Comparison of Diagnosis of Early Retinal Lesions of Diabetic Retinopathy Between a Computer System and Human Experts

Samuel C. Lee, PhD; Elisa T. Lee, PhD; Ronald M. Kingsley, MD; Yiming Wang, MS; Dana Russell, MPH; Ronald Klein, MD; Ann Warn, MD

Objective: To investigate whether a computer vision system is comparable with humans in detecting early retinal lesions of diabetic retinopathy using color fundus photographs.

Methods: A computer system has been developed using image processing and pattern recognition techniques to detect early lesions of diabetic retinopathy (hemorrhages and microaneurysms, hard exudates, and cotton-wool spots). Color fundus photographs obtained from American Indians in Oklahoma were used in developing and testing the system. A set of 369 color fundus slides were used to train the computer system using 3 diagnostic categories: lesions present, questionable, or absent (Y/Q/N). A different set of 428 slides were used to test and evaluate the system, and its diagnostic results were compared with those of 2 human experts—the grader at the University of Wisconsin Fundus Photograph Reading Center (Madison) and a general ophthalmologist. The experiments included comparisons using 3 (Y/Q/N) and 2 diagnostic categories (Y/N) (questionable cases excluded in the latter).

Results: In the training phase, the agreement rates, sensitivity, and specificity in detecting the 3 lesions between the retinal specialist and the computer system were all above 90%. The κ statistics were high (0.75-0.97), indicating excellent agreement between the specialist and the computer system. In the testing phase, the results obtained between the computer system and human experts were consistent with those of the training phase, and they were comparable with those between the human experts.

Conclusions: The performance of the computer vision system in diagnosing early retinal lesions was comparable with that of human experts. Therefore, this mobile, electronically easily accessible, and noninvasive computer system, could become a mass screening tool and a clinical aid in diagnosing early lesions of diabetic retinopathy.


Diabetic retinopathy has been identified as one of the leading causes of blindness. Persons with diabetic retinopathy are 29 times more likely to become blind than those without diabetes. Blindness owing to diabetes costs the US Government and general public $500 million annually. However, because diabetic retinopathy at its early stage is usually asymptomatic, an individual with diabetes may not be aware of the potential risk of developing retinopathy and consequently losing his or her vision. Regular retinal examinations are highly recommended by the National Eye Health Education Program of the National Eye Institute (Bethesda, Md). It is especially important among high-risk groups such as the American Indian population, which has a high prevalence and incidence of diabetic retinopathy. The high cost of examination and treatment and the shortage of ophthalmologists, especially in rural areas, are prominent factors that hinder patients from obtaining regular examinations. However, a low-cost mobile computer vision system that can make the initial diagnosis of diabetic retinopathy would be helpful to rural and underserved populations. Individuals who are diagnosed by the computer system as having early retinal lesions would be referred to an ophthalmologist or optometrist for further evaluation.

Diabetic retinopathy can be detected using ophthalmoscopy, fluorescein angiography, color fundus instant photographs, color fundus 35-mm slides, or real-time electronic imaging. Several studies comparing the image quality and effectiveness of these methods have
METHODS

The computer system (Figure 1) was designed to receive images that include the optic disc and the macula (Figure 2A). First, the image of the retina is digitized. Then the system checks the quality of the image. If the image does not show any retinal information, particularly retinal vessels, the system will not process it. These types of images can be detected by the computer system using color intensity histograms. Based on empirical data, an image that does not exhibit retinal information will have a color intensity histogram exhibiting a narrow-banded graph with a total intensity span lower than 30 on a 0 to 255 scale. These images may be caused by a variety of reasons such as insufficient pupil dilation, cataract, and problems in photograph and film development procedures (eg, patient blinks, effects of eyelashes or tears, severe film defects, and problems with the flash). If the retinal image has an intensity span higher than 30, image processing and pattern recognition techniques are then applied.

The image processing techniques employed were designed to achieve 3 purposes: image enhancement, noise removal, and image normalization. Following image processing, pattern recognition techniques were developed for recognizing various essential retinal features (the optic disc, macula, retinal background, and retinal blood vessels) and certain lesions of early nonproliferative diabetic retinopathy (HMA, HE, and CWS).

Other lesions such as macular edema, intraretinal microvascular abnormalities, or venous beading were not included. At the end of the diagnostic test, a computerized diagnostic report was provided, which listed all of the lesions (including questionable lesions), their sizes, and locations. The diagnostic report for the image in Figure 2A is shown in Figure 3. The lesions detected by the computer system are color coded and shown in Figure 2B. The diagnosis given by the Wisconsin Reading Center was moderate nonproliferative retinopathy with the presence of HMA, HE, and CWS. The development of the computer system consisted of the training phase and testing phase.

TRAINING PHASE

A total of 369 color fundus images obtained from American Indians who participated in the Vision Keepers Project were used as the training set to develop the system. The Vision Keepers Project, funded by the National Eye Institute, is an epidemiological study to determine the prevalence and incidence of eye disease in American Indians from Oklahoma. One thousand eighty-seven participants aged 49 to 83 years were examined by the ophthalmologist (A.W.) using a slitlamp biomicroscope with a 78-diopter (D) lens and indirect ophthalmoscopy. Duplicate fundus photographs of at least 1 eye of each of 1080 participants were obtained using a nonmydriatic 45 camera (Canon CR5-45 NM, Canon, Japan) through pharmacologically dilated pupils. Seven participants either refused or did not have fundus photographs taken for various reasons, including eye conditions that made it unsafe to dilate. Of the 1080 participants, fundus photographs were obtained for both eyes of 1052 participants and for 1 eye of 28 participants. The 2132 fundus photographs were sent to the University of Wisconsin Fundus Photograph Reading Center, Madison, for grading according to a modification of the Airlie House Classification Scheme.23-26 For 1973 of these photographs (92.3%), the entire field was considered gradable. One hundred four fundus slides (4.9%) could not be graded. Of the 2132 fundus slides processed by the computer system, 139 (7.7%) were rejected because of poor quality. Overall, 13.7% has been published.12-17 Recently there have been several articles published on automatic detection and quantification of diabetic retinopathy lesions from fluorescein angiograms.18-20 Fluorescein angiograms from individuals with diabetes were digitized for analysis using digital image-processing techniques.18-19 Computer algorithms were used to detect and count microaneurysms present in the fluorescein images.19 The accuracy, speed, and reproducibility of the computer techniques were assessed and compared with those of clinicians using both digitized and analog images. Manual counting procedures of microaneurysms used by the clinicians were laborious, time-consuming, and subject to human error.19 Digitization of fundus images enables the computer to discriminate microaneurysms from other features and to count the number of microaneurysms present. Computers are well suited for the extraction of quantitative information from such images because of their ability to process data in a fast and efficient manner with a high degree of reproducibility.19

Gardner et al21 recently tried to determine if neural networks could detect diabetic features in fundus images and compared the network with an ophthalmologist. They concluded that detection of normal vessels, exudates, and hemorrhages was possible, with success rates dependent on preprocessing and the number of images used in training.

The digital image processing method18-20 and the neural network method21-22 are 2 distinct methods used in pattern recognition. The image processing method is suitable for detecting and counting discrete objects such as hemorrhages and microaneurysms (HMA), hard exudates (HE), cotton-wool spots (CWS) (or soft exudates, etc. This method is particularly useful in deriving algorithms for recognizing small, vague objects, or like images in a nonuniformly illuminated medical image, such as the early lesions (HMA, HE, and CWS) of diabetic retinopathy in a color fundus image. On the other hand, the neural network method is suitable for solving pattern recognition problems involving general patterns such as the lesion patterns exhibited in fundus images, and for describing the various stages of the severity of diabetic retinopathy. Thus, the neural network method may be useful in grading diabetic retinopathy but not in detecting individual lesions.

Our computer system was designed to detect and quantify the following early retinal lesions from color fundus photographs: HMA, HE, and CWS. The system, which can be delivered using a diskette or via the Internet, was developed and tested using nearly 800 color
The computer system was well trained in diagnosing the early lesions of nonproliferative diabetic retinopathy, and the quality of these slides would be considered gradable by the Wisconsin Reading Center and acceptable by the computer system. These color retinal images were processed by the computer system to detect HMA, HE, and CWS. The sample size of 369 slides gave an 87% power at a significance level of .05 to detect a total disagreement rate of 10% with a 5% difference between the disagreement rates in the off-diagonal cells in the $3 \times 3$ or $2 \times 2$ tables.  

The system was refined by trial and error based on a retinal specialist's (R.M.K.) diagnosis (lesions present, questionable, or absent [Y/Q/N]). To determine whether the computer system was sufficiently trained, we first compared the diagnostic results of the retinal specialist with those of the computer system. If there was close agreement, we then compared the computer system with the grader and the ophthalmologist and compared these 2 human experts with the retinal specialist. If the agreement rates, sensitivity, specificity, and $\kappa$ statistics between the computer system and human experts were comparable with those obtained between the human experts (grader vs retinal specialist and general ophthalmologist vs retinal specialist), we concluded that the computer system had been well trained.

**TESTING PHASE**

A different set of 428 color fundus images (referred to as the testing set) were processed and examined by the computer system and its diagnostic results were compared with those made by the grader and the ophthalmologist. These testing slides were also obtained from the Vision Keeper participants with the same constraints as for the training slides.

In addition to the Y/Q/N diagnostic categories, we also made comparisons using just 2 categories (Y/N), excluding questionable lesions. We excluded the questionable category for 2 reasons. First, the ophthalmologist participating in this project did not use the questionable category in her diagnosis. Second, even though the grader used the questionable category, as did the retinal specialist, their agreement rates on questionable lesions were very low. Among the 369 slides, the grader diagnosed 31 questionable cases (including HMA, HE, and CWS), and the retinal specialist diagnosed 64 questionable cases. Only 2 cases were diagnosed as questionable by both experts.

The reasons for having a low agreement rate between the retinal specialist and the grader were (1) Different visual equipment might have been used to review the slides; therefore, differences in image size and quality might have affected the visibility of the lesions. (2) For vague lesions, the signal-to-noise ratio is low. To recognize a weak signal in a noisy environment is difficult, if not impossible, for humans and computers. (3) Somewhat different diagnostic criteria were used for questionable lesions.

For these reasons, it was also necessary to compare diagnostic agreement rates excluding questionable lesions. In addition to examining the agreement rates, $\kappa$ statistics, sensitivity, and specificity were computed for each comparison for the classification of Y/N. Values of $\kappa$ higher than 0.4 indicate moderate agreement, and values higher than 0.75 indicate excellent agreement.

**RESULTS**

**TRAINING PHASE**

*Table 1* presents the comparison between the computer system and the retinal specialist in diagnosing HMA, HE, and CWS, using Y/Q/N and Y/N. Eight of 369 training slides were considered nongradable by the specialist. The numbers in shaded cells are the questionable lesions diagnosed by either the specialist or the computer system. These questionable lesions were excluded when the Y/N classification was used. The agreement rates for detecting the 3 lesions between the computer system and the retinal specialist were excellent (*Table 2*). Sensitivity and specificity of the computer system were both high (range, 97%-100%). Therefore, we considered that the computer system was well trained in diagnosing the 3 early retinal lesions.

---

*Digital Retinal Image*  
*Image Quality Test*  
*Image Processing*  
*Pattern Recognition*  
*Essential Features (Optic Disc, Macula, Vessels)*  
*Early Lesions (HMA, HE, CWS)*  
*Early Diabetic Retinopathy*

**Figure 1.** The computer system for detecting early diabetic retinopathy. HMA indicates hemorrhages and microaneurysms; HE, hard exudates; and CWS, cotton-wool spots.

Results from comparing diagnoses among the retinal specialist, the grader, and the general ophthalmologist, and between the computer system and the same 2 human experts are presented in *Table 3*. Results showed that, when compared with the grader and oph-
thalmologist, the agreement rates of the computer system were comparable with those of the retinal specialist. Similarly, when using results from either the grader or ophthalmologist as gold standards, the sensitivity and specificity of the computer system were comparable with or exceeded those of the retinal specialist. Thus, it was concluded that the computer system was well trained.

**TESTING PHASE**

The results from the testing set of images were high as well (Tables 1 and 2). Using 2 human experts as standards, the computer system achieved high levels of sensitivity and specificity. It was clear from the comparisons that the agreement rates and $\kappa$ statistics were both higher when the classification consisted of only Y/N categories.

We developed a computer vision system designed to detect signs of early diabetic retinopathy. The system processed 2132 fundus slides that were also sent to the Wisconsin Reading Center for grading. The computer system rejected 159 slides due to poor quality while the Wisconsin Reading Center rejected 104 slides. The computer system was first trained using a set of 369
color fundus photographs. An additional set of 428 slides was then used to compare diagnoses made by the computer system with those made by a grader and a general ophthalmologist. The quality of these photographs was considered acceptable by both the grader and the computer system. In general, our experiments showed that the computer system was able to process all the photographs that the Wisconsin Reading Center considered gradable.

The 3 lesions (HMA, HE, and CWS) were chosen because the current computer system was developed to detect signs of early diabetic retinopathy. Work is still in progress to improve its detectability of these 3 lesions and to expand its capability to detect and quantify more severe lesions such as intraretinal microvascular abnormalities, venous beading, loops, and new retinal vessels.

A 78-D lens and indirect ophthalmoscope were used by the ophthalmologist to examine the patients. The only limitation of the computer system is that it does not have a good 3-dimensional stereoscopic view and may miss retinal thickening, which is not an issue in this article. It is not known whether there would be any difference in the results if a contact lens were used to detect these lesions, although most retinal experts agree that detection of any retinal lesion would be more accurate with a contact lens than a 78-D lens, even though these differences would be greater for macular edema compared with the lesions that were measured in this study.

Since the computer system was trained using the retinal specialist’s suggestions, it was expected that the agreement rates and other statistics between the retinal specialist and the computer system would be higher than those between the computer system and the other human experts. Besides the interobserver and intraobserver differences and different criteria used among human experts in diagnosing the 3 lesions, it was confirmed that the grader and the ophthalmologist often

---

### Table 1. Comparison Between Diagnoses Made by the Computer System and by the Retinal Specialist, the Grader, and the General Ophthalmologist*

<table>
<thead>
<tr>
<th>Computer System</th>
<th>Retinal Specialist</th>
<th>Testing Phase</th>
<th>Grader</th>
<th>General Ophthalmologist</th>
<th>Testing Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N Q Y Total</td>
<td>N Q Y Total</td>
<td>N Q Y Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>256 2 2 260</td>
<td>270 8 17 295</td>
<td>266 0 29 295</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>6 9 1 16</td>
<td>7 3 3 13</td>
<td>11 0 2 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>2 7 76 85</td>
<td>5 3 112 120</td>
<td>18 0 96 114</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>264 18 79 361</td>
<td>282 14 132 428</td>
<td>295 0 127 422</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>299 5 0 304</td>
<td>318 5 2 325</td>
<td>320 0 2 322</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>9 12 0 21</td>
<td>18 11 0 29</td>
<td>24 0 4 28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>6 7 23 36</td>
<td>12 6 56 74</td>
<td>24 0 48 72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>314 24 23 361</td>
<td>348 22 58 428</td>
<td>368 0 54 422</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>296 9 1 306</td>
<td>330 9 11 350</td>
<td>338 0 10 348</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td>3 5 1 9</td>
<td>5 3 2 10</td>
<td>7 0 3 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>9 8 29 46</td>
<td>4 6 57 67</td>
<td>24 0 40 64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>308 22 31 361</td>
<td>339 18 70 427</td>
<td>369 0 53 422</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Data are given as lesions diagnosed. HMA indicates hemorrhages and microaneurysms; N, lesions absent; Q, lesions questionable; Y, lesions present; HE, hard exudates; and CWS, cotton-wool spots.

### Table 2. Statistical Analyses Comparing the Computer System With the Human Experts*

<table>
<thead>
<tr>
<th></th>
<th>Retinal Specialist: Training Phase</th>
<th>Grader: Testing Phase</th>
<th>General Ophthalmologist: Testing Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agreement rate</td>
<td>Sensitivity†</td>
<td>Specificity†</td>
</tr>
<tr>
<td></td>
<td>k Statistic</td>
<td>Sensitivity‡</td>
<td>Specificity‡</td>
</tr>
<tr>
<td>HMA</td>
<td>0.911</td>
<td>97</td>
<td>99</td>
</tr>
<tr>
<td>HE</td>
<td>0.967</td>
<td>100</td>
<td>98</td>
</tr>
<tr>
<td>CWS</td>
<td>0.770</td>
<td>97</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td>0.875</td>
<td>0.837</td>
<td>0.745</td>
</tr>
<tr>
<td></td>
<td>0.745</td>
<td>0.710</td>
<td>0.633</td>
</tr>
<tr>
<td></td>
<td>0.837</td>
<td>0.749</td>
<td>0.665</td>
</tr>
<tr>
<td></td>
<td>0.749</td>
<td>0.710</td>
<td>0.749</td>
</tr>
<tr>
<td></td>
<td>0.837</td>
<td>0.749</td>
<td>0.665</td>
</tr>
</tbody>
</table>

*Data are given as percentage unless otherwise indicated. HMA indicates hemorrhages and microaneurysms; HE, hard exudates; CWS, cotton-wool spots; Y, lesions present; Q, lesions questionable; N, lesions absent; and ellipses, not applicable.

†Retinal specialist’s diagnosis was used as the standard.
‡Grader’s diagnosis was used as the standard.
§General ophthalmologist’s diagnosis was used as the standard.
used the presence of other lesions in different fields to
diagnose a specific lesion. This approach was not used in
the algorithm of the computer system. However, the
differences were not substantial.

The sensitivity and specificity of the computer sys-
tem when using the retinal specialist’s diagnosis as a
standard ranged from 97% to 100%, showing the potential
of the computer system to serve as a screening tool. When
using the other human experts as the standard, the com-
puter system was less sensitive, while specificity re-
ained high (range, 93%-100%). This demonstrated that
the computer system had a small false-positive rate be-
tween 0% and 7%.

The computer vision system is capable of learning
the diagnostic decision-making process of a human
expert with a range of accuracy between 91% and 94% when
the 3 diagnostic categories are used, and with a
range of accuracy between 97% and 99% when the
questionable lesions are excluded. At this time, since we
do not have a standard set of criteria in diagnosing
questionable lesions, the computer system has no standard
criteria to follow. Therefore, the sensitivity and
agreement rates between human experts and the com-
puter system, as well as between human experts, are
severely affected by the percentage of questionable
lesions diagnosed by either examiner. It is expected that
the higher the percentage of questionable lesions, the
lower the agreement rate.

The agreement rates, sensitivity, and specificity
between the computer system and the human experts
are comparable with those between human experts in
detecting early retinal lesions. The system is mobile,
electronically easily accessible (can be delivered by dis-
kette or via the Internet), and noninvasive. It displays
an enlarged image on the computer screen along with
the immediate diagnostic report (only a few seconds are
needed). For these reasons, this system could become a
useful clinical aid and a mass screening tool (when con-
ected to a real-time fundus camera) if other clinically
important lesions such as macular edema, intraretinal
microvascular abnormalities, and venous beading can
be detected. This would be especially useful in areas
where there is a lack of ophthalmologists. Further
research is needed.

The University of Oklahoma hopes to make this
computer system commercially available. While it
probably will not replace routine eye care, it may be
used as a clinical aid by displaying retinal images with
enhanced details and facilitating diagnosis. With
further development of the system, it is possible that it
could be used to detect progression of retinopathy in
clinical trials and epidemiological studies. We strongly
recommend that a set of standard diagnostic criteria
be established for computer detection and quantifica-
tion of retinal lesions, particularly the questionable
lesions.

Accepted for publication September 27, 2000.

This study was supported by grant U10EY09898 from
the National Eye Institute, Bethesda, Md.

We wish to thank the Vision Keepers Project partici-
ants for allowing us to use their fundus photographs in this
study and the reviewers of this article for their comments
and suggestions. We also thank Michelle Roberts, BA, and
Cathy Morales for their assistance.

Corresponding author and reprints: Samuel C. Lee,
PhD, School of Electrical and Computer Engineering, Univer-
sity of Oklahoma, 202 W Boyd, Norman, OK 73019
(e-mail: samlee@ou.edu).

REFERENCES

1. Centers for Disease Control and Prevention. Prevalence and incidence of diabe-
39:809-812.

Diabetes in America. 2nd ed. Bethesda, Md: National Institutes of Health and
National Institute of Diabetes and Digestive and Kidney Diseases; 1995:293-
339.

3. West KM, Erdreich LJ, Stober JA. A detailed study of risk factors for retinopathy

4. Newell SE, Tolbert B, Bennett J, Parsley TL. The prevalence and risk of diabetic
82:414-424.


6. Lee ET, Lee VS, Lum M, Russell D. Development of proliferative retinopathy in
41:359-367.

7. Leverton C. Grading diabetic retinopathy from a computer screen. J Audiov Me-

©2001 American Medical Association. All rights reserved.


