



REVIEW

Artificial Intelligence for Diabetic Retinopathy Screening Using Color Retinal Photographs: From Development to Deployment

Andrzej Grzybowski · Panisa Singhanetr · Onnisa Nanegrungsunk · Paisan Ruamviboonsuk

Received: January 9, 2023 / Accepted: February 14, 2023 / Published online: March 2, 2023
© The Author(s) 2023

ABSTRACT

Diabetic retinopathy (DR), a leading cause of preventable blindness, is expected to remain a growing health burden worldwide. Screening to detect early sight-threatening lesions of DR can reduce the burden of vision loss; nevertheless, the process requires intensive manual labor and extensive resources to accommodate the increasing number of patients with diabetes. Artificial intelligence (AI) has been shown to be an effective tool which can potentially lower the burden of screening DR and vision loss. In

this article, we review the use of AI for DR screening on color retinal photographs in different phases of application, ranging from development to deployment. Early studies of machine learning (ML)-based algorithms using feature extraction to detect DR achieved a high sensitivity but relatively lower specificity. Robust sensitivity and specificity were achieved with the application of deep learning (DL), although ML is still used in some tasks. Public datasets were utilized in retrospective validations of the developmental phases in most algorithms, which require a large number of photographs. Large prospective clinical validation studies led to the approval of DL for autonomous screening of DR although the semi-autonomous approach may be preferable in some real-world settings. There have been few reports on real-world implementations of DL for DR screening. It is possible that AI may improve some real-world indicators for eye care in DR, such as increased screening uptake and referral adherence, but this has not been proven. The challenges in deployment may include workflow issues, such as mydriasis to lower ungradable cases; technical issues, such as integration into electronic health record systems and integration into existing camera systems; ethical issues, such as data privacy and security; acceptance of personnel and patients; and health-economic issues, such as the need to conduct health economic evaluations of using AI in the context of the country. The

A. Grzybowski
Department of Ophthalmology, University of Warmia and Mazury, Olsztyn, Poland

A. Grzybowski (✉)
Institute of Research in Ophthalmology,
Foundation for Ophthalmology Development,
Mickiewicza 24/3B, 60-836 Poznań, Poland
e-mail: ae.grzybowski@gmail.com

P. Singhanetr
Mettapracharak Eye Institute, Mettapracharak (Wat Rai Khing) Hospital, Nakhon Pathom, Thailand

O. Nanegrungsunk
Retina Division, Department of Ophthalmology,
Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

P. Ruamviboonsuk
Department of Ophthalmology, College of Medicine, Rajavithi Hospital–Rangsit University, Bangkok, Thailand

deployment of AI for DR screening should follow the governance model for AI in healthcare which outlines four main components: fairness, transparency, trustworthiness, and accountability.

Keywords: Artificial intelligence; Deep learning; Deployment; Diabetic retinopathy; Diabetic retinopathy screening; Machine learning; Retinal photographs

Key Summary Points

The robust performance of some retrospective validation studies of deep learning (DL) for diabetic retinopathy (DR) screening on color retinal photographs has generated enthusiasm for using artificial intelligence (AI) in healthcare, not only in ophthalmology.

The steps of applying AI in healthcare, including DR screening, can be followed from *in silico* evaluation, offline retrospective evaluation, small-scale and large-scale prospective online evaluation, and post-market surveillance, comparable to the preclinical and subsequent phases of studies on new drugs.

Many factors, other than its diagnostic performance, should be in consideration for deployment of AI for DR screening. These challenges include workflow issues, such as mydriasis to lower ungradable cases; technical issues, such as integration in electronic health record system, ethical issues, such as data privacy and security, and acceptance of personnel and patients.

The deployment of AI for DR screening should follow the governance model for AI in healthcare which outlines four main components: fairness, transparency, trustworthiness, and accountability.

INTRODUCTION

Diabetes mellitus (DM) has been one of the major global public health problems for many decades, and the prevalence of DM is expected to increase continuously in all regions of the world in the coming decades [1]. The International Diabetes Federation (IDF) has projected that there will approximately 700 million patients with diabetes worldwide by 2045 [2]. Diabetic retinopathy (DR), one of the most common microvascular complications of DM, is one of the leading causes of preventable blindness, particularly in the working-age population. As such DR is expected to remain a growing burden on healthcare systems worldwide [3].

Not only is the global prevalence of DM and DR expected to increase, but concomitantly the global prevalence of vision-threatening DR (VTDR), which includes diabetic macular edema (DME), severe non-proliferative DR (NPDR), and proliferative DR (PDR), is also projected to increase. The global number of patients with VTDR is estimated to increase by 57.0% from approximately 28.5 million people in 2020 to approximately 44.8 million people in 2045 [4]. Screening to detect early sight-threatening lesions of DR for timely monitoring and treatment is an important strategy to reduce the burden of vision loss and blindness due to DR [5, 6].

The burden of DR disproportionately affects the poorer countries worldwide [1, 3], with low-to middle-income countries (LMICs) tending to have a higher prevalence of DM and DR due to lower healthcare resources compared to high-income countries. National screening programs for DR, in which all patients with DM in a country are targets, have shown to be an effective public health intervention to sufficiently reduce vision loss from DR [7–9]. Nevertheless, most of the screening programs worldwide are opportunistic, rather than national or systematic. Over the years, these screening programs have required increasingly intensive manual labor and extensive resources to accommodate the increasing number of patients with DM. Attempts have been made to lower the

screening burden while at the same time increase the screening rates. Examples of such attempts are the teleretinal imaging programs, which were proven to be cost-effective [9, 10].

Artificial intelligence (AI) algorithms have recently been shown to be effective tools for autonomous or assistive screening of DR. These algorithms can potentially lower the burden imposed on human personnel and improve patient access to care. Many retrospective studies have found that AI algorithms have a robust performance in terms of diagnostic accuracy of AI for detecting referable DR (rDR) or VTDR. However, in comparison, many fewer studies have been published on the prospective validation of AI algorithms for DR screening. While the results of these studies still supported the high performances of AI for DR screening, their performances were generally lower, but still acceptable, compared with the results from retrospective validation studies. Real-world studies on AI for DR screening, however, are rarely found in the literature.

Here, we review the use of AI for DR screening in these retrospective, prospective, and real-world studies with a focus on the analysis of AI on color retinal photographs, which are readily available, practical, and a ubiquitous tool for screening. We also reviewed the important issues of governance and challenges with AI deployment for those who are considering the implementation of AI for DR screening in clinical practice.

METHODS

We reviewed AI-based DR screening by searching the Google Scholar, PubMed, Medline, Scopus, and Embase databases for studies published in English up to 31 October 2022, using these keywords: “diabetes,” “DR screening,” “fundus photographs,” “artificial intelligence,” “automated DR system,” “machine learning,” and “deep learning”.

This article is based on previously conducted studies and does not contain any new studies with human participants or animals performed by any of the authors.

RESULTS

Brief Review of AI

Many methods can be used to develop algorithms for automated detection of DR, whether it is machine learning (ML) or deep learning (DL). ML is a method in which humans teach machines to detect specified patterns of features (feature extraction) in data for a specific task. This strategy allows the machine to detect only those features it has been taught. ML-based algorithms can therefore predict the presence of levels of DR that it has been taught. In comparison, DL is a method of ML in which an algorithm is engineered to detect the most predictive features directly from large sets of data labeled with certain information [11]. DL functions like a human brain and is based on artificial neural networks which learn features directly from data, and its performance increases with increasing data [12]. Since DL-based algorithms have no pre-specified patterns to learn from, there is no justification as to how the algorithm reaches certain outcomes, which is referred to as the “black box”. Although the process is unexplainable, and therefore a liability, at the same time, this gives AI the ability to offer new insights into certain diseases and to identify features that humans previously have not been able to recognize [13].

Development of AI Algorithms

The first phase in the development of AI algorithms for DR screening from retinal photographs is to establish a sufficiently large dataset of photographs of patients with DM, with each photograph marked for the severity of DR and the presence of DME. Then, approximately 80% of the data (photographs with labels) may be used to develop the DL algorithms, with 20% kept for evaluation of the algorithms. This phase is called the “in silico evaluation” or “internal validation,” and is comparable to the preclinical development program for new drugs. The algorithms are then validated with new datasets of retinal photographs of patients with DM. This phase is

Table 1 Comparison of publically available datasets of diabetic retinopathy

| Dataset | Total number of images (DR:normal) | Location | DR grading system | Proportion of mydriasis (%) | Camera | FOV (°) | Resolution (pixels) |
|------------|------------------------------------|----------|-------------------|-----------------------------|----------------|---------|--|
| Messidor | 1200 (660:540) | France | N/A | 66.7% | Topcon TRC-NW6 | 45 | 1440 × 960, 2240 × 1488, 2304 × 1536 |
| Messidor-2 | 1,748 (528:1217) | France | N/A | N/A | Topcon TRC-NW6 | 45 | Various |
| DIARETDB0 | 130 (110:20) | Finland | N/A | N/A | N/A | 50 | 1500 × 1152 |
| DIARETDB1 | 89 (84:5) | Finland | N/A | N/A | N/A | 50 | 1500 × 1152 |
| EyePACS | | USA | ICDR | N/A | Various | N/A | Various |
| IDRiD | 516 (168:348) | India | ICDR | 100% | Kowa VX 0 10a | 50 | 4288 × 2848 |
| APTOS 2019 | 5590 | India | ICDR | N/A | Various | N/A | N/A |
| DDR | 13,673 (6256:6266) | China | ICDR | N/A | Various | 45 | N/A |

DR Diabetic retinopathy, FOV field of view, ICDR International Clinical Diabetic Retinopathy Severity Scale, N/A not available

Table 2 Retrospective validation studies of artificial intelligence in diabetic retinopathy

| References | Algorithm | Approach | Dataset | Images (n) | Detection | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) | AUC (95% CI) |
|----------------------------|-------------|----------|---------------------------------------|------------|-----------|----------------------|----------------------|-------------------|-------------------|---------------------|
| Larsen et al. [27] | Retinalyze | ML | WCDRS | 400 | DR/no DR | 96.7% | 71.4% | N/A | N/A | 0.903 |
| Larsen et al. [28] | Retinalyze | ML | Steno Diabetes Center, Denmark | 365 | DR/no DR | 93.1% | 71.6% | N/A | N/A | 0.936 |
| Oliveira et al. [23] | Retmarker | DL | ARS-Centro | 21,544 | DR/no DR | 96.1% (94.4–97.9) | 51.7% (50.3–53.1) | N/A | N/A | 0.849 |
| Abramoff et al. [29] | IDP | ML | Messidor-2 | 1748 | rDR | 96.8% (94.4–99.3) | 59.4% (55.7–63.0) | 39.8% (35.2–44.3) | 98.5% (97.4–99.7) | 0.937 (0.916–0.959) |
| Hansen et al. [30] | IDP | ML | Nakuru Eye Study | 6788 | DR/no DR | 91.0% (88.0–93.4) | 69.9% (68.3–71.6) | 32.1% (29.6–34.7) | 98.0% (97.4–98.6) | 0.878 (0.850–0.905) |
| Abramoff et al. [21] | IDx-DR X2.1 | DL | Messidor-2 | 1748 | rDR | 96.8% (93.3–98.8) | 87.0% (84.2–89.4) | 67.4% (61.5–72.9) | 99.0% (97.8–99.6) | 0.980 (0.968–0.992) |
| Bhaskaranand et al. [33] | EyeArt v1.2 | ML | EyePACS | 40,542 | rDR | 100% (96.1–100) | 90.8% (88.5–92.7) | 56.4% (48.4–64.1) | 100% (99.5–100) | 0.989 (0.984–0.994) |
| Gulshan et al. [11] | ARDA | DL | Messidor-2 | 1748 | rDR | 87.0% (88.0–92.0) | 63.2% (61.7–64.6) | N/A | N/A | 0.879 (0.865–0.893) |
| Ting et al. [32] | SELENA+ | DL | EyePACS-1 | 9963 | rDR | 90.3% (87.5–92.7) | 98.1% (97.7–99.1) | N/A | N/A | 0.991 (0.988–0.993) |
| | | | SIDRP | 71,896 | rDR | 90.5% (87.3–93.0) | 91.2% (88.0–93.6) | N/A | N/A | 0.936 (0.925–0.943) |
| Li et al. [35] | DLA | DL | LabelMe, China | 35,201 | rDR | 100% (94.1–100) | 91.1% (90.7–91.4) | N/A | N/A | 0.958 (0.956–0.961) |
| Bhaskaranand et al. [34] | EyeArt v2.0 | DL | EyePACS | 850,908 | rDR | 91.3% (90.9–91.7) | 91.1% (90.9–91.3) | 72.5% (71.9–73.0) | 97.6% (97.5–97.7) | 0.965 (0.963–0.966) |
| Ruamvithoonsuk et al. [31] | ARDA | DL | Thailand's national screening program | 25,326 | rDR | 96.8% (89.3–99.3) | 95.6% (98.3–98.7) | N/A | N/A | 0.987 (0.977–0.995) |

Table 2 continued

| References | Algorithm | Approach | Dataset | Images (<i>n</i>) | Detection | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) | AUC (95% CI) |
|------------------------------|--------------------------|----------|----------------|------------------------|-----------|-------------------------|-------------------------|-----------------|-----------------|--------------|
| Grzybowski and Brona [36] | Retinalyze strategy 1 | ML | Poznan, Poland | 680 | rDR | 89.7% (78.8–96.1) | 71.8% (62.4–80.0) | 62.7% | 92.9% | 0.807 |
| | Retinalyze strategy 2 | ML | Poznan, Poland | 680 | rDR | 74.1% (61.0–84.7) | 93.6% (87.3–97.4) | 86.0% | 87.3% | 0.839 |
| | IDx-DR | DL | Poznan, Poland | 680 | rDR | 93.3% (83.8–98.2) | 95.5% (89.7–98.5) | 91.8% | 96.3% | 0.944 |

AI Artificial intelligence, *AUC* area under the curve, *CI* confidence interval, *DL* deep learning, *DLA* deep learning algorithm, *IDP* Iowa Detection Program, *ML* machine learning, *NPV* negative predictive value, *PPV* positive predictive value, *rDR* referable diabetic retinopathy, *VTRDR* vision threatening diabetic retinopathy

sometimes called the “external validation” or “offline validation”, and is unique for AI development since drug development programs do not include this phase. After this validation, the algorithms may be validated prospectively in the “early live clinical prospective validation” phase, which is comparable to phase 1 and phase 2 studies for small-scale safety and efficacy evaluation in drug development programs. The next phase for AI is the “comparative prospective evaluation” phase, which is comparable with phase 3 studies in drug development programs. The last post-market surveillance phase is similar for both AI algorithms and drug development programs [14].

Publicly Available Datasets for DR

The initial dataset for the development of AI for DR screening should be sufficiently large so that the algorithms can achieve robust performance in the in silico evaluation phase. Then, the algorithms can further be tuned to be more precise and generalizable with training data that is diverse in terms of patients’ demographics and ethnicity, image acquisition methods, and image qualities. Most open datasets were developed under the concept that the lack of large publicly available datasets to train the DL models with high-quality images is a reason for the present barriers in the development and application of automated DR detection programs in clinical practices. Public datasets provide researchers with invaluable information for use; several of such databases are shown in Table 1.

The Messidor database was created within the context of the Messidor project to facilitate studies on computer-assisted diagnoses of DR and has been available for public use since 2008. It includes not only the images but also the diagnoses of DR severity and risk of macular edema provided by medical experts; however, no annotations are provided [15].

DIARETDB0 and DIARETDB1 are public databases provided by Kauppi et al. [16, 17] with the purpose of creating a unified framework for evaluating and comparing automatic DR detection methods. Images were captured with

unknown camera settings, which the creators claim to correspond with practical situations. Together with images, the datasets include the “ground truth,” which are annotations provided by four medical professionals who are experts in medical education and ophthalmology. These four experts marked retinal areas containing microaneurysms, hemorrhages, and exudates in the images [16, 17]. The annotations ensure that the extracted DR findings are at the same location as those marked by the experts. The DIARETDB datasets were used for the development of an AI algorithm for automated segmentation and detection of retinal hemorrhages in retinal photographs [18].

EyePACS is a telescreening program that collects retinal photographs from many primary care clinics within the context of the EyePACS telescreening program for DR. This dataset comprises retinal photographs from both mydriatic and non-mydriatic conditions. DR severity grading for this dataset is provided based on the reading by trained graders.

The Asia-Pacific Teleophthalmology Society (APTOS) 2019 Challenge is another database. It is available at the APTOS 2019 Blindness Detection website, where retinal photographs were collected by the Aravind Eye Hospital in India. The images were gathered under different conditions and clinical environments and later labeled by trained ophthalmologists. Another public dataset from India is the IDRiD dataset from an ophthalmology clinic in Nanded. In addition to including data on DR severity and presence of DME, this dataset also provides annotations on DR lesions and the optic disc [19].

China’s public dataset, the DDR, provides three types of annotations, including image-level DR grading annotations, pixel-level annotations, and bounding-box annotations of lesions associated with DR, all labeled by ophthalmologists [20].

Retrospective Validation Studies from ML to DL

Table 2 summarizes selected studies on the retrospective validation of available AI models for

DR screening. It is evident that the early studies using ML in the earlier versions (Retalyze, Retmaker, EyeArt v1.2, and IDP in the period 2003–2015) could achieve high sensitivity, from 91.0% to 96.7%, but their specificities were relatively lower, from 51.7 to 71.6%. It was when DL was applied to IDX-DR X2.1 (new version of IDP; see Table 2) in 2016 that the retrospective validation for screening of rDR by IDX-DR could achieve robust sensitivity and specificity [21].

In the era of ML, prior to DL, three ML algorithms for DR screening, iGradingM, EyeArt v1.2, and Retmarker, were validated in one study in > 20,000 patients with DM in the National Health System (NHS) Diabetic Eye Screening Programme (DESP) in the UK. The investigators found that iGradingM classified the photographs as either having retinopathy or ungradable, which limited further analysis. The authors reported the comparable sensitivity of EyeArt and Retmarker to be 93.8% and 85.0% for detecting rDR, and 99.6% and 97.9% for detecting PDR. The sensitivity and false-positive rates for EyeArt were not affected by ethnicity, gender, nor camera type, but the values did decline with increasing patient age, whereas the performance of Retmarker was affected by patient age, ethnicity, and camera type [22].

Retmarker is an automated DR screening algorithm using ML that was developed in Portugal. This ML system extracts features, such as the presence of microaneurysms, and provides an output as disease or no disease. It also provides a co-registration component which combines images from two visits and compares them at the same location of the retina using the retinal vascular tree as landmarks; this allows the ML algorithm to be able to provide microaneurysm turnover rates [23]. The algorithm can be applied in the treatment of clinically significant macular edema (CSME) with intravitreal injections of ranibizumab by demonstrating a decrease in the absolute number of microaneurysms after the treatment [24]. Retmarker was found to be able to lower the grading burden by 48.8% in the DR screening program in Portugal [25].

Retalyze is another ML-based algorithm using feature extraction to detect DR. The sensitivity and specificity of this algorithm for

detecting DR or no DR was found to be approximately 95% and 71%, respectively, for validation under mydriasis [26] in populations of patients with DM from the Welsh Community DR study in the UK [27] and Steno Diabetes Center in Denmark [28].

IDx-DR (Digital Diagnostics, Coralville, IA, USA) was initially developed at the University of Iowa as an ML algorithm and given the name of the Iowa Detection Program (IDP). The later version of IDP, IDx-DR, became a combination of convolutional neural networks (CNN) and DL enhancement. The model detects different types of DR lesions and assesses image quality and imageability ratings. Validation of the IDP on the Messidor-2 dataset, which includes Caucasian populations, achieved a sensitivity of 96.8% and specificity of 59.4% in detecting rDR [29]. The IDP was later validated on a population from the Nakuru Study in Kenya and achieved a comparable sensitivity of 91.0% and specificity of 69.9%, implying that race might not affect its performance [30]. IDx-DR X2.1, the newer version, which is enhanced by DL components, was validated on the same Messidor-2 dataset but achieved the higher sensitivity of 96.8% and higher specificity of 87.0% for rDR, and a sensitivity of 100.0% and a specificity of 90.8% for VTDR [21].

ARDA (Automated Retinal Disease Assessment) is a DL algorithm developed by Verily Life Sciences LLC (South San Francisco, CA, USA). This algorithm was developed using datasets of approximately 130,000 retinal photographs of patients with DM from the USA and India. The initial validation of this algorithm was on both the Messidor-2 and EyePACS-1 datasets, with approximately 10,000 photographs, following which the algorithm was retrospectively validated on retinal photographs of patients from the national registry of diabetic patients in Thailand, which is a distinct dataset. In this validation, which was a comparison with human graders, the algorithm demonstrated a sensitivity of 96.8%, which was higher than that of the human graders (approx. 74%), while the specificity was comparable (96–97%) [31].

The Singapore Eye Research Institute and the National University Singapore developed a DL system called SELENA+ to detect rDR. These

algorithms were developed to have capabilities to detect a number of other vision-threatening eye diseases, such as glaucoma and age-related macular degeneration (AMD) from retinal photographs. The system was validated from datasets of > 70,000 photographs from ten countries and various races. The investigators of SELENA+ proposed two different scenarios of how AI could be integrated into clinical screening programs. The first scenario was the fully automated model, in which there were no human assistants; a sensitivity of 93.0% and a specificity of 77.5% was found in detecting referable cases, including glaucoma and AMD. The second scenario was a semi-automated model, in which there were human assistants working with AI. In this scenario, the specificity increased to 99.5%, with the sensitivity remaining relatively similar at 91.3% [32].

EyeArt v2.1 (Eyenuk Inc., Los Angeles, CA, USA) is another DL model developed for detecting rDR as a fully automated DR screening system that combines novel morphological image analysis with DL techniques. The earlier version of EyeArt, which was ML-based, was validated on 5084 diabetic patients from EyePACS and on another set of 40,542 images from an independent EyePACS dataset, achieving a sensitivity of 90.0% and specificity of 63.2% [33]. The REVERE 100k study demonstrated that EyeArt v2.1, which was DL-based, could achieve a higher sensitivity (91.3%) and higher specificity (91.1%) that were neither affected by patient ethnicity, gender, nor camera type in the validation of photographs of > 800,000 patients from the routine DR screening protocol of EyePACS [34].

A DL algorithm in China was developed from 71,043 de-identified retinal photographs from a web-based platform, LabelMe (Guangzhou, China). The photographs were provided by 36 ophthalmology departments, optometry clinics, and screening units in China. External validation of this algorithm was performed using over 35,000 images from population-based cohorts of Malaysians, Caucasian Australians, and Indigenous Australians. The sensitivity and specificity were found to be 92.5% and 98.5%, respectively [35].

Table 3 Prospective validation studies of AI in DR

| Algorithm | References | Population | Patients (n) | Dilation | DR grading system | Detection | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) |
|---|-----------------------------|-----------------------------------|--------------|-----------|----------------------------|-----------|----------------------|----------------------|-------------------|--------------------|
| <i>Conventional retinal photography cameras</i> | | | | | | | | | | |
| iGrading | Philip et al. [68] | Primary care, Scotland | 6722 | As needed | Scottish DR Grading Scheme | Any DR | 90.5% (89.3–91.6) | 67.4% (66.0–68.8) | N/A | N/A |
| Bosch | Soto-Pedre et al. [69] | Primary care, Spain | 5253 | Yes | MA detection | Any DR | 94.52% (92.6–96.5) | 68.77% (67.2–70.4) | 34.1% (31.7–36.5) | 98.66% (89.2–99.2) |
| IDx-DR X2.1 | Bawankar et al. [59] | India | 560 | No | AAO PPP 2019 | Any DR | 91.2% (86.4–94.7) | 96.9% (94.5–98.5) | 94.4% (90.4–96.8) | 95.0% (92.5–96.8) |
| | Abramoff et al. [38] | Ten primary care sites, USA | 819 | As needed | ETDRS | mtmDR | 87.2% (81.8–91.2) | 90.7% (88.3–92.7) | N/A | N/A |
| IDx-DR 2.0 | van der Heijden et al. [39] | Hoorn DCS center, The Netherlands | 898 | As needed | ICDR | rDR | 99.3% (96.1–99.9) | 68.8% (61.5–76.2) | 74.6% (68.4–80.8) | 99.1% (97.2–100) |
| | | | | | ICDR | VTDR | 99.1% (95.1–99.9) | 80.4% (73.9–85.9) | 75.3% (68.4–82.3) | 99.3% (96.3–100) |
| | | | | | ICDR | rDR | 68% (56–79) | 86% (84–88) | 30% (24–38) | 97% (95–98) |
| | | | | | EURODIAB | VTDR | 62% (32–85) | 95% (93–96) | 14% (7–27) | 99% (99–100) |
| | | | | | EURODIAB | rDR | 91% (69–98) | 84% (81–96) | 12% (8–18) | 100% (99–100) |
| SELENA + | Bellemo et al. [42] | Zambia, Africa | 1574 | N/A | ICDR | VTDR | 64% (36–86) | 95% (93–96) | 16% (8–29) | 99% (99–100) |
| | | | | | ICDR | rDR | 92.3% (90.1–94.1) | 89.0% (87.9–90.3) | N/A | N/A |
| | | | | | ICDR | VTDR | 99.4% (99.2–99.7) | N/A | N/A | N/A |
| | | | | | EURODIAB | DME | 97.2% (96.6–97.8) | N/A | N/A | N/A |
| VoxelCloud Retina | Zhang et al. [48] | Nationwide DR screening, China | 15,805 | No | ICDR | rDR | 83.3% (81.9–84.6) | 92.5% (92.1–92.9) | 61.8% (60.3–63.3) | 97.4% (97.2–97.7) |
| DLA | Scheetz et al. [47] | Australia | 203 | No | NHS | rDR | 96.9% (83.8–99.9) | 87.7% (81.8–92.2) | 59.6% (45.1–73.0) | 99.3% (96.4–100) |
| ARDA | Ruamviboonsuk et al. [43] | Nationwide DR screening, Thailand | 7651 | As needed | ICDR | VTDR | 91.3% (85.1–97.4) | 96.3% (95.1–97.4) | 79.2% (73.8–84.3) | 95.5% (92.8–97.9) |

Table 3 continued

| Algorithm | References | Population | Patients (#) | Dilation | DR grading system | Detection | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) |
|---|-------------------------|-------------------------------|--------------|----------|-------------------|-----------|----------------------|----------------------|-------------------|-------------------|
| EyeArt v2.1 | Heydon et al. [40] | NHS DESP, UK | 30,405 | N/A | ETDRS | rDR | 95.7% (94.8–96.5) | 54.0% (53.4–54.5) | N/A | N/A |
| EyeArt v2.1 | Ipp et al. [41] | Multicenter, USA | 893 | No | ETDRS | mtmDR | 95.5% (92.4–98.5) | 85.0% (82.6–87.4) | 59.5% (53.9–63.9) | 98.8% (98.2–99.4) |
| | | | | Yes | | mtmDR | 95.5% (92.6–98.4) | 85.3% (83.0–87.5) | 59.1% (53.8–64.4) | 98.8% (98.2–99.5) |
| | | | | No | | VTDR | 95.1% (90.1–100) | 89.0% (87.0–91.1) | 26.7% (19.5–33.0) | 99.8% (99.5–100) |
| | | | | Yes | | VTDR | 95.2% (90.4–100) | 89.5% (87.6–91.4) | 26.1% (19.6–32.6) | 99.8% (99.5–100) |
| <i>Smartphone-based retinal photography</i> | | | | | | | | | | |
| EyeArt v2.1 (FOP) | Rajalakshmi et al. [49] | Tertiary care hospital, India | 301 | Yes | ICDR | Any DR | 95.8% (92.9–98.7) | 80.2% (72.6–87.8) | 89.7% (85.5–93.8) | 91.4% (85.7–97.1) |
| EyeArt v2.0 (FOP) | Kim et al. [51] | Two university hospitals, USA | 72 | Yes | ICDR | rDR | 77.8% (67.3–85.7) | 71.5% (48.7–86.9) | N/A | N/A |
| Remidio | Natarajan et al. [50] | Dispensaries in Mumbai, India | 213 | Yes | ICDR | Any DR | 85.2% (66.3–95.8) | 92.0% (97.1–95.4) | N/A | N/A |
| | | | | | | rDR | 100% (78.2–100) | 88.4% (83.2–92.5) | N/A | N/A |

AJO PPPP American Academy of Ophthalmology Preferred Practice Pattern, *DCS* diabetic care system, *DESP* Diabetic Eye Screening Programme, *ETDRS* Early Treatment Diabetic Retinopathy Study Severity Scale, *MA* microaneurysm, *mtmDR* more than mild diabetic retinopathy, *NHS* National Health Screening guidelines

A recent study compared the performance of Retinalyze 1 and 2 and IDx-DR v2.1 for detecting rDR in retinal images captured without mydriasis in the same group of patients with DM. Both automated algorithms were able to analyze most of the images. However, the sensitivities and specificities of Retinalyze (89.7% and 71.8%, respectively, for version 1; 74.1% and 93.6%, respectively, for version 2) were lower than those of IDx-DR (93.3% and 95.5%, respectively). The investigators noted that Retinalyze's ability to annotate images is helpful for human verification, but concluded that the algorithm could not be used for diagnosing patients without direct clinician oversight [36].

Another retrospective validation study compared seven different DL algorithms for detecting rDR. The investigators found that most of the algorithms performed no better than human graders. The sensitivities varied widely (51.0–85.9%) although high negative predictive values (82.7–93.7%) were observed. Interestingly, one algorithm was significantly worse than human graders and would miss up to one fourth of advanced retinopathy cases (72.4% sensitivity for PDR), a limitation which could potentially lead to vision loss [37].

Prospective Validation Studies of DL

One of the first prospective validation studies on DL for DR screening was the trial for IDx-DR. This pivotal trial was then submitted for U.S. Food and Drug Administration (FDA) approval on the AI model. In this study, in which photographs of the Early Treatment Diabetic Retinopathy Study (ETDRS) were used as standards, IDx-DR was prospectively validated in primary care units in the USA and found to have a sensitivity of 87.2% and specificity of 90.7%, both of which were higher than the pre-specified superiority endpoints of a sensitivity of 85% and specificity of 82.5% [38]. However, these diagnostic parameters were lower than those found in the retrospective validation on IDx-DR (approx. 97% sensitivity and approx. 87% specificity) [21]. IDx-DR was then prospectively validated in the Hoorn Diabetes Care System in the Netherlands, for detecting

rDR and VTDR based on two DR classification systems, the International Clinical Diabetic Retinopathy Severity Scale (ICDR) and EURO-DIAB. The sensitivity of IDx-DR to detect VTDR was approximately 60% for both classifications, but its sensitivity to detect rDR was 91% when based on the EURODIAB and 68% when based on the ICDR. This discrepancy may have arisen from the different scores in different classification systems [39].

There have been a few prospective validation studies on EyeArt, another autonomous AI model approved by the U.S. FDA for DR screening. A large prospective validation study was conducted in > 30,000 patients in the NHS DESP in the UK using EyeArt v2.1 with DL incorporation. In the combined manual grading with EyeArt, this system achieved a sensitivity of 95.7% and a specificity of 54.0% for triaging rDR [40]. Another prospective study validated EyeArt in multiple primary care centers in the USA for the detection of more-than-mild DR (mtmDR) and VTDR; the referral rate in this study was found to be 31.3%. In the comparison on 2-field non-dilated retinal photographs with the standard dilated 4-wide-field stereoscopic photographs as standards, the investigators in this study found that 12.5% of images were classified as ungradable; however, this prevalence dropped to 2.7% under the dilate-if-needed protocol. Although imageability increased with pupillary dilation, EyeArt achieved similar sensitivities and specificities for detecting both mtmDR (95.5% sensitivity and 85.3% specificity) and VTDR (95.2% sensitivity and 89.5% specificity) in both the non-dilated and dilate-if-needed protocol [41].

In the prospective validation of SELENA+ in diabetic patients who attended the mobile screening in Zambia, upon grading 4504 images, the DL achieved a sensitivity of 92.3% and 99.4% for rDR and VTDR, respectively. This algorithm, which was developed in Singapore, showed excellent generalizability for patients from different races in both the retrospective and prospective validations [42].

Another prospective validation study was conducted for ARDA in the existing workflow of nine primary care centers in Thailand's national DR screening program. ARDA was still able to

achieve 91.4% sensitivity and 95.4% specificity for detecting VTDR. This prospective study applied a semi-automated approach using local retinal specialists to overread the results of DL. Various fundus cameras were used according to routine practice in the primary care sites; DL performance across the different camera models was not affected [43].

Prospective studies have been carried out on ML-based Retmarker to analyze microaneurysm turnover rates on retinal photographs as a biomarker of DR progression. The investigators found that higher microaneurysm turnover at the macula correlated with earlier development of CSME [44, 45]. Changes in microaneurysms in patients with mild NPDR in the first year were observed to be associated with the development of VTDR over 5 years [46].

The DL algorithm developed by Li et al. in China, which was previously trained with images collected from the Chinese dataset and later validated on populations of Indigenous Australians and Caucasians in Australia, was prospectively evaluated in Aboriginal Medical Services of Australian Healthcare settings. A sensitivity of 96.9% and specificity of 87.7% was found for detecting rDR in this population. In addition to the performance of the DL algorithm, this study also investigated the experience and acceptance of automated screening from patients, clinicians, and organizational stakeholders involved and found high acceptance for AI [47].

Another DL for automated DR screening, Voxel Cloud, was validated in > 150 diabetes centers in a screening program in China after training with private retinal image database comprising > 140,000 images and tested on both public and private datasets (one of them being the APTOS 2019 Blindness Detection dataset). This algorithm was able to achieve a sensitivity of 83.3% and a specificity of 92.5% for detecting rDR from 31,498 images [48].

AI Analysis on Smartphone Photographs

A few studies have been performed on retinal photographs taken using smartphone cameras for analysis by AI. These platforms were not

handheld but were smartphones attached to desktop cameras for use in low-resource settings. All published studies validating AI with smartphone photography were prospective since the retinal photographs were captured in real-time and analyses were made at the point of care. A pilot study conducted in India used the Remidio Fundus on Phone (FOP) application to capture retinal images after the pupils were dilated and graded by the offline EyeArt algorithm. This study achieved a sensitivity and specificity of 95.8% and 80.2%, respectively, for any DR, and a sensitivity and specificity of 99.1% and 80.4%, respectively, for VTDR [49]. Another smartphone-based AI for DR screening also conducted in India also used the Remidio FOP for image capture but used an offline automated analysis by Medios AI. The sensitivity and specificity for detecting rDR of this system was 100.0% and 88.4%, respectively [50]. Another study conducted in the USA used EyeArt for automated analysis on retinal photographs retrieved from a smartphone-based RetinaScope camera; the sensitivity and specificity for detecting rDR was 77.8% and 71.5%, respectively [51].

In summary, for these retrospective and prospective studies, including the studies on retinal photographs from smartphones presented in Tables 2 and 3, numerous algorithms have been developed and validated, both retrospectively and prospectively, with differences in approach, dataset, camera, images, or the level of DR in detection. The overall performances of AI was high, although performances were generally lower in some prospective validation studies.

Different Types of Retinal Photographs

Of all the retinal imaging modalities, color retinal photographs acquired from conventional retinal cameras using simple flash camera technology are ubiquitous and practical for screening DR in primary care settings, particularly in remote or underserved areas. Another type of retinal camera which is gaining in popularity recently uses white light-emitting diodes (LED) combined with confocal scanning laser

technology and enhanced color fidelity [52]. These white LED cameras provide color retinal photographs which differ in appearance from those obtained with conventional cameras. The differences between the photographs obtained from these two types of cameras appear to be the image viewing angle (45° for the conventional cameras and 60° for the LED cameras), the image resolutions, and color discriminations [52].

In one study, EyeArt V.2.1.0 was used to analyze retinal photographs obtained from both conventional cameras and the white LED cameras in 1257 patients with DM from the UK NHS DESP. The authors found that the diagnostic accuracy of detecting any retinopathy and PDR from the photographs obtained from both cameras was similar at 92.3% sensitivity and 100.0% specificity using human grading as the standard [53]. In the published article, it was not clear whether EyeART applied the same algorithm for the analysis of photographs from both cameras, and the authors stated that EyeART was not optimized for photographs from the LED cameras and also that reference patterns might not be properly recognized. Wongchaisuwat et al. performed a study in which they had to apply separate DL algorithms, one for the conventional photographs and another for the photographs from white LED cameras, to prospectively validate DL on photographs from both cameras for DR screening. These investigators achieved a sensitivity of 82% and 89%, and specificity of 92% and 84%, for conventional and LED photographs, respectively, to detect rDR using retinal examination from retinal specialists as the standard [54].

It would be an ideal situation if an AI model would be able to analyze the different domains of retinal photographs from conventional cameras and white LED cameras and achieve a robust performance. The effect of different domains of images on AI performance may be more pronounced for segmentation of optical coherence tomography (OCT) images. There has been an attempt to develop AI models that can perform well on OCT images from different domains of different manufacturers [55].

Different Manufacturers of Conventional Cameras

Although conventional retinal cameras provide similar photographs with the same viewing angle, the image resolutions and color discrimination can still differ among cameras from different manufacturers. A study by Srinivasan et al. aimed to address this issue on DL for DR screening. These authors found that the performance of the ARDA algorithm did not change with the different brands and versions of retinal camera. This analysis was conducted on 15,351 retinal photographs captured from seven different cameras. The sensitivity for detecting rDR varied slightly from 89.7% to 98.7% whereas the specificity varied from 93.3 to 98.2% [56].

However, for the two AI models which have been approved by the US FDA for DR screening, IDX-DR and EyeArt, the approval was based on the use for Topcon retinal cameras (Topcon Corp., Tokyo, Japan).

Mydriasis or Non-mydriasis

One of the major problems revealed by prospective validation studies has been ungradable photographs, the prevalence of which increases significantly without mydriasis [41]. Since AI tends to grade more ungradable photographs compared to human graders, screening using AI without mydriasis may result in significantly more patients with ungradable grading, many of whom may not have retinopathy [57]. The drawback of mydriasis is the long duration of up to 4–6 h caused by pharmacologic dilation. Another drawback is the fear of producing angle closure although this rarely occurs [58]. In a study conducted in India using the Bosch AI algorithm for DR screening, the investigators placed patients in a dark room for 2 min to achieve physiologic mydriasis before taking retinal photographs. Approximately 96% of the images had acceptable quality, and 91.2% sensitivity and 96.9% specificity were achieved from this prospective study [59].

DISCUSSION

Although DL for DR screening using color retinal photographs is at the forefront of studies of AI in ophthalmology, the majority of studies published to date were still *in silico* evaluation on new methods of applying AI or using AI for prediction tasks [60]. Reports on real-world implementation, which are comparable with phase 4 post-marketing surveillance in clinical studies on drugs, are scarce.

Challenges in Real-world Implementation

It is generally well accepted based on results from many prospective studies in various countries that AI is an effective tool for DR screening, with a high accuracy in detecting both rDR and VTDR for implementation in primary care settings. However, real-world implementation of AI for screening may not rest on only its accuracy but rather on the acceptability of personnel and patients, including their view on the integration of AI screening into existing clinical screening systems as an improvement step [61]. The indicators demonstrating that AI can improve healthcare systems can be the improvement in screening attendance rate, which is the proportion of patients with DM screened for DR [62] and the improvement in referral adherence, which is the proportion of patients detected as referrals by AI who receive eye care at referral centers. One study demonstrated that point-of-care delivery of screening results by AI may improve this adherence rate [63]. Finally, the indicator of reducing the rate of visual loss should be the goal for DR screening whether AI is to be deployed or not.

The deployment of AI, as either an autonomous or semi-autonomous screening, may affect the cost of DR screening. The cost-saving from AI may differ from one country to the other, depending on the healthcare resources. A study by the NHS DESP found that both the ML algorithms EyeArt and Retmarker were cost saving compared to human grading when they were deployed as semi-autonomous screening [64]. A threshold analysis testing demonstrated

that the highest cost of the ML models per patient, before which they became more expensive than human grading, was £3.82 per patient for Retmarker and £2.71 per patient for EyeArt [64]. A cost-minimization analysis from Singapore also found the semi-autonomous model to be more cost-saving than autonomous screening. In this study, the cost of human grading was US\$77 per patient per year, whereas the autonomous screening was \$66 per patient per year, and the semi-autonomous was \$62 per patient per year [65].

Technical challenges may also have influenced the successful implementation of AI. Integrating an AI model into the camera system is an important technical issue when a screening program requires the use of existing cameras. Integrating AI into DR screening system with an existing hospital management system or electronic health record system is another challenge to ensure sustainability [61].

In the bigger picture, the concept of AI governance should be applied in real-world implementation to ensure equality, privacy, fairness, inclusiveness, safety and security, robustness, transparency, explainability, accountability, and auditability [66]. There has been a proposition on the governance model for AI in healthcare which outlines four main components: fairness, transparency, trustworthiness, and accountability [67]. For fairness, there should be Data Governance Panels overseeing the collection and use of data; the design of AI models should ensure procedural and distributive justice. For transparency, there should be transparency for AI model decision-making and support for patient and clinician autonomy. For trustworthiness, educating patients and clinicians is important. Informed consent is required from patients and appropriate, and authorized use of patient data should be applied. For accountability, regulation and accountability at the approval, introduction, and development phases of AI in healthcare should be followed. Real-world implementation of AI for DR screening should undoubtedly follow this concept.

CONCLUSION

The breakthrough of AI for DR screening on color retinal photographs came with the advent of DL which carried the robustness results on diagnostic accuracy from in silico evaluation to off-line validation on various datasets from different populations. Although in subsequent phases of evaluation, many prospective validation studies supported the diagnostic accuracy, the results were generally lower, although acceptable for clinical practice. To implement AI for screening in the real world, many practical issues are at least as important as diagnostic accuracy, and they should be taken into high consideration. These include, for example, mydriasis, integration into existing camera systems, integration into electronic health record systems, and acceptability by patients and providers. AI for DR screening should follow the governance of AI in healthcare to ensure fairness, transparency, trustworthiness, and accountability for all the stakeholders including policy-makers, healthcare providers, AI developers, and patients.

ACKNOWLEDGEMENTS

Editorial assistance in the preparation of this article was provided by Methaphon Chainakul, MD and Niracha Arjkongharn, MD, of the Department of Ophthalmology, College of Medicine, Rajavithi Hospital, Rangsit University, Bangkok, Thailand; also Varis Ruamviboonsuk, MD, of Department of Biochemistry, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand. No funding or support was provided for any assistance given.

Funding. No funding or sponsorship was received for this study or publication of this article.

Author Contributions. All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Andrzej Grzybowski and Paisan Ruamviboonsuk. The first draft of the

manuscript was written by Panisa Singhanetr. The revision and correction of the manuscript were performed by Onnisa Nanegrungsunk. All authors commented on every version of the manuscript. All authors read and approved the final manuscript.

Disclosures. All authors declare that they have no competing interests. The views expressed in the publication are those of the authors. Andrzej Grzybowski has received grants from Alcon, Bausch&Lomb, Zeiss, Teleon, J&J, CooperVision, and Hoya; has received lecture fees from Thea, Polpharma, and Viatrix; and is a member of the advisory boards of Nevakar, GoCheckKids, and Thea. Panisa Singhanetr has nothing to disclose. Onnisa Nanegrungsunk has received reimbursement from Novartis for medical writing (for studies other than the present study; has received lecture fees from Allergan; has received financial support for attending educational meetings from Bayer and Novartis; and is on the advisory board of Roche. Paisan Ruamviboonsuk has received grants from Roche and Novartis; is a consultant for Bayer, Novartis, and Roche; and has received speaker fees from Novartis, Roche, Bayer, and Topcon.

Compliance with Ethics Guidelines. This article is based on previously conducted studies and does not contain any new studies with human participants or animals performed by any of the authors. No specific ethical approval was required for this article.

Data Availability. Data sharing is not applicable to this article as no data sets were generated or analyzed during the current study.

Open Access. This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons

licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of type 2 diabetes-global burden of disease and forecasted trends. *J Epidemiol Glob Health*. 2020;10(1):107–11.
2. Saeedi P, Petersohn I, Salpea P, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract*. 2019;157:107843.
3. GBD 2019 Blindness and Vision Impairment Collaborators. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *Lancet Glob Health*. 2021;9(2):e144–60.
4. Teo ZL, Tham YC, Yu M, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis. *Ophthalmology*. 2021;128(11):1580–91.
5. The Diabetic Retinopathy Study Research Group. Photocoagulation treatment of proliferative diabetic retinopathy. Clinical application of Diabetic Retinopathy Study (DRS) findings, DRS report number 8. *Ophthalmology*. 1981;88(7):583–600.
6. Early Treatment Diabetic Retinopathy Study Research Group. Early photocoagulation for diabetic retinopathy. ETDRS report number 9. *Ophthalmology*. 1991;98(5 Suppl):766–85.
7. Silpa-archa S, Ruamviboonsuk P. Diabetic retinopathy: current treatment and Thailand perspective. *J Med Assoc Thai*. 2017;100(Suppl 1):S136–47.
8. Liew G, Michaelides M, Bunce C. A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010. *BMJ Open*. 2014;4(2):e004015.
9. Nguyen HV, Tan GS, Tapp RJ, et al. Cost-effectiveness of a national telemedicine diabetic retinopathy screening program in Singapore. *Ophthalmology*. 2016;123(12):2571–80.
10. Murchison AP, Friedman DS, Gower EW, et al. A multi-center diabetes eye screening study in community settings: study design and methodology. *Ophthalmic Epidemiol*. 2016;23(2):109–15.
11. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402–10.
12. Tsiknakis N, Theodoropoulos D, Manikis G, et al. Deep learning for diabetic retinopathy detection and classification based on fundus images: a review. *Comput Biol Med*. 2021;135:104599.
13. Pareja-Ríos A, Ceruso S, Romero-Aroca P, Bonaque-González S. A new deep learning algorithm with activation mapping for diabetic retinopathy: back-testing after 10 years of tele-ophthalmology. *J Clin Med*. 2022;11(17):4945.
14. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med*. 2022;28:924–33.
15. Decencière E, Zhang X, Cazuguel G, et al. Feedback on a publicly distributed image database: the Mesidor database. *Image Anal Stereol*. 2014;33(3):231–4.
16. Kauppi T, Kalesnykiene V, Kämäräinen J-K, et al., editors. DIARETDB 0: evaluation database and methodology for diabetic retinopathy algorithms. 2007. https://www.it.lut.fi/project/imageret/diaretdb0/doc/diaretdb0_techreport_v_1_1.pdf. Accessed 7 Dec 2022.
17. Kauppi T, Kalesnykiene V, Kamarainen J-K, et al. DIARETDB1 diabetic retinopathy database and evaluation protocol. 2007. https://www.it.lut.fi/project/imageret/diaretdb1/doc/diaretdb1_techreport_v_1_1.pdf. Accessed 10 Dec 2022.
18. Aziz T, Ilesanmi AE, Charoenlarnopparut C. Efficient and accurate hemorrhages detection in retinal fundus images using smart window features. *Appl Sci*. 2021;11(14):6391.
19. Porwal P, Pachade S, Kamble R, et al. Indian diabetic retinopathy image dataset (IDRID): a database for diabetic retinopathy screening research. *Data*. 2018;3(3):25.

20. Li T, Gao Y, Wang K, Guo S, Liu H, Kang H. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Inf Sci.* 2019;501: 511–22.
21. Abràmoff MD, Lou Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning. *Invest Ophthalmol Vis Sci.* 2016;57(13):5200–6.
22. Tufail A, Kapetanakis VV, Salas-Vega S, et al. An observational study to assess if automated diabetic retinopathy image assessment software can replace one or more steps of manual imaging grading and to determine their cost-effectiveness. *Health Technol Assess.* 2016;20(92):1–72.
23. Oliveira CM, Cristóvão LM, Ribeiro ML, Abreu JR. Improved automated screening of diabetic retinopathy. *Ophthalmologica.* 2011;226(4):191–7.
24. Leicht SF, Kernt M, Neubauer A, et al. Microaneurysm turnover in diabetic retinopathy assessed by automated RetmarkerDR image analysis—potential role as biomarker of response to ranibizumab treatment. *Ophthalmologica.* 2014;231(4): 198–203.
25. Ribeiro L, Oliveira CM, Neves C, Ramos JD, Ferreira H, Cunha-Vaz J. Screening for diabetic retinopathy in the central region of Portugal. Added value of automated “disease/no disease” grading. *Ophthalmologica.* 2014;233:96–103.
26. Hansen AB, Hartvig NV, Jensen MS, Borch-Johnsen K, Lund-Andersen H, Larsen M. Diabetic retinopathy screening using digital non-mydratic fundus photography and automated image analysis. *Acta Ophthalmol Scand.* 2004;82(6):666–72.
27. Larsen M, Godt J, Larsen N, et al. Automated detection of fundus photographic red lesions in diabetic retinopathy. *Invest Ophthalmol Vis Sci.* 2003;44(2):761–6.
28. Larsen N, Godt J, Grunkin M, Lund-Andersen H, Larsen M. Automated detection of diabetic retinopathy in a fundus photographic screening population. *Invest Ophthalmol Vis Sci.* 2003;44(2): 767–71.
29. Abràmoff MD, Folk JC, Han DP, et al. Automated analysis of retinal images for detection of referable diabetic retinopathy. *JAMA Ophthalmol.* 2013;131(3):351–7.
30. Hansen MB, Abràmoff MD, Folk JC, Mathenge W, Bastawrous A, Peto T. Results of automated retinal image analysis for detection of diabetic retinopathy from the Nakuru study, Kenya. *PLoS ONE.* 2015;10(10):e0139148.
31. Ruamviboonsuk 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(1): 25.
32. 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(22):2211–23.
33. Bhaskaranand M, Ramachandra C, Bhat S, et al. Automated diabetic retinopathy screening and monitoring using retinal fundus image analysis. *J Diabetes Sci Technol.* 2016;10(2):254–61.
34. Bhaskaranand M, Ramachandra C, Bhat S, et al. The value of automated diabetic retinopathy screening with the EyeArt system: a study of more than 100,000 consecutive encounters from people with diabetes. *Diabetes Technol Ther.* 2019;21(11): 635–43.
35. 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(12):2509–16.
36. Grzybowski A, Brona P. Analysis and comparison of two artificial intelligence diabetic retinopathy screening algorithms in a pilot study: IDx-DR and Retalyze. *J Clin Med.* 2021;10(11):2532.
37. 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(5): 1168–75.
38. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med.* 2018;1:39.
39. van der Heijden AA, Abramoff MD, Verbraak F, van Hecke MV, Liem A, Nijpels G. Validation of automated screening for referable diabetic retinopathy with the IDx-DR device in the Hoorn Diabetes Care System. *Acta Ophthalmol.* 2018;96(1):63–8.
40. Heydon P, Egan C, Bolter L, et al. Prospective evaluation of an artificial intelligence-enabled algorithm for automated diabetic retinopathy screening of 30 000 patients. *Br J Ophthalmol.* 2021;105(5):723–8.
41. Ipp E, Liljenquist D, Bode B, et al. Pivotal evaluation of an artificial intelligence system for autonomous detection of referable and vision-threatening diabetic retinopathy. *JAMA Netw Open.* 2021;4(11): e2134254.

42. Bellemo V, Lim ZW, Lim G, et al. Artificial intelligence using deep learning to screen for referable and vision-threatening diabetic retinopathy in Africa: a clinical validation study. *Lancet Digit Health*. 2019;1(1):e35–44.
43. Ruamviboonsuk P, Tiwari R, Sayres R, et al. Real-time diabetic retinopathy screening by deep learning in a multisite national screening programme: a prospective interventional cohort study. *Lancet Digit Health*. 2022;4(4):e235–44.
44. Ribeiro ML, Nunes SG, Cunha-Vaz JG. Microaneurysm turnover at the macula predicts risk of development of clinically significant macular edema in persons with mild nonproliferative diabetic retinopathy. *Diabetes Care*. 2013;36(5):1254–9.
45. Pappuru RKR, Ribeiro L, Lobo C, Alves D, Cunha-Vaz J. Microaneurysm turnover is a predictor of diabetic retinopathy progression. *Br J Ophthalmol*. 2019;103(2):222–6.
46. Santos AR, Mendes L, Madeira MH, et al. Microaneurysm turnover in mild non-proliferative diabetic retinopathy is associated with progression and development of vision-threatening complications: a 5-year longitudinal study. *J Clin Med*. 2021;10(10):2142.
47. Scheetz J, Koca D, McGuinness M, et al. Real-world artificial intelligence-based opportunistic screening for diabetic retinopathy in endocrinology and indigenous healthcare settings in Australia. *Sci Rep*. 2021;11(1):15808.
48. Zhang Y, Shi J, Peng Y, et al. Artificial intelligence-enabled screening for diabetic retinopathy: a real-world, multicenter and prospective study. *BMJ Open Diabetes Res Care*. 2020;8(1):e001596.
49. Rajalakshmi R, Subashini R, Anjana RM, Mohan V. Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye*. 2018;32(6):1138–44.
50. Natarajan S, Jain A, Krishnan R, Rogye A, Sivaprasad S. Diagnostic accuracy of community-based diabetic retinopathy screening with an offline artificial intelligence system on a smartphone. *JAMA Ophthalmol*. 2019;137(10):1182–8.
51. Kim TN, Aaberg MT, Li P, et al. Comparison of automated and expert human grading of diabetic retinopathy using smartphone-based retinal photography. *Eye*. 2021;35(1):334–42.
52. Sarao V, Veritti D, Borrelli E, Satta SVR, Poletti E, Lanzetta P. A comparison between a white LED confocal imaging system and a conventional flash fundus camera using chromaticity analysis. *BMC Ophthalmol*. 2019;19(1):231.
53. Olvera-Barrios A, Heeren TF, Balaskas K, et al. Diagnostic accuracy of diabetic retinopathy grading by an artificial intelligence-enabled algorithm compared with a human standard for wide-field true-colour confocal scanning and standard digital retinal images. *Br J Ophthalmol*. 2021;105(2):265–70.
54. Wongchaisuwat N, Trinavarat A, Rodanant N, et al. In-person verification of deep learning algorithm for diabetic retinopathy screening using different techniques across fundus image devices. *Transl Vis Sci Technol*. 2021;10(13):17.
55. Wu Y, Olvera-Barrios A, Yanagihara R, et al. Training deep learning models to work on multiple devices by cross-domain learning with no additional annotations. *Ophthalmology*. 2023;130(2):213–22.
56. Srinivasan R, Surya J, Ruamviboonsuk P, Chotcomwongse P, Raman R. Influence of different types of retinal cameras on the performance of deep learning algorithms in diabetic retinopathy screening. *Life*. 2022;12(10):1610.
57. Beede E, Baylor EE, Hersch F, et al. A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. In: *Proceedings of the 2020 CHI conference on human factors in computing systems*; 2020. <https://doi.org/10.1145/3313831.3376718>
58. Patel KH, Javitt JC, Tielsch JM, et al. Incidence of acute angle-closure glaucoma after pharmacologic mydriasis. *Am J Ophthalmol*. 1995;120(6):709–17.
59. Bawankar P, Shanbhag N, Smitha KS, et al. Sensitivity and specificity of automated analysis of single-field non-mydratric fundus photographs by Bosch DR Algorithm-Comparison with mydratric fundus photography (ETDRS) for screening in undiagnosed diabetic retinopathy. *PLoS ONE*. 2017;12(12):0189854.
60. Bora A, Balasubramanian S, Babenko B, et al. Predicting the risk of developing diabetic retinopathy using deep learning. *Lancet Digit Health*. 2021;3(1):e10–9.
61. Yuan A, Lee AY. Artificial intelligence deployment in diabetic retinopathy: the last step of the translation continuum. *Lancet Digit Health*. 2022;4(4):e208–9.
62. Lawrenson JG, Graham-Rowe E, Lorencatto F, et al. Interventions to increase attendance for diabetic retinopathy screening. *Cochrane Database Syst Rev*. 2018;1(1):Cd12054.

63. Pedersen ER, Cuadros J, Khan M, et al. Redesigning clinical pathways for immediate diabetic retinopathy screening results. *NEJM Catalyst*. 2021;2:8. <https://doi.org/10.1056/CAT.21.0096>.
64. Tufail A, Rudisill C, Egan C, et al. Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. *Ophthalmology*. 2017;124(3):343–51.
65. Xie Y, Nguyen QD, Hamzah H, et al. Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. *Lancet Digit Health*. 2020;2(5):e240–9.
66. The Ministry of Economy Trade and Industry (METI). Governance guidelines for implementation of AI principles ver. 1.1. 2021. https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20210709_8.pdf. Accessed 20 Dec 2022.
67. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. *J Am Med Inform Assoc*. 2020;27(3):491–7.
68. Philip S, Fleming AD, Goatman KA, Fonseca S, McNamee P, Scotland GS, et al. The efficacy of automated “disease/no disease” grading for diabetic retinopathy in a systematic screening programme. *Br J Ophthalmol*. 2007;91(11):1512–7.
69. Soto-Pedre E, Navea A, Millan S, Hernaez-Ortega MC, Morales J, Desco MC, et al. Evaluation of automated image analysis software for the detection of diabetic retinopathy to reduce the ophthalmologists’ workload. *Acta Ophthalmol*. 2015;93(1):e52–6.