



Published in final edited form as:

*Invest Ophthalmol Vis Sci.* 2007 May ; 48(5): 2260–2267. doi:10.1167/iovs.06-0996.

## Automated Detection and Differentiation of Drusen, Exudates, and Cotton-wool Spots in Digital Color Fundus Photographs for Early Diagnosis of Diabetic Retinopathy IOVS-06-0996 accepted version

Meindert Niemeijer<sup>1</sup>, Bram van Ginneken, PhD<sup>1</sup>, Stephen R. Russell, MD<sup>2,3</sup>, Maria S.A. Suttorp-Schulten, MD, PhD<sup>4</sup>, and Michael D. Abràmoff, MD, PhD<sup>2,3,5</sup>

<sup>1</sup> Image Sciences Institute, University Medical Center Utrecht, Heidelberglaan 100, 3584CX, Utrecht, The Netherlands <sup>2</sup> Retina Service, Department of Ophthalmology and Visual Sciences, University of Iowa Hospitals and Clinics, 200 Hawkins Drive, Iowa City, IA 52242, USA <sup>3</sup> Department of Veterans Affairs, Iowa City VA Medical Center, 601 Highway 6 West, Iowa City, IA 55242, USA <sup>4</sup> Ophthalmology Service, OLVG, Amsterdam, The Netherlands <sup>5</sup> Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242, USA

### Abstract

**Purpose**—To describe and evaluate a machine learning based, automated system to detect exudates and cotton-wool spots in digital color fundus photographs, and differentiate them from drusen, for early diagnosis of diabetic retinopathy.

**Methods**—Three hundred retinal images from one eye of three hundred patients with diabetes were selected from a diabetic retinopathy tediagnosis database (non-mydratic camera two field photography); 100 with previously diagnosed ‘bright’ lesions, and 200 without. A machine learning computer program was developed that can identify and differentiate among drusen, (hard) exudates, and cotton-wool spots. A human expert standard for the 300 images was obtained by consensus annotation by two retinal specialists. Sensitivities and specificities of the annotations on the 300 images by the automated system and a third retinal specialist were determined.

**Results**—The system achieved an area under the ROC curve of 0.95 and sensitivity/specificity pairs of 0.95/0.88 for the detection of ‘bright’ lesions of any type, and 0.95/0.86, 0.70/0.93 and 0.77/0.88 for the detection of exudates, cotton-wool spots and drusen, respectively. The third retinal specialist achieved pairs of 0.95/0.74 for ‘bright’ lesions, and 0.90/0.98, 0.87/0.98 and 0.92/0.79 per lesion type.

**Conclusions**—An machine learning based, automated system capable of detecting exudates and cotton-wool spots and differentiating them from drusen in color images obtained in community based diabetic patients has been developed and approaches the performance level of that of retinal experts. If the machine learning can be improved with additional training datasets, it may be useful to detect clinically important ‘bright’ lesions, enhance early diagnosis and reduce suffering from visual loss in patients with diabetes.

---

Correspondence to: Michael D. Abràmoff.

Disclosure: Meindert Niemeijer, Bram van Ginneken, Michael D. Abràmoff: P

## Introduction

Diabetic retinopathy is the most common cause of blindness in the working population of the United States (1). Early diagnosis and timely treatment have been shown to prevent visual loss and blindness in patients with diabetes(2–4). For patients recently diagnosed with diabetes, a high proportion of normal appearing fundi are expected, and only 5%–20% may demonstrate fundoscopically visible diabetic retinopathy(5). Digital photography of the retina examined by expert readers has been shown to be sensitive and specific in detecting the early signs of retinopathy(6). Early diabetic retinopathy lesions may be classified into ‘red lesions’, such as microaneurysms, hemorrhages and intraretinal microvascular abnormalities, and ‘bright lesions’, such as lipid or lipoprotein exudates and superficial retinal infarcts (cotton-wool spots) (6;7). We and others have described machine learning computer systems capable of detecting ‘red lesions’ and blood vessels in retinal color photographs with high accuracy(8–13). Because retinal exudates can represent the only visible sign of early diabetic retinopathy in some patients, computer-based systems that can detect exudates have been proposed(14–18). However, to diagnose the ‘bright lesions’ associated with diabetic retinopathy, namely exudates and cotton-wool spots, the lesions must be differentiated from drusen, the ‘bright lesions’ associated especially with age-related macular degeneration (AMD), which can have similar appearance(19), as well as from posterior hyaloid reflexes and flash artifacts, which can sometimes mimic bright lesions in appearance. A computer-based system that can detect ‘bright’ lesions must therefore be capable of differentiating among different lesion types, as they have different diagnostic importance and management implications. Extending our previous work on machine learning automated detection of blood vessels and red lesions, we have developed a machine learning algorithm that can detect bright lesions in retinal color photographs and can differentiate among exudates, cotton-wool spots and drusen.

The purpose of this study was to describe and evaluate this machine learning based computer algorithm to detect exudates and cotton-wool spots in digital color fundus photographs, and differentiate them from drusen. The evaluation was performed on a representative sample of images of patients with diabetes drawn from a telediagnosis project in the Netherlands, with 100 color fundus images containing ‘bright lesions’ and 200 images with no abnormalities. Because the purpose was to compare the machine learning algorithm to that of human experts, a human expert reference standard was created by having three masked retinal specialists annotate the sample photographs.

## Methods

### Subjects

A total of 430 retinal images, of adequate quality as originally read by an ophthalmologist for clinical reading, were selected from the EyeCheck project in the Netherlands for this study (20). 230 images were originally read as containing one or more of exudates, cotton-wool spots and/or drusen, and 200 were originally read as not containing any of these bright lesions. 130 of the images, all originally read as containing bright lesions, were used to train the machine learning algorithm, and 300 images, from 300 subjects, were used to perform diagnostic validation. The EyeCheck project database contains images for over 18,000 exams (over 72,000 images) of patients without previously known diabetic retinopathy (20). The study was performed according to the tenets of the declaration of Helsinki, and Investigational Review Board approval was obtained. The researchers only had access to the anonymous images and their original diagnosis. Because of the retrospective nature of the study and patient anonymity, informed consent was judged not to be necessary by the IRB. The EyeCheck project ([www.eyecheck.nl](http://www.eyecheck.nl)) performs early diagnosis of diabetic retinopathy over the internet in a community based population using digital cameras in primary care physicians’ offices, so the images were obtained at multiple sites.

## Fundus photography

Images were obtained at multiple sites with three different ‘non-mydriotic’ cameras: the Topcon NW 100, the Topcon NW 200 (Topcon, Tokyo, Japan) and the Canon CR5-45NM (Canon, Tokyo, Japan). The imaging protocol has been described previously: briefly, digital color photographs were obtained under natural dilation in a dark room, and if natural dilation did not suffice, the pupil was dilated pharmacologically using one drop of Tropicamide 0.5% per eye as per protocol(20). One disc centered and one fovea centered image was obtained for each eye, both at 45° field of view. For the Topcon cameras, spatial resolution was approximately 8×8µm per pixel (2048×1536), while for the Canon type cameras, it was approximately 15×15µm (1024×768pixels). All camera types have 1-layer CCD RGB sensors, and images were JPEG compressed at the lowest loss compression setting, resulting in image files of approximately 0.15–0.5MB per image.

## Machine learning training images

The machine learning algorithm is a so-called supervised algorithm, and therefore needs a set of annotated lesions to ‘learn’ how to detect bright lesions and differentiate among them. For this purpose, 130 anonymous images originally read as containing bright lesions were selected. All pixels in all of these images were segmented by retinal specialist A as to whether they were (part of) an exudate, cotton-wool spot, drusen or background retina. Vessels, disc and red lesions, if present, were treated as background retina. The images utilized to create the training set were not included in the test set (see below). The training set contained 1113 exudates (93067 pixels), 45 cotton-wool spots (33959 pixels) and 2030 drusen (287186 pixels).

## Image data set for testing performance

Three hundred anonymous images were selected as the testing set. One hundred images were selected at random from all images that were originally read clinically as containing one or more bright lesions, and two hundred images were selected from all images originally read as containing no lesions (not containing any exudates, cotton-wool spot, or drusen). Three masked retinal specialists (A, B, C) performed annotation on all images in random order indicating whether one or more exudates, cotton-wool spots or drusen or any combination thereof was present. A consensus annotation was then obtained, using a teleconference discussion format, by asking two of the retinal specialists (A, B) to reach consensus on all images where their independent annotations had differed. The consensus annotation was used as the reference standard, and contained 105 images with bright lesions and 195 without. In other words, some images originally thought to be without bright lesions did contain one or more bright lesions according to the consensus reference standard and vice versa). There were 42 images with exudates, 30 with cotton-wool spots and 52 with drusen in this test set. Some images also contained ‘red’ lesions (microaneurysms, hemorrhages, microvascular abnormalities or neovascularizations), and the distribution of the presence of lesions is given in Table 1.

## Machine learning algorithm

The machine learning algorithm is based on our earlier work using retinal pixel and lesion classification(8–10). To perform detection and differentiation of bright lesions, if any, in a previously unseen image, the following steps were performed:

1. Each pixel was classified, resulting in a so-called lesion probability map that indicates the probability of each pixel to be part of a bright lesion
2. Pixels with high probability were grouped into probable lesion pixel clusters
3. Based on cluster characteristics each probable lesion pixel cluster was assigned a probability indicating the likelihood that the pixel cluster was a true bright lesion.

4. Each bright lesion cluster likely to be a bright lesion was classified as exudate, cotton-wool spot or drusen.

Figure 1 illustrates the above steps. The system performs the classification steps (1, 3, and 4) using a statistical classifier that, given a training set of labeled examples, can differentiate among different types of pixels or clusters based on so-called *features* or numerical characteristics, such as pixel color, cluster area, etc. The features used depend on the type of objects that are to be classified. For each classification step several statistical classifiers were tested on the training set and the one demonstrating the best performance was used on the test set. A k-nearest neighbor classifier was selected for steps 1 and 3, and a linear discriminant classifier for the third classification step(17). When presented with an unseen image in the test set, the automated algorithm gave two outputs: whether bright lesions were present or not, and which class each lesion was: exudate, cotton-wool spot, or drusen. A more extensive description of the algorithm is given in the Appendix.

### Outcome parameters and data analysis

The machine learning algorithm annotated all images in the test set repeatedly while varying the threshold for normal/abnormal cutoff (steps 1 and 3). Sensitivity, specificity, and kappa of the annotation compared to the consensus standard was calculated at each threshold setting. Sensitivity, specificity, and kappa, of retinal specialist C compared to the consensus standard were also determined. For the machine learning algorithm, these sensitivity/specificity pairs can be used to create Receiver Operator Characteristic (ROC) curves for all bright lesions, and for the differentiation of exudates, cotton-wools and drusen. The ROC curve shows the sensitivity and specificity at various thresholds, and the system can be set for a specific sensitivity/specificity by selecting the corresponding threshold. For a screening system, sensitivity is more important than specificity, so a threshold could be chosen that maximizes sensitivity while still maintaining enough specificity. Because human experts cannot (consciously) adjust their lesion detection threshold, only a single sensitivity/specificity pair was obtained for each human grader, and plotted in the ROC curve as a point (see Figure 2). The area under the ROC curve is regarded as the most accurate and 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. In addition, the number of images where the original annotations of all three retinal specialists (A,B,C) and the machine learning algorithm agreed, were also determined for all bright lesions as a group, and for the three classes of lesions individually.

## Results

Table 2 shows sensitivity, specificity, and kappa values of the machine learning algorithm and retinal specialist C, compared to the consensus standard.

Figure 2 shows the ROC curves for the system and the ROC points for retinal specialist C, for both overall bright lesion detection and bright lesion differentiation per image. The area under the curve of the automated system for the detection of bright lesions was 0.95, and for the detection of exudates, cotton-wool spots and drusen it was 0.94, 0.85 and 0.88 respectively.

The automated system achieved sensitivity/specificity of 0.95/0.88 for the detection all bright lesions, and 0.95/0.86, 0.70/0.93 and 0.77/0.88 for the detection of exudates, cotton-wool spots and drusen respectively. The third retinal specialist achieved 0.95/0.74 for bright lesions, and 0.90/0.98, 0.87/0.98 and 0.92/0.79 for the detection of exudates, cotton-wool spots and drusen, respectively.

In total, 1739 bright lesions were detected at this optimal threshold setting. Of these 1739 lesions, 1513 lesions were classified correctly, and 226 were misclassified: of these 226 confusions, 124 drusen were misclassified as exudates, 49 exudates were misclassified as drusen, 5 drusen were misclassified as cotton-wools spots, 2 exudates were misclassified as cotton-wool spots and there were 45 other confusions, either misclassifications or non-lesions. Because the consensus standard was determined by two independent retinal specialists, it was useful to determine the kappas of their (independent) annotations before the consensus process, and these were 0.80, 0.65, 0.73 and 0.65, respectively.

In 225 of 300 images, the consensus standard, the automated system and retinal specialist C were all in full agreement on the presence of bright lesions. In 167 of 300 cases, the consensus standard, the automated system and retinal specialist C were in full agreement on the type(s) of bright lesion. Examples are shown in Figure 3A and 3B. Two examples of cases where human experts did not agree among themselves and the automated system also did not agree with the human experts are shown in Figures 3C and 3D.

## Discussion

Our results demonstrate that a machine learning algorithm as described herein is capable of detecting exudates, cotton-wool spots and drusen and differentiating among them on color images of the retina obtained in a community-based population of patients with diabetes. The area-under-the-ROC-curve of 0.95 shows that this proof-of-concept system has sensitivity and specificity approaching that of retinal experts.

Differentiating among the three types of bright lesions from color photographs can be challenging, as illustrated by Figures 3A–D. The lesions in these images are subtle, and correct differentiation may be improved with knowledge of patient age, consideration of contextual lesions and the size of lesion classes. Despite the complexity of this task, in 1513/1739 (87%) of cases, the automated system and the three retinal specialists agreed on the presence and type of lesion. As might be expected, drusen and exudates were easily confused because they are often similar in size, while cotton-wool spots are much less often confused with the other two types of lesions.

The existing literature has focused almost exclusively on the automatic detection of exudates, only a single study included the detection of cotton-wool spots, and no studies took account of drusen in this context (14–16;18;19;21). Our results indicate that of all three bright lesion types, exudates present the easiest lesion to detect for both automated system and retinal specialist C. The capacity of our system algorithm to detect and differentiate all three types of lesions distinguishes it from prior studies.

Despite the encouraging performance achieved by the present system, several major issues remain. One disadvantage of this study is its simple application. We compare the machine learning system and human experts' performance on digital fundus photographs obtained in a telediagnosis setting with non-mydratiac cameras. The accepted standard for detection of diabetic retinopathy is 7-field stereo fundus photography by certified photographers and read by certified readers(22). However, the reference standard utilized here is the consensus reading of the photographs by retinal specialists. This study, and the trained algorithm, may be biased and some bright lesions may have been missed by both human experts and the automated system because of the limitations of digital 2-field non-stereo photography (6). Because we set out to compare the performance of the system to human experts on the same photographs, our results are only valid in that context. To achieve a more comprehensive evaluation of this or a similar system, a comparison of the machine learning algorithm on non-mydratiac non-stereo images, to 7-field non-mydratiac images evaluated by human experts, would be necessary.

A second limiting factor that might limit optimal performance of the machine learning algorithm could be a constraint of the quality of the annotations of the training images. In other words, the ‘better’ this training set, the better the theoretical performance of the algorithm. Improving the quality of the training data using multiple experienced clinicians and larger numbers of training pixels, may improve system performance, and also lessen the problem of clinician inter-observer variability as evidenced in Figure 3C and 3D.

Thirdly, both human readers and algorithms have difficulty detecting retinal thickening especially in the absence of context lesions such as exudates, beading or hemorrhages. Judging retinal thickening is especially difficult when using non-stereo fundus photography for early diagnosis(2;6;22). Retinal thickening without context lesions, if present, will be missed by the algorithm. This is a downside of non-stereo fundus photography, whether read by human experts or an algorithm, though the relative frequency of retinal thickening without context signs may be low (23;24).

Fourthly, the system has been tested on a small number of patients. If tested on a larger prospective dataset, performance may not be comparable to these results. The ROCs in Figure 2 are interesting because they suggest that professional experience or training differences may affect the performance of human experts. Retinal specialist C is an expert on AMD research and drug trials, in addition to being an expert in diabetic retinopathy, while retinal specialists A and B are primarily diabetic retinopathy specialists. As Table 2 depicts, the performance of C compared to the consensus standard on cotton-wool spots and exudates is comparable, but C is much more sensitive to drusen: one possible explanation is C’s heightened awareness of subtle AMD lesions.

An important question that cannot be answered by a preliminary study such as the present one is how the current algorithm, capable of detecting and differentiating exudates, cotton-wool spots and drusen, might fit into a complete system for diabetic retinopathy classification and health care delivery. We envision several different approaches, one of which would be a system for early detection of diabetic retinopathy in patients with diabetes that currently are not receiving the recommended regular dilated eye exams. Such a system may require the following subcomponents, many of which we and others have presented previously:

- quality assurance: an algorithm capable of detecting images that are not adequate, such as flash artifacts, blinking, blur, cataract (25)
- vessel segmentation: an algorithm capable of segmenting the retinal blood vessels even where pathology is present(8;9;26)
- segmentation of optic disc, and possibly fovea: an algorithm capable of localizing and segmenting these substructures(27–30)
- microaneurysm, retinal hemorrhage, neovascularization and other retinal microvascular abnormalities: an algorithm for detecting these ‘red lesions’(10)
- exudate and cotton-wool spot detection, and differentiation from non-DR bright lesions: the algorithm presented herein
- combination of the outputs of these subcomponents over multiple images of a single patient into a simple ‘diabetic retinopathy probability’ statistic

In conclusion, we have developed and tested a machine learning system capable of detecting exudates, cotton-wool spots and drusen, and differentiating amongst these, on color images of the retina obtained in a population of patients with diabetes. The performance of this proof-of-concept system has sensitivity and specificity that approaches retinal experts. If such a system can be improved by better quality training data and tested on larger datasets, it has the potential to help prevent visual loss and blindness in patients suffering from diabetes.

## Acknowledgments

MN: Dutch Ministry of Economic Affairs IOP IBVA02016;KNAW Van Walree Fund travel grant MDA: NEI R01-EY017066; DoD A064-032-0056; Research to Prevent Blindness, NY, NY; Wellmark Foundation; US Department of Agriculture; University of Iowa; Netherlands Organization for Health Related Research.

## References

1. Klonoff DC, Schwartz DM. An economic analysis of interventions for diabetes. *Diabetes Care* 2000;23:390–404. [PubMed: 10868871]
2. Kinyoun J, Barton F, Fisher M, Hubbard L, Aiello L, Ferris F III. Detection of diabetic macular edema. Ophthalmoscopy versus photography--Early Treatment Diabetic Retinopathy Study Report Number 5. The ETDRS Research Group. *Ophthalmology* 1989;96:746–750. [PubMed: 2740076]
3. Early Treatment Diabetic Retinopathy Study Research Group. Early photocoagulation for diabetic retinopathy. ETDRS report number 9. *Ophthalmology* 1991;98:766–785. [PubMed: 2062512]
4. 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* 2000;107:19–24. [PubMed: 10647713]
5. Wilson C, Horton M, Cavallerano J, Aiello LM. The Use of Retinal Imaging Technology Increases the Rates of Screening and Treatment of Diabetic Retinopathy. *Diabetes Care* [IN PRINT, GET FROM MEDLINE]. 2004
6. 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* 2002;134:204–213. [PubMed: 12140027]
7. Olson JA, Strachan FM, Hipwell JH, Goatman KA, McHardy KC, Forrester JV, Sharp PF. A comparative evaluation of digital imaging, retinal photography and optometrist examination in screening for diabetic retinopathy. *Diabet Med* 2003;20:528–534. [PubMed: 12823232]
8. Niemeijer M, Staal JS, van Ginneken B, Loog M, Abramoff MD. Comparative study of retinal vessel segmentation on a new publicly available database. *Proceedings of the SPIE* 2004;5370–5379.
9. Staal JS, Abramoff MD, Niemeijer M, Viergever MA, van Ginneken B. Ridge based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging* 2004;23:501–509. [PubMed: 15084075]
10. Niemeijer M, vanGinneken B, Staal J, Suttrop-Schulten MS, Abramoff MD. Automatic detection of red lesions in digital color fundus photographs. *IEEE Trans Med Imaging* 2005;24:584–592. [PubMed: 15889546]
11. Abramoff MD, Kim CY, Fingert JH, Shuba LM, Greenlee EC, Alward WLM, Kwon YH. Automatic Segmentation of the Cup and Rim From Optic Disc Stereo Color Images: Correlation Between Glaucoma Fellows, Glaucoma Experts, and Pixel Feature Classification [ARVO abstract 3635]. *Inv Ophthalm Vis Sci* 2006;47(4)
12. 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* 2003;44:761–766. [PubMed: 12556411]
13. 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* 1996;29:284–302. [PubMed: 8812075]
14. Gardner GG, Keating D, Williamson TH, Elliott AT. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool [see comments]. *Br J Ophthalmol* 1996;80:940–944. [PubMed: 8976718]
15. 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* 2002;19:105–112. [PubMed: 11874425]
16. Osareh A, Mirmehdi M, Thomas B, Markham R. Automated identification of diabetic retinal exudates in digital colour images. *Br J Ophthalmol* 2003;87:1220–1223. [PubMed: 14507751]
17. Duda, RA.; Hart, PE.; Stork, DG. *Pattern classification*. Wiley-Interscience; New York: 2001.

18. Walter T, Klein JC, Massin P, Erginay A. A contribution of image processing to the diagnosis of diabetic retinopathy--detection of exudates in color fundus images of the human retina. *IEEE Trans Med Imaging* 2002;21:1236–1243. [PubMed: 12585705]
19. 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. [PubMed: 14706060]
20. Abramoff MD, Suttorp-Schulten MS. Web-based screening for diabetic retinopathy in a primary care population: the EyeCheck project. *Telemed J E Health* 2005;11:668–674. [PubMed: 16430386]
21. Lee SC, Lee ET, Kingsley RM, Wang Y, Russell D, Klein R, Warn A. Comparison of diagnosis of early retinal lesions of diabetic retinopathy between a computer system and human experts. *Arch Ophthalmol* 2001;119:509–515. [PubMed: 11296016]
22. Early Treatment Diabetic Retinopathy Study Research Group. Grading diabetic retinopathy from stereoscopic color fundus photographs--an extension of the modified Airlie House classification. ETDRS report number 10. *Ophthalmology* 1991;98:786–806. [PubMed: 2062513]
23. Schiffman RM, Jacobsen G, Nussbaum JJ, Desai UR, Carey JD, Glasser D, Zimmer-Galler IE, Zeimer R, Goldberg MF. Comparison of a digital retinal imaging system and seven-field stereo color fundus photography to detect diabetic retinopathy in the primary care environment. *Ophthalmic Surg Lasers Imaging* 2005;36:46–56. [PubMed: 15688971]
24. Welty CJ, Agarwal A, Merin LM, Chomsky A. Monoscopic versus stereoscopic photography in screening for clinically significant macular edema. *Ophthalmic Surg Lasers Imaging* 2006;37:524–526. [PubMed: 17152553]
25. Niemeijer M, van Ginneken B, Abramoff MD. Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening. *Medical Image Analysis Journal*. 2006in press
26. 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* 2000;19:203–210. [PubMed: 10875704]
27. Abramoff, MD.; Niemeijer, M. Automatic detection of the optic disc location in retinal images using optic disc location regression. New York, NY: IEEE EMBC; 2006. p. 4432-4435.2006
28. Niemeijer M, Abramoff MD, van GB. Segmentation of the optic disc, macula and vascular arch in fundus photographs. *IEEE Trans Med Imaging*. 2006in press
29. Foracchia M, Grisan E, Ruggeri A. Detection of optic disc in retinal images by means of a geometrical model of vessel structure. *IEEE Trans Med Imaging* 2004;23:1189–1195. [PubMed: 15493687]
30. Abramoff MD, Alward WL, Greenlee EC, Shuba LM, Kim CY, Fingert JH, Kwon YH. Automated segmentation of the optic nerve head from stereo color photographs. *Inv Ophthalm Vis Sci*. 2006in press
31. ter Haar Romeny, BM. *Front End Vision and Multi-scale Image Analysis*. Kluwer Academic Publishers; Dordrecht: 2003.

## Appendix

### Appendix A. Machine learning system – feature classification

#### Probability of pixel being part of a bright lesion

To determine the probability for all pixels in an image (in the test set) to be part of a bright lesion, the green channel of the RGB camera image is convolved with a set of 14 digital filters. These filters are based on Gaussian derivatives and are invariant to rotation and translation of the image. The specific image filters used were selected from a larger set of second order irreducible invariants(31) using the training images and a feature selection algorithm. If a bright lesion is present, the combination of filter responses for pixels in or close to that bright lesion will be different from the filter responses for non-lesion pixels. The classifier used to classify the pixels on the basis of the filter responses, was a k-Nearest Neighbor (kNN) classifier(17). Such a classifier is capable of assigning probabilities to pixels in an image, based on the



combination of filter responses, and can generate a lesion probability map (example: second row of Figure 1), that indicates the probability that each pixel is part of a bright lesion. The kNN classifier is a machine learning algorithm and is trained by offering it a set of training pixels from the training image set where the classification has been established (bright lesion or non-bright lesion).

### Candidate bright lesion pixel clusters

By setting the threshold at 60% (pixels with a probability higher than 60% are considered part of a bright lesion are retained and by grouping connected pixels above this threshold), a set of bright lesion pixel clusters is obtained (example: third row of Figure 3). Because their output is required for further processing, algorithms that perform red lesion classification, optic disc segmentation and vessel segmentation were applied to the images as we have previously described(8–10;27). That part of a bright lesion that overlaps with the optic disc is removed at this step, using the segmented optic disc. From here on, such a bright lesion pixel cluster will be termed 'potential lesion.'

### Bright lesion detection

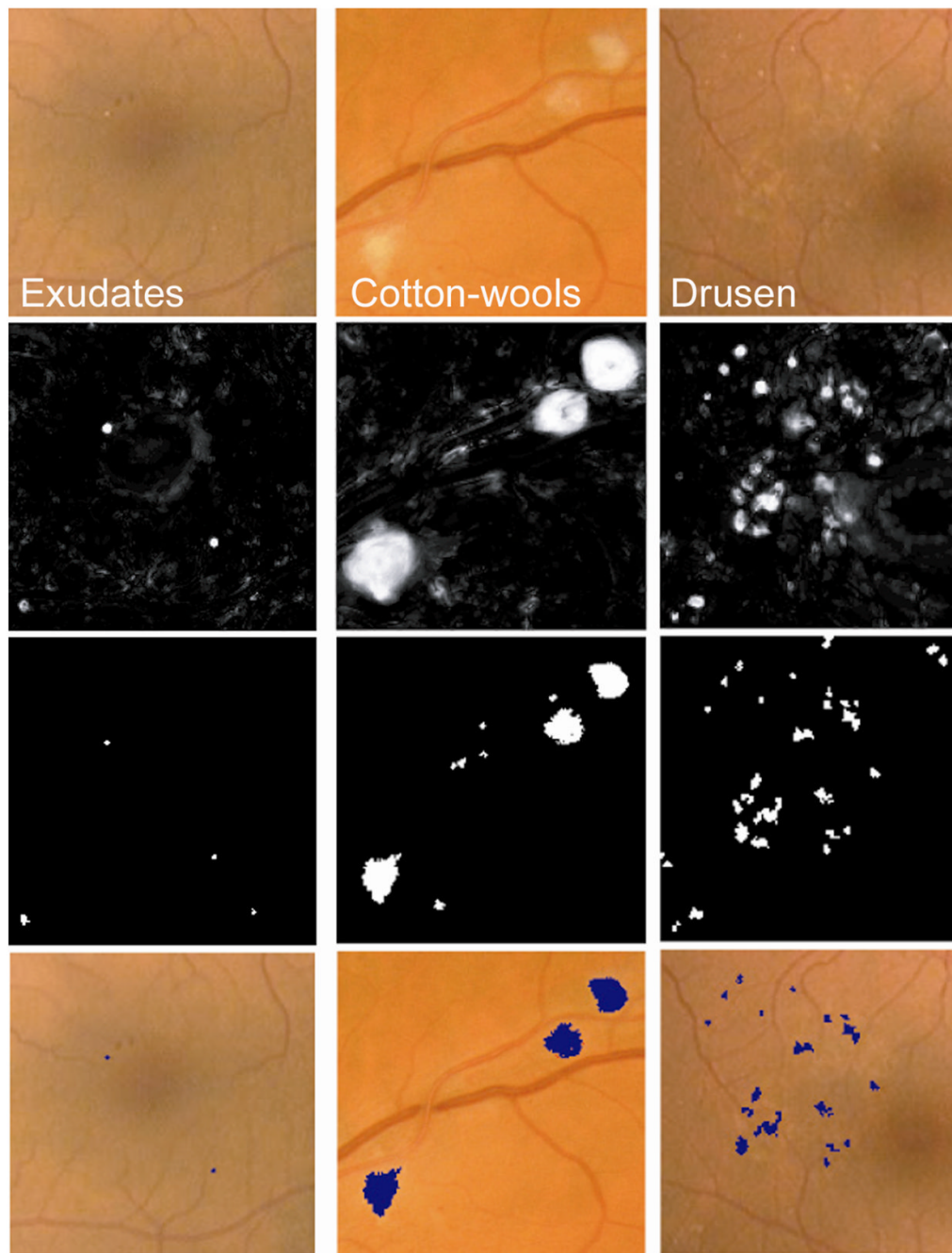
The potential lesions include potential spurious responses, most of which occur along the major vessels, and some of which are from posterior hyaloid reflexes. A second kNN classifier was trained using a set of example potential lesions extracted from the training set to suppress spurious bright lesion clusters (see Table 3 for the features used). The contrast features measure the contrast of the cluster in multiple image RGB color planes, and other features provide information about the size, shape and contrast of a potential lesion, its proximity to the closest vessel, and proximity to the closest red lesion, as potential lesions close to a red lesion are more likely to be true bright lesions. Each potential lesion is thereby assigned a probability indicating the likelihood that it is a true bright lesion. The probability that the image contains any type of bright lesion is now given by the maximum probability assigned to any of the potential lesions in the image. For the final classification step all potential lesions with a probability below 70% were discarded. The value for this threshold was determined on the training set, and the results were not very sensitive to small variations in this threshold (see the bottom row of Figure 3).

### Bright lesion classification

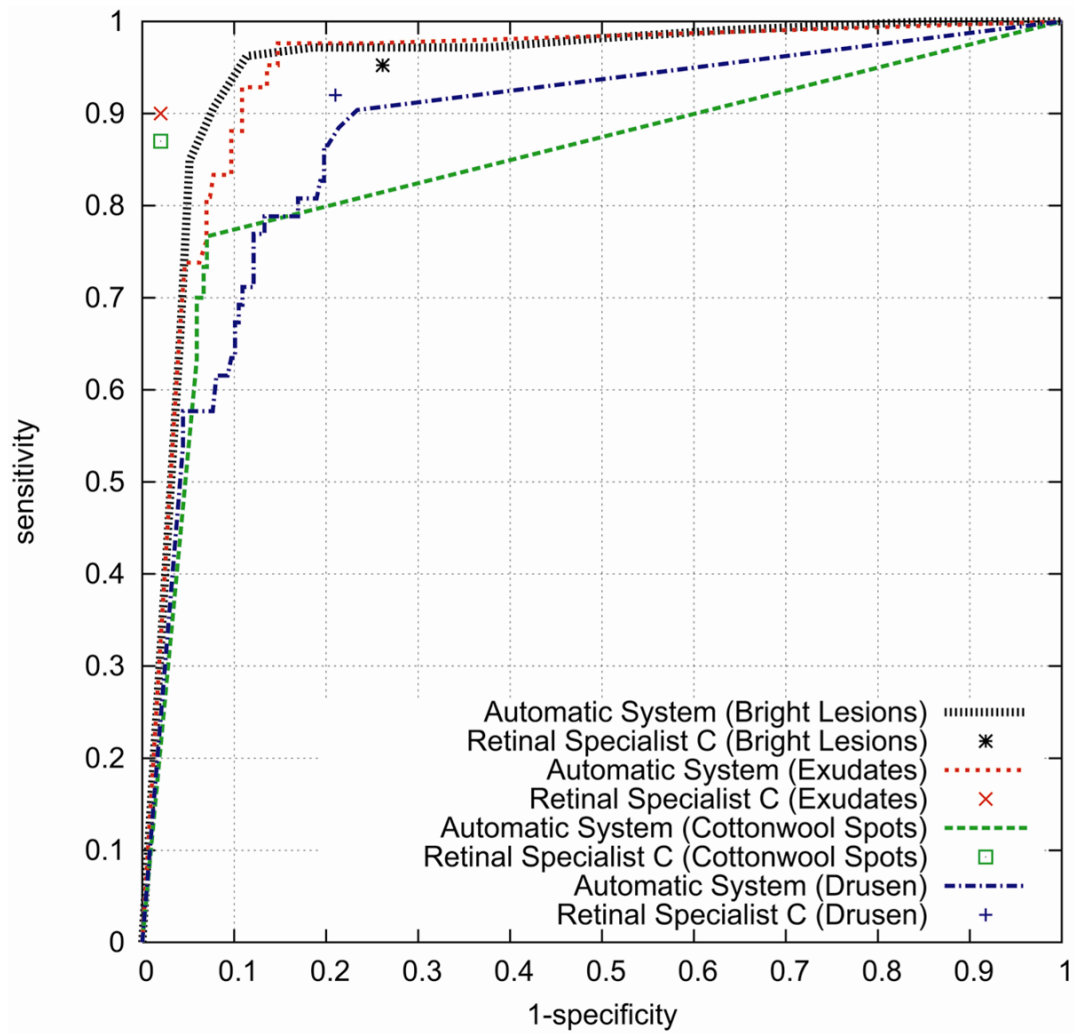
A third classifier was trained using the bright lesion types from the training set. The features described in Table 3, and in addition, the following features were used:

- the number of red lesions in the image
- the number of detected bright lesions from the previous step
- the probability for the cluster

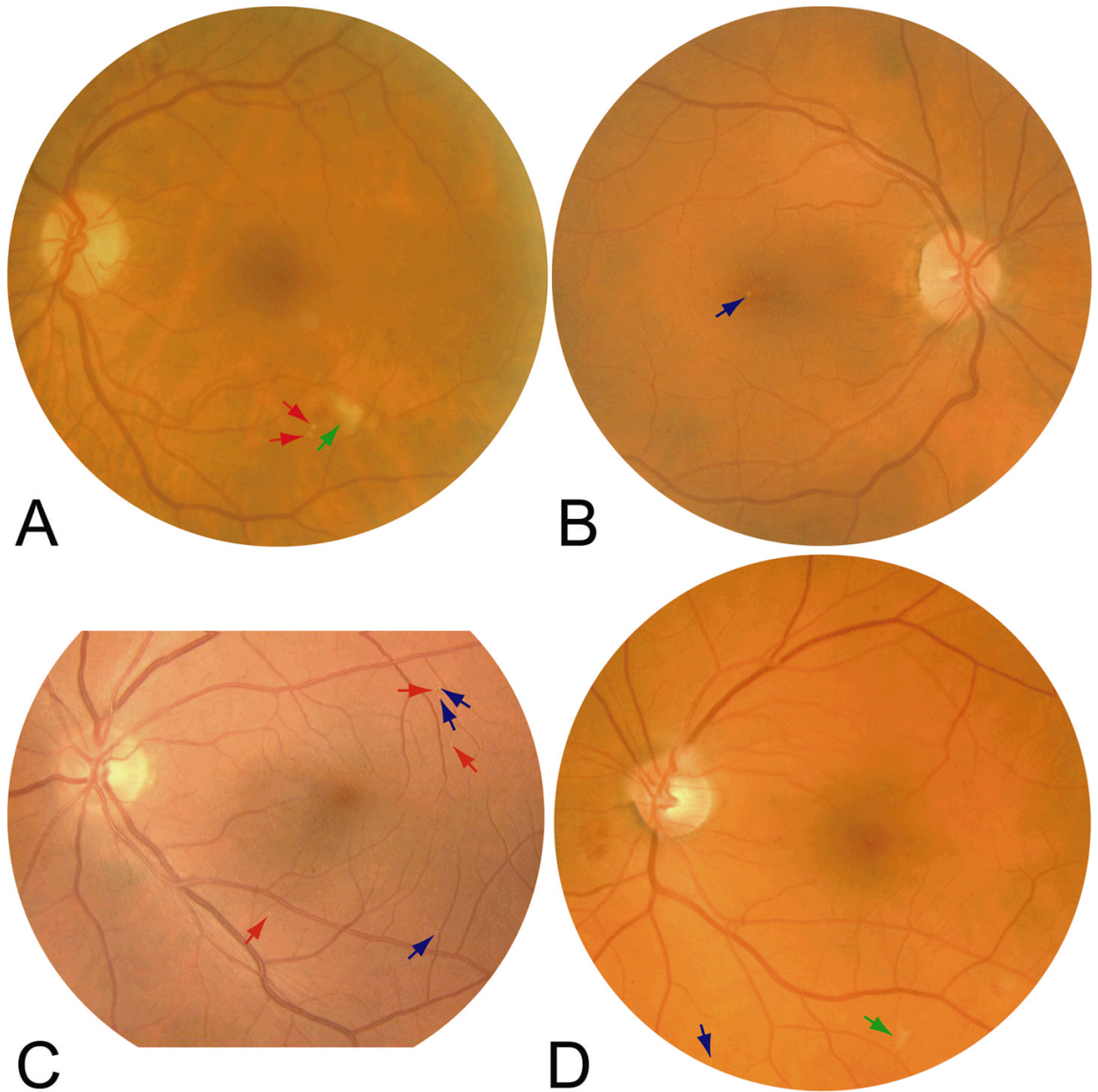
A Linear Discriminant Analysis classifier labeled all found lesions in the test set as to whether they were exudates, cotton-wool or drusen(17).



**Figure 1.** Machine learning algorithm steps performed to detect and differentiate 'bright lesions'. From left to right column, exudates, cotton-wool spots, and drusen. From top to bottom, first row shows the relevant region in the retinal color image (all at the same scale), second row shows the posterior probability map after the first classification step, third row shows the pixel clusters that are probable bright lesions (potential lesions), and the bottom row shows those objects which the system classified as true bright lesions overlaid on the original image.



**Figure 2.** ROC curves of the automatic system for the different detection tasks. Sensitivity/specificity pairs of Retinal Specialist C are also plotted as points in the graph.



**Figure 3.**

Figure 3A. Example image where experts and automated system agreed on the presence of exudates and cotton-wool spots. As identified by the automated system, the red arrows denote exudates and the green arrow a cotton-wool spot.

Figure 3B. Example image where experts and automated system agreed on the presence of drusen. As identified by the automated system, the blue arrows denote drusen.

Figure 3C. Example image where experts and automated system did not agree on the presence of drusen or exudates. As identified by the automated system, the blue arrows denote drusen (retinal specialist A and automated system) or exudates (retinal specialist B and C), and the red arrows denote exudates (retinal specialists B and C and automated system) or drusen (retinal specialist A).

Figure 3D. Example image where experts and automated system did not agree on the presence of drusen/exudate/cotton-wool spot. As identified by the automated system, the blue arrow denotes a drusen (automated system), exudate (retinal specialists B and C) or no abnormality (retinal specialist A); the green arrow denotes a cotton-wool spot (retinal specialist C and automated system), exudate (retinal specialist B) or no abnormality (retinal specialist A).

**Table 1**

Distribution of presence of 'red' and 'bright' lesions in the test images

	red +	red -	
bright +	34	71	105
bright -	4	191	195
	38	262	300

Sensitivity, specificity and kappa of the machine learning algorithm and retinal specialist C compared to the reference standard

**Table 2**

	Sensitivity		Specificity		k	
	Machine learning algorithm	Retinal specialist C	Machine learning algorithm	Retinal specialist C	Machine learning algorithm	Retinal specialist C
Bright lesions	0.95	0.95	0.88	0.74	0.80	0.63
Exudates	0.95	0.90	0.86	0.98	0.61	0.88
Cotton-wool spots	0.70	0.87	0.93	0.98	0.57	0.82
Drusen	0.77	0.92	0.88	0.79	0.56	0.52

**Table 3**

Overview of the features used to determine whether a potential lesion is a true bright lesion or spurious

Feature number	Description
1	Area in pixels of potential lesion.
2	Length of the perimeter in pixels of the potential lesion.
3	Compactness of the potential lesion.
4,5	Length and width in pixels of the potential lesion.
6	Mean gradient value at scale of 2 pixels at the potential lesion border
7/8	Mean/standard deviation of all green channel pixels within the potential lesion after shade correction (shade correction: a 48 pixel Gaussian blurred version of the image is subtracted from the image itself)
9/10	Mean/standard deviation of all green channel pixels in a square region outside the potential lesion after shade correction. Square region is 50×50 pixels unless the potential lesion is larger than 25 pixels wide or high, when the width or height or both are calculated according to the following formula: $dist=(dist/25.0f)*50$
11/12/13	Mean l,u,v intensities of all pixels within the potential lesion
14/15/16	Standard deviation of l,u,v intensities of all pixels within the potential lesion
17/18/19	Mean l,u,v intensities of all pixels in the square region outside the potential lesion (see above).
20/21/22	Standard deviation of l,u,v intensities of all pixels in the square region outside the potential lesion (see above).
23	Local pixel contrast in the green channel, by subtracting the average pixel intensity within the potential lesion from the average pixel value in a three pixel wide border around the potential lesion. The border is obtained by dilating the potential lesion three times with a single pixel.
24	Local pixel variance contrast in the green channel, similar to 23 but the variance is used instead of the average.
25/26	Same as 7/8 but using a locally normalized image
27/28	Same as 9/10 but using a locally normalized image
29/30	Same as 9/10 but pixels of other potential lesions are excluded from the outside area pixel means and standard deviations
31/32	Same as 9/10 but pixels of vessels are excluded from the outside area pixel means and standard deviations
33/80	Mean/standard deviation of filter outputs. Gaussian filterbank: I, Ix, Iy, Ixy, Ixx, Iyy at scales 1,2,4,8 pixels
81	Mean vessel probability at the potential lesion boundary, obtained from vessel probability map
82	Standard deviation of vessel probability at the potential lesion boundary, obtained as before
83	Distance in pixels to the closest red lesion.