BMJ Open Diabetes Research & Care

# Artificial intelligence-enabled screening for diabetic retinopathy: a real-world, multicenter and prospective study

Yifei Zhang,[1] Juan Shi,[1] Ying Peng,[1] Zhiyun Zhao,[1] Qidong Zheng,[2] Zilong Wang,[3] Kun Liu,[4] Shengyin Jiao,[3] Kexin Qiu,[3] Ziheng Zhou,[3,5] Li Yan,[6] Dong Zhao,[7] Hongwei Jiang,[8] Yuancheng Dai,[9] Benli Su,[10] Pei Gu,[11] Heng Su,[12] Qin Wan,[13] Yongde Peng,[14] Jianjun Liu,[15] Ling Hu,[16] Tingyu Ke,[17] Lei Chen,[18] Fengmei Xu,[19] Qijuan Dong,[20] Demetri Terzopoulos,[21,22] Guang Ning,[1] Xun Xu,[4] Xiaowei Ding,[3,5] Weiqing Wang [1]

**To cite:** Zhang Y, Shi J, Peng Y, *et al*. Artificial intelligence-enabled screening for diabetic retinopathy: a real-world, multicenter and prospective study. *BMJ Open Diab Res Care* 2020;**8**:e001596. doi:10.1136/bmjdrc-2020-001596

► Supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjdrc-2020-001596).

YZ, JS, YiP, ZhZ, QZ and ZW are joint first authors.

Received 21 May 2020
Revised 16 July 2020
Accepted 13 August 2020

Check for updates

© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

**Correspondence to**
Dr Weiqing Wang; wqingw61@163.com and Mr Xiaowei Ding; dingxiaowei@sjtu.edu.cn

## ABSTRACT

**Introduction** Early screening for diabetic retinopathy (DR) with an efficient and scalable method is highly needed to reduce blindness, due to the growing epidemic of diabetes. The aim of the study was to validate an artificial intelligence-enabled DR screening and to investigate the prevalence of DR in adult patients with diabetes in China.

**Research design and methods** The study was prospectively conducted at 155 diabetes centers in China. A non-mydriatic, macula-centered fundus photograph per eye was collected and graded through a deep learning (DL)-based, five-stage DR classification. Images from a randomly selected one-third of participants were used for the DL algorithm validation.

**Results** In total, 47 269 patients (mean (SD) age, 54.29 (11.60) years) were enrolled. 15 805 randomly selected participants were reviewed by a panel of specialists for DL algorithm validation. The DR grading algorithms had a 83.3% (95% CI: 81.9% to 84.6%) sensitivity and a 92.5% (95% CI: 92.1% to 92.9%) specificity to detect referable DR. The five-stage DR classification performance (concordance: 83.0%) is comparable to the interobserver variability of specialists (concordance: 84.3%). The estimated prevalence in patients with diabetes detected by DL algorithm for any DR, referable DR and vision-threatening DR were 28.8% (95% CI: 28.4% to 29.3%), 24.4% (95% CI: 24.0% to 24.8%) and 10.8% (95% CI: 10.5% to 11.1%), respectively. The prevalence was higher in female, elderly, longer diabetes duration and higher glycated hemoglobin groups.

**Conclusion** This study performed, a nationwide, multicenter, DL-based DR screening and the results indicated the importance and feasibility of DR screening in clinical practice with this system deployed at diabetes centers.

**Trial registration number** NCT04240652.

## Significance of this study

### What is already known about this subject?

- Previous studies have indicated a high prevalence of diabetes in China; however, the prevalence of diabetes retinopathy (DR) varied and nationwide program for DR screening is lacking.
- A potential value of automated deep learning (DL) algorithm in DR screening was indicated; however, its feasibility in clinical application in population with great heterogeneity needs further investigation.

### What are the new findings?

- We currently validated an artificial intelligence (AI)-enabled DR screening in real-world practice at 155 diabetes centers with comparable performance to human specialists.
- Our study is a large-scale nationwide DR screening program using data from representative cohorts and offered evidence of DR prevalence in patients with diabetes in China.
- It provided evidence of efficiency and accuracy in DL-based DR screening in clinical practice through a comprehensive survey.

### How might these results change the focus of research or clinical practice?

- DL-based DR screening at diabetes centers is feasible, and with a high prevalence of DR detected, it may provide an optional solution to this public health problem in the future.

## INTRODUCTION

According to recent estimates, there were 451 million people with diabetes, aged 18–99 years worldwide in 2017, and the number will increase to 693 million by 2045.[1] The diabetes epidemic is worse in China.[2 3] Per the 2013 national survey, 10.9% of Chinese adults were estimated to suffer from diabetes, and among them, only 36.5% were aware of this diagnosis and 32.2% were treated.[3] The higher prevalence and lower treatment rate of diabetes in China will lead to a higher incidence of diabetes related complications nationwide.[4 5]

Diabetic retinopathy (DR) is one of the common chronic complications of diabetes, which is the leading cause of blindness,



BMJ Open Diab Res Care: first published as 10.1136/bmjdrc-2020-001596 on 21 October 2020. Downloaded from http://drc.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

although preventable in the working age group.[6–8] Early screening and timely referral can delay its progress and effectively prevent vision loss.[9] However, relative to the high prevalence of diabetes in China, the ability to screen for DR is inadequate and a nationwide program for DR screening is scarce. The reasons are multifaceted, including the shortage of eye care specialists, the lack of efficient screening methods and the multidisciplinary process from image acquisition to the diagnosis of DR. In real-world clinical settings, a large portion of patients with diabetes receive their first DR diagnosis during their independent ophthalmologist visits in the symptomatic stage of DR, instead of an earlier diagnosis at diabetes centers or referral visits to ophthalmologists in the non-symptomatic stage.[10–12] In addition, strategies for managing DR in China are difficult to reproduce due to regional economic barriers and living habit differences. Therefore, it is essential to establish a standardized system for early DR detection and management that is feasible for the whole country.

Deep learning (DL), a form of artificial intelligence (AI), has emerged and shown convincing performance in several areas, including medical science.[13–15] A recent study by Ting *et al*[16] has revealed a potential value of automated DL system in DR grading using images from multiethnic cohorts of patients with diabetes, together with several other studies has shown a high sensitivity and specificity in identifying DR (especially referable DR), indicating that the proper use of DL technology in clinical settings may help deliver data-driven analytics for better patient outcome.[16–23]

However, the evidence to confirm the clinical value of DL for DR screening in large-scale healthcare settings is insufficient and most studies have been performed on high-quality image datasets that could hardly represent the variety of image quality and other operational limitations of real-world DR screening applied at diabetes centers.[17 18] There are few reports regarding the practical application of AI in clinic-based DR screening, with patient cohorts of 3049 and 1415, respectively.[20 24] Its feasibility and quality in real-world use must be further explored using datasets with larger sample sizes and demographic variations.

Therefore, in the present study, we conducted a prospective, nationwide DR screening, using a DL algorithm, with a cohort of 47 269 patients at 155 diabetes centers in China. The operational feasibility and accuracy of the DL algorithm was validated and the prevalence of DR, referable DR (moderate non-proliferative DR (NPDR) or worse), and vision-threatening DR (VTDR, severe NPDR or worse, and/or clinically significant macular edema (CSME)) was reported.

## METHODS

### Population

The National Metabolic Management Center (MMC) is a pilot diabetes care system in China, founded in 2016. It aims at establishing a nationwide, standard and reproducible platform based on advanced medical equipment and Internet of Things technology for the diagnosis and management of diabetes and its complications.[25] The Diabetic Retinopathy Screening and Prevention Program is an MMC branch project. Its purpose is to develop an efficient workflow for the early detection, timely follow-up and management of DR, and to establish a referral system for future treatment and long-term follow-up.

Between June 2018 and August 2019, a total of 47 269 consecutive patients with diabetes aged 18 years or older from 155 MMCs in China were enrolled in the present study. The involved MMCs were in the hospitals with different levels according to tiered medical service system throughout 26 provinces in China. All the participants were screened for DR by the DL-based system, which labeled the fundus images as DR stage or ungradable due to image quality issues. Fundus images obtained from one-third of randomly selected participants were reviewed offline by a two-stage reading performed by a panel of specialists for the purposes of DL algorithm validation on both DR grading and image quality assessment (figure 1).

All the participants underwent a full medical examination at the local MMCs.

### Baseline data collection

The eligible participants were those with a diagnosis of diabetes according to the WHO criteria.[26] Detailed inclusion and exclusion criteria are summarized in the online supplemental methods. At baseline, all data (including a standardized questionnaire and comprehensive clinical and laboratory examinations) were collected from each participant through an MMC specialized electronic medical record system.[25]

Data collection was conducted by trained staff according to a standard protocol. Social demographic characteristics, medical history and lifestyle factors were recorded. Height and body weight were measured by a height-weight scale with participants in light clothes without shoes, and body mass index (BMI) was calculated as the weight in kilograms divided by height in meters squared. Blood pressure and heart rate were measured with electronic blood pressure monitors after at least a 5 min rest in the seated position. Waist circumference was measured on standing participants midway between the lower edge of the costal arch and the upper edge of the iliac crest. The participants were required to undergo a standard steamed bread meal test after an overnight fasting, and blood samples were collected at 0 and 2 hours during the test. Detailed data collection procedures are listed in the online supplemental methods.

### Fundus photography acquisition

One standard, non-mydriatic, 45° field of view, macula-centered color and non-stereoscopic retinal fundus image was acquired from each eye of each participant. Various models of fundus cameras were used. Topcon



BMJ Open Diab Res Care: first published as 10.1136/bmjdrc-2020-001596 on 21 October 2020. Downloaded from http://drc.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

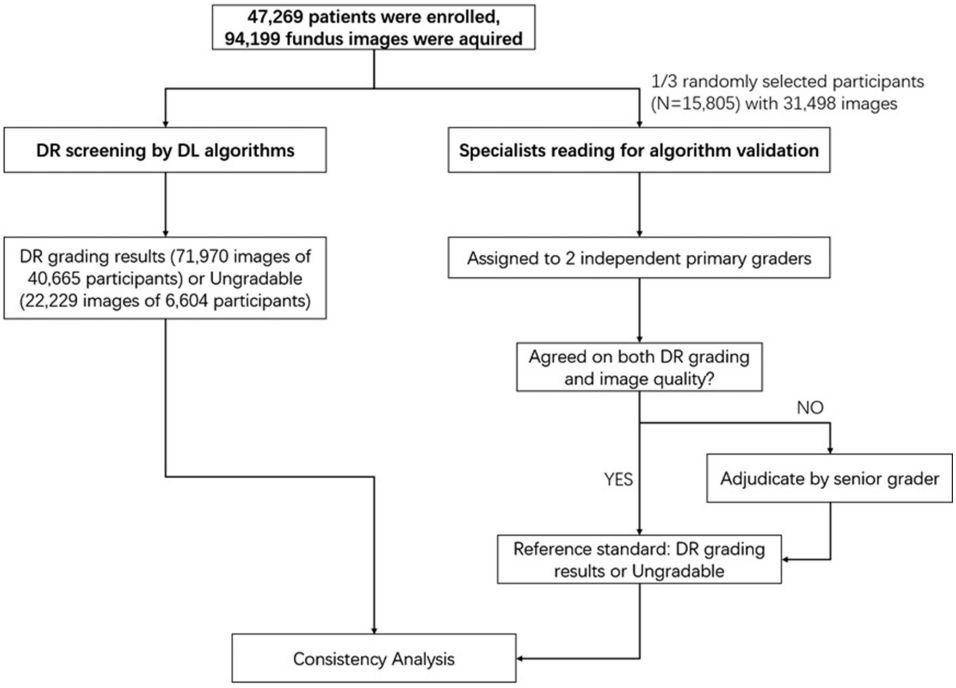



**Figure 1** Fundus image grading work flow and adjudication. DL, deep learning; DR, diabetic retinopathy.

TRC-NW400, MiiS DSC-200, Canon CR-2 PLUS AF, Canon CR-2 AF and Zeiss VISUCAM200 cameras were used in >80% of all the centers (online supplemental table 1). At all the centers, trained technicians took only non-mydriatic images, and no pupillary dilation images were additionally acquired. All the participants' images were anonymized before grading.

## Development of the DL algorithms

VoxelCloud Retina, an automated retinal disease screening system, was used to grade fundus images. The VoxelCloud Retina DR system was developed using DL techniques.

Two sets of data were used to train the different deep learning networks that form the final ensemble of DR and diabetic macular edema (DME) severity classification modules. The first dataset comprises 143 626 fundus photographs of 37 231 patients obtained from 2005 to 2015 from a large private retinal image database (online supplemental tables 2 and 3).

The second dataset comprises 1184 color fundus images from a public hospital in China, which were assigned a DR severity grade based on consensus from three ophthalmologists (online supplemental table 4). These data were chosen to help improve the model performance on confusing cases that could fall on the boundary between two grades.

The DR and DME models are an ensemble of six neural networks (online supplemental figure 1). All the six neural networks use the state-of-the-art Inception-ResNet v2 architecture;[27] however, several design differences among them are critical for the effective performance of the model ensemble. Details are presented in the online supplemental methods.

In addition to the DR and DME models, the system also includes trained independent lesion models that detect the presence of lesions that contribute to DR grade, including fundus hemorrhage, hard exudates and laser scars. These independent lesion models are used to achieve improvements in the DR prediction performance, based on the DR classification rules listed in online supplemental table 5.

All color fundus images are normalized to pixel intensity values between 0 and 1 and are resized to a standard resolution of 800 by 800 pixels before being processed by the system.

The system was also tested on various private and public datasets. The testing results on the APTOS 2019 Blindness Detection dataset, a public dataset also collected in a real-world scenario close to that of the present study, is reported in the online supplemental methods (https://www.kaggle.com/c/aptos2019-blindness-detection/overview).

## DL-based DR grading

The system specified for DR screening comprises the following three modules:

### Quality control Module

The quality control (QC) module evaluates the quality of fundus images before the five-stage DR grading. The quality of fundus images is classified as gradable or ungradable. Those assessed as ungradable (low quality) are not sent for further DR grading. Gradable images are further sorted into excellent and adequate quality, while ungradable images are sorted into insufficient information and non-fundus images. The gradeability criteria in the model training phase were: 1) the image must cover at least 45° of the retinal area with the macula and the optic disc visible; 2) at least 80% of the retinal area must



BMJ Open Diab Res Care: first published as 10.1136/bmjdrc-2020-001596 on 21 October 2020. Downloaded from http://drc.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

be recognizable and 3) no overexposure, underexposure or blur caused by focusing failure and motion.

### DR severity classification module

The DR severity classification module provides each fundus image a five-stage DR severity classification that can be further transferred to multiple binary classifications to meet different demands. The severity classification mainly follows the International Clinical Diabetic Retinopathy (ICDR) severity scale,[28] which is developed by the International Council of Ophthalmology and adopted by the American Academy of Ophthalmology.[29] Slight modifications were made to adapt to the situation, considering that only a single non-mydriatic fundus image was acquired from each eye, covering the posterior pole, instead of seven mydriatic images covering all four quadrants (online supplemental table 5).[30]

The patient-level DR grade was based on the worse DR grade of the two eyes. If both eye images of a patient are classified as ungradable, then the patient is classified as ungradable. If only one eye image is classified as gradable, then the patient-level DR grading is based on this eye. If a patient has only one eye image, and it is classified as ungradable, then this patient is classified as ungradable.

### DME severity classification module

The DME severity classification module provides subjects each fundus image to a three-stage DME severity estimation that can be further transferred to multiple binary classifications. As DME assessment, which requires retinal thickness information is not possible in non-mydriatic fundus images, the presence of hard exudates is regarded as a presumptive diagnosis of DME (online supplemental table 6).

DR was defined as presence of mild NPDR or worse; referable DR, moderate NPDR or worse and VTDR, severe NPDR or worse and/or CSME.

## Expert ground truth grading

The ground truth for fundus image diagnosis was provided by a two-stage reading by specialist graders. The grading team was led by the Ophthalmology Center of the Shanghai General Hospital (National Clinical Research Center for Eye Diseases). All graders were ophthalmologists from tertiary hospitals with 3 years or more of work experience. Each grader finished two rounds of training and passed a qualification test following ICDR guidelines. Graders were divided into primary graders and reviewers (senior graders) based on their seniority and performance. The grading was conducted in two stages.

### Stage 1

Two primary graders read the fundus image and gave image quality grades and DR grades independently. If the two primary graders reached a consensus on both the image quality and DR grades, the grading of this fundus image ended in stage 1 and the grades served as the ground truth.

### Stage 2

A reviewer (senior grader) who could access the assessments of both primary graders' was added to the grading process if the two primary graders disagreed on either the image quality or DR grades. The reviewer's sole opinion served as the final grade for such cases (figure 1).

## Statistical analysis

Statistical analyses were performed with the use of SPSS V.22.0 (Chicago, Illinois, USA). Data were provided in the form of the mean and SD for continuous variables, or the number with the percentage for categorical variables. The prevalence (95% CIs) of DR, referable DR and VTDR were estimated overall and compared within subgroups of sex, age, categories of diabetes duration and glycated hemoglobin (HbA1c) with the $\chi^2$ test. The demographic and clinical characteristics were assessed and compared by sex with the $\chi^2$ test for categorical variables, and with the Student's t-test for continuous variables.

Fundus images from one-third of randomly selected participants were used for DL algorithm validation. The ground truth of fundus image diagnosis provided by the expert panel is considered as the reference standard. The accuracy of DR grading, image quality and two-category derivatives (one DR grading or worse) of patients with diabetes were evaluated. The consistency and the accuracy among the DL algorithm and reference standard, the primary graders in the expert panel and the primary grader and reference standard were analyzed; 2×2 tables were generated to analyze the sensitivity, specificity, negative predictive value and positive predictive value of the DL algorithm in detecting DR, referable DR and severe NPDR or worse, as well as the image quality compared with the reference standard at the individual eye level. Consistency evaluations of the five-stage grading confusion matrix by kappa index and quadratic weighted kappa scores were also calculated. All p values were two-tailed and a p value <0.05 was considered statistically significant.

# RESULTS

## Clinical characteristics of all the participants

In total, 47 269 participants with diabetes from 155 centers were enrolled in the present study, among which 27 110 (57.4%) were men (table 1 and figure 2). The mean (SD) age of all the participants was 54.29 (11.60) years, the mean diabetes duration was 6.80 (6.71) years and the mean HbA1c was 9.06 (2.27) % or 75.45 (24.85) mmol/mol. Since 97.92% of the participants had type 2 diabetes (1.61% type 1 diabetes, 0.32% gestational diabetes and 0.14% others, totaling 99.99% due to rounding), no further analysis was performed based on the diabetes classification.

## DL algorithm validation

A total of 31 498 images from one-third (No.=15 805) of the randomly selected participants were used for DL algorithm validation (figure 1).



BMJ Open Diab Res Care: first published as 10.1136/bmjdrc-2020-001596 on 21 October 2020. Downloaded from http://drc.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

**Table 1** Clinical characteristics of the study participants

| | Total (n=47 269) | By sex stratification | | P value |
|---|---|---|---|---|
| | | Male (n=27 110) | Female (n=20 159) | |
| Age, years | 54.29±11.60 | 52.76±11.58 | 56.35±11.29 | <0.001 |
| High school education and above | 17 661 (43.0) | 12 301 (51.5) | 5360 (31.2) | <0.001 |
| Han Chinese ethnicity | 37 457 (96.1) | 21 788 (95.9) | 15 669 (96.4) | 0.010 |
| Family history of diabetes | 16 294 (40.1) | 9271 (39.3) | 7023 (41.3) | <0.001 |
| Duration of diabetes, years | 6.80±6.71 | 6.31±6.50 | 7.47±6.93 | <0.001 |
| History of hypertension | 16 266 (39.8) | 8835 (37.2) | 7431 (43.4) | <0.001 |
| History of dyslipidemia | 11 361 (27.9) | 6825 (28.8) | 4536 (26.5) | <0.001 |
| Body mass index, kg/m$^2$ | 25.62±3.77 | 25.77±3.63 | 25.42±3.95 | <0.001 |
| Waist circumference, cm | 91.07±10.29 | 92.73±9.83 | 88.80±10.49 | <0.001 |
| Systolic blood pressure, mm Hg | 131.40±19.06 | 130.36±18.23 | 132.81±20.04 | <0.001 |
| Diastolic blood pressure, mm Hg | 77.34±11.34 | 78.70±11.28 | 75.51±11.16 | <0.001 |
| HbA1c, % | 9.06±2.27 | 9.14±2.32 | 8.94±2.20 | <0.001 |
| HbA1c, mmol/mol | 75.45±24.85 | 76.36±25.34 | 74.18±24.09 | <0.001 |
| Fasting blood glucose, mmol/L | 9.34±3.74 | 9.38±3.80 | 9.27±3.66 | 0.005 |
| Postprandial blood glucose, mmol/L | 16.16±5.44 | 16.10±5.33 | 16.24±5.58 | 0.016 |
| Triglyceride, mmol/L | 2.27±2.35 | 2.39±2.59 | 2.09±1.96 | <0.001 |
| Total cholesterol, mmol/L | 4.83±1.36 | 4.73±1.36 | 4.98±1.33 | <0.001 |
| LDL cholesterol, mmol/L | 2.79±1.00 | 2.74±0.97 | 2.87±1.03 | <0.001 |
| Serum creatine, μmol/L | 66.67±30.43 | 73.69±31.45 | 56.88±25.96 | <0.001 |
| Urinary acid, μmol/L | 317.83±96.50 | 338.38±96.70 | 289.15±88.57 | <0.001 |

Data are given as a mean±SD or as a number and percentage in parentheses. Comparisons of mean values and proportions by sex were performed using the Student's t-test and $\chi^2$ tests, respectively.
HbA1c, glycated hemoglobin; LDL, low-density lipoprotein.

For image quality assessment, from these images, 26 698 (84.8%) images were assessed as gradable by the reference standard (online supplemental table 7). Compared with the reference standard, the QC module had a 63.3% (95% CI: 61.9% to 64.7%) sensitivity and 85.0% (95% CI: 84.6% to 85.4%) specificity, with positive predictive value 43.2% (95% CI: 42.0% to 44.3%) and negative predictive value 92.8% (95% CI: 92.4% to 93.1%), respectively. The interobserver variability (setting one grader as reference standard) between two primary expert graders had a sensitivity of 69.6% (95% CI: 68.4% to 70.8%) and a specificity of 86.8% (95% CI: 86.4% to 87.2%), with positive predictive value 53.1% (95% CI: 51.9% to 54.2%) and negative predictive value 93.0% (95% CI: 92.7% to 93.3%).


**Figure 2** Geographic distribution of the 155 metabolic management centers in China involved in this study.



BMJ Open Diab Res Care: first published as 10.1136/bmjdrc-2020-001596 on 21 October 2020. Downloaded from http://drc.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

For DR grading, the concordance between the DL algorithm and reference standard was 83.0% for the five-stage DR grading. The corresponding quadratic weighted kappa were 0.72 (95% CI: 0.72 to 0.72) (online supplemental table 8 and online supplemental figure 2). The DL algorithm had an 83.3% (95% CI: 81.9% to 84.6%) sensitivity and 92.5% (95% CI: 92.1% to 92.9%) specificity for detecting referable DR. The positive and negative predictive values were 61.8% (95% CI: 60.3% to 63.3%) and 97.4% (95% CI: 97.2% to 97.7%), respectively. The Youden index was 75.8%. For two-stage manual grading, the concordance for the five-stage DR grading between the two primary graders, and between the primary graders and the reference standard were 84.3% and 91.0%, respectively. The corresponding quadratic weighted kappa were 0.74 (95% CI: 0.74 to 0.74) and 0.87 (95% CI: 0.87 to 0.87), respectively. The concordance between the DL algorithm and primary grader 1, primary grader 2 or one primary grader (combined two primary graders) were 82.8%, 81.8% and 82.3%, respectively. The corresponding quadratic weighted kappa were 0.66 (95% CI: 0.66 to 0.66), 0.67 (95% CI: 0.67 to 0.67) and 0.67 (95% CI: 0.67 to 0.67), respectively (online supplemental table 8). Confusion matrices of the five-stage DR evaluation between the two primary graders, and between the primary graders and the reference standard are reported in online supplemental figures 3 and 4.

Typical examples of false negative and false positive cases of DL QC and the grading module are shown in online supplemental figures 5 and 6.

### AI-enabled DR screening

In total, 94 199 fundus images from all the participants were graded by the DL algorithm. Among all the images, 22 404 (23.8%) images were assessed as high quality, 49 566 (52.6%) as medium quality and 22 229 (23.6%) as low quality (ungradable) by the QC module (online supplemental table 9). Thus, a total of 71 970 (76.4%) images from 40 665 (86.0%) participants were finally qualified for DR grading by the DL algorithm (online supplemental tables 9 and 10). The ungradable images were mainly due to small pupil size or the presence of cataracts or other rare eye diseases and camera operation problems (online supplemental table 11).[19 21 22 31–33] Participants with ungradable images were recommended to the ophthalmology department for further examination.

Among the 40 665 gradable participants, the estimated prevalence of DR was 28.8% (95% CI: 28.4% to 29.3%), referable DR was 24.4% (95% CI: 24.0% to 24.8%) and VTDR was 10.8% (95% CI: 10.5% to 11.1%) (table 2). When analyzed by risk factor stratifications, the estimated prevalence of DR was higher in women 29.6% (95% CI: 28.9% to 30.3%), than in men, 28.3% (95% CI: 27.7% to 28.8%) (p=0.0029). The estimated prevalence of DR increased with age and duration of diabetes (both p values for trend <0.0001). Similar results were found in referable DR and VTDR in the stratification of these risk factors. Furthermore, by the HbA1c stratification, when HbA1c was <10.0% (85.77 mmol/mol), the prevalence of DR and referable DR increased with the raise of HbA1c (both p values for trend <0.0001), but decreased slightly without statistical significance when the HbA1c was 10.0% or higher (both p values >0.05). The prevalence of VTDR increased constantly with the raise of HbA1c (p value for trend <0.0001) (table 2, and online supplemental tables 12 and 13, and online supplemental figure 7).

The five-stage DR grading and corresponding DME classification results by the DL algorithm for 40 665 gradable participants are shown in online supplemental table 14. The percentage of ungradable images and the DR grading results based on different types of cameras were listed in online supplemental tables 15 and 16.

## DISCUSSION

In this large multicenter, real-world DR screening program, a DL-based AI system was deployed at 155 diabetes centers. Our study demonstrated that, in Chinese adults with diabetes, the estimated prevalence for any DR, referable DR and VTDR was 28.8%, 24.4% and 10.8%, respectively. The high prevalence of DR in various stages indicated the importance and urgency of early detection of DR in China. A DL system with comparable sensitivity and specificity to a panel of specialists enabled the efficient screening for DR at diabetes centers nationwide, and it may provide a solution to this problem.

Screening for DR in daily clinical work has not yet been well established at diabetes centers in China due to resource, infrastructure and retinal specialist limitations. Therefore, a comprehensive survey on DR prevalence and its actual burden in the whole country remains unaddressed.[6] Highly demanded at every diabetes center is the timely diagnosis and treatment of DR in order to achieve better outcomes over the widest diabetic population regardless of geographic and economic barriers.

Epidemiological studies published in the recent 10 years have demonstrated the prevalence of DR in China ranged from 5.4% to 44.8% in patients with diabetes.[6–8 34] The variability of DR prevalence in different studies was mainly due to the heterogeneity among the studies, including sample size, study design, clinical characteristics of participants, geographic region and DR classification criteria. A recent meta-analysis, which collected data from 31 community-based studies, showed that the pooled prevalence of any DR in DM participants was 18.45%, for NPDR it was 15.06% and for PDR it was 0.99%.[6] However, a single survey that reports the actual prevalence of DR in the whole country is lacking. In the present study, a large multicenter DR screening program, implemented with the aid of AI technology, was conducted in 26 provinces in China. The survey has provided the most up-to-date information on DR characteristics in adults with diabetes and has indicated a high prevalence of DR in China. In addition, through stratification, the crude prevalence of DR was higher in older age groups and, together with the societal aging, it increases the burden



BMJ Open Diab Res Care: first published as 10.1136/bmjdrc-2020-001596 on 21 October 2020. Downloaded from http://drc.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

Table 2 Prevalence of diabetic retinopathy (DR), referable DR and vision-threatening DR (VTDR) in total and among different risk factor stratification

| | Prevalence % (95% CI) | | | |
|---|---|---|---|---|
| | DR | Referable DR | VTDR (including CSME) | No. of patients |
| Total | 28.8 (28.4 to 29.3) | 24.4 (24.0 to 24.8) | 10.8 (10.5 to 11.1) | 40 665 |
| Gender | | | | |
| Male | 28.3 (27.7 to 28.8) | 23.5 (23.0 to 24.1) | 10.0 (9.6 to 10.4) | 23 686 |
| Female | 29.6 (28.9 to 30.3) | 25.5 (24.9 to 26.2) | 11.9 (11.4 to 12.4) | 16 979 |
| Age groups, years | | | | |
| 18–29 | 18.7 (16.6 to 20.8) | 12.7 (10.9 to 14.4) | 6.2 (4.9 to 7.5) | 1375 |
| 30–39 | 22.9 (21.6 to 24.3) | 17.0 (15.8 to 18.2) | 6.8 (6.0 to 7.6) | 3775 |
| 40–49 | 27.9 (26.9 to 29.0) | 22.2 (21.4 to 23.1) | 8.9 (8.3 to 9.5) | 8650 |
| 50–59 | 30.2 (29.5 to 31.0) | 25.6 (24.8 to 26.3) | 11.0 (10.5 to 11.5) | 14 231 |
| 60–69 | 30.2 (29.3 to 31.1) | 27.1 (26.3 to 28.0) | 12.5 (11.9 to 13.2) | 10 304 |
| ≥70 | 33.5 (31.6 to 35.4) | 31.8 (29.9 to 33.7) | 18.0 (16.5 to 19.6) | 2330 |
| Diabetic duration, years | | | | |
| <5 | 20.0 (19.4 to 20.6) | 15.6 (15.0 to 16.1) | 5.9 (5.5 to 6.2) | 17 175 |
| 5–10 | 30.8 (29.9 to 31.8) | 25.7 (24.7 to 26.7) | 10.7 (10.0 to 11.4) | 7246 |
| 10–15 | 41.4 (40.1 to 42.7) | 35.9 (34.6 to 37.1) | 16.8 (15.8 to 17.8) | 5403 |
| 15–20 | 49.1 (47.1 to 51.1) | 44.3 (42.3 to 46.2) | 22.1 (20.4 to 23.7) | 2426 |
| ≥20 | 52.5 (50.0 to 54.9) | 48.0 (45.5 to 50.4) | 10.7 (9.9 to 11.4) | 1618 |
| HbA1c, % | | | | |
| <6.5 | 18.4 (17.1 to 19.6) | 14.9 (13.7 to 16.0) | 6.4 (5.6 to 7.1) | 3778 |
| 6.5–6.9 | 21.2 (19.8 to 22.7) | 17.2 (15.8 to 18.5) | 7.3 (6.4 to 8.2) | 3046 |
| 7.0–7.9 | 26.5 (25.4 to 27.6) | 21.7 (20.7 to 22.8) | 9.5 (8.7 to 10.2) | 6208 |
| 8.0–8.9 | 32.9 (31.7 to 34.2) | 27.3 (26.1 to 28.4) | 11.2 (10.4 to 12.0) | 5685 |
| 9.0–9.9 | 34.0 (32.6 to 35.3) | 29.4 (28.2 to 30.7) | 12.4 (11.5 to 13.3) | 4941 |
| ≥10.0 | 33.3 (32.5 to 34.2) | 28.4 (27.6 to 29.2) | 13.0 (12.4 to 13.6) | 11 338 |

CSME, clinically significant macular edema; DR, diabetic retinopathy; HbA1c, glycated hemoglobin; VTDR, vision-threatening diabetic retinopathy.

to the healthcare system. However, since the prevalence of DR was decreased in subgroups with lower degrees of HbA1c, it may predict a better glycemic control with the lessening of eye complications.

Most DL-based DR grading studies have focused on the methodology development and validation using high-quality, curated public datasets.[17–19] The implementation of automated DL algorithms for DR screening in real-world practice was rare.[20 23 24 35] One example was the large community-based, nationwide DR screening program using DL algorithm in Thailand.[23] Another two examples in its use in clinical settings were performed by Gulshan *et al* and van der Heijden *et al*, respectively.[20 24] The former study involved 3049 patients with diabetes in two eye care clinics in India.[20] The results demonstrated 88.9% and 92.1% sensitivities, and 92.2% and 95.2% specificities for the detection of moderate or worse DR in the two clinics, respectively. The latter was performed in the Hoorn diabetes center including 1415 patients which reported a 68.0% sensitivity and 86.0% specificity for detecting referable DR by the IDx-DR device based on ICDR standard, compared with adjudicated reference standard by a panel of three experts; the averaged sensitivity and specificity of the three experts against the adjudicated reference standard were 74.7% and 99.7%, respectively; however, the quality of the fundus images collected was unsatisfactory, which may be due to the implementation of the study in the non-ophthalmic specialized clinical setting.[24] These studies offered good examples and indicated the feasibility and validity of DL implementation in real-world clinical work flows. However, in these studies, the DL algorithms were deployed only at individual centers with small or moderate sample size. The wide deployment of DL-based systems to multiple non-ophthalmic specialized medical centers or healthcare systems with different resources remains unclear.

Therefore, in the present study we applied a DL algorithm for DR screening at 155 diabetes care centers involving 47 269 patients with diabetes in China. A variety of fundus cameras meeting the base requirements



BMJ Open Diab Res Care: first published as 10.1136/bmjdrc-2020-001596 on 21 October 2020. Downloaded from http://drc.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

for photograph acquisition were used. The DL algorithms provided a five-stage DR severity grading and DME detection in a real-time manner. The DL system was integrated with various fundus camera models used in MMCs, allowing seamless, push-button image QC and DR staging onsite. None of the deep neural networks in the DL system was trained or fine tuned using any MMC images, demonstrating strong domain transfer and generalization capability, as well as robustness and reproducibility on unseen images. The sensitivity for detecting referable DR was 83.3%, and the specificity was 92.5%, with an Youden index of 75.8%. The performance of the DL system is comparable to the interobserver variability of specialists who are limited in availability (1.1 hour/day on average) and have a long response time (1.5 days on average) in real-world practice. The high specificity (92.5%) performance of the DL system in detecting referable DR may be used as a safe and low-false-alarm autonomous referral decision, that is, all patients classified as referable DR by the algorithms are referred to specialists without further manual review. The algorithms were trained on datasets collected from different populations and scenarios, and they show good generalization characteristics. Furthermore, in order to evaluate the effects of the QC module of the DL system, the quality assessment results obtained by the algorithm and by the reference standard were compared. Although low-quality images were inevitable in non-ophthalmic clinical settings, by enabling AI QC feedback in the image acquisition phase, the proportion of qualified images could reach 92.8% of all the fundus images acquired according to the negative predictive value of the QC model, together with strengthening the training process on technician's operation skills (ie, distinguishing patients with small pupil or cataracts, and improving image contrast or focus issues), the percentage of low-quality images will reduce to the least extent and lead to more reliable subsequent DR grading in the future work.

There are several strengths in the present study. First, it was conducted at 155 diabetes centers in China. The study results were representative because the involved MMCs were in the hospitals with different levels according to tiered medical service system and in the regions with different economic and culture background. Furthermore, the study sample size was large and enrolled consecutive patients with proper sex ratio, wide distribution of age, diabetes duration and metabolic control situation which mimics the characteristics of diabetes in the real-world situation. Second, it was a large AI-enabled DR screening program, with comparable performance to specialists. The automated DL system proved to be a scalable solution given the markedly increased diabetes prevalence and relatively inadequate medical resources in China, so as to perform effective screening of patients at diabetes centers that diagnose and manage the majority of patients with diabetes. In addition, the image QC module has significantly increased the validity and accuracy of DR screening, which enables the regular screening of DR in non-ophthalmic clinical settings.

The study has several limitations. First, since the study was conducted at multiple clinical centers, even with the large sample size, the DR prevalence was not commensurate to that of the general population. Second, the estimated prevalence of DR (27.57%) and referable DR (16.59%) by the reference standard in one-third of the randomly selected participants were relatively lower than those by the AI screening. The higher negative predictive values, but the lower positive predictive value might lead to an overestimate of DR prevalence by the DL algorithm. While, the other factors, including only one single non-mydriatic fundus photography instead of multifield fundus photography were obtained might underestimate the DR prevalence by the DL algorithm. In addition, there were disagreements between the human graders and the DL QC model. The typical example of false negative result was the out of focus image judged as ungradable by the algorithm but gradable by the graders, while the false positive result was the too dark image judged as ungradable by graders but gradable by the algorithm (online supplemental figure 5). For all the above reasons, one should be cautious in interpreting the current findings.

In conclusion, in the present study, we validated the feasibility and accuracy of an automated DL algorithm in DR screening and surveyed the prevalence of DR, referable DR and VTDR at 155 diabetes centers in China. With comparable performance to human specialists and scalability, the automated system may offer an effective, cost-efficient and practical screening in routine diabetes follow-up and retinal complication management. More diabetes centers and primary care facilities are now joining the program to improve and validate the screening and referral procedures, thereby endeavoring to mitigate the public health problem.

**Author affiliations**
[1]Department of Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; Shanghai National Clinical Research Center for metabolic Diseases, Key Laboratory for Endocrine and Metabolic Diseases of the National Health Commission of the PR China, Shanghai National Center for Translational Medicine, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[2]Department of Internal Medicine, The Second People's Hospital of Yuhuan, Yuhuan, China
[3]Department of Research, VoxelCloud, Shanghai, China
[4]Department of Ophthalmology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[5]Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China
[6]Department of Ophthalmology, The Third People's Hospital of Datong, Datong, China
[7]Center for Endocrine Metabolism and Immune Diseases, Beijing Luhe Hospital, Capital Medical University, Beijing, China
[8]Department of Endocrinology and Metabolism, The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology; Luoyang City Clinical Research Center for Endocrinology and Metabolism, Luoyang, China
[9]Department of Internal Medicine of Traditional Chinese Medicine, Sheyang Diabetes Hospital, Yancheng, China



BMJ Open Diab Res Care: first published as 10.1136/bmjdrc-2020-001596 on 21 October 2020. Downloaded from http://drc.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

[10]Department of Endocrinology, The Second Affiliated Hospital Dalian Medical University, Dalian, China
[11]Department of Endocrinology, Datong Coal Group Ltd. General Hospital, Datong, China
[12]Department of Endocrine and Metabolic Diseases, The First People's Hospital of Yunnan Province, Kunming, China
[13]Department of Endocrinology and Metabolism, The Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan, China
[14]Department of Endocrinology and Metabolism, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[15]Department of Endocrinology, Longkou People's Hospital, Yantai, China
[16]Department of Endocrinology, The Third Affiliated Hospital of Nanchang University, Nanchang, China
[17]Department of Endocrinology, The Second Affiliated Hospital of Kunming Medical University, Kunming, China
[18]Department of Endocrinology and Metabolism, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou, Jiangsu, China
[19]Department of Endocrinology and Metabolism, Hebi Coal (group) Ltd. General Hospital, Hebi, China
[20]Department of Endocrinology and Metabolism, People's Hospital of Zhengzhou, Zhengzhou, China
[21]Department of Computer Science, Computer Graphics & Vision Laboratory, University of California Los Angeles, Los Angeles, California, USA
[22]Department of Research, VoxelCloud, Los Angeles, California, USA

**Acknowledgements** The authors would like to thank all study participants and participating centers. We thank Drs Yufan Wang (Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China), Weijiang Chu (Laizhou Municipal Hospital, Shandong Province, China), Lin Zhang (Bayannur Hospital, Inner Mongolia Autonomous Region, China), Yanmei Yu (Mudanjiang Cardiovascular Hospital, Heilongjiang Province, China), Xingjian Zhou (Xiangyang No.1 People's Hospital, Hubei Province, China), Hongmei Qiu (People's Hospital of Yuxi City, Yunnan Province, China), Wenbing Ai (Yiling Hospital of Yichang, Hubei Province, China), Xueqin Wang (First People's Hospital of Nantong, Jiangsu Province), Zhiqiang Kang (ZhengZhou Central Hospital, Henan Province, China), Xiaowen Chen (Huangshi Central Hospital, Hubei Province, China), Chunlei Deng (People's Hospital of Ningxiang, Hunan Province, China), Mingfu Ma (The Fifth People's Hospital of Qinghai Province, Qinghai Province, China), Hong Yang (Ruian People's Hospital, Zhejiang Province, China), Huige Shao (Changsha Central Hospital, Hunan Province, China), Shen Qu (Shanghai Tenth People's Hospital, Shanghai, China), Feixia Shen (The First Affiliated Hospital of Wenzhou Medical University, Zhejiang Province, China), Bangqun Ji (Xingyi People's Hospital, Guizhou Province, China), Jianling Du (The First Affiliated Hospital of Dalian Medical University, Liaoning Province, China), Riqiu Chen (Lishui City People's Hospital, Zhejiang Province, China), Wei Tang (Geriatric Hospital of Nanjing Medical University, Jiangsu Province, China), Xueyuan Jia (Dujiangyan Medical Center, Sichuan Province, China), Laixiang Li (Qinan County Hospital of Traditional Chinese Medicine, Gansu Province, China), Lin Yuan (Zhuhai People's Hospital, Guangdong Province, China), Yongjun Chen (Hongze Huaian District People's Hospital, Jiangsu Province, China), Ning Xu (The First People's Hospital of Lianyungang, Jiangsu Province, China), Lin Gao (Affiliated Hospital of Zunyi Medical University, Guizhou Province, China), Canli Gu (Ruyang People's Hospital, Henan Province, China), Zhaoli Yan (The Affiliated Hospital of Inner Mongolia Medical University, Inner Mongolia Autonomous Region, China), Wenzhi Zhang (Qixia People's Hospital), Shandong Province, China), Bingyin Shi (The First Affiliated Hospital of Xi'an JiaoTong University, Shanxi Province, China), Hongyan Deng (Puai Hospital, Hubei Province, China), Jie Shen (The Third Affiliated Hospital of Southern Medical University, Guangdong Province, China), Anhua Huang (Renhuai City People Hospital, Guizhou Province, China), Feng Wei (Binzhou People's Hospital, Shandong Province, China), Yufang Gao (Zuozhou City Hospital, Hebei Province, China), Jinhong Chen (People's Hospital of Hengxian County, Nanning, Guangxi, China), Yu Zhao (Baoan Central Hospital of Shenzhen, Guangdong Province, China), Xinhua Ye (Changzhou No.2 People's Hospital, Jiangsu Province, China), Weici Xie (The First People's Hospital of Tianmen in Hubei Province, China), Jinsong Kuang (The Fourth Hospital of People, Shenyang, Liaoning Province, China), Yan Feng (Mudanjiang City Second People's Hospital, Heilongjiang Province, China), Yunping Zhang (The People's Hospital of Xiangyun, Yunnan Province, China), Wei Zhu (Sucheng Cai Town Hospital, Jiangsu Province, China), Shiwei Cui (Affiliated Hospital of Natong University, Jiangsu Province, China), Zunhai Zhou (Yangpu Hospital,Tongji University, Shanghai, China), Xiaoqing Su (Jiangxi Pingxiang People's Hospital, Jiangxi Province, China), Yu Shi (Qidong People's Hospital, Jiangsu Province, China), Xiaoju Qi (The People's Hospital of Yunxian, Yunnan Province, China), Jie Yang (Fujian Jianou Hospital, Fujian Province, China), Jianying Pu (Shanghai Gonghui Hospital, Shanghai, China), Ping Tang (Shenzhen Luohu Hospital Group, Guangdong Province, China), Rongyue Chen (Third People's Hospital of Xuchang, Henan Province, China), Yingli Pan (Fangda Medical Yingkou People's Hospital, Liaoning Province, China), Jinhua Qiu (People's Hospital of Xinfeng County, Jiangxi Province, China), Liwei Qiu (Anyang Hospital of Traditional Chinese Medicine, Henan Province, China), Hui Cao (First People's Hospital of Shangqiu, Henan Province, China), Xurong Jia (People's Hospital of Rongshui Miao Autonomous County, Guangxi, China), Shaofang Wang (The People's Hospital of Anyang City, Henan Province, China), Jun Liao (People's Hospital of Ruijin City, Jiangxi Province, China), Xiaomin Xie (The First People's Hospital of Yinchuan, Ningxia Hui Autonomous Region, China), Zhihong Zeng (Longquan People's Hospital, Zhejiang Province, China), Yiyuan Yao (First People's Hospital of Xiushui County, Jiangxi Province, China), Xiaoshu Wang (West China-Guang'an Hospital, Sichuan University, Sichuan Province, China), Huiju Zhong (Xiangya Changde Hospital, Hunan Province, China), Jialin Xia (Guixi People's Hospital, Jiangxi Province, China), Xiujun Yan (First People's Hospital of Guannan County, Jiangsu Province, China), Sha Gan (Fengqing County People's Hospital of Yunnan, Yunnan Province, China), Lianzeng Sun (Shandong Energy Zibo Mining Group Co., Ltd Central Hospital, Shandong Province, China), Bo Zhang (The People's Hospital of Shimen County, Hunan Province, China), Dadong Fei (Zaozhuang Municipal Hospital, Shandong Province, China), Lianhuan Zhang (Shaoxing Hospital of Traditional Chinese Medicine, Zhejiang Province, China), Hui Zheng (People's Hospital of Wulian County, Shandong Province, China), Shan Dong (First People's Hospital of Qingzhen, Guizhou, Guizhou Province, China), Bin Liu (Nuclear industry Beijing 401 Hospital, Beijing, China), Xianchen Liu (Chifeng City Center Hospital Ningcheng County, Inner Mongolia Autonomous Region, China), Bi Lu (Aoyang Hospital, Jiangsu Province, China), Ling Gao (Xiangyang Central Hospital, Hubei Province, China), Xuejian Ni (Taiping Street Community Health Service Center, Suzhou Xiangcheng District, Jiangsu Province, China), Xiangning Sun (Central Hospital of Qinghe County, Hebei Province, China), Qian Zhang (The Second Affiliated Hospital of Guizhou Medical University, Guizhou Province, China), Laijun Qiao (Gongyi City People's Hospital, Henan Province, China), Hongjun Fu (Taizhou Enze Medical Center Luqiao Hospital, Zhejiang Province, China), Jingwen Gan (Liyuan Community Health Service Center, Tongzhou District, Beijing, China), Haiying Niu (Luquan People's Hospital, Hebei Province, China), Cuirong Wu (Shenzhen Zhonghai Hospital, Guangdong Province, China), Libo Chen (Shenzhen Nanshan Hospital, Guangdong Province, China), Zhiyuan Yang (Luoyang Central Hospital Affiliated to Zhengzhou University, Henan Province, China), Mingjun Gu (Shanghai Pudong Gongli Hospital, Shanghai, China), Xiaoyan Shi (PKUCare Luzhong Hospital, Shandong Province, China), Rong Li (Chongzhou People's Hospital, Sichuan Province, China), Chunxiao Shi (People's Hospital of Anshun City, Guizhou Province, China), Xu Lian (Hongqi Hospital Affiliated to Mudanjiang Medical University, Heilongjiang Province, China), Fengshi Tian (Tianjin 4th Centre Hospital, Tianjin, China), Yugang Hu (Chaozhou People's Hospital, Guangdong Province, China), Lingling Xu (Shenzhen Hospital of Southern Medical University, Guangdong Province, China), Jianbo Yun (Weitang Health Center, Xinbei District, Changzhou, Jiangsu Province, China), Wangjun Chen (Taicang Shaxi People's Hospital, Jiangsu Province, China), Weiyuan Huang (Jiangyin Harbour Hospital, Jiangsu Province, China), Yun Liang (People's Hospital of Fengdu County, Chongqing, China), Tao Yang (Jiangsu Province Hospital, Jiangsu Province, China), Angui Yang (Zhongxiang People's Hospital, Hubei Province, China), Weiping Tu (Shaoxing Shangyu People's Hospital, Zhejiang Province, China), Yaoming Xue (Nanfang Hospital, Southern Medical University, Guangdong Province, China), Zheng Li (Daxing Xihongmen Hospital, Beijing, China), Pengqiu Li (Sichuan Academy of Medical Sciences·Sichuan Provincial People's Hospital, China), Xiaopang Rao (Qingdao Chengyang People's Hospital, Shandong Province, China), Li Yan (Sun Yat-sen Memorial Hospital,Sun Yat-sen University, Guangdong Province, China), Guiyang Liu (People's Hospital of Renshou County, Meishan, Sichuan Province, China), Junqiang Ba (The First People's Hospital of Zunyi, Guizhou Province, China), Yezi Sun (Zhangjiagang First People's Hospital, Jiangsu Province, China), Zhewu Yin (Fu Ning People's Hospital, Jiangsu Province, China), Wenbing Xu (People's Hospital of Yungang District, Datong City, Shanxi Province, China), Xiongwei Dong (Fangsong Street Community Health Service Center, Songjiang District, Shanghai, China), Wei Wang (Xiang'an Hospital of Xiamen University, Fujian Province, China), Xiaotai Jin (Xinrui hospital, Wuxi New District, Jiangsu Province, China), Binbin Tian (Yuzhou City People's Hospital, Henan Province, China), Zhigang Zhao (Zhengzhou Yihe Hospital, Henan Province, China), Zuhua Gao (Taizhou Hospital of Zhejiang Province, China), Chunlong Mei (Longhua County Hospital, Hebei Province, China), Qiaoyun Qian (Dongtai Hospital of Traditional Chinese Medicine, Jiangsu Province, China), Yunxia Chen (Cangzhou People's Hospital, Hebei Province, China), Peng Su (Tongnan District People's Hospital, Chongqing, China), Jingze Huang (Pingtan Comprehensive Experimental



BMJ Open Diab Res Care: first published as 10.1136/bmjdrc-2020-001596 on 21 October 2020. Downloaded from http://drc.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

Area Hospital, Fujian Province, China), Hongxia Tang (Zhangjiakou First Hospital, Hebei Province, China), Tongfu Bian (Yancheng Bufeng Central Health Center, Jiangsu Province, China), Xuefeng Li (Affiliated Taihe Hospital of Hubei University of Medicine, China), Guiying Wang (The Fifth People's Hospital of Datong, Shanxi Province, China), Ziqi Zhao (Liaoning Lida Diabetes Hospital, Liaoning Province, China), Guoqi Yang (Second People's Hospital of Yandu District, Jiangsu Province, China), Chunfang Qian (Chedun Town Community Health Service Center, Songjiang District, Shanghai, China), Yong Dai (People's Hospital of QingXian, Hebei Province, China), Yaxiong Shi (The Second Affiliated Hospital of Fujian Medical University, Fujian Province, China), Fuzai Yin (First Hospital of Qinhuangdao, Hebei Province, China), Xuemin Li (Handan Seventh hospital, Hebei Province, China), Wei Wang (Xinbang Town Community Health Service Center, Songjiang District, Shanghai, China), Yane Liu (Shanxian Central Hospital, Shandong Province, China), Xiaohua Li (Shanghai Seventh People's Hospital, Shanghai, China), Yanling Feng (The First People's Hospital of Jinzhong, Shanxi Province, China) for their collection of data and taking care of patients.

**Contributors** WW, XX, XD and GN conceived and design the study. YZ, JS, YiP, ZhZ, ZW and SJ analyzed the data. JS, YiP, QZ, LY, DZ, HJ, YD, BS, PG, HS, QW, YoP, JL, LH, TK, LC, FX and QD contributed to data collection. KQ organized the expert panel for manual grading. KL and XX directed the fundus image diagnosis by specialist graders. XD, ZiZ, KQ, ZW and SJ were involved in the development, optimization and verification of the DL Algorithms. YZ, JS, YiP, ZhZ, QZ, ZW, XD and DT drafted and revised the manuscript. WW, XD and YZ approved the final version of the manuscript. WW and XD are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the decision to submit for publication.

**Funding** This research was supported by grants from National Key R&D Program of China (2016YFC0901200, 2018YFC1314800); Chinese Academy of Engineering (2019-XZ-42); the National Natural Science Foundation of China (81670797); the Program for Shanghai Outstanding Medical Academic Leader (2019LJ07); the Youth Program of Shanghai Municipal Health and Family Planning Commission (20174Y0081) and the Yang Fan Project of Shanghai Science and Technology Committee (19YF1442700).

**Map disclaimer** The depiction of boundaries on the map(s) in this article does not imply the expression of any opinion whatsoever on the part of BMJ (or any member of its group) concerning the legal status of any country, territory, jurisdiction or area or of its authorities. The map(s) are provided without any warranty of any kind, either express or implied.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** The study protocol was approved by the ethics committees at Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, the leading MMC center (KY2018-103-3) and at the other participating centers subsequently if necessary. This study complied with the provisions of the Declaration of Helsinki. All study participants provided written informed consent.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available on reasonable request.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

**ORCID iD**

Weiqing Wang http://orcid.org/0000-0001-6027-3084

## REFERENCES

1 Cho NH, Shaw JE, Karuranga S, *et al*. IDF diabetes atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res Clin Pract* 2018;138:271–81.
2 Xu Y, Wang L, He J, *et al*. Prevalence and control of diabetes in Chinese adults. *JAMA* 2013;310:948–59.
3 Wang L, Gao P, Zhang M, *et al*. Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013. *JAMA* 2017;317:2515–23.
4 Ma RCW. Epidemiology of diabetes and diabetic complications in China. *Diabetologia* 2018;61:1249–60.
5 Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* 2018;14:88–98.
6 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.
7 Yang Q-H, Zhang Y, Zhang X-M, *et al*. Prevalence of diabetic retinopathy, proliferative diabetic retinopathy and non-proliferative diabetic retinopathy in Asian T2DM patients: a systematic review and meta-analysis. *Int J Ophthalmol* 2019;12:302–11.
8 Zhang G, Chen H, Chen W, *et al*. Prevalence and risk factors for diabetic retinopathy in China: a multi-hospital-based cross-sectional study. *Br J Ophthalmol* 2017;101:1591–5.
9 Vujosevic S, Aldington SJ, Silva P, *et al*. Screening for diabetic retinopathy: new perspectives and challenges. *Lancet Diabetes Endocrinol* 2020;8:337–47.
10 Abramoff MD, Niemeijer M, Russell SR. Automated detection of diabetic retinopathy: barriers to translation into clinical practice. *Expert Rev Med Devices* 2010;7:287–96.
11 Andonegui J, Zurutuza A, de Arcelus MP, *et al*. Diabetic retinopathy screening with non-mydriatic retinography by general practitioners: 2-year results. *Prim Care Diabetes* 2012;6:201–5.
12 Sapkota R, Chen Z, Zheng D, *et al*. The profile of sight-threatening diabetic retinopathy in patients attending a specialist eye clinic in Hangzhou, China. *BMJ Open Ophthalmol* 2019;4:e000236.
13 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.
14 Carin L, Pencina MJ. On deep learning for medical image analysis. *JAMA* 2018;320:1192–3.
15 Esteva A, Robicquet A, Ramsundar B, *et al*. A guide to deep learning in healthcare. *Nat Med* 2019;25:24–9.
16 Ting DSW, Cheung CY-L, 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.
17 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.
18 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;57:5200–6.
19 Li Z, Keel S, Liu C, *et al*. An automated grading system for detection of Vision-Threatening Referable diabetic retinopathy on the basis of color fundus Photographs. *Diabetes Care* 2018;41:2509–16.
20 Gulshan V, Rajan RP, Widner K, *et al*. Performance of a Deep-Learning algorithm vs manual grading for detecting diabetic retinopathy in India. *JAMA Ophthalmol* 2019;137:987–93.
21 Abràmoff 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.
22 Natarajan S, Jain A, Krishnan R, *et al*. Diagnostic accuracy of community-based diabetic retinopathy screening with an Offline artificial intelligence system on a smartphone. *JAMA Ophthalmol* 2019;137:1182–8.
23 Raumviboonsuk P, Krause J, Chotcomwongse P, *et al*. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *NPJ Digit Med* 2019;2:25.
24 van der Heijden AA, Abramoff MD, Verbraak F, *et al*. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn diabetes care system. *Acta Ophthalmol* 2018;96:63–8.
25 Zhang Y, Wang W, Ning G. Metabolic management center: an innovation project for the management of metabolic diseases and complications in China. *J Diabetes* 2019;11:11–13.
26 Gabir MM, Hanson RL, Dabelea D, *et al*. The 1997 American diabetes association and 1999 World Health organization criteria for hyperglycemia in the diagnosis and prediction of diabetes. *Diabetes Care* 2000;23:1108–12.
27 Szegedy C, Ioffe S, Vanhoucke V. *Inception-v4, inception-resnet and the impact of residual connections on learning*. In ICLR Workshop, 2016.
28 Wilkinson CP, Ferris FL, Klein RE, *et al*. Proposed International clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003;110:1677–82.
29 AAoORV P. *Preferred practice Pattern® guidelines. diabetic retinopathy*. San Francisco, CA: American Academy of Ophthalmology, 2017.
30 Bresnick GH, Mukamel DB, Dickinson JC, *et al*. A screening approach to the surveillance of patients with diabetes for the



BMJ Open Diab Res Care: first published as 10.1136/bmjdrc-2020-001596 on 21 October 2020. Downloaded from http://drc.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

presence of vision-threatening retinopathy. *Ophthalmology* 2000;107:19–24.
31 He J, Cao T, Xu F, *et al*. Artificial intelligence-based screening for diabetic retinopathy at community hospital. *Eye* 2020;34:572–6.
32 Gupta V, Bansal R, Gupta A, *et al*. Sensitivity and specificity of nonmydriatic digital imaging in screening diabetic retinopathy in Indian eyes. *Indian J Ophthalmol* 2014;62:851–6.
33 Scanlon PH, Foy C, Malhotra R, *et al*. The influence of age, duration of diabetes, cataract, and pupil size on image quality in digital photographic retinal screening. *Diabetes Care* 2005;28:2448–53.
34 Liu L, Wu X, Liu L, *et al*. Prevalence of diabetic retinopathy in mainland China: a meta-analysis. *PLoS One* 2012;7:e45264.
35 Kanagasingam Y, Xiao D, Vignarajan J, *et al*. Evaluation of artificial Intelligence–Based grading of diabetic retinopathy in primary care. *JAMA Netw Open* 2018;1:e182665.



BMJ Open Diab Res Care: first published as 10.1136/bmjdrc-2020-001596 on 21 October 2020. Downloaded from http://drc.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

# Artificial intelligence-enabled screening for diabetic retinopathy: a real-world, multi-center, and prospective study

## Supplemental Material

### Supplementary Methods

The eligible participants were those with a diagnosis of diabetes according to the World Health Organization criteria.[1] Detailed inclusion and exclusion criteria are as follows:

**Inclusion criteria:**

Male and female adults diagnosed with diabetes aged 18 years and older who undergo fundus photography at a National Metabolic Management Center (MMC) in China. Diabetes was defined as 1) self-report of having been previous diagnosed of diabetes by a clinician; 2) newly diagnosed diabetes according to the 1999 World Health Organization (WHO) criteria (1); diabetes types were classified by a qualified physician on each site. While the classification information was not available, type 1 diabetes (T1DM) was defined as glutamic acid decarboxylase antibody (GADA) positive and serum C-peptide concentrations of less than 0.8 ng/mL. Besides T1DM, type 2 diabetes (T2DM), and gestational diabetes mellitus (GDM), rare cases including single gene mutation diabetes, maturity onset diabetes of the young (MODY), secondary diabetes caused by pancreatic damage, Cushing's syndrome, thyroid dysfunction, or acromegaly etc were also included.

**Exclusion criteria:**

(1) Those who have a history of drug abuse.

(2) Sexually transmitted diseases such as AIDS and syphilis, and infectious diseases such as viral hepatitis and tuberculosis which are in the active phase.

(3) Any situation that influences the inclusion by the researcher's judgment.

**Data collection:**

Data collection was conducted by trained staff according to a standard protocol. Social demographic characteristics, medical history and lifestyle factors were recorded. Education attainment was categorized as lower than high school, and high school or higher. Height and body weight were measured by a height-weight scale with participants in light clothes without shoes, and body mass index (BMI) was calculated as the weight in kilograms divided by height in meters squared. Blood pressure and heart rate was measured with an electronic blood pressure monitor after at least 5-minute rest in the seated position. Waist circumference was measured on standing participants midway between the lower edge of the costal arch and the upper edge of the iliac crest.

The participants were required to undergo a standard steamed bread meal test after an overnight fasting, and blood samples were collected at 0 and 2 hours during the test. Fasting blood glucose (FBG), postprandial blood glucose (PBG), fasting serum C peptide, postprandial serum C peptide, glycated hemoglobin (HbA1c), serum creatinine, uric acid, and lipid profiles were tested in local MMCs.

**DL model training and design details of diabetic retinopathy (DR) and diabetic macular edema (DME) classification**

The DR and DME models are an ensemble of 6 neural networks. All the 6 neural networks use the state-of-the-art Inception ResNet v2 architecture; however, several design differences among them are critical to the effective performance of the model ensemble. Firstly, they differ in the data used to train them and with respect to fine-tuning iterations and initializations, such that the model ensemble performs well on multiple data domains. Secondly, a multi-stage training scheme is adopted; the later networks in the ensemble are trained based on handpicked confusing cases (images that fall on the boundary between two classes) and cases that the previous networks commonly misclassify. This enables the model ensemble to perform more reliably on difficult cases. Thirdly, the first network in our ensemble contains multiple classification branches, for both DR grading and the classifications of individual lesions that contribute to the DR grade. The individual lesion classifications serve as an additional supervision, so that the network learns to pick up the features of the lesions that contribute to DR, hence improving the DR performance. The first network in the ensemble is trained with a focal loss, while the remaining networks are trained with a 5-way sigmoid cross entropy loss to classify the input image into 5-stage DR grades.

**DL algorithm testing results on the APTOS 2019 Blindness Detection dataset:**

The APTOS 2019 Blindness Detection dataset (https://www.kaggle.com/c/aptos2019-blindness-detection/overview) is a large set of

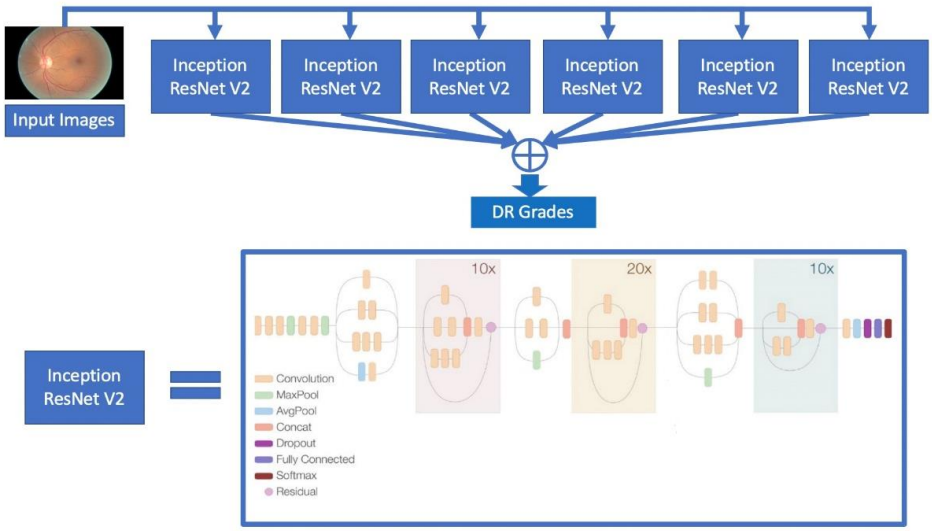
retinal images acquired using fundus photography under a variety of imaging conditions, collected by the Aravind Eye Hospital in India. It is a real-world dataset that includes noise in both the images and labels. The images may contain artifacts and be out of focus, underexposed, or overexposed. The images were gathered from multiple clinics using a variety of cameras over an extended period of time, which introduced further variation. A clinician has rated each image for the severity of DR on a scale of 0 to 4: 0 - No DR, 1 – Mild DR, 2 – Moderate DR, 3 – Severe DR, 4 - Proliferative DR.

For DR grading, the concordance between the DL system and reference standard was 70.4% for the 5-stage DR grading. The corresponding quadratic weighted kappa was 0.86. The DL system had a 98.9% sensitivity and 85.0% specificity for detecting referable DR. The positive and negative predictive values were 96.8% and 98.5%, respectively. The Youden index was 95.2%. The algorithms were trained on datasets collected from different populations under different senarios, showing good generalization characteristics.

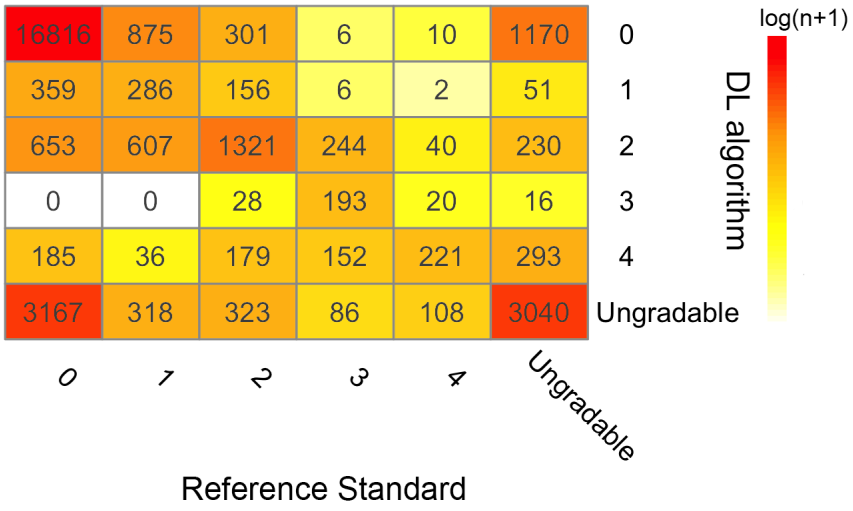
## REFERENCE

1 Gabir MM, Hanson RL, Dabelea D, et al. The 1997 American Diabetes Association and 1999 World Health Organization criteria for hyperglycemia in the diagnosis and prediction of diabetes. *Diabetes Care* 2000;23:1108-12.

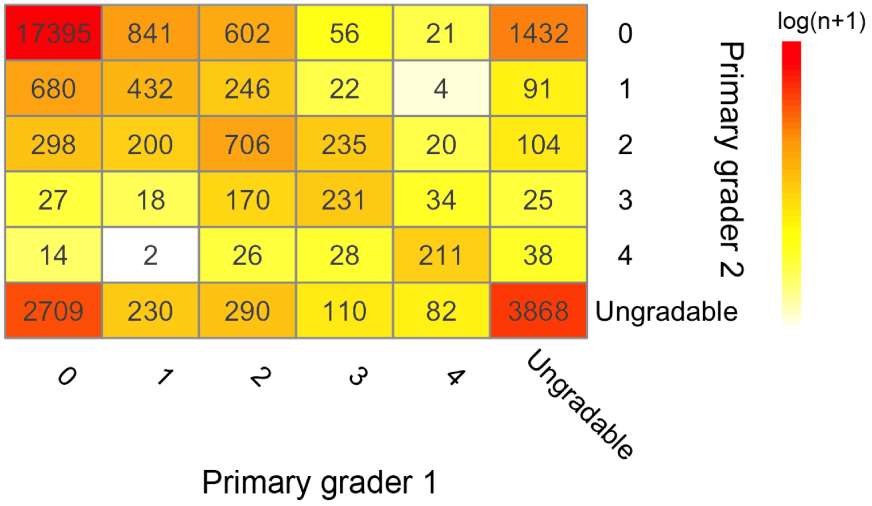
**Supplementary Figure 1. The overall architecture of the diabetic retinopathy DL model.** The model is an ensemble of 6 neural networks. The input color fundus images are normalized and resized. The results from the 6 neural networks are merged to produce the DR grades.



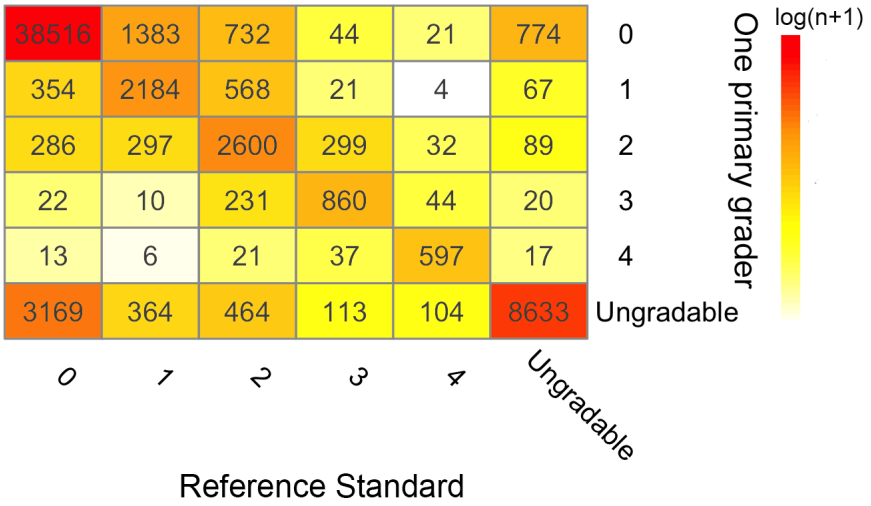
**Supplementary Figure 2. Confusion matrices for 5-stage DR grading between the DL system and Reference Standard.**



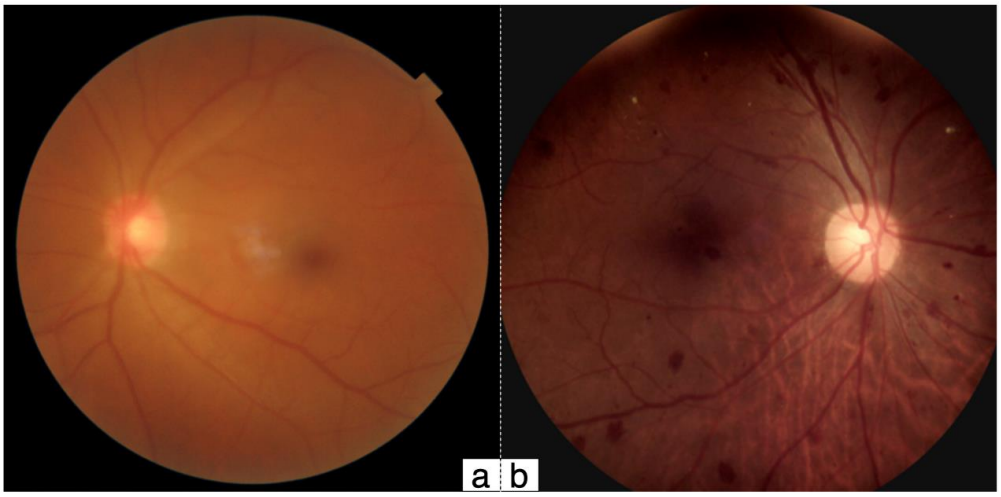
**Supplementary Figure 3. Confusion matrix for 5-stage DR grading between two primary graders.**



**Supplementary Figure 4. Confusion matrix for 5-stage DR grading between a primary grader and the reference standard.**

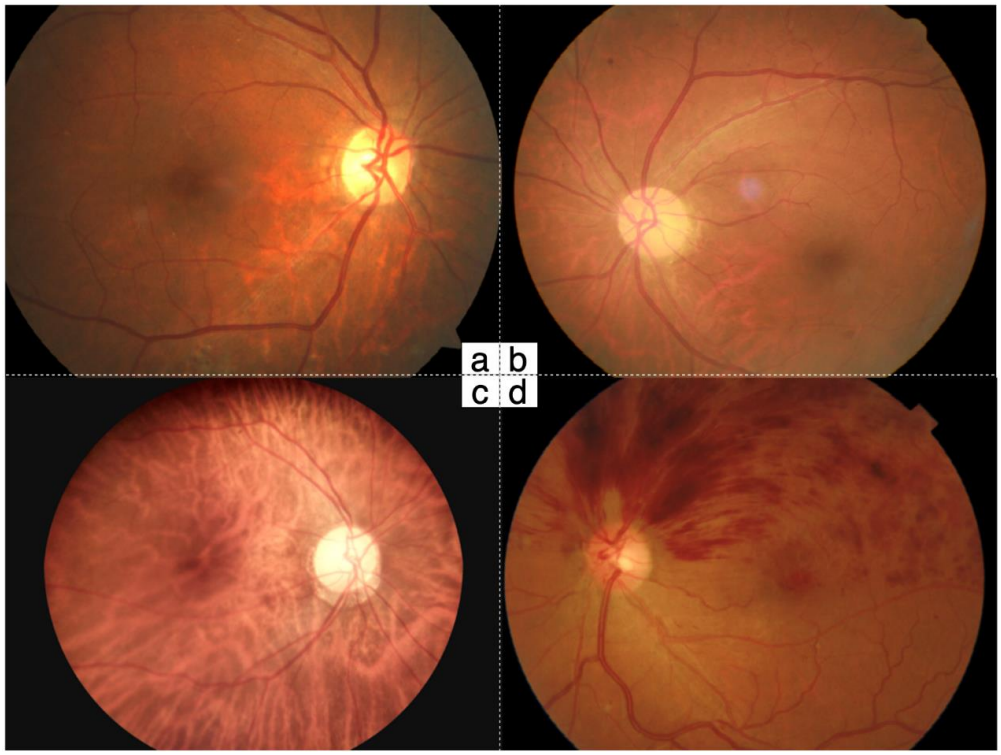


**Supplementary Figure 5. Typical examples of false negative and false positive results by the DL system's Quality Control (QC) module.**



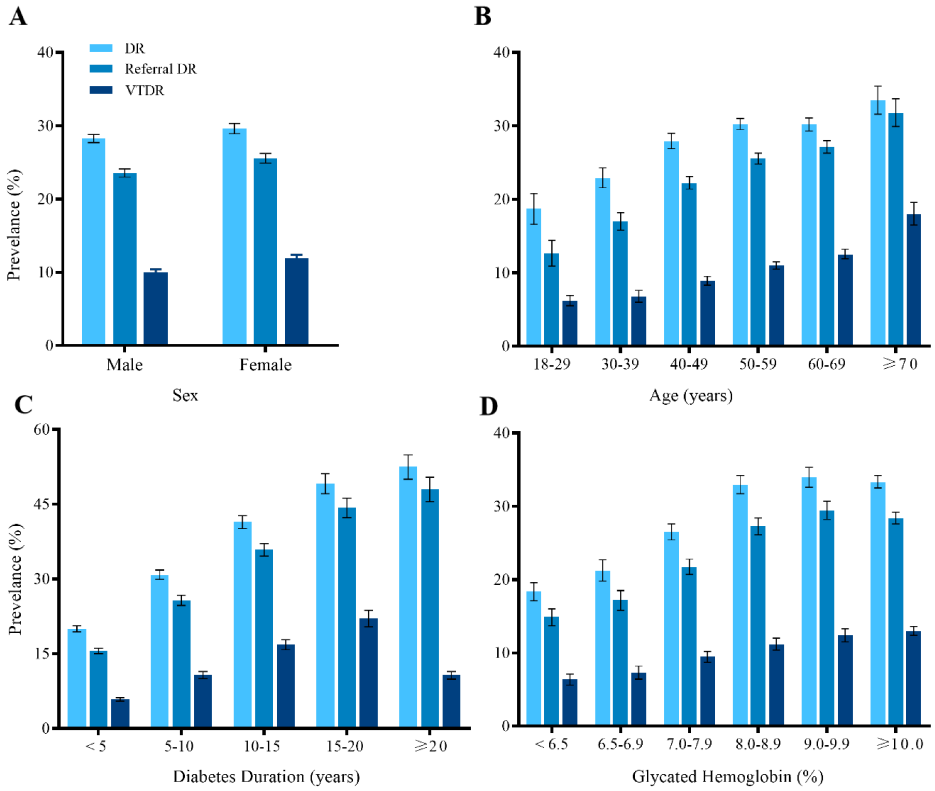

a: False negative, the image is out of focus and judged ungradable by the QC module but gradable by graders; b: False positive, the image is judged gradable by the QC module but too dark and ungradable by the graders.

**Supplementary Figure 6. Typical examples of false negative and false positive results by the DL system's grading module.**




a) False negative, unrecognized laser scars; b) False negative, unrecognized intraretinal microvascular abnormality; c) False positive, combined pathologic myopia; d) False positive, combined hypertensive retinopathy

**Supplementary Figure 7. Prevalence of diabetic retinopathy (DR), referable DR and vision-threatening DR (VTDR) among different risk factor stratifications.**




Error bars indicate 95% confidence intervals

Table S1. Vendors, types, and models of fundus cameras and their corresponding numbers of deployments in MMCs

| **Camera vendor*** | **Camera types** | **Camera model*** | **No. of centers deployed** |
|---|---|---|---|
| Topcon | Desktop | 3D OCT-1 | 1 |
| | | TRC-50DX | 1 |
| | | TRC-NW200 | 2 |
| | | TRC-NW300 | 5 |
| | | TRC-NW400 | 48 |
| | | TRC-NW8 | 2 |
| MiiS | Hand-held with fixed base | DSC-200 | 46 |
| Canon | Desktop | CR-2 1-0.4A | 1 |
| | | CR-2 AF | 9 |
| | | CR-2 PLUS | 3 |
| | | CR-2 PLUS AF | 17 |
| Zeiss | Desktop | VISUCAM200 | 6 |
| Kang Huaruiming | Desktop | APS-AER | 3 |
| Optomed | Hand-held without fixed base | M5 | 2 |
| | | FI-90100 | 1 |
| KOWA | Desktop | nonmyd7 | 1 |
| | | rtx1-e | 1 |

| | | | |
|---|---|---|---|
| NIDEK | Desktop | AFC-210 | 1 |
| SYSEYE | Desktop | RetiCam3100 | 1 |
| Crystalvue | Desktop | Fundusvue | 1 |
| DRS | Desktop | DRS | 1 |
| DRS + Canon | Desktop | DRS/CR-2 AF | 1 |
| Canon + Optomed | Desktop + hand-held without fixed base | CR-2 AF/M5 | 1 |

* Two centers deployed cameras from two vendors. Therefore, total number of vendors were 11 and corresponding models were 21.

Table S2. The demographic characteristics of patients for dataset 1 used for developing DL algorithm

| | **Patient** | **No.** | **Percentage** |
|---|---|---|---|
| Age, years | Total | 37,231 | |
| | Mean ± SD | 54.85±10.96 | |
| | <18 | 36 | 0.10% |
| | 18-29 | 611 | 1.64% |
| | 30-39 | 2,641 | 7.09% |
| | 40-49 | 7,573 | 20.34% |
| | 50-59 | 13,156 | 35.34% |
| | 60-69 | 10,655 | 28.62% |
| | ≥70 | 2,552 | 6.85% |
| | Unknown | 7 | 0.02% |
| Ethnicity | | | |
| | Latin American | 18,701 | 50.23% |
| | Caucasian | 2,843 | 7.64% |
| | Asian | 1,863 | 5.00% |
| | African Descent | 1,615 | 4.34% |
| | Other or Unknown | 12,209 | 32.79% |
| Gender | | | |
| | Male | 14,867 | 39.93% |
| | Female | 19,413 | 52.14% |
| | Other or Unknown | 2,951 | 7.93% |
| Diabetic duration, years | | | |
| | <5 | 11,857 | 31.85% |
| | 5-10 | 11,983 | 32.19% |
| | 10-15 | 5,852 | 15.72% |
| | 15-20 | 3,144 | 8.44% |
| | >=20 | 2,668 | 7.17% |
| | BorderLine | 470 | 1.26% |
| | Gestational | 34 | 0.09% |
| | Not Diabetic | 528 | 1.42% |
| | Unknown | 695 | 1.87% |
| Insulin Dependent | | | |
| | FALSE | 26,328 | 70.72% |
| | TRUE | 10,903 | 29.28% |
| Concomitant Disease | | | |
| | Cataract | 1,490 | 4.00% |
| | Glaucoma | 1,237 | 3.32% |
| | Maculopathy | 1,022 | 2.75% |
| | Occlusion | 248 | 0.67% |

Table S3. The characteristics of eyes for dataset 1 used for developing DL algorithm

| Eye | No. | Percentage |
|---|---|---|
| Total | 57,289 | |
| DR Grade | | |
| No DR | 20,035 | 34.97% |
| Mild-NPDR | 10,634 | 18.56% |
| Moderate-NPDR | 21,900 | 38.23% |
| Severe-NPDR | 3,308 | 5.77% |
| PDR | 1,412 | 2.46% |
| Positive Lesion Label | | |
| Pure Microangioma | 10,642 | 18.58% |
| Hemorrhage and Microangioma | 25,077 | 43.77% |
| Cotton Wool | 12,090 | 21.10% |
| Hard Exudates | 11,521 | 20.11% |
| DME | 4,934 | 8.61% |

Table S4. The DR grades characteristics of patients for dataset 2 used for developing DL algorithm

| | **Image** | **No.** | **Percentage** |
|---|---|---|---|
| DR | Total | 1,184 | |
| | No DR | 105 | 8.87% |
| | Mild-NPDR | 312 | 26.35% |
| | Moderate-NPDR | 393 | 33.19% |
| | Severe-NPDR | 274 | 23.14% |
| | PDR | 100 | 8.45% |

Table S5. 5-stage diabetic retinopathy severity classification

| **Disease Severity Level** | **Findings observable from single nonmydriatic fundus image** | **ICDR findings from mydriatic fundus images** |
|---|---|---|
| No apparent retinopathy | No abnormalities | No abnormalities |
| Mild NPDR | Microaneurysms only | Microaneurysms only |
| Moderate NPDR | More than just microaneurysms but less than severe NPDR | More than just microaneurysms but less than severe NPDR |
| Severe NPDR | Any of the following and no signs of proliferative retinopathy:<br>• More than 20 intraretinal hemorrhages<br>• Definite venous beading<br>• Prominent IRMA | Any of the following and no signs of proliferative retinopathy:<br>• More than 20 intraretinal hemorrhages in each of **four** quadrants<br>• Definite venous beading in **two** or more quadrants<br>• Prominent IRMA in **one** or more quadrants |
| PDR | Any of the following:<br>• Neovascularization<br>• Vitreous/preretinal hemorrhage<br>• Laser surgery scars | Any of the following:<br>• Neovascularization<br>• Vitreous/preretinal hemorrhage |

DR, diabetic retinopathy; NPDR, non-proliferative diabetic retinopathy; PDR, proliferative diabetic retinopathy; IRMA, intraretinal microvascular abnormality.

Table S6. Diabetic macular edema severity classification

| **Severity Level** | **Findings Observable upon single nonmydriatic fundus image** |
|---|---|
| No apparent DME | No apparent hard exudates (HEs) |
| Clinically insignificant diabetic macular edema | HEs outside one disc diameter of the foveola |
| Clinically significant macular edema (CSME) | HEs within one disc diameter of the foveola |

DME, diabetic macular edema; HEs, hard exudates; CSME, clinically significant macular edema.

Table S7. Image quality assessment from the DL system compared to those of the reference standard

| | | **Reference standard** | |
|---|---|---|---|
| | **Image quality** | Gradable | Ungradable |
| **DL algorithm quality control module** | Gradable | 22,696 | 1,760 |
| | Ungradable | 4,002 | 3,040 |

Table S8. Concordance and quadratic weighted kappa scores for the 5-stage diabetic retinopathy grading

| | **Concordance** | **quadratic weighted kappa scores (95%CI)** |
|---|---|---|
| DL algorithm vs. reference standard | 83.0% | 0.72 (0.72-0.72) |
| Intergrader (between two primary graders) | 84.3% | 0.74 (0.74-0.74) |
| One primary grader vs. reference standard | 91.0% | 0.87 (0.87-0.87) |
| DL algorithm vs primary grader 1 | 82.8% | 0.66 (0.66-0.66) |
| DL algorithm vs primary grader 2 | 81.8% | 0.67 (0.67-0.67) |
| DL algorithm vs one primary grader (combining two primary graders) | 82.3% | 0.67 (0.67-0.67) |

Table S9. Quality analysis of the fundus photographs by individual eye

| **Images** | **Total** | **High quality** | **Medium quality** | **low quality (Ungradable)** |
|---|---|---|---|---|
| Number | 94,199 | 22,404 | 49,566 | 22,229 |
| % | 100% | 23.8% | 52.6% | 23.6% |

Table S10. Quality analysis of fundus photographs by participant

| Participants | Total | Both eyes have images | | | Single eye has image | |
|---|---|---|---|---|---|---|
| | | Both eyes gradable | One eye gradable | Both eyes Ungradable | One eye gradable | One eye Ungradable |
| Number | 47,269 | 31,305 | 9,156 | 6,469 | 204 | 135 |
| % | 100% | 66.2% | 19.4% | 13.7% | 0.4% | 0.3% |

Table S11. Main causes for ungradable images assessed in 592 randomly selected low quality images

| **Causes for ungradable** | **No. of images** | **Percentage** | **Total Images** |
|---|---|---|---|
| Overexposure, underexposure, false focus (mainly including small pupil) | 285 | 48.14% | 592 |
| Occlusion of optical path (include cataract) | 217 | 36.66% | |
| Ring artifact (because Outside light enters the film from the interspace of eye and the objective lens) | 62 | 10.47% | |
| Halo | 28 | 4.73% | |

Table S12. Comparisons of prevalence of DR, referable DR or VTDR between adjacent groups by HbA1c levels

| **Category** | **DR** | **Referable DR** | **VTDR (including CSME)** |
|---|---|---|---|
| (6.5%-6.9%) vs. (< 6.5%) | 0.003 | 0.010 | 0.114 |
| (7.0%-7.9%) vs. (6.5%-6.9%) | < 0.001 | <0.001 | < 0.001 |
| (8.0%-8.9%) vs. (7.0%-7.9%) | < 0.001 | <0.001 | 0.002 |
| (9.0%-9.9%) vs. (8.0%-8.9%) | 0.269 | 0.013 | 0.052 |
| (≥10.0%) vs. (9.0%-9.9%) | 0.446 | 0.174 | 0.295 |

P values between subgroups of HbA1c calculated by $\chi^2$-test.

Table S13. The association between risk factors and DR or referable DR by logistic regression

| **Variables** | **DR** | | | **Referable DR** | | |
|---|---|---|---|---|---|---|
| | OR | 95% CI | P | OR | 95% CI | P |
| Age, years | 0.993 | 0.991 to 0.996 | < 0.001 | 1.001 | 0.999 to 1.004 | 0.282 |
| Sex | 1.012 | 0.961 to 1.065 | 0.660 | 1.024 | 0.970 to 1.081 | 0.391 |
| Duration of diabetes, years | 1.094 | 1.089 to 1.099 | < 0.001 | 1.094 | 1.089 to 1.099 | < 0.001 |
| HbA1c, % | 1.111 | 1.098 to 1.123 | < 0.001 | 1.123 | 1.110 to 1.136 | < 0.001 |

The ORs were adjusted for age, sex, duration of diabetes and HbA1c (except the current variable)

Table S14. 5-stage diabetic retinopathy and corresponding diabetic macular edema classification by the DL system for 40,665 gradable participants

| | | **CSME** | | |
|---|---|---|---|---|
| | | **0** | **1** | **Ungradable** |
| **5-stage diabetic retinopathy grading** | **0** | 28,939 | 0 | |
| | **1** | 1,813 | 0 | |
| | **2** | 5,519 | 1,151 | |
| | **3** | 120 | 335 | |
| | **4** | 1,849 | 939 | |
| | **Ungradable** | | | 6,604 |

CSME, clinically significant macular edema.

Table S15. Percentage of ungradable images with different cameras

| **Types of cameras** | **No. of ungradable images** | **Percentage** |
|---|---|---|
| Topcon | 8,113 | 25.2% |
| MiiS | 3,277 | 18.7% |
| Canon | 8,400 | 24.8% |
| Other types | 2,439 | 23.0% |

Table S16. Prevalence of DR, referable DR, and VTDR based on the different types of cameras

| **Prevalence % (95% CI)** | | | |
|---|---|---|---|
| **Types of cameras** | **DR** | **Referable DR** | **VTDR (including CSME)** |
| Topcon | 27.0 (26.3-27.8) | 21.4 (20.7-22.1) | 7.8 (7.3-8.2) |
| MiiS | 29.6 (28.6-30.6) | 27.3 (26.3-28.3) | 16.1 (15.3-16.9) |
| Canon | 31.2 (30.4-31.9) | 26.3 (25.6-27.1) | 11.2 (10.7-11.7) |
| Other types | 25.7 (24.5-27.0) | 22.0 (20.8-23.2) | 9.1 (8.3-9.9) |