



Performance of the AIDRScreening system in detecting diabetic retinopathy in the fundus photographs of Chinese patients: a prospective, multicenter, clinical study

Yao Yang¹, Jianying Pan¹, Miner Yuan¹, Kunbei Lai¹, Huirui Xie¹, Li Ma¹, Suzhong Xu², Ruzhi Deng², Mingwei Zhao³, Yan Luo¹, Xiaofeng Lin¹

¹State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Guangzhou, China; ²Eye Hospital, Wenzhou Medical University, Wenzhou, China; ³Department of Ophthalmology, Peking University People's Hospital, Beijing, China

Contributions: (I) Conception and design: X Lin; (II) Administrative support: X Lin, Y Luo, S Xu, M Zhao; (III) Provision of study materials or patients: X Lin, Y Luo, S Xu, M Zhao, R Deng, Y Yang, M Yuan, K Lai, L Ma; (IV) Collection and assembly of data: Y Yang, J Pan, M Yuan, K Lai, H Xie, L Ma, R Deng, S Xu, M Zhao; (V) Data analysis and interpretation: Y Yang, J Pan, Y Luo, X Lin; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Yan Luo; Xiaofeng Lin. State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou 5100600, China. Email: luoyan2@mail.sysu.edu.cn; linxiaof@mail.sysu.edu.cn.

Background: Diabetic retinopathy (DR) is the leading cause of blindness in the working-age population worldwide, and there is a large unmet need for DR screening in China. This observational, prospective, multicenter, gold standard-controlled study sought to evaluate the effectiveness and safety of the AIDRScreening system (v. 1.0), which is an artificial intelligence (AI)-enabled system that detects DR in the Chinese population based on fundus photographs.

Methods: Participants with diabetes mellitus (DM) were recruited. Fundus photographs (field 1 and field 2) of 1 eye in each participant were graded by the AIDRScreening system (v. 1.0) to detect referable DR (RDR). The results were compared to those of the masked manual grading (gold standard) system by the Zhongshan Image Reading Center. The primary outcomes were the sensitivity and specificity of the AIDRScreening system in detecting RDR. The other outcomes evaluated included the system's diagnostic accuracy, positive predictive value, negative predictive value, diagnostic accuracy gain rate, and average diagnostic time gain rate.

Results: Among the 1,001 enrolled participants with DM, 962 (96.1%) were included in the final analyses. The participants had a median age of 60.61 years (range: 20.18–85.78 years), and 48.2% were men. The manual grading system detected RDR in 399 (41.48%) participants. The AIDRScreening system had a sensitivity of 86.72% (95% CI: 83.39–90.05%) and a specificity of 96.09% (95% CI: 94.14–97.54%) in the detection of RDR, and a false-positive rate of 3.91%. The diagnostic accuracy gain rate of the AIDRScreening system was 16.57% higher than that of the investigator, while the average diagnostic time gain rate was –37.32% lower.

Conclusions: The automated AIDRScreening system can detect RDR with high accuracy, but cannot detect maculopathy. The implementation of the AIDRScreening system may increase the efficiency of DR screening.

Keywords: Fundus photographs; diabetic retinopathy (DR); screening; artificial intelligence; multicenter

Submitted Jan 20, 2022. Accepted for publication Jul 15, 2022.

doi: 10.21037/atm-22-350

View this article at: <https://dx.doi.org/10.21037/atm-22-350>

Introduction

Diabetic retinopathy (DR) is the leading cause of blindness in the working-age population worldwide (1-3). DR-induced vision loss can be prevented by early detection and effective treatment (1,3-5), which can be achieved by employing DR screening programs (6,7). However, in reality, there is a large unmet need for DR screening in China. The reported prevalence of DR in patients with diabetes mellitus (DM) is 24.7–37.5% in China, wherein over one-fifth of the global population with a high prevalence of DM (up to 11.7%) lives (8-11). It is estimated that over 30 million patients with DM in China are at a risk of vision loss due to DR (8). However, up to 85% of patients with DM are not informed of the risk of DR or do not undergo any DR screening (9-12).

An extreme scarcity of ophthalmic medical resources and a lack of public awareness of DR have constrained the implementation of DR prevention programs in China (8,12,13). According to data from the Chinese Ophthalmic Society (COS), there are only around 35,000 registered eye doctors in China, and only 1/10 of these are retinal specialists and mainly work in urban tertiary hospitals (14). It is impossible to screen the full population of patients with DM (which is estimated to be >110 million) in China for DR using these limited eye care resources, especially as 87% of the DM patients at high risk for DR live in rural China where medical care resources are often unavailable (15).

An increasing number of artificial intelligence (AI)-based algorithms for DR detection from retinal images are demonstrating higher diagnostic accuracy than the manual grading system (16-23). The AI-based ID_x-DR was first approved for automated DR screening by the US Food and Drug Administration (FDA); however, it could only be applied to fundus photographs captured using the Topcon NW400 camera at that time (24). Recently, a multicenter, head-to-head, real-world, validation study indicated that most AI-based DR screening systems perform no better than do the manual grading system, with only 2 having a rather high sensitivity (92.71% and 92.71%) and 1 having comparable sensitivity (80.47%) and specificity (81.28%) (25). Previous research suggests that the contrast between the retinal background and DR lesions may vary considerably across different ethnicities (19). Given the ethnic variation and the dearth of DR screening in China, there is an urgent need to develop a fully automated DR screening system specifically designed for Chinese patients.

An AI-based automated DR screening system (AIDRScreening v. 1.0, Shenzhen SiBright CO. Ltd.,

China) was recently approved as the first innovative medical device by the Chinese National Medical Product Administration in May 2019. AIDRScreening is also the first AI-based DR screening system to obtain a certificate from the Chinese National Institute for Food and Drug Control after the completion of design validation for software requirement specifications, software quality evaluation, and cybersecurity. Thus, AIDRScreening was allowed to be used in this study to detect referable DR (RDR), which was defined according to the COS guidelines.

We performed a prospective, multicenter, gold standard-controlled study—the first of its kind—to assess the sensitivity, specificity, and diagnostic accuracy of the AIDRScreening system in the detection of RDR in China using multiple fundus camera types to improve RDR detection. We presented the following article in accordance with the STARD reporting checklist (available at <https://atm.amegroups.com/article/view/10.21037/atm-22-350/rc>).

Methods

Study design

This observational, prospective, multicenter, gold standard-controlled study was sponsored by Shenzhen SiBright Co. Ltd. In China. This study (clinical trial registration No. NCT03602989) was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the institutional review boards of Zhongshan Ophthalmic Center (ZOC, 01 center, 2018QXPJ001), Peking University People's Hospital (PKUPH, 02 center, 2018PHA047), and the Eye Hospital of Wenzhou Medical University (EHWMU, 03 center, 2018MD03). Written informed consent was obtained from each participant.

Study population and the inclusion and exclusion criteria

For this prospective study, we enrolled 1,001 consecutive participants at 3 centers from July 2018 to January 2019. The data collection was planned before the study, and reference standards were implemented. To be eligible for inclusion in this study, participants had to meet the following inclusion criteria: (I) be aged ≥ 18 years and (II) have a diagnosis of type 1 or type 2 DM in accordance with the criteria established by either the World Health Organization or the American Diabetes Association (i.e., a fasting plasma glucose level ≥ 126 mg/dL or a hemoglobin A1c level $\geq 6.5\%$). Participants were excluded from the study

if they met any of the following exclusion criteria: (I) had an unqualified fundus image, and/or (II) had a history of retinal surgeries. More detailed information about the inclusion and exclusion criteria can be found at ClinicalTrials.Gov (Identifier: NCT03602989).

Study protocol

Each participant underwent a comprehensive clinical examination, including assessments of visual acuity (Snellen E Chart) and intraocular pressure, slit-lamp biomicroscopy, retinal photography, and anterior chamber gonioscopy as determined by the investigators. All the investigators were ophthalmologists with >10 years of clinical experience.

Field 1, which centered on the optic disc, and Field 2, which centered on the macula, were the standardized 2-field fundus photographs that were acquired by certified photographers in this study (26,27). In this study, 3 types of fundus cameras were used to capture the retinal images. The Zeiss Visucam FF450 and Topcon TRC-50DX were used at the center 01, while the Zeiss Visucam FF450 was used at center 02. As these fundus cameras require pupil dilation for assessment, all the participants at these 2 centers underwent pupil dilation. At center 03, a type of nonmydriatic fundus camera, the Topcon NW-400, was used. If a participant's pupil diameter was >4 mm, the fundus photography was performed without mydriasis; otherwise, mydriasis was required. Before pupil dilation, a slit-lamp examination was performed by the investigators to rule out contraindications for mydriasis.

Standard operating procedures for fundus photograph image acquisition (26,27) and certified photographers were used to control the quality of the images. The quality control coordinator of the Zhongshan Image Reading Center (ZIRC) in Guangzhou, China, evaluated the final quality of the uploaded images based on location, focus, sharpness, and legibility (26). Images were excluded when interference factors, such as overexposure, an unclear focus, insufficient brightness, and excessive artifacts, were identified. Images with sufficient quality were submitted to the AI system and graders at the ZIRC for further grading.

We randomly selected 1 eye of each participant. If satisfactory fundus photographs could only be obtained from 1 eye, the randomized enrollment principle was no longer followed, and each eye with qualified images was included.

The output of the AI system was either detected RDR or nondetected RDR. RDR was defined as stage II or

more severe DR in accordance with the COS guidelines. Under the International Council of Ophthalmology (ICO) Guidelines for Diabetic Eye Care, stage I, stage II, and stage III DR in China corresponds to mild, moderate, and severe non-proliferative DR (NPDR) (3). However, the staging of proliferative DR (PDR) in China is somewhat different to that of the ICO guidelines. Under the COS guidelines, stage IV to stage VI PDR are defined as follows: stage IV refers to neovascularization of the optic disc or elsewhere (when accompanied by vitreous/preretinal hemorrhage, it is defined as high-risk PDR); stage V refers to the presence of fibrous membrane, which may be accompanied by preretinal hemorrhage or vitreous hemorrhage; and stage VI refers to traction retinal detachment, combined with fibrous membrane, combined with/without vitreous hemorrhage, and neovascularization of the iris and the anterior chamber angle.

The diagnostic procedures employed by the AI system, ZIRC, and the investigator are displayed in the flowchart of the study design in *Figure 1*. The diagnostic outputs of the AI system, ZIRC, and investigator were masked from each other at all times until the data were locked for the statistical analysis.

Development of a deep learning algorithm for the AIDRScreening system

The deep learning system was an ensemble of 3 convolutional neural networks (CNNs). The convolutional layers were used to extract features from the input images at varying spatial scales, while the fully connected layers were used to generate the predictions. The prediction output was a binary of a diagnosis, which was obtained by comparing the probability output of the neural network to the threshold value. The CNNs were not used for the assessment of image quality.

For each image under consideration, the fundus region was first detected and cropped, and the cropped image was subsequently rescaled to 512×512 pixels as the input for the CNNs. All the networks were independently trained by the adaptive moment estimation method using the 73,849 images in the training set and the 4,928 images in the test set; the images had been cross-labeled by dozens of ophthalmologists. The images were collected from different units, such as the Eye Institute, Endocrinology Department, Eye Examination Center, and DR Screening Project. Extensive testing was conducted of the AI system to validate the effectiveness of the algorithm for DR diagnosis, and the

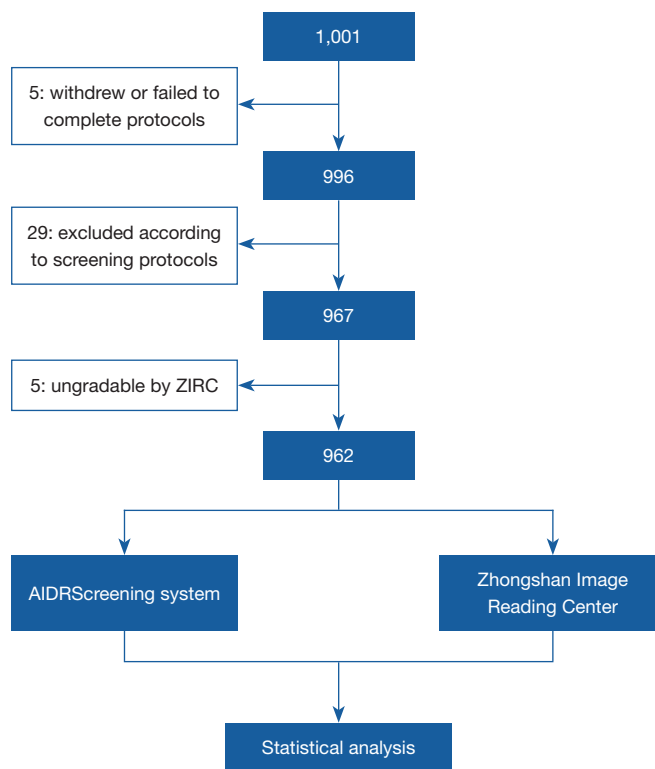


Figure 1 Waterfall diagram. ZIRC, Zhongshan Image Reading Center; DR, diabetic retinopathy.

following factors were considered: different distributions of different retinal diseases (different DR severity levels and other retinal diseases), different camera types, different image perturbations, and multiple data sets collected from different units.

Reference standards

The grading output of the ZIRC was used as a gold standard control. All the graders at the ZIRC were certified by the Doheny Eye Institute Reading Center in the United States and the National Health System of Diabetic Eye Screening Program in the United Kingdom. In the present study, the ZIRC graders used Windows Photo Viewer as the grading system. The manual grading was performed independently by 2 graders. If the independent grading of the 2 graders was consistent, the manual grading was concluded. If discrepancies existed among the 2 graders, a panel (comprising the first 2 graders and a third senior grader) discussion was conducted to reach a final conclusive grading. The graders were blinded to all the participant

information. The reading center outputs included the determination of the image quality, DR severity, and other retinal diseases, including age-related macular degeneration (AMD), retinal vein occlusion, and hypertensive retinopathy. The grading time was recorded during the reading process. For the internal control, 10% of the retinal images were randomly selected for reassessment. The intergrader and intrareader consistency were both >95%.

Primary and secondary outcomes

The primary outcomes were the sensitivity (i.e., the proportion of the positive sample of the AI system to the positive sample of the ZIRC) and specificity of the AI system in the detection of RDR. The secondary outcomes included 1) the diagnostic accuracy of the AI system in detecting RDR; 2) the positive predictive value (PPV) and negative predictive value (NPV); 3) the average time for the AI-based diagnosis; 4) the diagnostic accuracy gain rate of the AI system relative to that of the investigator—calculated as follows: $(AI_{\text{accu}} - \text{Investigator}_{\text{accu}}) / \max(AI_{\text{accu}},$

Table 1 Disease distribution of the included retinal images

Retinal diseases	Number	Percentage (%)
No apparent DR	431	44.80
DR		
Stage I DR	18	1.87
Stage II DR	318	33.06
Stage III DR	55	5.72
Stage IV DR	18	1.87
Stage V DR	7	0.73
Stage VI DR	1	0.10
Other retinal diseases (without DR)		
Hypertensive retinopathy	1	0.10
AMD	80	8.32
RVO	0	0
Others	33	3.43
Total	962	100

DR, diabetic retinopathy; AMD, age-related macular degeneration; RVO, retinal vein occlusion.

Investigator_{accu}); and 5) the average diagnostic time gain rate—calculated as follows: $(AI_{time} - \text{Investigator}_{time}) / \max(AI_{time}, \text{Investigator}_{time})$. In this study, the definite threshold was preset; thus, a receiver operating characteristic curve was not used as an outcome of the AI screening.

Statistical analysis

The statistical analysis was performed by an independent third party statistical agency, the Peking University Health Science Center. All the analyses were conducted using SAS (version 9.4) statistical software. The Chinese National Medical Products Administration has not set a benchmark for the sensitivity and specificity of AI-assisted DR screening products but requires that the values should be close to a sensitivity of 85% and a specificity of 90% according to the international level (7,24). Study recruitment was guided by a 22% prevalence rate of stage II or above DR in the diabetic population and an estimated 85% positive coincidence rate (sensitivity) and 90% negative coincidence rate (specificity). The sample size was calculated using PASS 13 software (NCSS). The quantitative data are presented as the mean [standard deviation (SD)], median, minimum (min), and

maximum (max). The categorical data are presented as the number of cases and percentages. Cases with missing data were excluded.

The impact of age, sex, DM type, diabetes duration, camera type, and mydriatic status on sensitivity and specificity for the detection of RDR were analyzed using logistic regression analysis. A two-sided P value <0.05 was considered statistically significant.

Results

Study objects

A total of 1,001 participants were enrolled in this study, 996 (99.50%) of whom underwent all procedures. Among the participants, 29 were excluded according to the study protocols: 15 due to the presence of unqualified images, 10 due to having no proof of a diagnosis of DM other than their own claims of having a history of DM, and 4 due to having a supplementary history that met 1 of the exclusion criteria. An additional 5 participants were excluded due to an inability to undergo ZIRC grading. Thus, a subset of 962 (96.10%) participants underwent the full analysis (see *Figure 1*).

The median age of the participants was 60.61 years (range: 20.18–85.78 years), and 48.2% of the participants were men. Among the participants, 41 (4.26%) had type 1 DM, 920 (95.63%) had type 2 DM, and 1 (0.1%) was not sure of the DM type. The average DM duration was 8.9 ± 6.8 years, and the median was 8.0 (min =1.0, max =45.0) years. Among the participants, 273 (28.4%) were treated with insulin.

According to the grading results from the ZIRC, the prevalence rates of DR and RDR were 43.34% (417/962) and 41.48% (399/962), respectively (see *Supplementary Table S1*). A total of 81 participants (81/962, 8.42%) had vision-threatening DR (DR stage \geq III).

A total of 114 (11.85%) participants had other retinal diseases. Among them, 8.32% (80/962), 0.1% (1/962), and 3.43% (33/962) had AMD, hypertensive retinopathy, and other fundus diseases (e.g., glaucoma and myelinated nerve fibers), respectively (see *Table 1*).

AI system characteristics

AI system performance

In the detection of RDR, the sensitivity and specificity of the AI system were 86.72% (95% CI: 83.39–90.05%) and 96.09% (95% CI: 94.14–97.54%), respectively, relative to

Table 2 Performance of AIDRScreening in the detection of RDR

RDR detection	Value	95% CI
Sensitivity	86.72%	83.39–90.05%
Specificity	96.09%	94.14–97.54%
Accuracy	92.20%	90.33–93.82%
PPV	94.02%	91.09–96.22%
NPV	91.08%	88.49–93.25%

RDR, referable diabetic retinopathy (stage II or more severe diabetic retinopathy); CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.

the results of the ZIRC gold standard grading system. The accuracy, PPV, and NPV of the RDR detection system were 92.20%, 94.02%, and 91.08% respectively (see *Table 2*). In the detection of vision-threatening DR (DR stage \geq III according to the COS guidelines, which corresponds to severe NPDR or PDR in the ICO guidelines), the AI system had an accuracy of 97.53%.

The average diagnostic times of the investigator, AI system, and ZIRC were 38 ± 32 seconds (s), 24 ± 8 s, and 69 ± 24 s, respectively. The accuracy gain rate of the AI system was 16.57% higher than that of the investigators, while the average time gain rate was -37.32% lower.

AI system generalizability

The AI system performed similarly well across the 3 trial centers (see *Table 3*). With a sensitivity of 85.32–87.86%, a specificity of 95.21–97.33%, and an accuracy of 91.61–92.91%, all the primary outcomes exceeded the prespecified noninferiority end points of the 3 research centers.

Retinal images of 110 participants were acquired without mydriasis using the Topcon NW400. All 110 participants were from center 03. As stated above, participants at centers 01 and 02 obtained their fundus photographs under mydriasis. In the 110 patients imaged using the Topcon NW400 without mydriasis, the sensitivity and specificity were 84.00% and 93.33%, respectively. For those 176 patients imaged by the Topcon NW400 after mydriasis, the sensitivity and specificity were 90.00% and 96.51%, respectively. The sensitivity, specificity, and accuracy of the AIDRScreening system for RDR detection were 87.86%, 95.21%, and 91.61%, respectively, with the Topcon NW400 images; 86.38%, 96.42%, and 92.65%, respectively, with the Zeiss Visucam FF450 images; and 83.33%, 96.15%, and 90.00%, respectively, with the Topcon TRC-50DX images (see *Table 4*).

Safety analysis

The AI system overdiagnosed RDR in 22 eyes (false-positive rate: 22/563, 3.91%) and failed to detect RDR in 53 eyes (false-negative rate: 53/399, 13.28%). Among the 53 false-negative cases, 2 eyes (2/962, 0.21%) were diagnosed with DR stage III (severe NPDR) by the ZIRC with intraretinal microvascular abnormalities detected, and 51 eyes (51/962, 5.30%) were diagnosed with DR stage II (moderate NPDR). Among the 22 false-positive eyes, 3 eyes were diagnosed with DR stage I, 13 eyes with no abnormality, and 6 eyes with early or intermediate AMD but no DR by the ZIRC. In the present study, no adverse effects or severe adverse effects, such as acute angle closure glaucoma during or after mydriasis, were recorded.

Discussion

This study is the first prospective, multicenter, gold standard-controlled study on an AI-based DR screening system in China. Our findings showed that the AIDRScreening system (v. 1.0) can safely and effectively detect RDR automatically, with a sensitivity of 86.72% and a specificity of 96.09% relative to those of the gold standard system, which exceeds the prespecified noninferiority thresholds that require a sensitivity $\geq 85\%$ and a specificity $\geq 90\%$. The false-negative rate of the AIDRScreening for RDR detection was 13.28%. Leopard fundus, overexposure, or underexposure can result in relatively low-contrast images, which may explain the failure of the AIDRScreening system to diagnose RDR in some cases. The presence of drusen, retinal pigment abnormality, or retinal hemorrhage caused by other diseases may explain why the system overdiagnosed RDR in some cases.

For screening purposes, sensitivity is important to ensure that patients with RDR receive the necessary care, while specificity is also practically important to ensure that patients are not unnecessarily referred to retinal specialists and that no resources are unnecessarily allocated. A balance between sensitivity and specificity is important for AI-assisted DR screening. The higher specificity of our AIDRScreening system is particularly important in China, where there is a shortage of retinal specialists. Our results are also comparable to those of other previous studies on AI-based DR screening systems (16,19,20,24).

Severe NPDR and PDR are considered types of vision-threatening DR (1,3). In such cases, both the COS and ICO guidelines recommend that patients receive an immediate ophthalmology evaluation (3,8). The AIDRScreening

Table 3 Performance of AIDRScreening in different sites among the included eyes

RDR detection	Center 01 (ZOC)		Center 02 (PKUPH)		Center 03 (EHWMU)	
	Value	95% CI	Value	95% CI	Value	95% CI
Sensitivity	86.67%	81.23–92.11%	85.32%	78.68–91.96%	87.86%	82.45–93.27%
Specificity	95.65%	93.02–98.29%	97.33%	93.87–99.13%	95.21%	91.74–98.67%
Accuracy	92.11%	89.39–94.82%	92.91%	89.98–95.83%	91.61%	88.40–94.82%
PPV	92.86%	88.59–97.12%	94.90%	88.49–98.32%	94.62%	90.74–98.50%
NPV	91.67%	88.17–95.16%	91.92%	88.12–95.72%	89.10%	84.21–93.99%
PLR	19.93	10.84–36.66	31.91	13.39–76.03	18.32	8.87–37.85
NLR	0.14	0.09–0.21	0.15	0.10–0.24	0.13	0.08–0.20
OR	142.36	–	212.73	–	140.92	–

ZOC, Zhongshan Ophthalmic Center; PKUPH, Peking University People's Hospital; EHWMU, Eye Hospital of Wenzhou Medical University; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; PLR, positive likely ratio; NLR, negative likely ratio; OR, odds ratio.

Table 4 Performance of AIDRScreening with different camera types among the included eyes

RDR detection	Topcon TRC-50DX		Zeiss Visucam FF450		Topcon TRC-NW400	
	Value	95% CI	Value	95% CI	Value	95% CI
Sensitivity	83.33%	62.62–95.26%	86.38%	82.00–90.77%	87.86%	82.45–93.27%
Specificity	96.15%	80.36–99.90%	96.42%	94.58–98.26%	95.21%	91.74–98.67%
Accuracy	90.00%	78.19–96.67%	92.65%	90.61–94.70%	91.61%	88.40–94.82%
PPV	95.24%	76.18–99.88%	93.55%	90.28–96.82%	94.62%	90.74–98.50%
NPV	86.21%	68.34–96.11%	92.18%	89.57–94.78%	89.10%	84.21–93.99%
PLR	21.67	3.14–149.31	24.13	14.39–40.45	18.32	8.87–37.85
NLR	0.17	0.07–0.43	0.14	0.10–0.19	0.13	0.08–0.20
OR	127.47	–	172.36	–	140.92	–

CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; PLR, positive likely ratio; NLR, negative likely ratio; OR, odds ratio.

system (v. 1.0) had an accuracy of 97.53% in the detection of these severe forms of DR, and high accuracy is important in ensuring that patients receive timely care and thus avoid blindness. Our findings indicate that AIDRScreening is safe and effective; thus, this system could alleviate the current burden placed on the extremely scarce ophthalmic medical resources in China. In future applications, it may be better to employ the AIDRScreening system in a semiautomated fashion to prevent missing RDR diagnoses.

In our study, 3 different types of fundus cameras were used to achieve high specificity (95.21–96.42%) and accuracy (90.00–92.65%). The 2018 IDx-DR study only

used the Topcon NW400 fundus camera (24), which had received US FDA approval for clinical application at that time. However, the Idx-DR can now use images from other types of fundus cameras in Europe and other regions worldwide.

It should be noted that the sensitivity of the AIDRScreening system for RDR detection was 83.33% when the images were captured by the Topcon TRC-50DX. Some fundi photographs from the Topcon TRC-50DX that exhibited a small dark spot due to a dirty lens were not excluded, the quality control coordinator at the ZIRC judged that this small dark spot did not interfere

with DR grading. However, this spot might have affected the diagnostic results of the AIDRScreening, resulting in relatively low sensitivity. For the participants without mydriasis imaged by the Topcon NW400, the sensitivity and specificity of the AIDRScreening system were 84.00% and 93.33% respectively, which were similar to the sensitivity (86.7%) and specificity (90.7%) of the Idx-DR (24). However, the sensitivity and specificity of the AIDRScreening system increased to 90.00% and 96.51% in the participants imaged by the same fundus camera after mydriasis. We found that some of the images taken without mydriasis were darker than those obtained from participants with mydriasis, especially in the peripheral part of the images, which affected the diagnostic results of the AIDRScreening system and yielded lower sensitivity. However, the specificities of the AIDRScreening system for RDR detection in the participants imaged by the Topcon TRC-50DX and Topcon NW400 without mydriasis were 96.15% and 93.33%, respectively. This high specificity can help in preventing unnecessary referrals, which is particularly important in China.

Additionally, in this study, all the investigators were ophthalmologists, specializing in different fields, with >10 years of clinical experience. The image grading time of the AIDRScreening system was 24 s/eye, which is much shorter than the time required for the manual image grading (38 s/eye). Similarly, the accuracy gain rate of the AIDRScreening system was 16.57%, and the average time gain rate was -37.32%, relative to those of the investigators, which suggests that the AIDRScreening system can improve the efficiency of DR screening. Given the extremely large number (over 114 million) of DM patients (15) and the limited and uneven distribution of ophthalmology resources (mainly in big cities) (14), such an efficient and reliable automated system would have significant benefits for DR screening in China and similar areas.

The current study had some limitations. First, the AIDRScreening system (v. 1.0) used in this study could not automatically assess image quality, but the AIDRScreening system (v. 2.0) has this ability. Second, diabetic macular edema (DME) was not evaluated in this study. Diabetic maculopathy is also considered a major cause of visual impairment in patients with DM. The AIDRScreening system was not designed to grade DME in combination with optical coherence tomography (OCT) images for several reasons. First, DME is easily identified in patients and can be treated in a timely fashion. Second, DME is difficult to detect on fundus photographs and is commonly detected

using OCT in clinical settings. Third, the use of OCT has been limited due to its high costs, poor mobility, and the imbalance of primary medical resources in China. Fundus photographs are not the ideal method for maculopathy detection. A deep learning model for predicting center-involved DME from fundus photographs trained by Varadarajan *et al.* had a sensitivity of 85% and a relatively low specificity of 80% (28). An AI algorithm combining fundus photographs with OCT images had a very high sensitivity (>95%) and appeared to be more effective than the deep learning model in screening DME (29). Recently, Katz *et al.* (30) introduced a new segmentation network for fundus photographs to evaluate the presence of DME. The overall sensitivity of this classifier was 95.5% with a specificity of 81.2% and a false-positive rate of 31.4%. Thus, in our opinion, AI algorithms combining fundus photographs with OCT images may be the best choice for DME screening.

In conclusion, this study showed that the automated AIDRScreening system can reliably detect RDR with high accuracy in adult Chinese patients with DM, independent of camera type and mydriatic status. Due to its good performance in this prospective study, the AIDRScreening system was first approved by the National Medical Products Administration of China in August 2020 after the completion of a registration review to evaluate the benefits and risks of the product. The code of the AIDRScreening system cannot be accessed, but the service can be purchased in China at this time. The implementation of the AIDRScreening system may increase the efficiency of DR screening, decrease the demand for retinal specialists, and thereby greatly reduce vision loss caused by DR.

Acknowledgments

The authors would like to thank Professor Yan Xiaoyan for her help with the statistical analysis.

Funding: The study was supported by funding from Shenzhen SiBright Co. Ltd., China.

Footnote

Reporting Checklist: The authors have completed the STARD reporting checklist. Available at <https://atm.amegroups.com/article/view/10.21037/atm-22-350/rc>

Data Sharing Statement: Available at <https://atm.amegroups.com/article/view/10.21037/atm-22-350/dss>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://atm.amegroups.com/article/view/10.21037/atm-22-350/coif>). All authors report that the study was funded by Shenzhen SiBright Co. Ltd, and the authors have received the written approval from Shenzhen SiBright Co. Ltd. regarding this publication. The authors have no other conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study (clinical trial registration No. NCT03602989) was conducted in accordance with the Declaration of Helsinki (as revised in 2013) and was approved by the institutional review boards of Zhongshan Ophthalmic Center (ZOC, 01 center, 2018QXPJ001), Peking University People's Hospital (PKUPH, 02 center, 2018PHA047), and the Eye Hospital of Wenzhou Medical University (EHWMU, 03 center, 2018MD03). Written informed consent was obtained from each participant.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Cheung N, Mitchell P, Wong TY. Diabetic retinopathy. *Lancet* 2010;376:124-36.
- Yau JW, Rogers SL, Kawasaki R, et al. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care* 2012;35:556-64.
- Wong TY, Sun J, Kawasaki R, et al. Guidelines on Diabetic Eye Care: The International Council of Ophthalmology Recommendations for Screening, Follow-up, Referral, and Treatment Based on Resource Settings. *Ophthalmology* 2018;125:1608-22.
- Wong TY, Cheung CM, Larsen M, et al. Diabetic retinopathy. *Nat Rev Dis Primers* 2016;2:16012.
- Early photocoagulation for diabetic retinopathy. ETDRS report number 9. Early Treatment Diabetic Retinopathy Study Research Group. *Ophthalmology* 1991;98:766-85.
- Bragge P, Gruen RL, Chau M, et al. Screening for presence or absence of diabetic retinopathy: a meta-analysis. *Arch Ophthalmol* 2011;129:435-44.
- Scanlon PH. The English National Screening Programme for diabetic retinopathy 2003-2016. *Acta Diabetol* 2017;54:515-25.
- Association OSoCM. The image collection and film reading for screening diabetic retinopathy guidelines of China (2014). *Chinese Journal of Ophthalmology* 2014;50:851-65.
- Wang FH, Liang YB, Zhang F, et al. Prevalence of diabetic retinopathy in rural China: the Handan Eye Study. *Ophthalmology* 2009;116:461-7.
- Jin G, Xiao W, Ding X, et al. Prevalence of and Risk Factors for Diabetic Retinopathy in a Rural Chinese Population: The Yangxi Eye Study. *Invest Ophthalmol Vis Sci* 2018;59:5067-73.
- Song P, Yu J, Chan KY, et al. Prevalence, risk factors and burden of diabetic retinopathy in China: a systematic review and meta-analysis. *J Glob Health* 2018;8:010803.
- Xie XW, Xu L, Wang YX, et al. Prevalence and associated factors of diabetic retinopathy. *The Beijing Eye Study 2006. Graefes Arch Clin Exp Ophthalmol* 2008;246:1519-26.
- Pan CW, Wang S, Qian DJ, et al. Prevalence, Awareness, and Risk Factors of Diabetic Retinopathy among Adults with Known Type 2 Diabetes Mellitus in an Urban Community in China. *Ophthalmic Epidemiol* 2017;24:188-94.
- Zhang X. *China Health Statistics Yearbook*. Beijing: China Union Medical College Press; 2018. 23-71 p.
- Xu Y, Wang L, He J, et al. Prevalence and control of diabetes in Chinese adults. *JAMA* 2013;310:948-59.
- 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;316:2402-10.
- Wong TY, Bressler NM. Artificial Intelligence With Deep Learning Technology Looks Into Diabetic Retinopathy Screening. *JAMA* 2016;316:2366-7.
- Gargeya R, Leng T. Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology* 2017;124:962-9.
- Ting DSW, Cheung CY, Lim G, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal

- Images From Multiethnic Populations With Diabetes. *JAMA* 2017;318:2211-23.
20. Xu M, Li A, Kong H, et al. Endogenous endophthalmitis caused by a multidrug-resistant hypervirulent *Klebsiella pneumoniae* strain belonging to a novel single locus variant of ST23: first case report in China. *BMC Infect Dis* 2018;18:669.
 21. Sayres R, Taly A, Rahimy E, et al. Using a Deep Learning Algorithm and Integrated Gradients Explanation to Assist Grading for Diabetic Retinopathy. *Ophthalmology* 2019;126:552-64.
 22. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol* 2019;103:167-75.
 23. Keel S, Wu J, Lee PY, et al. Visualizing Deep Learning Models for the Detection of Referable Diabetic Retinopathy and Glaucoma. *JAMA Ophthalmol* 2019;137:288-92.
 24. Abramoff MD, Lavin PT, Birch M, et al. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 2018;1:39.
 25. Lee AY, Yanagihara RT, Lee CS, et al. Multicenter, Head-to-Head, Real-World Validation Study of Seven Automated Artificial Intelligence Diabetic Retinopathy Screening Systems. *Diabetes Care* 2021;44:1168-75.
 26. Association OSocCM. Guidelines for Image Collection and Film Reading of Diabetic Retinopathy Screening in China (2017). *Chinese Journal of Ophthalmology* 2017;53:6.
 27. Liegl R, Liegl K, Ceklic L, et al. Nonmydriatic ultra-wide-field scanning laser ophthalmoscopy (Optomap) versus two-field fundus photography in diabetic retinopathy. *Ophthalmologica* 2014;231:31-6.
 28. Varadarajan AV, Bavishi P, Ruamviboonsuk P, et al. Predicting optical coherence tomography-derived diabetic macular edema grades from fundus photographs using deep learning. *Nat Commun* 2020;11:130.
 29. Hwang DK, Chou YB, Lin TC, et al. Optical coherence tomography-based diabetic macula edema screening with artificial intelligence. *J Chin Med Assoc* 2020;83:1034-8.
 30. Katz O, Presil D, Cohen L, et al. Evaluation of a New Neural Network Classifier for Diabetic Retinopathy. *J Diabetes Sci Technol* 2021. [Epub ahead of print]. doi: 10.1177/19322968211042665.

(English Language Editors: L. Huleatt and J. Gray)

Cite this article as: Yang Y, Pan J, Yuan M, Lai K, Xie H, Ma L, Xu S, Deng R, Zhao M, Luo Y, Lin X. Performance of the AIDRScreening system in detecting diabetic retinopathy in the fundus photographs of Chinese patients: a prospective, multicenter, clinical study. *Ann Transl Med* 2022;10(20):1088. doi: 10.21037/atm-22-350

Table S1 Outcome classification of the ZIRC and AI system

DR Grading	ZIRC (Gold standard, recruited eyes)		
	RDR	Non-RDR	Total
AI system			
RDR	346	22	368
Non-RDR	53	541	594
Total	399	563	962

DR, diabetic retinopathy; RDR, referable diabetic retinopathy (stage II or more severe diabetic retinopathy); non-RDR, nonreferable diabetic retinopathy; ZIRC, Zhongshan Ophthalmic Image Reading Center; AI, artificial intelligence.