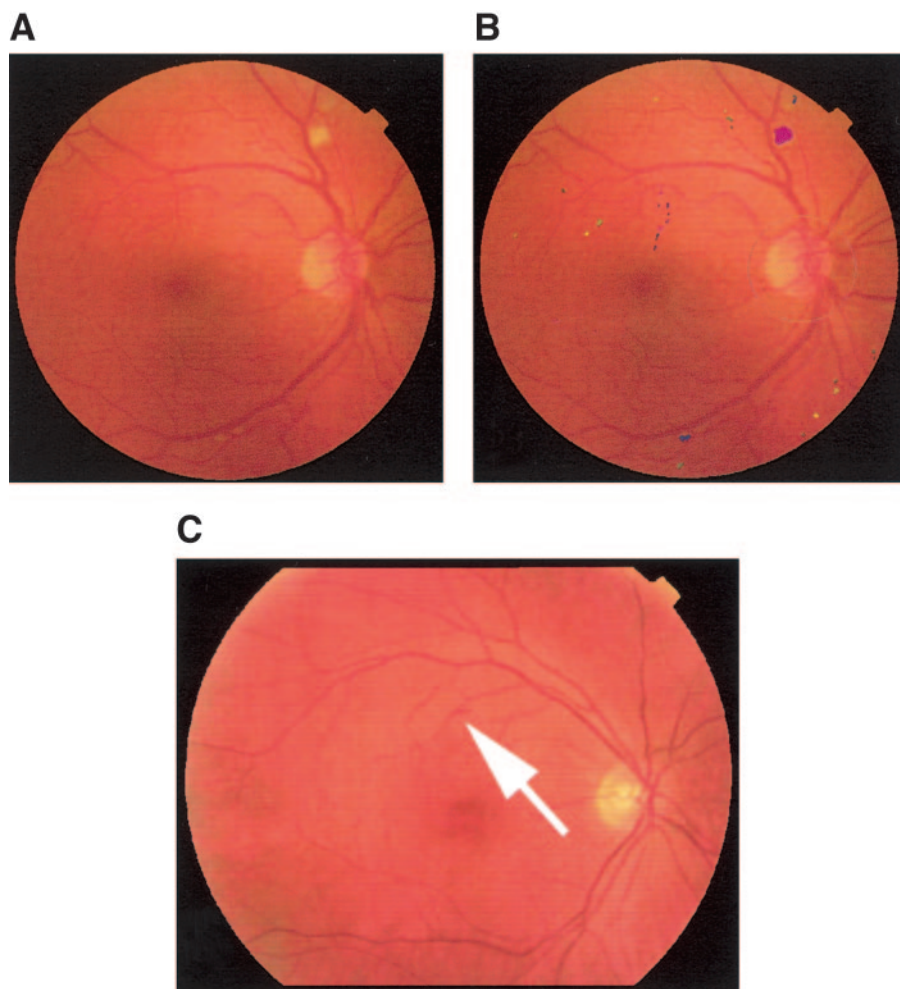

Figure 2— A: *Example of one image from an exam where both expert and automated system agree on the presence referable retinopathy.* B *shows where the automated system detected exudates and cotton wool spots (shades of red and purple) and hemorrhages (shades of yellow-green) and the location of the optic disc (white circle).* C: *Example of false negative exam showing evidence of diabetic retinopathy (large hemorrhage attached to artery, indicated with white arrow), detected by the human expert but missed by the system.*

NMM could be increased up to a maximum of 127, at which point the system had a (very low) specificity of 0.22.

We performed a preliminary, qualitative analysis of a sample of 87 false negative exams obtained at a specific point in the ROC curve to understand where the system was failing and the potential for improvement. Of these 87 false negative exams, 24 of 87 (27%) contained large hemorrhages and/or neovascularizations of the disc or elsewhere, 23 contained one or a few small hemorrhages, 18 contained exudates or cottonwool spots only, and 22 exams contained other lesions that the system was not designed to detect, including nevi and retinal scars, that were not deemed to be diabetic retinopathy by a second expert.

The effect of camera type and pharmacologic dilation on system perfor-

mance at the optimal threshold was as follows: in the subgroup that was pharmacologically dilated, sensitivity/specificity was 0.76/0.39, while in the group that was not pharmacologically dilated, sensitivity/specificity was 0.80/0.43. Across camera types, they were camera *1*) 0.86/0.22, *2*) 0.88/0.56, *3*) 0.79/0.83, *4*) 0.78/0.73, respectively.

CONCLUSIONS — To our knowledge, this is the first time that an automated diabetic retinopathy screening system, based exclusively on previously published algorithms, has been tested on an unselected set of exams of this size obtained from a representative diabetic population. The data used in this study were collected over a 3-year period and represent real, unselected screening data obtained in a diabetic retinopathy screening

program. The area under the curve of 0.85 on the first visit and an optimal sensitivity of 0.84 and specificity of 0.64 show the potential of such a system. That the system is relatively stable is shown by the fact that the area under the curve for the second visit is 0.84. The limited validation on a sample of 500 exams shows that the original single human expert reading corresponds quite closely to the independent readings by three masked retinal specialists. However, because independent expert readings on the full dataset are not available, it is impossible to determine how the system compares with human readers on this aspect.

However, the system cannot yet be recommended for clinical practice because the NNM of diabetic retinopathy was 80 at the optimal threshold and was never higher than 127. Another more important reason is that of the 87 false negatives at the optimal threshold, ~25% were isolated neovascularizations or large hemorrhages not accompanied by any other lesion such as exudates or microaneurysms, and such isolated lesions still require urgent referral to an ophthalmologist. Therefore, there is a compelling need for algorithm improvement, including better detection of neovascularizations, for which no specific algorithm has yet been published. In comparison with these results, Usher et al. (39) previously tested a complete system on a set of 733 patients with a high prevalence of diabetic retinopathy of 38% and found a maximum sensitivity of 95.1% and a specificity of 46.3%. Larsen et al. (40) tested a previously commercially available system on 100 patients, 63% with diabetic retinopathy, and found a sensitivity of 96.7% and a specificity of 71.4%. These smaller test populations, with a higher prevalence of diabetic retinopathy, should not be considered representative of a true, large scale screening population, in which an automated system is expected to be deployed.

The results do show that performance is close enough to that of human experts to warrant additional studies on high-quality, validated datasets such as that of the ETDRS (41). Additionally, the results show that reporting only sensitivity and specificity or area under the ROC curve may not be sufficient to evaluate a system for screening for diabetic retinopathy, as our preliminary analysis of false negatives shows. Finally, the results show that a measured performance comparable with human experts on limited tasks, such as

red lesion detections, does not translate directly to comparable results on much larger, unselected datasets. This caveat, to our knowledge, also applies to all other published diabetic retinopathy lesion detection algorithms. There is a large difference between the rate of exams with images of insufficient quality in this study (4.85%) and in most previous studies, including our own study, where we reported a rate of ungradeable photographs of 12% (33). The EyeCheck population is a partially closed population and not a cohort. Patients, once photographed, remain in the population until they either fail to show up for their repeat exam or are referred because of insufficient image quality or referable abnormalities including diabetic retinopathy. Thus, over time, patients with initially insufficient image quality leave the population, while patients in the population accrue insufficient image quality much more slowly, for example, from cataract formation. Patients that had preexisting abnormalities leave the population at their first exam, while the accrual of diabetic retinopathy is much slower. On the other hand, new screening sites are added almost monthly, so there is a continuous influx of new patients. We expect that the observed change in rates of insufficient quality exams reflects the above balance. There seemed to be a slight effect of increasing camera resolution resulting in improved system performance, meriting additional study.

There are several important limitations with this study. Most importantly, the estimate by the system was compared with a single reading by a human expert, and because of the nature of the study could not be compared with the gold standard seven-field stereo photography read by experts (41). It is therefore impossible to state whether the performance of the system would have been different if compared with the gold standard.

The automated system is also slow, taking 28 min per patient (four images). This was a nonoptimized version running in debug mode, and meanwhile we have more than halved this time. For screening projects online, response time can be increased simply by running the system on multiple PCs. However, for real-time diagnosis directly in the camera, an option we currently see as less attractive, this is likely to still be too long.

In summary, this study indicates that automated detection of diabetic retinopathy using a combination of published al-

gorithms cannot yet be recommended for clinical practice, based on this test on a true population of patients screened for diabetic retinopathy. On the other hand, the performance is such that evaluation on validated, publicly available datasets should be pursued. Individual algorithms for lesion detection, as well as the manner in which lesion detection algorithms are combined, merit additional research before they can be considered for automated detection of diabetic retinopathy in patients with diabetes, where they may aid in the prevention of blindness and vision loss in these patients.

References

1. Fong DS, Aiello L, Gardner TW, King GL, Blankenship G, Cavallerano JD, Ferris FL II, Klein R: Diabetic retinopathy. *Diabetes Care* 26:226–229, 2003
2. Centers for Disease Control and Prevention: Data & Trends: National Diabetes Surveillance System: Preventive care practices, 1994–2004. Available from http://www.cdc.gov/diabetes/statistics/preventive/tX.htm. Accessed 8 July 2007
3. Fong DS, Aiello LP, Ferris FL III, Klein R: Diabetic retinopathy. *Diabetes Care* 27: 2540–2553, 2004
4. Fong DS, Aiello L, Gardner TW, King GL, Blankenship G, Cavallerano JD, Ferris FL III, Klein R: Retinopathy in diabetes. *Diabetes Care* 27 (Suppl. 1):S84–S87, 2004
5. Chia DS, Yap EY: Comparison of the effectiveness of detecting diabetic eye disease: diabetic retinal photography versus ophthalmic consultation. *Singapore Med J* 45:276–279, 2004
6. Early Treatment Diabetic Retinopathy Study Research Group: Early photocoagulation for diabetic retinopathy: ETDRS report 9. *Ophthalmology* 98:766–785, 1991
7. Aiello LM, Bursell SE, Cavallerano J, Gardner WK, Strong J: Joslin Vision Network Validation Study: pilot image stabi-

lization phase. *J Am Optom Assoc* 69:699–710, 1998

8. Bresnick GH, Mukamel DB, Dickinson JC, Cole DR: A screening approach to the surveillance of patients with diabetes for the presence of vision-threatening retinopathy. *Ophthalmology* 107:19–24, 2000

9. Brechner RJ, Cowie CC, Howie LJ, Herman WH, Will JC, Harris MI: Ophthalmic examination among adults with diagnosed diabetes mellitus. *JAMA* 270:1714–1718, 1993

10. Wilson C, Horton M, Cavallerano J, Aiello LM: Addition of primary care-based retinal imaging technology to an existing eye care professional referral program increased the rate of surveillance and treatment of diabetic retinopathy. *Diabetes Care* 28:318–322, 2005

11. American Academy of Ophthalmology Retina Panel: Preferred practice pattern: diabetic retinopathy (article online), 2003. Available from www.aao.org/ppp. Accessed 10 October 2006

12. Lin DY, Blumenkranz MS, Brothers RJ, Grosvenor DM: The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *Am J Ophthalmol* 134:204–213, 2002

13. Williams GA, Scott IU, Haller JA, Maguire AM, Marcus D, McDonald HR: Single-field fundus photography for diabetic retinopathy screening: a report by the American Academy of Ophthalmology. *Ophthalmology* 111:1055–1062, 2004

14. Lawrence MG: The accuracy of digital-video retinal imaging to screen for diabetic retinopathy: an analysis of two digital-video retinal imaging systems using standard stereoscopic seven-field photography and dilated clinical examination as reference standards. *Trans Am Ophthalmol Soc* 102:321–340, 2004

15. Hoover A, Goldbaum M: Locating the optic nerve in a retinal image using the fuzzy convergence of the blood vessels. *IEEE Trans Med Imaging* 22:951–958, 2003

16. Patton N, Aslam TM, MacGillivray T, Deary IJ, Dhillon B, Eikelboom RH, Yogesan K, Constable IJ: Retinal image analysis: concepts, applications and potential. *Prog Retin Eye Res* 25:99–127, 2006

17. Hoover A, Kouznetsova V, Goldbaum M: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans Med Imaging* 19:203–210, 2000

18. Goldbaum MH, Katz NP, Nelson MR, Haff LR: The discrimination of similarly colored objects in computer images of the ocular fundus. *Invest Ophthalmol Vis Sci* 31:617–623, 1990

19. Sleightholm MA, Arnold J, Aldington SJ, Kohner EM: Computer-aided digitisation of fundus photographs. *Clin Phys Physiol Meas* 5:295–301, 1984

20. Abràmoff MD, Alward WL, Greenlee EC, Shuba LM, Kim CY, Fingert JH, Kwon YH: Automated segmentation of the optic nerve head from stereo color photographs using physiologically plausible feature detectors. *Inv Ophthalm Vis Sci* 48:1665–1673, 2007

21. Spencer T, Olson JA, McHardy KC, Sharp PF, Forrester JV: An image-processing strategy for the segmentation and quantification of microaneurysms in fluorescein angiograms of the ocular fundus. *Comput Biomed Res* 29:284–302, 1996

22. Gardner GG, Keating D, Williamson TH, Elliott AT: Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *Br J Ophthalmol* 80:940–944, 1996 [See comments.]

23. Cree MJ, Olson JA, McHardy KC, Sharp PF, Forrester JV: A fully automated comparative microaneurysm digital detection system. *Eye* 11:622–628, 1997

24. Hipwell JH, Strachan F, Olson JA, McHardy KC, Sharp PF, Forrester JV: Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool. *Diabet Med* 17:588–594, 2000

25. Sinthanayothin C, Boyce JF, Williamson TH, Cook HL, Mensah E, Lal S, Usher D: Automated detection of diabetic retinopathy on digital fundus images. *Diabet Med* 19:105–112, 2002

26. Osareh A, Mirmehdi M, Thomas B, Markham R: Automated identification of diabetic retinal exudates in digital colour images. *Br J Ophthalmol* 87:1220–1223, 2003

27. Li H, Chutatape O: Automated feature extraction in color retinal images by a model based approach. *IEEE Trans Biomed Eng* 51:246–254, 2004

28. Abràmoff MD, Niemeijer M: Automatic detection of the optic disc location in retinal images using optic disc location regression. In *Proceedings of IEEE–EMBC*. New York NY, Engineering in Medicine and Biology Society, Institute of Electronic and Electrical Engineers, p. 4432–4435, 2006

29. Niemeijer M, Staal JS, van Ginneken B, Loog M, Abràmoff MD: Comparative study of retinal vessel segmentation on a new publicly available database. *Proc Soc Photo Opt Instrum Eng* 44:5370–5379, 2004

30. Niemeijer M, vanGinneken B, Staal J, Suttorp-Schulten MS, Abramoff MD: Automatic detection of red lesions in digital color fundus photographs. *IEEE Trans Med Imaging* 24:584–592, 2005

31. Niemeijer M, van Ginneken B, Russell SR, Suttorp-Schulten MS, Abramoff MD: Automated detection and differentiation of drusen, exudates, and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Invest Ophthalmol Vis Sci* 48:2260–2267, 2007

32. Niemeijer M, Abramoff MD, van Ginneken B: Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. *Med Image Anal* 10:888–898, 2006

33. Abramoff MD, Suttorp-Schulten MS: Web-based screening for diabetic retinopathy in a primary care population: the EyeCheck project. *Telemed J E Health* 11:668–674, 2005

34. Cavallerano J, Lawrence MG, Zimmer-Galler I, Bauman W, Bursell S, Gardner WK, Horton M, Hildebrand L, Federman J, Carnahan L, Kuzmak P, Peters JM, Darkins A, Ahmed J, Aiello LM, Aiello LP, Buck G, Cheng YL, Cunningham D, Goodall E, Hope N, Huang E, Hubbard L, Janczewski M, Lewis JW, Matsuzaki H, McVeigh FL, Motzno J, Parker-Taillon D, Read R, Soliz P, Szirth B, Vigersky RA, Ward T: Telehealth practice recommendations for diabetic retinopathy. *Telemed J E Health* 10:469–482, 2004

35. Robbins AS, Hurley LD, Dudenhoefer EJ, Chao SY: Performance characteristics of digital fundus photography as a screening test for diabetic retinopathy in a low-risk population. *Diabetes Technol Ther* 3:193–200, 2001

36. Early Treatment Diabetic Retinopathy Study Research Group: Fundus photographic risk factors for progression of diabetic retinopathy: ETDRS report number 12. *Ophthalmology* 98:823–833, 1991

37. Wilkinson CP, Ferris FL III, Klein RE, Lee PP, Agardh CD, Davis M, Dills D, Kampik A, Pararajasegaram R, Verdaguer JT: Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 110:1677–1682, 2003

38. Staal J, Abramoff MD, Niemeijer M, Viergever MA, van Ginneken B: Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging* 23:501–509, 2004

39. 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* 21:84–90, 2004

40. Larsen M, Godt J, Larsen N, Lund-Andersen H, Sjolie AK, Agardh E, Kalm H, Grunkin M, Owens DR: Automated detection of fundus photographic red lesions in diabetic retinopathy. *Invest Ophthalmol Vis Sci* 44:761–766, 2003

41. Kinyoun J, Barton F, Fisher M, Hubbard L, Aiello L, Ferris F III: Detection of diabetic macular edema: ophthalmoscopy versus photography: ETDRS report number 5: the ETDRS Research Group. *Ophthalmology* 96:746–750, 1989