JAMA | Original Investigation

# Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes

Daniel Shu Wei Ting, MD, PhD; Carol Yim-Lui Cheung, PhD; Gilbert Lim, PhD; Gavin Siew Wei Tan, FRCSEd; Nguyen D. Quang, BEng; Alfred Gan, MSc; Haslina Hamzah, BSc; Renata Garcia-Franco, MD; Ian Yew San Yeo, FRCSEd; Shu Yen Lee, FRCSEd; Edmund Yick Mun Wong, FRCSEd; Charumathi Sabanayagam, MD, PhD; Mani Baskaran, MD, PhD; Farah Ibrahim, MB, BCh, BAO; Ngiap Chuan Tan, MCI, FAMS; Eric A. Finkelstein, MHA, PhD; Ecosse L. Lamoureux, PhD; Ian Y. Wong, FRCOph; Neil M. Bressler, MD; Sobha Sivaprasad, FRCOph; Rohit Varma, MD, MPH; Jost B. Jonas, MD, PhD; Ming Guang He, MD, PhD; Ching-Yu Cheng, MD, PhD; Gemmy Chui Ming Cheung, FRCOph; Tin Aung, MD, PhD; Wynne Hsu, PhD; Mong Li Lee, PhD; Tien Yin Wong, MD, PhD

**IMPORTANCE** A deep learning system (DLS) is a machine learning technology with potential for screening diabetic retinopathy and related eye diseases.

**OBJECTIVE** To evaluate the performance of a DLS in detecting referable diabetic retinopathy, vision-threatening diabetic retinopathy, possible glaucoma, and age-related macular degeneration (AMD) in community and clinic-based multiethnic populations with diabetes.

**DESIGN, SETTING, AND PARTICIPANTS** Diagnostic performance of a DLS for diabetic retinopathy and related eye diseases was evaluated using 494 661 retinal images. A DLS was trained for detecting diabetic retinopathy (using 76 370 images), possible glaucoma (125 189 images), and AMD (72 610 images), and performance of DLS was evaluated for detecting diabetic retinopathy (using 112 648 images), possible glaucoma (71 896 images), and AMD (35 948 images). Training of the DLS was completed in May 2016, and validation of the DLS was completed in May 2017 for detection of referable diabetic retinopathy (moderate nonproliferative diabetic retinopathy or worse) and vision-threatening diabetic retinopathy (severe nonproliferative diabetic retinopathy or worse) using a primary validation data set in the Singapore National Diabetic Retinopathy Screening Program and 10 multiethnic cohorts with diabetes.

**EXPOSURES** Use of a deep learning system.

**MAIN OUTCOMES AND MEASURES** Area under the receiver operating characteristic curve (AUC) and sensitivity and specificity of the DLS with professional graders (retinal specialists, general ophthalmologists, trained graders, or optometrists) as the reference standard.

**RESULTS** In the primary validation dataset (n = 14 880 patients; 71 896 images; mean [SD] age, 60.2 [2.2] years; 54.6% men), the prevalence of referable diabetic retinopathy was 3.0%; vision-threatening diabetic retinopathy, 0.6%; possible glaucoma, 0.1%; and AMD, 2.5%. The AUC of the DLS for referable diabetic retinopathy was 0.936 (95% CI, 0.925-0.943), sensitivity was 90.5% (95% CI, 87.3%-93.0%), and specificity was 91.6% (95% CI, 91.0%-92.2%). For vision-threatening diabetic retinopathy, AUC was 0.958 (95% CI, 0.956-0.961), sensitivity was 100% (95% CI, 94.1%-100.0%), and specificity was 91.1% (95% CI, 90.7%-91.4%). For possible glaucoma, AUC was 0.942 (95% CI, 0.929-0.954), sensitivity was 96.4% (95% CI, 81.7%-99.9%), and specificity was 87.2% (95% CI, 86.8%-87.5%). For AMD, AUC was 0.931 (95% CI, 0.928-0.935), sensitivity was 93.2% (95% CI, 91.1%-99.8%), and specificity was 88.7% (95% CI, 88.3%-89.0%). For referable diabetic retinopathy in the 10 additional datasets, AUC range was 0.889 to 0.983 (n = 40 752 images).

**CONCLUSIONS AND RELEVANCE** In this evaluation of retinal images from multiethnic cohorts of patients with diabetes, the DLS had high sensitivity and specificity for identifying diabetic retinopathy and related eye diseases. Further research is necessary to evaluate the applicability of the DLS in health care settings and the utility of the DLS to improve vision outcomes.

*JAMA*. 2017;318(22):2211-2223. doi:10.1001/jama.2017.18152

Related article page 2199

Supplemental content

**Author Affiliations:** Author affiliations are listed at the end of this article.

**Corresponding Author:** Tien Yin Wong, MD, PhD, Singapore National Eye Center, 11 Third Hospital Ave, Singapore 168751 (wong.tien.yin@snec.com.sg).

By 2040, it is projected that approximately 600 million people will have diabetes, with one-third expected to have diabetic retinopathy.[1-3] Screening for diabetic retinopathy, coupled with timely referral and treatment, is a universally accepted strategy for blindness prevention.[2] However, programs for screening diabetic retinopathy are challenged by issues related to implementation, availability of human assessors, and long-term financial sustainability.[2,4-7]

A deep learning system (DLS) uses artificial intelligence and representation learning methods to process large data and extract meaningful patterns.[8,9] A few DLSs have recently shown high sensitivity and specificity (>90%) in detecting referable diabetic retinopathy from retinal photographs, primarily using high-quality images from publicly available databases from homogenous populations of white individuals.[10-12] The performance of a DLS in screening for diabetic retinopathy should ideally be evaluated in clinical or population settings in which retinal images from patients of different races and ethnicities (and therefore with varying fundi pigmentation) have varying qualities (eg, due to poor pupil dilation, media opacity, poor contrast or focus).[13,14] Furthermore, in screening programs for diabetic retinopathy, the detection of incidental but related vision-threatening eye diseases, such as glaucoma and age-related macular degeneration (AMD), should be incorporated because missing such cases is clinically unacceptable.[15]

The primary aim of this study was to train and validate a DLS to detect referable diabetic retinopathy, vision-threatening diabetic retinopathy, and related eye diseases (referable possible glaucoma and referable AMD) by evaluating retinal images obtained primarily from patients with diabetes in an ongoing community-based national diabetic retinopathy screening program in Singapore, with further external validation on referable diabetic retinopathy in 10 additional multiethnic datasets from different countries with diverse community- and clinic-based populations with diabetes. The secondary aim was to determine how the DLS could fit in 2 potential models of diabetic retinopathy screening—a fully automated model for communities with no existing screening programs and a semiautomated model in which referable cases from the DLS undergo a secondary assessment by human graders.

### Key Points

**Question** How does a deep learning system (DLS) using artificial intelligence compare with professional human graders in identifying diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes?

**Findings** In the primary validation dataset (71 896 images; 14 880 patients), the DLS had a sensitivity of 90.5% and specificity of 91.6% for detecting referable diabetic retinopathy; 100% sensitivity and 91.1% specificity for vision-threatening diabetic retinopathy; 96.4% sensitivity and 87.2% specificity for possible glaucoma; and 93.2% sensitivity and 88.7% specificity for age-related macular degeneration, compared with professional graders.

**Meaning** The DLS had high sensitivity and specificity for identifying diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes.

## Methods

This study was approved by the centralized institutional review board (IRB) of SingHealth, Singapore (protocol SHF/FG648S/2015) and conducted in accordance with the Declaration of Helsinki. Information on race/ethnicity was collected to evaluate the consistency of DLS diagnostic performance across races/ethnicities. Patients' informed consent was exempted by the IRB because of the retrospective nature of study using fully anonymized retinal images.

### Training Datasets of the DLS

The DLS for referable diabetic retinopathy was developed and trained using retinal images of patients with diabetes who participated in the ongoing Singapore National Diabetic Retinopathy Screening Program (SIDRP) between 2010 and 2013 (SIDRP 2010-2013; **Table 1** and **Table 2**). The SIDRP was established from 2010, progressively covered all 18 primary care clinics across Singapore, and screened half of the diabetes population by 2015.[16] SIDRP uses digital retinal photography, a tele-ophthalmology platform, and assessment of diabetic retinopathy by a team of trained professional graders. For each patient, 2 retinal photographs (optic disc and fovea) were taken of each eye. All trained graders received 3 to 6 months of training before certification and underwent annual reaccreditation. Specifically for this study, in the training set (SIDRP 2010-2013), each retinal image was analyzed by 2 trained senior certified nonmedical professional graders (>5 years' experience)[17]; if there were discordant findings between the nonmedical professional graders, arbitration was performed by a retinal specialist (PhD-trained with >5 years' experience in conducting diabetic retinopathy assessment) to generate final grading.

For referable possible glaucoma and AMD, the DLS was trained using images from SIDRP 2010-2013 and several additional population- and clinic-based studies of patients with glaucoma and AMD (Table 1; eTable 1 in the Supplement).[17-20,22,23,25-29]

### Architecture of the DLS

The DLS consisted of a convolutional neural network to implicitly recognize characteristics of referable diabetic retinopathy, possible glaucoma, and AMD from the appearance in retinal images. Training of the DLS entailed exposure of multiple examples of retinal images (with and without each of the 3 conditions) to the neural networks, allowing the networks to gradually adapt their weight parameters to model and differentiate between conditions. Once the training was complete, the DLS could be used to classify unseen images. Technical details are shown in eFigure 1 in the Supplement.

### Validation Datasets

Details of validation datasets are described in Table 1. For diabetic retinopathy, the primary validation dataset was the

Table 1. Summary of the Training and Validation Datasets for Referable Diabetic Retinopathy, Referable Possible Glaucoma, and Referable Age-Related Macular Degeneration

| Source Datasets (Location) | Race/Ethnicity | Cohort | Camera | Assessor | No. Images | No. Eyes | No. Patients |
|---|---|---|---|---|---|---|---|
| **Referable Diabetic Retinopathy** | | | | | | | |
| Training | | | | | | | |
| Singapore National Diabetic Retinopathy Screening Program 2010-2013[16] (Singapore) | Chinese, Malay, Indian | Community-based | Topcon | 2 Professional senior graders; arbitration by 1 retinal specialist | 76 370 | 38 185 | 13 099 |
| Primary validation | | | | | | | |
| Singapore National Diabetic Retinopathy Screening Program 2014-2015[16] (Singapore) | Chinese, Malay, Indian | Community-based | Topcon | 1 Retinal specialist (criterion standard); 2 professional senior graders | 71 896 | 35 948 | 14 880 |
| External validation | | | | | | | |
| Guangdong (China) | Chinese | Community-based | FundusVue | 2 Graders; arbitration by 1 retinal specialist | 15 798 | 7899 | 3970 |
| Singapore Malay Eye Study[17-20] (Singapore) | Malay | Population-based | Canon | 1 Professional senior grader; 1 retinal specialist | 3052 | 1526 | 763 |
| Singapore Indian Eye Study[17-20] (Singapore) | Indian | Population-based | Canon | 1 Professional senior grader; 1 retinal specialist | 4512 | 2256 | 1128 |
| Singapore Chinese Eye Study[17-20] (Singapore) | Chinese | Population-based | Canon | 1 Professional senior grader; 1 retinal specialist | 1936 | 968 | 484 |
| Beijing Eye Study[21] (China) | Chinese | Population-based | Canon | 2 Board-certified ophthalmologists | 1052 | 526 | 263 |
| African American Eye Disease Study[22] (United States) | African American | Population-based | Topcon | 2 Retinal specialists | 1968 | 968 | 484 |
| Royal Victoria Eye and Ear Hospital[23] (Australia) | White | Clinic-based | Topcon | 2 Graders | 2302 | 1151 | 588 |
| Mexican (Mexico) | Hispanic | Clinic-based | Topcon | 2 Retinal specialists | 1172 | 586 | 343 |
| Chinese University of Hong Kong[24] (Hong Kong) | Chinese | Clinic-based | Topcon | 2 Retinal specialists | 1254 | 627 | 314 |
| University of Hong Kong (Hong Kong) | Chinese | Clinic-based | Carl Zeiss | 2 Optometrists | 7706 | 3853 | 1932 |
| Categorical total, validation | | | | | 112 648 | 56 324 | 15 157 |
| Categorical total, training and validation | | | | | 189 018 | 94 509 | 38 256 |
| **Referable Possible Glaucoma** | | | | | | | |
| Training | | | | | | | |
| Singapore National Diabetic Retinopathy Screening Program 2010-2013[16] (Singapore) | Chinese, Malay, Indian | Community-based | Topcon | 2 Professional senior graders; arbitration by 1 retinal specialist | 76 108 | 38 185 | 13 099 |
| Singapore Malay Eye Study[25] (Singapore) | Malay | Population-based | Canon | 1 Professional senior grader; 1 retinal specialist | 10 114 | 6560 | 3280 |
| Singapore Indian Eye Study[25] (Singapore) | Indian | Population-based | Canon | 1 Professional senior grader; 1 retinal specialist | 10 819 | 6800 | 3400 |
| Singapore Chinese Eye Study[25] (Singapore) | Chinese | Population-based | Canon | 1 Professional senior grader; 1 retinal specialist | 26 731 | 6706 | 3353 |
| Singapore National Eye Centre, Glaucoma (Singapore) | Chinese, Malay, Indian | Clinic-based | Topcon | 2 Glaucoma specialists | 1417 | 1365 | 846 |
| Categorical total, training | | | | | 125 189 | 59 616 | 23 978 |
| Primary validation | | | | | | | |
| Singapore National Diabetic Retinopathy Screening Program 2014-2015[16] (Singapore) | Chinese, Malay, Indian | Community-based | Topcon | 1 Retinal specialist; 2 professional senior graders | 71 896 | 35 948 | 14 880 |
| Categorical total, training and validation | | | | | 197 085 | 95 564 | 38 858 |

(continued)

Table 1. Summary of the Training and Validation Datasets for Referable Diabetic Retinopathy, Referable Possible Glaucoma, and Referable Age-Related Macular Degeneration (continued)

| Source Datasets (Location) | Race/Ethnicity | Cohort | Camera | Assessor | No. | | |
|---|---|---|---|---|---|---|---|
| | | | | | Images | Eyes | Patients |
| Referable Age-Related Macular Degeneration | | | | | | | |
| Training | | | | | | | |
| Singapore National Diabetic Retinopathy Screening Program 2010-2013[16] (Singapore) | Chinese, Malay, Indian | Community-based | Topcon | 2 Professional senior graders; arbitration by 1 retinal specialist | 38 185 | 38 185 | 13 099 |
| Singapore Malay Eye Study[26] (Singapore) | Malay | Population-based | Canon | 1 Professional senior grader; 1 retinal specialist | 8616 | 6560 | 3280 |
| Singapore Indian Eye Study[26] (Singapore) | Indian | Population-based | Canon | 1 Professional senior grader; 1 retinal specialist | 7447 | 6800 | 3400 |
| Singapore Chinese Eye Study[26] (Singapore) | Chinese | Population-based | Canon | 1 Professional senior grader; 1 retinal specialist | 16 182 | 6706 | 3353 |
| Singapore National Eye Centre, AMD[27-29] (Singapore) | Chinese, Malay, Indian | Clinic-based | Topcon | 2 retinal specialists | 2180 | 348 | 174 |
| Categorical total, training | | | | | 72 610 | 58 599 | 23 306 |
| Primary validation | | | | | | | |
| Singapore National Diabetic Retinopathy Screening Program 2014-2015[16] (Singapore) | Chinese, Malay, Indian | Community-based | Topcon | 1 retinal specialist; 2 trained professional senior graders | 35 948 | 35 948 | 14 880 |
| Categorical total, training and validation | | | | | 108 558 | 94 547 | 38 189 |
| **Total images for referable diabetic retinopathy, referable possible glaucoma, and referable age-related macular degeneration training and validation** | | | | | **494 661** | **111 538** | **46 934** |

same SIDRP among patients seen between 2014 and 2015 (SIDRP 2014-2015). The primary analysis was to determine if the DLS was equivalent or better than 2 trained senior nonmedical professional graders (>5 years' experience) currently employed in the SIDRP in detecting referable diabetic retinopathy and vision-threatening diabetic retinopathy, with reference to a retinal specialist (>5 years' experience in diabetic retinopathy grading).

The DLS was then externally validated using 10 additional multiethnic cohorts of participants with diabetes from different settings (community, population-based, and clinic-based). A range of retinal cameras were used, and assessment of diabetic retinopathy was facilitated by retinal specialists, general ophthalmologists, trained nonmedical professional graders, or optometrists across the cohorts (Table 1). All retinal images were captured with JPEG compression format (resolutions 5-7 megapixels, except for images of eyes in the Hispanic cohort [<1 megapixel]).

## Training, Experience, and Credentials of the Grading Team for External Validation Datasets

Guangdong: 5 nonmedical United Kingdom–certified professional graders (>2 years' experience), supervised by 1 retinal specialist (>10 years' experience). Singapore Malay Eye Study, Singapore Indian Eye Study, and Singapore Chinese Eye Study: 1 certified professional senior grader (>7 years' experience), supervised by 2 senior retinal specialists from Australia (>15 years' experience). Beijing Eye Study: 4 Chinese board-certified ophthalmologists (>5 years' experience), supervised by 2 retinal specialists (>20 years' experience). African American Eye Study: 2 retinal specialists (>5 years' experience). Royal Victorian Eye and Ear Hospital: 4 professional senior graders (>7 years' experience). Mexican study: 2 retinal specialists (>5 years' experience). Chinese University of Hong Kong: 3 retinal specialists (>6 years' experience). The University of Hong Kong: 6 optometrists (>4 years' experience). Singapore National Eye Center Glaucoma Study: 3 glaucoma specialists (>5 years' experience). Singapore National Eye Center AMD Phenotyping Study: 10 retinal specialists (>5 years' experience).

## Definition of Referable Diabetic Retinopathy, Vision-Threatening Diabetic Retinopathy, Referable Possible Glaucoma, and Referable AMD

Diabetic retinopathy levels from all retinal images were defined using the International Classification Diabetic Retinopathy Scale.[30] Referable diabetic retinopathy was defined as a diabetic retinopathy severity level of moderate nonproliferative diabetic retinopathy or worse, diabetic macular edema, and/or ungradable image. Vision-threatening diabetic retinopathy was defined as severe nonproliferative diabetic retinopathy and proliferative diabetic retinopathy. Diabetic macular edema was assessed as present if hard exudates were detected at the posterior pole of the retinal images. If more than one-third of the photograph was obscured, it was considered ungradable and the individual was considered referable. Referable possible glaucoma was defined as a ratio of vertical cup to disc diameter of 0.8

Table 2. Training and Validation Datasets for Diabetic Retinopathy on Training, Primary Validation, and 10 External Validation Datasets[a]

| | No. | | | No. (%) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Nonreferable Eyes | | Referable Eyes[b] | | | | |
| Datasets | Patients | Images | Eyes | No Diabetic Retinopathy | Mild Nonproliferative Diabetic Retinopathy | Moderate Nonproliferative Diabetic Retinopathy | Severe Nonproliferative Diabetic Retinopathy[c] | Proliferative Diabetic Retinopathy[c] | Diabetic Macular Edema | Ungradable |
| **Training** | | | | | | | | | | |
| Singapore National Diabetic Retinopathy Screening Program 2010-2013[16,d] | 13 099 | 76 370 | 38 185 | 33 709 (88.3) | 3310 (8.7) | 597 (1.6) | 478 (1.3) | 70 (0.2) | 2026 (5.3) | 21 (0.1) |
| **Primary Validation** | | | | | | | | | | |
| Singapore National Diabetic Retinopathy Screening Program 2014-2015[16,e] | 14 880 | 71 896 | 35 948 | 33 087 (92.0) | 1808 (5.0) | 455 (1.3) | 170 (0.5) | 24 (0.1) | 320 (0.9) | 404 (1.1) |
| **External Validation** | | | | | | | | | | |
| Community-based | | | | | | | | | | |
| Guangdong[f] | 3970 | 15 798 | 7899 | 5665 (71.7) | 1235 (15.6) | 737 (9.3) | 0 | 154 (1.9) | 0 | 108 (1.4) |
| Population-based | | | | | | | | | | |
| Singapore Malay Eye Study[17-20] | 763 | 3052 | 1526 | 1143 (74.9) | 215 (14.1) | 113 (7.4) | 18 (1.2) | 9 (0.6) | 53 (3.5) | 28 (1.8) |
| Singapore Indian Eye Study[17-20] | 1128 | 4512 | 2256 | 1639 (72.7) | 422 (18.7) | 125 (5.5) | 5 (0.2) | 17 (0.8) | 71 (3.1) | 48 (2.1) |
| Singapore Chinese Eye Study[17-20] | 484 | 1936 | 968 | 759 (78.4) | 131 (13.5) | 60 (6.2) | 1 (0.1) | 7 (0.7) | 17 (1.8) | 10 (1) |
| Beijing Eye Study[21] | 263 | 1052 | 526 | 493 (93.7) | 4 (0.8) | 11 (2.1) | 4 (0.8) | 0 | 12 (2.3) | 2 (0.4) |
| African American Eye Disease Study[22] | 492 | 1968 | 984 | 807 (82.0) | 50 (5.1) | 37 (3.8) | 5 (0.5) | 16 (1.6) | 28 (2.85) | 41 (4.17) |
| Clinic-based | | | | | | | | | | |
| Royal Victoria Eye and Ear Hospital[23] | 588 | 2302 | 1151 | 432 (37.5) | 121 (10.5) | 159 (13.8) | 123 (10.7) | 191 (16.6) | 249 (21.6) | 125 (10.9) |
| Mexican | 343 | 1172 | 586 | 38 (6.5) | 284 (48.5) | 192 (32.8) | 51 (8.7) | 18 (3.1) | 223 (38.1) | 3 (0.5) |
| Chinese University of Hong Kong[24] | 314 | 1254 | 627 | 224 (35.7) | 114 (18.2) | 235 (37.5) | 43 (6.9) | 11 (1.8) | 96 (15.3) | 0 |
| University of Hong Kong[g] | 1932 | 7706 | 3853 | 1984 (51.5) | 1485 (38.5) | 155 (4.0) | 14 (0.4) | 0 | 214 (5.55) | 1 (0.03) |
| **Total** | **38 253** | **189 018** | **94 509** | **79 980** (84.9) | **9179** (9.74) | **2876** (3.04) | **912** (0.97) | **517** (0.55) | **3309** (3.50) | **791** (0.41) |

[a] For study locations and race/ethnicity data, see Table 1.

[b] Referable diabetic retinopathy is defined as moderate nonproliferative diabetic retinopathy, severe nonproliferative diabetic retinopathy, proliferative diabetic retinopathy, and diabetic macular edema.

[c] Vision-threatening diabetic retinopathy is defined as severe nonproliferative diabetic retinopathy and proliferative diabetic retinopathy.

[d] Nine patients had only 1 eye.

[e] In the Singapore Diabetic Retinopathy Screening Program 2014-2015, there were 6291 patients who were repeats from the Singapore Diabetic Retinopathy Screening Program 2010-2013, and 8589 were unique patients.

[f] Forty-one patients had only 1 eye.

[g] Eleven patients had only 1 eye.

Table 3. Demographics, Diabetes History, and Systemic Risk Factors of Patients Attending the Singapore National Diabetes Retinopathy Screening Program Between 2010 to 2013 (Training Dataset) and 2014 to 2015 (Primary Validation Dataset)

| Demographics and Vascular Risk Factors | Primary Training Dataset (SIDRP 2010-2013) | Primary Validation Dataset (SIDRP 2014-2015) |
|---|---|---|
| No. of retinal images | 76 370 | 71 896 |
| No. of eyes | 38 185 | 35 948 |
| No. of patients | 13 099 | 14 880[a] |
| Age, mean (SD), y | 62.77 (11.32) | 60.16 (12.19) |
| Men, No. (%) | 6518 (49.76) | 4334 (51.02) |
| Race/ethnicity, No. (%) | | |
| Chinese | 9615 (73.79) | 6160 (72.51) |
| Indian | 1427 (10.95) | 1037 (12.21) |
| Malay | 1582 (12.14) | 1020 (12) |
| Other | 407 (3.12) | 278 (3.27) |
| Systemic risk factors, mean (SD) | | |
| Body mass index[b] | 26.54 (4.69) | 27.22 (4.99) |
| Diabetes duration, median (IQR), y | 6.4 (1.6-8.7) | 3.7 (0.4-6.1) |
| Blood pressure, mm Hg | | |
| Systolic | 129.9 (16.85) | 132.05 (17.57) |
| Diastolic | 70.46 (10.06) | 72.77 (10.78) |
| $HbA_{1c}$, % | 7.25 (1.41) | 7.54 (1.88) |
| Lipids, mg/dL | | |
| Total cholesterol | 81.90 (17.28) | 83.70 (19.26) |
| HDL cholesterol | 24.12 (6.48) | 23.58 (6.30) |
| LDL cholesterol | 45.54 (14.94) | 46.98 (15.66) |
| Triglycerides | 27.36 (16.20) | 30.24 (39.96) |
| Creatinine, mg/dL | 0.92 (0.42) | 0.85 (0.41) |

Abbreviations: $HbA_{1c}$, glycated hemoglobin; HDL, high-density lipoprotein; IQR, interquartile range; LDL, low-density lipoprotein; SIDRP, Singapore National Diabetic Retinopathy Screening Program.

SI conversion factors: To convert values for creatinine to µmol/L, multiply by 88.4; total cholesterol, HDL cholesterol, and LDL cholesterol values to mmol/L, multiply by 0.0259; triglyceride values to mmol/L, multiply by 0.0113.

[a] In the Singapore National Diabetic Retinopathy Screening Program 2014-2015, a total of 14 880 patients visited the primary eye care clinics for diabetic retinopathy screening. Of those, 6291 were follow-up patients who attended the Singapore National Diabetic Retinopathy Screening Program 2010-2013 and were excluded for analysis in eTable 3 in the Supplement to eliminate the risk of overfitting diagnostic performance of the deep learning system.

[b] Body mass index was calculated as weight in kilograms divided by height in meters squared.

or greater, focal thinning or notching of the neuroretinal rim, optic disc hemorrhages, or localized retinal nerve fiber layer defects—features sometimes referred to as glaucoma suspects. Referable AMD was defined as the presence of intermediate AMD (numerous medium-sized drusen, 1 large drusen ≥125 µm in greatest linear diameter, noncentral geographical atrophy, and/or advanced AMD [central geographical atrophy or neovascular AMD]) according to the Age-Related Eye Disease Study grading system.[31]

## Reference Standards

For the primary validation dataset (SIDRP 2014-2015), the reference standard was grading by a retinal specialist (>5 years' experience in conducting diabetic retinopathy assessment) who was masked to the grading of the trained nonmedical professional graders. For all other retinal images from the 10 external validation datasets, reference standards were based on individual studies' assessment of diabetic retinopathy, which was based on retinal specialists, general ophthalmologists, trained nonmedical professional graders, or optometrists (Table 1). The DLS performance for identifying referable diabetic retinopathy in the 10 external validation datasets was compared against these reference standards. For the analysis on referable possible glaucoma and referable AMD, the reference standard was the retinal specialist (Table 1).

## Statistical Analysis

Initially the area under the curve (AUC) of the receiver operating characteristic (ROC) curve of DLS was calculated on the training dataset of the SIDRP 2010-2013 across a range of classification thresholds, and one was selected that achieved a predetermined optimal sensitivity of 90% for detecting referable diabetic retinopathy, vision-threatening diabetic retinopathy, referable possible glaucoma, and referable AMD. For diabetic retinopathy screening, international guidelines recommended a minimum sensitivity of 60% (Australia) to 80% (United Kingdom).[32,33] In Singapore, the DLS sensitivity was preset at 90% based on the trained professional graders' past performances and criteria set by the Ministry of Health, Singapore. The hypothesis determined was that the DLS was at least comparable to the professional graders' performance.

Primary analysis was to evaluate the performance of the DLS in the setting of the ongoing SIDRP 2014-2015 (the primary validation set) by determining whether the DLS was equivalent or superior to professional graders in the screening program. Thus, the AUC, sensitivity, and specificity of the DLS vs the professional graders in detecting referable diabetic retinopathy and vision-threatening diabetic retinopathy was computed to the reference standard (retinal specialist) at individual-eye levels.

Next, the following subsidiary analyses were performed: (1) the analyses were repeated excluding patients who appeared in both the SIDRP 2010-2013 training set and the primary validation set of SIDRP 2014-2015 (n = 6291 seen more than once in SIDRP), with the patient treated as having referable diabetic retinopathy if either eye had referable diabetic retinopathy; (2) performance of the DLS was evaluated using higher-quality images with no media opacity (eg, cataracts) as noted by professional graders; (3) AUC subgroups were computed stratified by age, sex, and glycemic control; and (4) the analysis was repeated by calculating the AUC, sensitivity, and specificity of the DLS and the proportion of concordant and discordant eyes on the 10 external validation datasets, compared with the reference standards in these studies (retinal specialists, general ophthalmologists, trained graders, or optometrists; Table 1).

The DLS performance was then evaluated in detection of referable possible glaucoma and referable AMD, with reference to a retinal specialist, using the primary validation dataset (SIDRP 2014-2015).

Table 4. Primary Validation Dataset Showing the Area Under the Curve, Sensitivity, and Specificity of the Deep Learning System vs Trained Professional Graders in Patients With Diabetes, SIDRP 2014-2015, With Reference to a Retinal Specialist's Grading

| | Value (95% CI)[a] | | |
|---|---|---|---|
| | Deep Learning System | Trained Professional Graders | *P* Value[b] |
| Referable diabetic retinopathy[c] | | | |
| Area under the curve[d] | 0.936 (0.925-0.943) | | |
| Sensitivity, % | 90.5 (87.3-93.0) | 91.2 (88.0-93.6) | .68 |
| Specificity, % | 91.6 (91.0-92.2) | 99.3 (99.2-99.4) | <.001 |
| Vision-threatening diabetic retinopathy[e] | | | |
| Area under the curve[d] | 0.958 (0.956-0.961) | | |
| Sensitivity, % | 100 (94.1-100.0)[f] | 88.5 (75.3-95.1) | <.001 |
| Specificity, % | 91.1 (90.7-91.4) | 99.6 (99.6-99.7) | <.001 |

[a] Eyes were the units of analysis (n = 35 948). Asymptotic 95% CI was computed for the logit of each proportion and using the cluster sandwich estimator of standard error to account for possible dependency of eyes within each individual (exception, sensitivity calculation for the deep learning system).

[b] *P* value was calculated between the deep learning system vs trained professional graders using the McNemar test.

[c] Referable diabetic retinopathy was defined as moderate nonproliferative diabetic retinopathy, severe nonproliferative diabetic retinopathy, proliferative diabetic retinopathy, diabetic macular edema, and ungradable eye.

[d] Cluster-bootstrap, biased-corrected 95% CI was computed for each area under the curve, with individual patients as the bootstrap sampling clusters.

[e] Vision-threatening diabetic retinopathy was defined as severe nonproliferative diabetic retinopathy and proliferative diabetic retinopathy.

[f] Exact Clopper-Pearson left-sided 97.5% CI was calculated owing to estimate being at the boundary.

For a secondary aim, an examination of how the DLS could fit in 2 potential diabetic retinopathy screening models was performed: a fully-automated model for communities with no existing screening programs, vs a semiautomated model in which referable cases from the DLS have a secondary assessment by human graders—a method currently used in some communities and countries (eg, United States, United Kingdom, and Singapore) (eFigure 2 in the Supplement).[32-35] For this analysis, in the fully-automated model, eyes were considered referable if any one of the 3 conditions (referable diabetic retinopathy, referable possible glaucoma, or referable AMD) were present. In the semiautomated model, eyes classified as referable by the DLS would undergo a secondary assessment by trained professional graders to reclassify eyes if necessary. For semiautomated models, evaluation was made of the proportion of images requiring secondary assessment when presetting the DLS sensitivity threshold at 90%, 95%, and 99% in detection of referable status.

Cluster-bootstrap, biased-corrected, asymptotic 2-sided 95% CIs adjusted for clustering by patients were calculated and presented for proportions (sensitivity, specificity) and AUC, respectively. In a few exceptional cases with estimate of sensitivity at the boundary of 100%, the exact Clopper-Pearson method was used instead to obtain CI estimates.[36]

All hypotheses tested were 2-sided, and a *P* value of less than .05 was considered statistically significant. No adjustment for multiple comparisons was made because the study was restricted to a small number of planned comparisons. All analyses were performed using Stata version 14 (StataCorp).

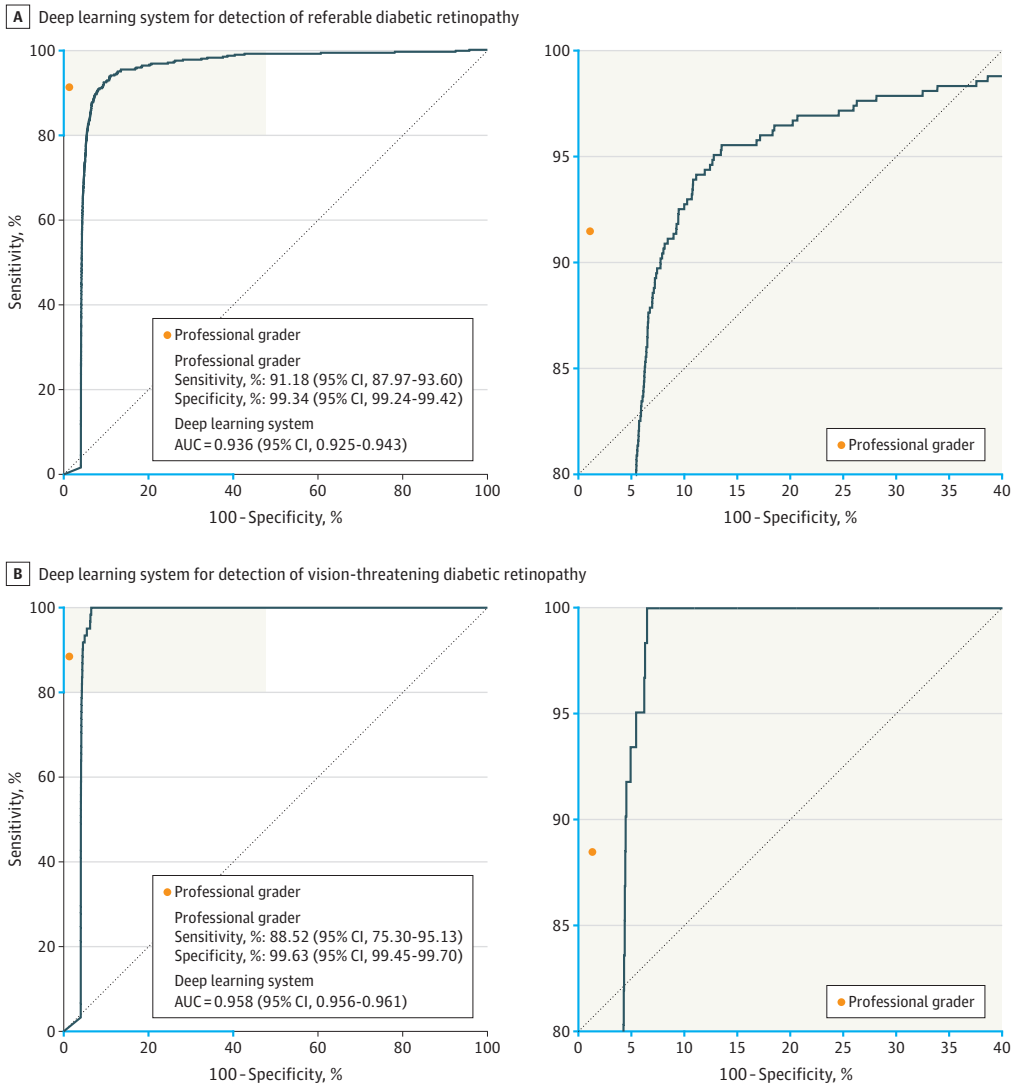
## Results

From a total of 494 661 retinal images, the DLS was trained for detection of referable diabetic retinopathy (using 76 370 images), referable possible glaucoma (using 125 189 images), and referable AMD (using 72 610 images); performance of the DLS was evaluated using 112 648 images for detection of referable diabetic retinopathy, 71 896 images for referable possible glaucoma, and 35 948 images for referable AMD. All images were assembled between January 2016 and March 2017 (Table 1), the DLS training was completed in May 2016, and validation was completed in May 2017. Among 76 370 images in the training dataset, 11.7% demonstrated any diabetic retinopathy, 5.3% referable diabetic retinopathy, and 1.5% vision-threatening diabetic retinopathy. In the primary validation dataset, estimates were 8.0% for having any diabetic retinopathy, 3.0% for referable diabetic retinopathy, and 0.6% for vision-threatening diabetic retinopathy (n = 71 896 images). In the 10 external validation datasets, estimates were 35.3% for any diabetic retinopathy, 15.4% for referable diabetic retinopathy, and 3.4% for vision-threatening diabetic retinopathy (n = 40 752 images; Table 2). For possible glaucoma, 2630 images (1907 eyes) were considered referable; for AMD, 2900 images (1017 eyes) were considered referable (eTable 1 in the Supplement).

The overall patients demographics, diabetes history, and systemic risk factors of the training and validation datasets are listed in **Table 3** (SIDRP 2010-2013 and SIDRP 2014-2015, primary validation set) and eTable 2 in the Supplement (10 external validation datasets for referable diabetic retinopathy and training datasets for referable possible glaucoma and referable AMD).

The diagnostic performance of the DLS as compared with trained professional graders, both with reference to the retinal specialist standard using this primary validation dataset, is shown in **Table 4**. The AUC of the DLS was 0.936 for referable diabetic retinopathy and 0.958 for vision-threatening diabetic retinopathy **(Figure 1)**. Sensitivity of the DLS in detecting referable diabetic retinopathy was comparable with that of trained graders (90.5% vs 91.1%; *P* = .68), although the

Figure 1. Receiver Operating Characteristic Curve and Area Under the Curve of the Deep Learning System for Detection of Referable Diabetic Retinopathy and Vision-Threatening Diabetic Retinopathy in the Singapore National Diabetic Retinopathy Screening Program (SIDRP 2014-2015; Primary Validation Dataset), Compared with Professional Graders' Performance, With Retinal Specialists' Grading as Reference Standard



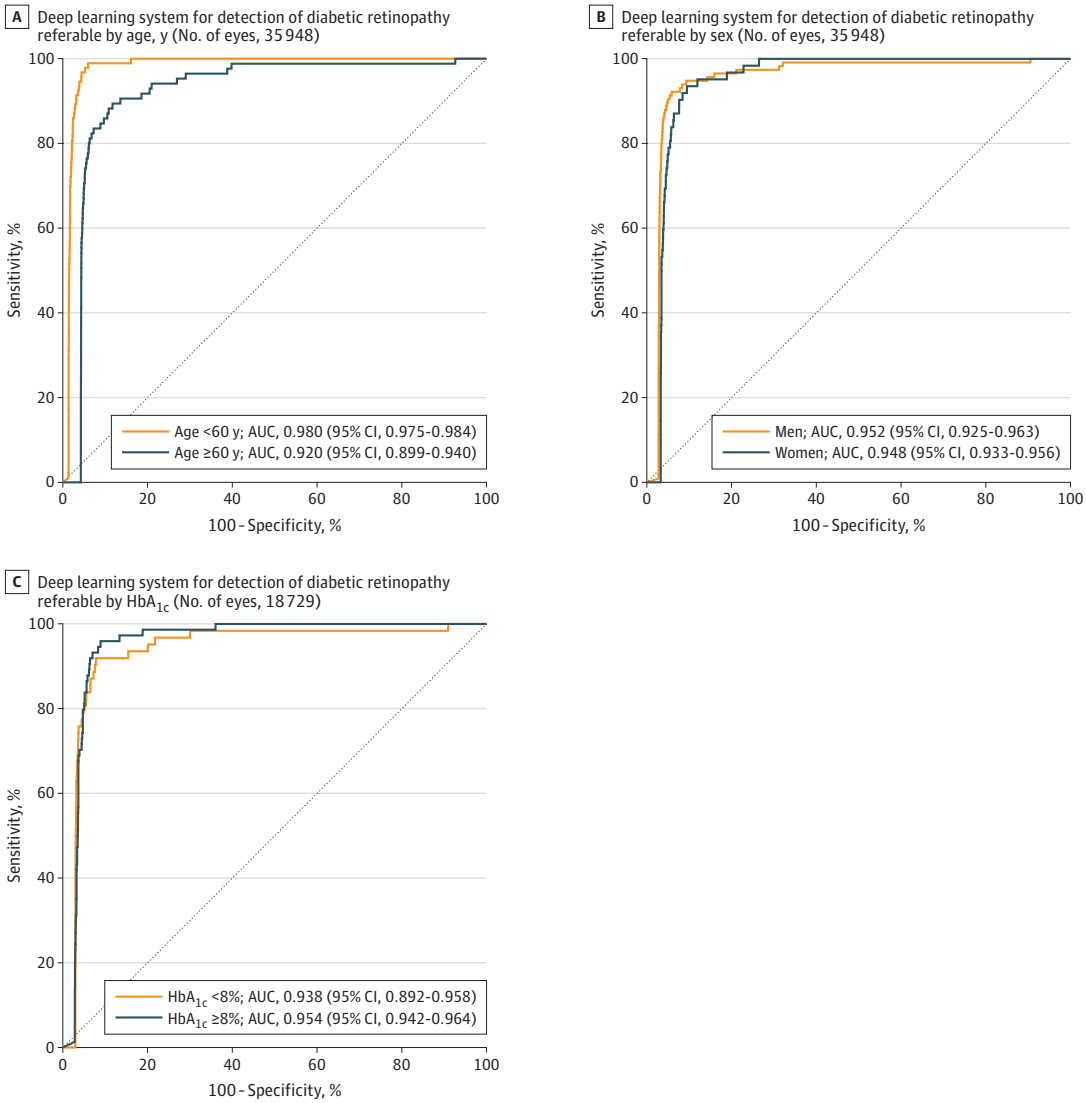

AUC indicates area under the receiver operating characteristic curve; SIDRP, Singapore National Diabetic Retinopathy Screening Program.

graders had higher specificity (91.6% vs 99.3%; $P < .001$) (Table 4; Figure 1). For vision-threatening diabetic retinopathy, the DLS had higher sensitivity compared with trained graders (100% vs 88.5%; $P < .001$), but lower specificity (91.1% vs 99.6%; $P < .001$). Among eyes with referable diabetic retinopathy, the sensitivity of diabetic macular edema was 92.1% for the DLS and 98.2% for professional graders.

Five subsidiary analyses were performed. First, the DLS showed similar diagnostic performance in 8589 unique patients of SIDRP 2014-2015 (with no overlap with training set) as in the primary analysis (eTable 3 in the Supplement). Second, in a subset of 97.4% eyes (n = 35 055) with excellent retinal image quality (no media opacity), the AUC of the DLS for referable diabetic retinopathy increased to 0.949 (95% CI, 0.940-0.957); for vision-threatening diabetic retinopathy, it increased to 0.970 (0.968-0.973). Third, the DLS showed comparable performance in different subgroups of patients stratified by age, sex, and glycemic control (**Figure 2**).

Fourth, the DLS showed clinically acceptable performance (sensitivity ≥90%) for referable diabetic retinopathy with respect to multiethnic populations of different communities, clinics, and settings (**Table 5**). Among the 10 external validation datasets, the AUC of referable diabetic retinopathy ranged from 0.889 to 0.983. The DLS showed clinically acceptable AUCs of greater than 0.90 for different cameras (eg, FundusVue, Canon, Topcon, and Carl Zeiss). Most datasets (except for Singapore Chinese, Malay, and Indian patients) had more than 80% concordance between the DLS and trained professional graders, with sensitivity of more

Figure 2. Receiver Operating Characteristic Curve and Area Under the Curve of the Deep Learning System for Detection of Referable Diabetic Retinopathy in SIDRP 2014-2015 (Primary Validation Set) by Age, Sex, and $HbA_{1c}$ Level




Eyes are the units of analysis. Glycated hemoglobin ($HbA_{1c}$) levels were available for only 52.1% of patients. Cluster-bootstrap biased-corrected 95% CI was computed for each area under the receiver operating characteristic curve (AUC), with individual patients as the bootstrap sampling clusters. See Methods for defintions of referable conditions. A, $P < .001$. B, $P = .74$. C, $P = .34$. SIDRP indicates Singapore National Diabetic Retinopathy Screening Program.

than 91% in the eyes classified as referable by retinal specialists, general ophthalmologists, trained graders, or optometrists (Table 5).

Fifth, for referable possible glaucoma, the AUC of the DLS was 0.942 (95% CI, 0.929-0.954), sensitivity was 96.4% (95% CI, 81.7%-99.9%), and specificity was 87.2% (86.8%-87.5%); for referable AMD, the AUC was 0.931 (95% CI, 0.928-0.935), sensitivity was 93.2% (95% CI, 91.1%-99.8%) and specificity was 88.7% (95% CI, 88.3%-89.0%) (**Figure 3**).

For the secondary aim, we evaluated the performance of the DLS in 2 diabetic retinopathy screening models (eFigure 2 in the Supplement): the fully- automated model had sensitivity of 93.0% (95% CI, 91.5%-94.3%) and specificity of 77.5% (95% CI, 77.0%-77.9%) to detect overall referable cases (referable diabetic retinopathy, possible glaucoma, or AMD), while the semiautomated model (DLS followed by graders) had sensitivity of 91.3% (95% CI, 89.7%-92.8%) and specificity of 99.5% (95% CI, 99.5%-99.6%) to detect overall referable status. The performance of different semiautomated models with a preset sensitivity threshold of 90%, 95%, and 99% are shown in eTable 4 in the Supplement.

## Discussion

In this evaluation of nearly half a million of images from multiethnic community, population-based and clinical datasets, the DLS had high sensitivity and specificity for

Table 5. External Validation Datasets Showing the Area Under the Curve, Sensitivity, Specificity, Concordant and Discordant Rates of the Deep Learning System in Detecting Referable Diabetic Retinopathy Among Populations With Diabetes, With Comparison to Retinal Specialists, General Ophthalmologists, Trained Graders, or Optometrists[a]

| | | % (95% CI) | | Concordance Between DLS and Grader, No. (%)[d] | | | | |
|---|---|---|---|---|---|---|---|---|
| Datasets (No. of Images) | AUC (95% CI)[b] | Sensitivity[c] | Specificity[c] | DLS+ Graders+ | DLS+ Graders− | DLS− Graders+ | DLS− Graders− | Total Concordant Images |
| Community-based | | | | | | | | |
| Guangdong (N = 15 798) | 0.949 (0.943-0.955) | 98.7 (97.7-99.3) | 81.6 (80.7-82.5) | 1785 (11.3) | 2575 (16.3) | 16 (0.1) | 11 422 (72.3) | 13 207 (83.6) |
| Population-based | | | | | | | | |
| Singapore Malay Eye Study,[17-20] (N = 3052) | 0.889 (0.863-0.908) | 97.1 (92.5-98.9) | 82.0 (79.4-84.4) | 282 (9.2) | 611 (20.0) | 3 (0.1) | 2156 (70.6) | 2438 (79.9) |
| Singapore Indian Eye Study,[17-20] (N = 4512) | 0.917 (0.899-0.933) | 99.3 (95.1-99.9) | 73.3 (70.9-75.5) | 298 (6.6) | 1543 (34.2) | 0 | 2671 (59.2) | 2969 (65.8) |
| Singapore Chinese Eye Study[17-20] (N = 1936) | 0.919 (0.900-0.942) | 100 (92.5-100.0)[e] | 76.3 (72.7-79.6) | 138 (7.1) | 560 (28.9) | 0 | 1239 (64.0) | 1377 (71.1) |
| Beijing Eye Study,[21] (N = 1052) | 0.929 (0.903-0.955) | 94.4 (72.7-99.9) | 88.5 (85.4-91.2) | 35 (3.3) | 117 (11.1) | 1 (0.1) | 899 (85.5) | 934 (88.8) |
| African American Eye Disease Study[22] (N = 1968) | 0.980 (0.971-0.989) | 98.8 (93.5-100.0) | 86.5 (84.1-88.7) | 171 (8.7) | 242 (12.3) | 2 (0.1) | 1553 (78.9) | 1724 (87.6) |
| Clinic-based | | | | | | | | |
| Royal Victoria Eye and Ear Hospital[8] (N = 2302) | 0.983 (0.972-0.991) | 98.9 (97.5-99.6) | 92.2 (89.5-94.3) | 1066 (46.3) | 198 (8.6) | 5 (0.2) | 1034 (44.9) | 2100 (91.2) |
| Mexican (N = 1172) | 0.950 (0.934-0.966) | 91.8 (88.4-94.4) | 84.8 (80.4-88.5) | 571 (48.7) | 83 (7.1) | 52 (4.4) | 466 (39.8) | 1037 (88.5) |
| Chinese University of Hong Kong (N = 1254) | 0.948 (0.921-0.972) | 99.3 (97.3-99.8) | 83.1 (77.9-87.3) | 576 (45.9) | 165 (13.2) | 4 (0.3) | 509 (40.6) | 1085 (86.5) |
| University of Hong Kong (N = 7706) | 0.964 (0.958-0.970) | 100 (99.0-100)[e] | 81.3 (80.0-82.6) | 701 (9.1) | 1310 (17.0) | 0 | 5695 (73.9) | 6396 (83.0) |

Abbreviations: AUC, area under the receiver operating characteristic curve; DLS, deep learning system.

[a] For study locations and race/ethnicity data, see Table 1. Referable diabetic retinopathy was defined as moderate nonproliferative diabetic retinopathy, severe, proliferative diabetic retinopathy, and ungradable images.

[b] Cluster-bootstrap, biased-corrected 95% CI was computed for each area under the curve, with individual patients as the bootstrap sampling clusters.

[c] Asymptotic 95% CI was computed for the logit of each proportion and using the cluster sandwich estimator of standard error to account for possible dependency of eyes within each individual.

[d] DLS+ and grader+ indicates positive concordance; DLS− and grader−, negative concordance. Last column reports total concordance (sum of these 2 values).

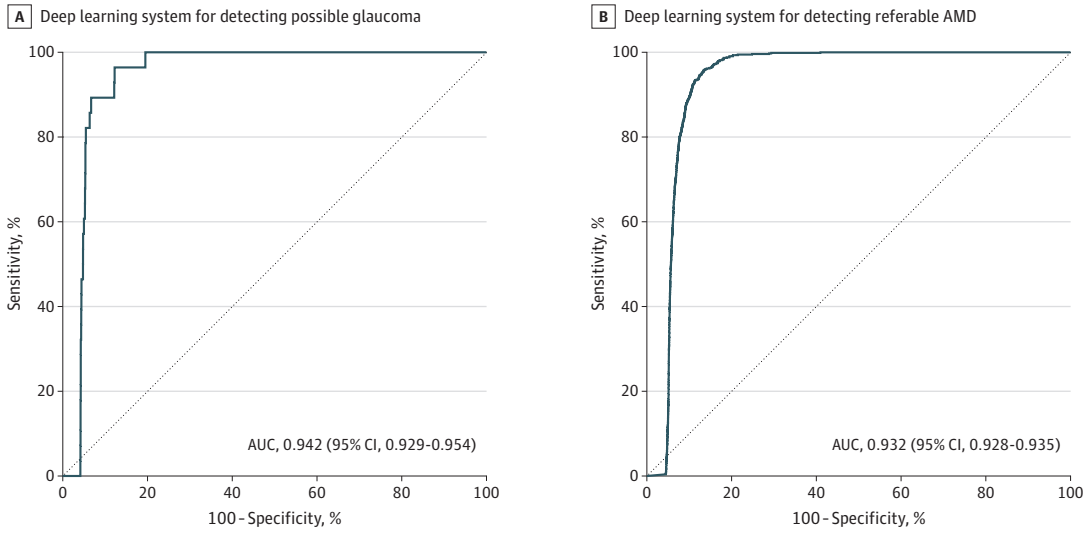
[e] Exact Clopper-Pearson left-sided 97.5% CI was calculated owing to estimate being at the boundary.

identifying referable diabetic retinopathy and vision-threatening diabetic retinopathy, as well as for identifying related eye diseases, including referable possible glaucoma and referable age-related macular degeneration. The performance of the DLS was comparable and clinically acceptable to the current model based on assessment of retinal images by trained professional graders and showed consistency in 10 external validation datasets of multiple ethnicities and settings, using diverse reference standards in assessment of diabetic retinopathy by professional graders, optometrists, or retinal specialists. This study also examined how the DLS could be deployed in 2 common diabetic retinopathy screening models: a "fully-automated" screening model that showed clinically acceptable performance to detect all 3 conditions, useful in communities without any existing diabetic retinopathy screening programs; and a "semi-automated" model in which diabetic retinopathy screening programs using trained professional graders already exist, and the DLS could be incorporated.

There have been previous studies of automated software for diabetic retinopathy screening[14,37,38]; most recent ones used a DLS.[10-12] Gulshan et al[10] reported a DLS with high sensitivity and specificity (>90%) and an AUC of 0.99 for referable diabetic retinopathy using approximately 10 000 images retrieved from 2 publicly available databases (EyePAC-1 and Messidor-2). Similarly, Gargeya and Leng[11] showed optimal DLS diagnostic performance in detecting any diabetic retinopathy using 2 other public databases (Messidor-2 and E-Ophtha). To facilitate translation, it is important to develop and test the DLS in clinical scenarios using diverse retinal images of varying quality from different camera types and in representative diabetic retinopathy screening populations.[13] The current study therefore substantially added to other current studies.

First, the DLS was trained to also detect other related eye diseases including referable possible glaucoma and referable AMD in addition to diabetic retinopathy. Second, the training and validation data sets were substantially larger (nearly

Figure 3. Primary Validation Dataset and Area Under the Curve of the Deep Learning System in Detecting Referable Possible Glaucoma and Referable Age-Related Macular Degeneration (AMD) Among Patients With Diabetes, SIDRP 2014-2015, With Reference to a Retinal Specialist




Eyes are the units of analysis. Cluster-bootstrap biased-corrected 95% CI was computed for each area under the receiver operating characteristic curve (AUC), with individual patients as the bootstrap sampling clusters. Referable possible glaucoma defined as ratio of vertical cup to disc diameter of 0.8 or greater, focal thinning or notching of the neuroretinal rim, optic disc hemorrhages, or localized retinal nerve fiber layer defects. Referable acute macular degeneration (AMD) defined as the presence of intermediate AMD (numerous intermediate drusens, 1 large drusen >125um) and/or advanced AMD, geographic atrophy, or neovascular AMD, using the Age-Related Eye Disease Study grading system.[30] Repeats from the Singapore National Diabetes Retinopathy Screening Program (SIDRP) 2014-2015 were excluded from the analysis. Asymptotic 95% CI was computed for the logit of each proportion and using the cluster sandwich estimator of standard error to account for possible dependency of eyes within each individual. Cluster-bootstrap biased-corrected 95% CI was computed for each AUC, with individual patients as the bootstrap sampling clusters.

500 000 images) and included images from patients of diverse racial and ethnic groups (ie, darker fundus pigmentation in African American and Indian individuals to lighter fundus in white individuals). The DLS showed consistent diagnostic performance across images of varying quality and different camera types, and across patients with varying systemic glycemic control level.

Third, primary validation of the DLS was conducted in an ongoing diabetic retinopathy screening program in which there were poorer quality images, including ungradable ones. This results in somewhat lower performance of the DLS (AUC, 0.936) than the system by Gulshan et al that used higher-quality images.[10] Fourth, this study also had fewer cases of severe disease (eg, vision-threatening diabetic retinopathy, referable possible glaucoma, and referable AMD), but this is more representative of populations for routine diabetic retinopathy screening.[39]

To ensure no degradation in health outcomes, a threshold was set to ensure false-negative rates were no worse than human assessment by trained professional graders. Although the results suggest that professional nonmedical graders may outperform the DLS (with high specificity of 99% for referable diabetic retinopathy and vision-threatening diabetic retinopathy), given the very low marginal cost of the DLS, the low prevalence rate of the conditions in the target screening population (<5%), and equality in health outcomes, the DLS could be used with a semiautomated model in which first-line screening with the DLS is followed by human assessment for patients who test positive. This will allow increasing screening episodes with lower cost and no degradation in health outcomes.

## Limitations

This study has several limitations. First, the training set was not developed entirely based on the retinal specialists' grading for all images. Although the reference standard in the primary validation dataset used grading by a retinal specialist, reference standards for the external datasets were based on varying assessment by retinal specialists, general ophthalmologists, trained graders, or optometrists. The performance of the DLS may potentially be further improved if all images in the training and validation data sets had criterion standard references evaluated by the retinal specialists. Nevertheless, the diagnostic performance of the DLS remained clinically acceptable and highly reproducible in both the primary validation data set and in the 10 external datasets in which the reference standards vary depending on whether the images were evaluated by retinal specialists (African American, Mexican, Hong Kong Chinese), general ophthalmologists (Beijing Chinese), optometrists (Hong Kong Chinese) or professional nonmedical graders (the remaining datasets) from the different countries (Table 5).

Second, the DLS uses multiple levels of representation to analyze each retinal image without showing the actual

diabetic retinopathy lesions (eg, microaneurysms, retinal hemorrhages). These data points can possibly be the shape or contour of the optic disc or tortuosity or caliber of the retinal vessels. Such black-box issues may have an effect on physicians' acceptance for clinical use.[13]

Third, identification of diabetic macular edema from fundus photographs may not identify all cases appropriately without clinical examination and optical coherence tomography.

## Conclusions

In this evaluation of retinal images from multiethnic cohorts of patients with diabetes, the DLS had high sensitivity and specificity for identifying diabetic retinopathy and related eye diseases. Further research is necessary to evaluate the applicability of the DLS in health care settings and the utility of the DLS to improve vision outcomes.

### ARTICLE INFORMATION

**Accepted for Publication:** October 31, 2017.

**Author Affiliations:** Singapore Eye Research Institute, Singapore National Eye Center, Singapore (Ting, C.-L. Cheung, G. S. Tan, Quang, Gan, Hamzah, San Yeo, S. Y. Lee, E. Y. Wong, Sabanayagam, Baskaran, Lamoureux, Cheng, G. C. Cheung, Aung, T. Y. Wong); Duke-NUS Medical School, National University of Singapore, Singapore (Ting, G. S. Tan, San Yeo, S. Y. Lee, E. Y. Wong, Sabanayagam, Baskaran, Ibrahim, N. C. Tan, Lamoureux, Cheng, G. C. Cheung, Aung, T. Y. Wong); Department of Ophthalmology and Visual Sciences, Chinese University of Hong Kong, Hong Kong SAR, China (C.-L. Cheung); School of Computing, National University of Singapore (Lim, Hsu, M. L. Lee); Instituto Mexicano De Oftalmologia, IAP, Queretaro, Mexico. (Garcia-Franco); SingHealth Polyclinic, Singapore Health Service, Singapore (N. C. Tan); Lien Center for Palliative Care, Health Services and Systems Research Program, Duke-NUS Graduate Medical School, Singapore (Finkelstein); Department of Ophthalmology, The University of Hong Kong, Hong Kong SAR, China (I. Y. Wong); Johns Hopkins Wilmer Eye Institute, Baltimore, Maryland (Bressler); Moorfields Eye Hospital National Health Service Foundation Trust, London, United Kingdom (Sivaprasad); University of Southern California Gayle and Edward Roski Eye Institute, Los Angeles, California (Varma); Department of Ophthalmology, Ruprecht-Karls University of Heidelberg, Heidelberg, Germany (Jonas); State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Centre, Sun Yatsen University, Guangzhou, China (He).

**Author Contributions:** Drs D. S. Ting and T. Y. Wong had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs He, Cheng, G. C. M. Cheung, Tin, Hsu, Lee, and T. Y. Wong contributed equally.
*Concept and design:* D. S. Ting, C. Y-L. Cheung, Lim, G. S. W. Tan, Garcia-Franco, Finkelstein, Lamoureux, Sivaprasad, Varma, Jonas, He, Cheng, Hsu, M. Lee, T. Y. Wong.
*Acquisition, analysis, or interpretation of data:* D. S. Ting, C. Y-L. Cheung, Lim, G. S. W. Tan, Nguyen, Gan, Hamzah, Yeo, S. Y. Lee, E. Y. M. Wong, Sabanayagam, Baskaran, Ibrahim, N. C. Tan, Finkelstein, Lamoureux, I. Y. Wong, Bressler, Varma, Jonas, He, C. Y-L. Cheung, Tin, T. Y. Wong.
*Drafting of the manuscript:* D. S. Ting, Gan, Hamzah, Lamoureux, T. Y. Wong.
*Critical revision of the manuscript for important intellectual content:* D. S. Ting, C. Y. Cheung, Lim, G. S. W. Tan, Nguyen, Garcia-Franco, Yeo, S. Y. Lee, E. Y. M. Wong, Sabanayagam, Baskaran, Ibrahim, N. C. Tan, Finkelstein, I. Y. Wong, Bressler, Sivaprasad, Varma, Jonas, He, Cheng, C. Y. Cheung, Tin, Hsu, M. L. Lee, T. Y. Wong.
*Statistical analysis:* D. S. Ting, Lim, Nguyen, Gan, Sabanayagam, Ibrahim, Lamoureux, I. Y. Wong, Tin, T. Y. Wong.
*Obtained funding:* D. S. Ting, C. Y. Cheung, Tin, T. Y. Wong.
*Administrative, technical, or material support:* D. S. Ting, C. Y. Cheung, Lim, G. S. W. Tan, Hamzah, Garcia-Franco, Baskaran, N. C. Tan, Sivaprasad, Jonas, Cheng, Tin, T. Y. Wong.
*Supervision:* D. S. Ting, C. Y. Cheung, G. S. W. Tan, Yeo, S. Lee, Lamoureux, Varma, He, C. Y. Cheung, Hsu, M. L. Lee, T. Y. Wong.

**Conflict of Interest Disclosures:** All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Drs D. S. Ting, Lim, M. L. Lee, Hsu, and T. Y. Wong reported that they are coinventors on a patent for the deep learning system used in this study; potential conflicts of interests are managed according to institutional policies of the Singapore Health System (SingHealth) and the National University of Singapore (NUS). Dr Bressler reported holding a patent on a system and method for automated detection of age-related macular degeneration and other retinal abnormalities, unrelated to the deep learning system in this paper. Dr He reported holding a patent on an automated image analysis system and retinal camera for retinal diseases, unrelated to the deep learning system in this article. No other disclosures were reported.

**Funding/Support:** This project received funding from National Medical Research Council (NMRC), Ministry of Health (MOH), Singapore National Health Innovation Center (NHIC), Innovation to Develop Grant (NHIC-I2D-1409022); SingHealth Foundation Research Grant (SHF/FG648S/2015), and the Tanoto Foundation; unrestricted donations to the Retina Division, Johns Hopkins University School of Medicine. For the Singapore Epidemiology of Eye Diseases (SEED) study, we received funding from NMRC, MOH (grants 0796/2003, IRG07nov013, IRG09nov014, STaR/0003/2008 and STaR/2013; CG/SERI/2010) and Biomedical Research Council (grants 08/1/35/19/550 and 09/1/35/19/616). The Singapore Diabetic Retinopathy Program received funding from the MOH, Singapore (grants AIC/RPDD/SIDRP/SERI/FY2013/0018 and AIC/HPD/FY2016/0912).

**Role of the Funders/Sponsors:** The NMRC (Singapore), MOH (Singapore), NHIC (Singapore), and Tanoto Foundation, and Johns Hopkins School of Medicine had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

### REFERENCES

1. Yau JW, Rogers SL, Kawasaki R, et al; Meta-Analysis for Eye Disease (META-EYE) Study Group. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care*. 2012;35(3):556-564.

2. Ting DS, Cheung GC, Wong TY. Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. *Clin Exp Ophthalmol*. 2016;44(4):260-277.

3. Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet*. 2010;376(9735):124-136.

4. Wang LZ, Cheung CY, Tapp RJ, et al. Availability and variability in guidelines on diabetic retinopathy screening in Asian countries. *Br J Ophthalmol*. 2017;101(10):1352-1360.

5. Burgess PI, Msukwa G, Beare NA. Diabetic retinopathy in sub-Saharan Africa: meeting the challenges of an emerging epidemic. *BMC Med*. 2013;11:157.

6. Hazin R, Colyer M, Lum F, Barazi MK. Revisiting Diabetes 2000: challenges in establishing nationwide diabetic retinopathy prevention programs. *Am J Ophthalmol*. 2011;152(5):723-729.

7. Ting DS, Ng JQ, Morlet N, et al. Diabetic retinopathy management by Australian optometrists. *Clin Exp Ophthalmol*. 2011;39(3):230-235.

8. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-444.

9. Lim G, Lee ML, Hsu W, Wong TY. Transformed representations for convolutional neural networks in diabetic retinopathy screening. In: *MAIHA, Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*. 2014; 34-38.

10. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410.

11. Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning. *Ophthalmology*. 2017;124(7):962-969.

12. Abràmoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci*. 2016;57(13):5200-5206.

13. Wong TY, Bressler NM. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. *JAMA*. 2016;316(22):2366-2367.

14. Abramoff MD, Niemeijer M, Russell SR. Automated detection of diabetic retinopathy:

barriers to translation into clinical practice. *Expert Rev Med Devices*. 2010;7(2):287-296.

**15**. Chew EY, Schachat AP. Should we add screening of age-related macular degeneration to current screening programs for diabetic retinopathy? *Ophthalmology*. 2015;122(11):2155-2156.

**16**. Nguyen HV, Tan GS, Tapp RJ, et al. Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in singapore. *Ophthalmology*. 2016;123(12):2571-2580.

**17**. Huang OS, Tay WT, Ong PG, et al. Prevalence and determinants of undiagnosed diabetic retinopathy and vision-threatening retinopathy in a multiethnic Asian cohort: the Singapore Epidemiology of Eye Diseases (SEED) study. *Br J Ophthalmol*. 2015;99(12):1614-1621.

**18**. Wong TY, Cheung N, Tay WT, et al. Prevalence and risk factors for diabetic retinopathy: the Singapore Malay Eye Study. *Ophthalmology*. 2008;115(11):1869-1875.

**19**. Shi Y, Tham YC, Cheung N, et al. Is aspirin associated with diabetic retinopathy? the Singapore Epidemiology of Eye Disease (SEED) study. *PLoS One*. 2017;12(4):e0175966.

**20**. Chong YH, Fan Q, Tham YC, et al. Type 2 diabetes genetic variants and risk of diabetic retinopathy. *Ophthalmology*. 2017;124(3):336-342.

**21**. Jonas JB, Xu L, Wang YX. The Beijing Eye Study. *Acta Ophthalmol*. 2009;87(3):247-261.

**22**. Varma R. African American Eye Disease Study. National Institutes of Health website. http://grantome.com/grant/NIH/U10-EY023575-03. 2017. Accessed September 25, 2017.

**23**. Lamoureux EL, Fenwick E, Xie J, et al. Methodology and early findings of the Diabetes Management Project: a cohort study investigating the barriers to optimal diabetes care in diabetic patients with and without diabetic retinopathy. *Clin Exp Ophthalmol*. 2012;40(1):73-82.

**24**. Tang FY, Ng DS, Lam A, et al. Determinants of quantitative optical coherence tomography angiography metrics in patients with diabetes. *Sci Rep*. 2017;7(1):2575.

**25**. Chua J, Baskaran M, Ong PG, et al. Prevalence, risk factors, and visual features of undiagnosed glaucoma: the Singapore Epidemiology of Eye Diseases study. *JAMA Ophthalmol*. 2015;133(8):938-946.

**26**. Cheung CM, Li X, Cheng CY, et al. Prevalence, racial variations, and risk factors of age-related macular degeneration in Singaporean Chinese, Indians, and Malays. *Ophthalmology*. 2014;121(8):1598-1603.

**27**. Cheung CM, Bhargava M, Laude A, et al. Asian age-related macular degeneration phenotyping study: rationale, design and protocol of a prospective cohort study. *Clin Exp Ophthalmol*. 2012;40(7):727-735.

**28**. Ting DSW, Yanagi Y, Agrawal R, et al. Choroidal remodeling in age-related macular degeneration and polypoidal choroidal vasculopathy: a 12-month prospective study. *Sci Rep*. 2017;7(1):7868.

**29**. Ting DS, Ng WY, Ng SR, et al. Choroidal thickness changes in age-related macular degeneration and polypoidal choroidal vasculopathy: a 12-month prospective study. *Am J Ophthalmol*. 2016;164:128-136.

**30**. Wilkinson CP, Ferris FL III, Klein RE, et al; Global Diabetic Retinopathy Project Group. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology*. 2003;110(9):1677-1682.

**31**. Klein R, Davis MD, Magli YL, Segal P, Klein BE, Hubbard L. The Wisconsin age-related maculopathy grading system. *Ophthalmology*. 1991;98(7):1128-1134.

**32**. Chakrabarti R, Harper CA, Keefe JE. Diabetic retinopathy management guidelines. *Expert Rev Ophthalmol*. 2012;7(5):417-439. doi:10.1586/eop.12.52

**33**. National Health Service (NHS) Diabetic Eye Screening Programme and Population Screening Programmes. Diabetic eye screening: commission and provide. https://www.gov.uk/government/collections/diabetic-eye-screening-commission-and-provide. 2015. Accessed September 24, 2017.

**34**. Tufail A, Rudisill C, Egan C, et al. Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. *Ophthalmology*. 2017;124(3):343-351.

**35**. Ting DSW, Tan GSW. Telemedicine for diabetic retinopathy screening. *JAMA Ophthalmol*. 2017;135(7):722-723.

**36**. Ren S, Lai H, Tong W, Aminzadeh M, Hou X, Lai S. Nonparametric bootstrapping for hierarchical data. *J Appl Stat*. 2010;37(9):1487-1498. doi:10.1080/02664760903046102

**37**. Abràmoff MD, Niemeijer M, Suttorp-Schulten MS, Viergever MA, Russell SR, van Ginneken B. Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes. *Diabetes Care*. 2008;31(2):193-198.

**38**. Abràmoff MD, Reinhardt JM, Russell SR, et al. Automated early detection of diabetic retinopathy. *Ophthalmology*. 2010;117(6):1147-1154.

**39**. Sivaprasad S, Gupta B, Crosby-Nwaobi R, Evans J. Prevalence of diabetic retinopathy in various ethnic groups: a worldwide perspective. *Surv Ophthalmol*. 2012;57(4):347-370.

**40**. Bhargava M, Cheung CY, Sabanayagam C, et al. Accuracy of diabetic retinopathy screening by trained non-physician graders using non-mydriatic fundus camera. *Singapore Med J*. 2012;53(11):715-719.