The Evolution of Diabetic Retinopathy Screening Programmes: A Chronology of Retinal Photography from 35 mm Slides to Artificial Intelligence

Josef Huemer 1,2
Siegfried K Wagner 1
Dawn A Sim 1

1 NIHR Biomedical Research Center at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; 2 Vienna Institute for Research in Ocular Surgery, A Karl Landsteiner Institute, Hanusch Hospital, Vienna, Austria

Abstract: As a third of people with diabetes mellitus (DM) will suffer the microvascular complications of diabetic retinopathy (DR) and therapeutic options can effectively prevent visual impairment, systematic screening has substantially reduced disease burden in developed countries. In an effort to tackle the rising incidence of DM, screening programmes have modernized in synchrony with technical and infrastructural advancements. Patient evaluation has shifted from face-to-face ophthalmologist-based review delivered through community grassroots to asynchronous store-and-forward modern telemedicine platforms commissioned on a nationwide scale. First pioneered with primitive 35-mm slide film retinal photography, the last decade has seen an emergence of high resolution and widefield imaging devices, which may reveal extents of DR indiscernible to the clinician but with implications of potential earlier identification. Similar progress has been seen in image analysis approaches—automated image analysis of retinal photographs of DR has evolved from qualitative feature detection to rules-based algorithms to autonomous artificial intelligence-powered classification. Such models have, relatively rapidly, been validated and are now receiving approval from health regulation authorities with deployment into the clinical sphere. In this review, we chart the evolution of global DR screening programmes since their inception highlighting major milestones in healthcare infrastructure, telemedicine approaches and imaging devices that have shaped the robust and effective frameworks recognised today. We also provide an outlook for the future of DR screening in the context of recent technological advancements with respect to their limitations in current times.

Keywords: telemedicine, artificial intelligence, imaging, diabetes, retina, photography

Introduction

Despite the first description of diabetic macular changes by Eduard Jaeger and Albert von Graefe in the 1850s, the relationship between cystoid macular oedema and diabetes mellitus (DM) would remain disputed until Edward Nettleship published histopathological evidence of diabetic macular oedema in 1872. 1–3 Whereas the original descriptions by Jaeger were documented by ocular fundus paintings, Jackman and Webster were the first to publish a human retinal photograph nearly 30 years later in 1886. 4,5 In 1968, the Airlie House classification of diabetic retinopathy (DR) established the first standards defining formal quantitative and qualitative features of disease revealed through standardised stereoscopic color
fundus photography (CFP), as will be further explained in the review of imaging modalities.\textsuperscript{6,7} This provided the basic foundation for subsequent landmark studies as the Diabetic Retinopathy Study (DRS) and Early Treatment for Diabetic Retinopathy Study (ETDRS).\textsuperscript{8}

DM is a global pandemic, estimated to affect 642 million people in 2040. A third of these will suffer from some degree of retinopathy.\textsuperscript{9,10} The impacts of DR on health-related quality of life, psychological well being and visual functioning are well established.\textsuperscript{11–13} The rationale for DR screening is based on the rules set forth by the WHO for screening in medicine by Wilson and Junger in 1968, including among others that DR is an important public health problem, with suitable testing, clear benefits of early treatment, and cost-efficiency.\textsuperscript{14} Hence, systematic DR screening programs in Iceland, the United Kingdom (UK), Singapore and Ireland as well as regional programs in the US and various European countries have been established.\textsuperscript{15–20}

In this article, we evaluate the features and health economic impacts of different DR screening programs based on retinal photography.

**History**

Iceland, being cited as a pioneer of DR screening, started a screening program for patients with insulin-dependent diabetes in 1980.\textsuperscript{19–21} Screening was undertaken by an ophthalmologist in conjunction with the annual acquisition of CFP. Regional programs in Sweden were also among the first established screening programs in this era, systematically screening all DM patients in a tertiary center in Lund with a follow-up of 5 years.\textsuperscript{22} In Stockholm, Sweden’s capital, with more than a million inhabitants, a screening program started in 1990 based on four field 45° CFP in more than 100 primary healthcare centers.\textsuperscript{23} This program was based on a national training program of doctors and nurses, which had started 10 years earlier and had led to improved DM control and awareness. Analysing referral letters for patients with DM and severe sight loss, Bäcklund et al could show that new blindness in DM was reduced by more than 30% following the implementation of DR screening.\textsuperscript{23} A consensus statement by the British Diabetes Association declared that DR screening should be based on methods that can provide a minimum sensitivity of 80% and 95% of specificity for referable DR.\textsuperscript{24} The UK is the largest country to have adopted nationwide screening for all patients with DM above the age of 12 years based on telemedicine. Screening commenced in 2003, reaching national coverage in 2008. The four constituent countries, England, Scotland, Wales and Northern Ireland provide screening programs through an asynchronous store-and-forward telemedicine approach. Dilated CFP are acquired in a community setting by trained technicians and images are assessed using hierarchical graders with subsequent referral to the hospital eye services of the National Health Service (NHS) when indicated.\textsuperscript{25} This program has an uptake of above 80% of the diabetic population, having screened more than 2,250,000 patients with DM from 2016 to 2017 in England alone.\textsuperscript{26} The value of this effort was recognised in 2010 when, for the first time in 50 years, DR was no longer the primary cause for blindness certification in the working-age population of the UK.\textsuperscript{27} Singapore was the latest nation to establish nationwide telemedicine-based DR screening. Whereas previously patients had attended ad hoc screening in the primary care clinics for CFP, now standardised two field 45° CFPs are taken by trained nurses and directly sent to a reading center, providing same-day reports and, if necessary, referrals to hospital eye services.\textsuperscript{28} Screening for DR is not limited to the nationwide programs covered in this review; many countries have adopted opportunistic screening contingent on national guidelines. As these programs, however, are neither structured nor audited, and therefore are not comparable due to the lack of publicly available data, they will not be covered in this review. There are multiple examples of regional telemedicine screenings, often covering rural areas in Australia, the US, Canada, Europe, and India.\textsuperscript{18,29–38} A graphical overview of population and selected regional DR screening programs, their implementation and outcome reporting, precisely instant or synchronous versus asynchronous store-and-forward reporting is represented in Table 1 and displayed in Figure 1.

There is evidence that patient acceptance of telemedicine services may be superior to traditional face-to-face consultations. When conducting a randomised trial comparing telemedicine screening to direct face-to-face screening in the US, Mansberger et al demonstrated that attendance was higher when DR screening was based on a telemedicine-based methodology. Interestingly, the authors were able to show that when both groups were combined to be screened via telemedicine, the attendance proportions in both groups were comparable, showing that the acceptance of telemedicine screening is higher compared to face-to-face examinations.\textsuperscript{39}
It seems conceivable that increased attendance and registration of large screening programs, therefore, is best achieved via telemedical screening programs.\textsuperscript{40}

**Different Imaging Approaches**

The Arlie House Symposium in 1968, an international meeting of leaders in ophthalmology, internal medicine and neurosurgery sought to classify and quantify the emerging global problem of DR. Consensus of participants was established on the definition, assessment and recording of disease features identified through standardised fundus photography.\textsuperscript{8}

Only three years later, a modified version of the Arlie House Classification was used for grading images based on the seven field 30° stereoscopic CFP in the first landmark trial comparing the impact of panretinal photocoagulation or observation on severe vision loss in DR.\textsuperscript{41} Over a 5-year study period, scatter photocoagulation treatment reduced the risk of severe vision loss by around 50% compared to observation, and showed a reduction of DR progression.\textsuperscript{42} The ETDRS study, another landmark trial, based its assessment on the Arlie House Classification using the seven field 30° stereoscopic CFP protocol.\textsuperscript{43} The investigators however elaborated further on the grading criteria by introducing additional granular features, such as microaneurysms and intraretinal microvascular abnormalities. Indeed, this detailed grading system has remained in use for many decades and is still considered to be the gold standard. Whereas this modality has proven to be of great value for research, it is not feasible in a screening setting.

When comparing single, two field and three field 45° CFP with the ETDRS gold standard, differences in the sensitivity and specificity have been found. The sensitivity and specificity for the detection of referrable disease was highest in three field CFP (92% and 96%), followed by two field CFP (96% and 89%) and was lowest in single-field CFP (78% and 86%),\textsuperscript{44–47} The sensitivity and specificity of two mydriatic 45° CFP, the first centered at the macula and the second on the optic disc, has been shown to be 80.2% and 96.2%, respectively.\textsuperscript{48} Despite the superior diagnostic accuracy of multiple field CFP, acquisition needs to be balanced with the greater focus on patient collaboration, resource and training. Accordingly, most screening services have adopted a two field CFP strategy as a compromise.

Due to the technological advances of digital photography and its achievements regarding resolution, the UK DR screening programs initially utilized 45° digital CFP with non-mydriatic fundus cameras on dilated patients.\textsuperscript{25} The cameras must provide a resolution of at least 30 pixels per degree on a photograph of 45° in width and 40° in height and be able to accommodate ±15 D of refractive error.\textsuperscript{25} Agreement between CFP on 35 mm

**Table 1 Selected Population Screenings and Selected Regional Screening Programs**

<table>
<thead>
<tr>
<th>Name</th>
<th>Starting Date (and End)</th>
<th>(n) Screened Per Annum</th>
<th>Country (r)</th>
<th>(i) National Reporting</th>
<th>(a) Synchronous Reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHS DESP\textsuperscript{151}</td>
<td>2003</td>
<td>2,250,000</td>
<td>England</td>
<td>n</td>
<td>a</td>
</tr>
<tr>
<td>NHS Scotland DRS\textsuperscript{152}</td>
<td>2003</td>
<td>264,000</td>
<td>Scotland</td>
<td>n</td>
<td>a</td>
</tr>
<tr>
<td>DESW\textsuperscript{153}</td>
<td>2003</td>
<td>170,000</td>
<td>Wales</td>
<td>n</td>
<td>a</td>
</tr>
<tr>
<td>NIDESP\textsuperscript{154}</td>
<td>2008</td>
<td>66,000</td>
<td>Northern Ireland</td>
<td>n</td>
<td>a</td>
</tr>
<tr>
<td>SiDRP\textsuperscript{155}</td>
<td>2010</td>
<td>600,000</td>
<td>Singapore</td>
<td>n</td>
<td>i</td>
</tr>
<tr>
<td>Iceland\textsuperscript{156}</td>
<td>1980</td>
<td>5000</td>
<td>Iceland</td>
<td>n</td>
<td>i</td>
</tr>
<tr>
<td>Aboriginal controlled community</td>
<td>2000–2004</td>
<td>1138 Aboriginals and 271 non</td>
<td>Australia</td>
<td>r</td>
<td>a</td>
</tr>
<tr>
<td>Health services (ACCHS)\textsuperscript{157}</td>
<td></td>
<td>38,596 over 5 years (7719 p/a)</td>
<td>France</td>
<td>r</td>
<td>a</td>
</tr>
<tr>
<td>Ophdiat\textsuperscript{158}</td>
<td>2004–2009</td>
<td>17,471 over 7 years (2496 p/a)</td>
<td>Finland</td>
<td>r</td>
<td>a</td>
</tr>
<tr>
<td>South-Ostrobothnia digital mobile screening\textsuperscript{31}</td>
<td>2006</td>
<td>140000\textsuperscript{160}</td>
<td>USA</td>
<td>r</td>
<td>i</td>
</tr>
<tr>
<td>EyePACS\textsuperscript{159}</td>
<td>2007</td>
<td>42,000 (in 2015)</td>
<td>Spain</td>
<td>r</td>
<td>a</td>
</tr>
</tbody>
</table>

**Abbreviations:** NHS, National Health Service; DESP, diabetic eye screening program; DRS, diabetic retinopathy screening; DESW, Diabetic Eye Screening Wales; NIDESP, Northern Ireland diabetic eye screening program; SiDRP, Singapore diabetic retinopathy screening program.
Figure 1 Timeline bubble chart displaying selected regional screenings with either synchronous (green) or asynchronous (blue) reporting of screening outcomes. The bubble diameter is an approximation of the number of screened patients with diabetes per annum (yellow), references can be found in Table 1.
film and digital CFP has been shown to be substantial to almost perfect regarding DR severity levels as well as moderate to substantial for diabetic macular edema severity levels.\(^4^9\) In a systematic review assessing the detection of any level of DR with digital retinal imaging, Piyasena et al could show that both non-mydriatic and mydriatic fundus imaging delivered satisfactory levels of sensitivity with 86% after exclusion of ungradable images. The proportion of ungradable images in non-mydriatic settings was 18.6% compared to 6.2% in mydriatic settings.\(^4^7\) With respect to the heterogeneity of the studies the authors concluded, that specificity results would depend on whether ungradable images were counted as test positive and causing difficulties in comparing specificity and proposing standardised reporting of ungradable images. In a randomised trial, Scanlon et al found age to be the strongest predictor of ungradable images, mostly related to cataracts and smaller pupil size.\(^5^0\) As the application of mydriatic eye drops also has potential side effects, such as mydriasis-induced acute angle closure with an incidence of 6 in 20,000 in a Caucasian population, a fine balance between image quality, gradability and patient risk needs to be emphasised.\(^5^1\) The Scottish DR screening is based on a single-field macular centered 45° digital CFP, only using mydriatic eye drops if the image quality is poor.\(^5^2\)

The approach of two 45° digital CFP is also used in the first autonomous artificial intelligence (AI)-based device with FDA approval.\(^5^3\) In contrast to the UK, where trained technicians conduct image acquisition, for this study, a device was introduced in combination with a non-mydriatic fundus camera and AI system operator with no prior experience of fundus imaging and a standardised 4-hour training program. The AI would detect the image quality and operators would improve quality and/or perform pupil dilation for the patients if necessary. The algorithm succeeded the preset superiority endpoints for detecting referrable, in this study defined as more-than-mild DR (ETDRS level ≥ 35) with a sensitivity of 87% and a specificity of 90.7% as well as imageability rate of 96.1% hence becoming the first FDA autonomous artificial intelligence device in any field of medicine.\(^5^3\)

Another increasingly popular method to conduct color fundus imaging without pupil dilation is confocal scanning ophthalmoscopy (cSO). Beams emitted by lasers or light-emitting diodes (LED) of different wavelengths scan the retina generating a composite pseudocolor image with varying fields of view. Optos (Optos plc, Scotland), a Scottish company launching their first device 25 years ago, provides various camera models consisting of a 200-degree ultra-widefield based on three lasers with different wavelengths, namely blue (488 nm), green (532 nm) and red (635 nm).
The potential benefits of incorporating ultra-widefield imaging were shown in a screening setting, where the Optos revealed peripheral features of the disease in 20% of patients without discernible DR. The Clarus ultra-widefield camera by Carl Zeiss Meditec is able to capture a single-field image of 133°, which can be montaged to 200° on two fields with LED-based cSO. Another example is the multicolor scanning laser imaging module by Heidelberg Engineering, where a special lens can be mounted onto the OCT/cSO camera, providing a 105° pseudocolor image. Centervue Eido combines a confocal scanning laser ophthalmoscopy technique with a confocal white light imaging to obtain true color CFP covering 60° per image. These technical advances may be of great interest even in a screening setting due to the advances of non-mydriatic imaging; however, further studies are warranted.

However, the aforementioned devices enabling DR screening may be costly and require technical support and infrastructure limiting their use in less affluent settings, such as the developing world. One potential option is CFP acquisition using smartphone-based camera systems in combination with AI. The Remidio Fundus on Phone camera system consists of a slit lamp-based non-mydriatic smartphone camera, which provides 45° images to detect DR and was validated against high-end fundus cameras, showing on-par performance for any DR and referable DR, thus having received FDA approval.

Automated Image Analysis: Before and After AI

Telemedical assessment of CFP for DR screening requires a large amount of trained graders in reading centers to cope with the rising volume of images. For this and other reasons, the prospect of automated analysis of retinal images has garnered interest from ophthalmologists and computer scientists alike. The first publication analysing computer-aided detection (CADe) in ophthalmology was published in 1973, focusing on the detection of contour lines of retinal vessel images. In 1984, a French group was the first to describe microaneurysms in fluorescein angiography. The development of digital photography and technical advances in computing power led to computer-aided diagnosis (CADx) systems, which not only incorporate qualitative detection of features but also enable calculation of disease probability and risk stratification. To provide guidance, the American Telemedicine Association has therefore released statements on the quality, namely to validate against the gold standard ETDRS 30° seven-field stereoscopic CFP.

Traditional machine learning techniques for image analysis were based on thresholding, edge detection, processing and filters to detect disease features with more advanced models leveraging ensemble-based approaches, multi lesion approach or content-based image retrieval. Already in 2008, Abramoff et al published the results of an automated retinal image analysis system (ARIAS), reporting 84% sensitivity and 64% specificity for the detection of referable DR in a retrospective analysis of 7689 CFP. In 2013, the same group reported the results of the Iowa Detection Program (IDP) for detection of referable DR and reached 96.8% sensitivity and 59.4% specificity. Other examples of ARIAS based on using classic machine learning techniques have shown to provide image grades without the help of human graders. Tufail et al conducted a study in 2013 to assess the safety and cost-efficiency of three ARIAS, namely iGradingM, Retmarker and EyeArt with images from the UK National Health Service Diabetic Eye Screening Program (NHS DESP), comparing it to manual grading. They showed that EyeArt would provide a 94.7% sensitivity for any DR, 93.8% for referable DR and 99.6% for PDR. Retmarker would achieve 73.0% sensitivity for any DR, 85% for referable DR and 97.9% for PDR. iGradingM classified all images as either showing disease or being ungradable; hence, iGradingM could not be included in further analysis. Cost efficiency compared to human graders was shown for Retmarker and EyeArt. iGradingM is an ARIAS that provides binary grades for disease/no disease as well as for image quality. It has been used as a level 1 grading in the Scottish DR screening program after extensive validation since 2010. In a retrospective evaluation, the rate of DR was 6.6% with a sensitivity of 97.8% for referable DR, the specificity was calculated as 41.2%. The above mentioned traditional machine learning techniques have been reviewed in detail elsewhere, their relevance today has, however, been surpassed by the recent developments in deep learning (DL).

Although the principles of DL were described decades ago, only recently has its potential to healthcare applications become popular. The main reasons are due to improvements in computing power, especially by graphical processor units (GPU), large quantities of digital data and publically available pre-trained convolutional neural network (CNN) models like AlexNet, VGGNet or GoogleNet. DL models have outperformed traditional
methods in speech recognition, computer vision, and have also been shown to be equivalent to humans in disease detection from medical imaging. One of the main differences of CNNs compared to traditional techniques is that raw data are processed by identifying features of interest through extensive training.

With the vast amount of data collected from screening services, DL development may be ideally afforded by DR screening. The group of Abramoff et al, who had extensively reported on automated image analysis systems mentioned previously, enhanced their algorithm by incorporating DL, showing an increased, although not statistically significant sensitivity of 96.8% for referable disease; however, they reported a steep increase from 59.4% to 87% specificity validating on the same publicly available MESSIDOR-2 dataset with an area under the receiver operator curve (AUC) of 0.98 in the detection of referable disease. The algorithm, now called IDx-DR, has also been validated on a Dutch dataset labelled by three retina specialists using the International classification of diabetic retinopathy (ICDR) and the EUODIAB grading. The reported sensitivity and specificity using ICDR grades were 68% and 86%, respectively, and 91% and 84% using the EUODIAB grades. A key evaluation aspect in prediction models validation is that of the prospective setting. The same group subsequently published results from a prospective trial in a primary care setting, exceeding the pre-specified primary endpoint goals of a >85% sensitivity and >82.5% specificity with a sensitivity/specificity of 87.2%/90.7% and an imageability rate of 96.1%, receiving FDA approval for diagnosing more than mild DR autonomously. It is conceivable that the metrics of the IDx-DR algorithm could represent benchmark metrics for future market authorisation from the FDA.

One concern of DL models, which are typically trained for the classification of a single disease, is their limited ability to generalise to unseen pathology. Although a given DL model may demonstrate clinical effectiveness in the detection of referable DR, what of the scenario where the image may demonstrate features of other diseases, such as glaucoma, which are known to occur with a higher incidence in people with DM? Ting et al have reported on a DL model to analyse DR, glaucoma and age-related macular degeneration (AMD) using CFP. This algorithm, now called SELENA+ was trained on almost 500,000 images, the authors reported a sensitivity of 90.5%, a specificity of 91.6% and an AUC of 0.936 for detecting referable disease and was tested on the Singapore Integrated DR programme (SiDRP). To test the algorithm in multi-ethnicity environments, it was externally validated on datasets from six different countries, including Singapore, Hong Kong, China, Australia, USA and Mexico, achieving an AUC of 0.889 to 0.983 in these datasets. Of importance, this algorithm could detect possible glaucoma with a sensitivity/specificity of 96.4%/87.2% and 93.2%/88.7% for AMD when compared to human graders. Only recently, the Singaporean Health Minister announced that this algorithm will be deployed across the Singapore DR Screening program by 2022.

In 2016, Gulshan et al published the results of a study sponsored by Google Inc, about DR detection by a CNN trained on about 128,000 CFP graded by 54 US licensed ophthalmologists and ophthalmology residents. Similar to IDx-DR the model could detect referable DR with an AUC of 0.991 and 0.990 for the EyePACS-1 and MESSIDOR-2 datasets, sensitivity/specificity for EyePACS-1 were 90.3%/98.5% and 87.0%/98.3% for MESSIDOR-2.

The EyeART system owned by Eyenuk is another example of a DL system used for DR grading, that has improved its performance by combining their traditional machine learning approach with CNN. EyeART v1, as aforementioned, has been evaluated in a UK-based study. EyeART v2.0, however, is a cloud-based technology and has been tested on >800,000 images of about 100,000 consecutive patients. It achieved a 91.3% sensitivity and 91.1% specificity for referable DR with an AUC of 0.965 compared to the EyePACS graders. Furthermore, the algorithm was tested on the above mentioned Remidio Fundus on phone - smartphone-based system. Albeit a small study with only 296 patients included, EyeART achieved a sensitivity/specificity of 95.8%/80.2% for detecting any DR and 99.1%/80.4% for detection of sight-threatening DR. Remidio, the company behind the smartphone-based CFP, has developed a DL system called MEDIOS AI. This offline system on the same smartphone, which takes the photograph, was trained to detect more-than-mild DR achieving an impressive performance of 100% sensitivity and 87% specificity in 231 patients compared to the grading from ophthalmologists using the same images. Further studies with larger cohorts are ongoing at multiple sites.

Another, also commercially available web-based solution is called RetCAD, owned by a dutch company named Thirona. This algorithm is trained to detect DR and AMD, and has been validated on the MESSIDOR dataset with 1200 images and the AREDS dataset for AMD with
The performance was
In a systematic review recently pub-
So
Almost 80% of respon-
The same group has also developed
Dove
Gulshan et al conducted a prospective vali-
The screening for DR with the
in this year.
To
and have
Poplin
Furthermore, the
The costs of DR have been
Deep
€
able in one dataset.
From CFP, however, were limited to the metadata from the
et al reported to have also trained models to detect HbA1c
Deep learning, of course, is not limited to detect DR, and,
depending on the labels of the images during the training
of the model, other outcomes can be evaluated. Poplin et al
reported in 2017 a model that was trained with the images
and labels of the UK biobank study for prediction of
cardiovascular risk factors from CFP using DL. The DL
system could detect age with a mean absolute error within
3.26 years, smoking status (AUC 0.71) and sex with an
astonishing AUC of 0.97, task retinologists were not able
to conquer in 150 years of looking at the fundus.102 Poplin
et al reported to have also trained models to detect HbA1c
from CFP, however, were limited to the metadata from
the datasets included in the study, as HbA1c was only avail-
able in one dataset.102 The same group has also developed
a model to detect anemia from CFP with an AUC of 0.89,
again training on images of the UK biobank study.103
Although only providing proof of concept and lacking
robust evaluation, these two examples show the potential
benefit of telemedicine screenings for systemic disease
evaluation in patients with DR. Oculomics, a term intro-
duced for describing ocular biomarkers for systemic dis-
ease, could be of great interest in these systematic and
ideally nationwide screenings.104 Clinicians should, however, be aware, that many of the
AI-related publications lack reporting standards.105 To
address this scarcity, the CONSORT-AI and SPIRIT-AI
steering groups are planning to extend the current state-
ment, complementary to current reporting standards as the
TRIPOD-ML initiative. It is also important to understand
the metrics of the compared methods, and to judge the
validation data sets on multiple ethnicities, image quality
as well as being aware of flaws in methodology.106–108
So far, only the IDx-DR device has been granted FDA
approval and although the promise of DL is big, the
clinical value has yet to be robustly evaluated. Few studies
in medical images follow a prospective and/or randomised
design, and are at higher risk for bias; data and code
availability are not given in most studies.

It is worth mentioning that these so-called narrow AI
models are trained for specific tasks, and therefore will not
be able to detect retinal comorbidities like choroidal mel-
anomas or retinal detachments. These limitations will
make human adjudication in ungradable CFPs for the
foreseeable future irreplaceable.

With respect to the global distribution of diabetes, it is
of paramount importance for the latter, that ARIAS are not
only validated in high-income countries to provide further
certainty before a potential implementation can be envi-
sioned. This was, for example, undertaken by the groups
of Abramoff et al with 6788 images from the Nakuru Eye
Study in Kenya, and Bellemo et al in Zambia, validating
their model on 76,370 images of 13,099 patients with
diabetes.109,110 Gulshan et al conducted a prospective vali-
dation of the above mentioned DL system developed by
Google inc., and evaluated the performance on data from
two eye care centers in India.82,111 The performance was
compared to human graders, in a binary fashion for detect-
ing more-than-mild DR (ETDRS level ≥35) or referable
diabetic macular edema. With an 88.9%/92.2% sensitivity/
specificity and an AUC of 0.963 in one center and a
92.1%/95.2% sensitivity/specificity and an AUC of
0.980 at a second center, the automated DR detection
was equal to or succeeded human grading.112

Health Economics
The global pandemic of DR requires efforts on different
levels, and epidemiological as well as technical stand-
points need to be considered.113 Almost 80% of respon-
ders in the DR Barometer study replied that the sight
impairment due to their DR and diabetic macular edema
make daily life activities including work difficult and
sometimes impossible.114 The costs of DR have been
shown to rise with the progression of the disease as pub-
lished in a study by Woung et al in Taiwan,115 and have
been calculated to account for 1.5% of the entire health-
care costs in Germany in 2002, estimated to reach 3.51
Billion € in this year.116 The screening for DR with the
aim of early detection and timely as well as appropriate
referral according to the guidelines has been shown to be
cost-effective.117,118 In a systematic review recently pub-
lished, Lee et al could show, that teleophthalmology is the
most cost-effective intervention among all telemedicine
strategies in diabetes management.119 Furthermore, the
implementation of telemedicine screening when compared
to family physician-based screening, both based on fundus imaging has also been shown to be cost-effective. The authors concluded, that firstly the costs of grading in a central reading center are lower, and secondly due to higher specificity on the grades less unnecessary referrals were generated. The US Department of Veteran Affairs concluded in a retrospective study that in order to be cost-effective, a DR screening service has to include more than 3500 patients, or patients, who are less than 80 years of age. However, offering annual screening to all patients with DM might not be cost-effective, as the number of patients with DM is increasing. The risk for patients with DM and no evidence of DR to develop sight-threatening DR over a period of 2 years are relatively low, independent of single-field or two field CFP screenings. This finding was supported by the analysis of 116,134 patients undergoing DR screening in the US by Kaiser Permanente, the largest private healthcare provider in the US, showing that patients with no or mild DR rarely required retinal intervention in the 2 years after retinal screening. By stratifying patients into low risk and high-risk groups due to their repeated findings in the screening services, and subsequent adaptation of the screening intervals, cost reduction could be achieved. It is, however, unclear how patient behaviour would change under extended screening intervals, giving the patient a wrong impression of the potential damage that can be caused by DR.

The UK national diabetic eye screening program (NDESP) is by far the largest existing program, having screened 2,257,124 out of 3,175,121 eligible patients alone between 2016 and 2017. These images were all graded by trained graders, undergoing thorough training to be eligible, with various levels of adjudication. In 2011 the total amount of workload for the human graders in the NDESP was calculated to be greater than 300,000 hours per year for retinal images of roughly 1,700,000 patients. This labour-intensive work also contains further potential for reduction of cost. Strategies to improve the cost-effectiveness could include, of course, automated image processing using ARIAS. Tufail et al presented two strategies for calculating cost-effectiveness of two ARIAS, Retmarker and EyeArt. The grading system in the NDESP contains three levels of graders, with adjudication for different levels of DR. In strategy 1, the ARIAS would replace the level 1 grader, in strategy 2 the ARIAS would act as a filter before the level 1 grader. The authors concluded that, if properly implemented, ARIAS could help the impending challenge of DR in both developed and developing countries. In a preprint issued by the Lancet, Xie et al have conducted an economic analysis comparing a semi-automated DL system, a fully automated DL system and the current full human grading model of the SiDRP in Singapore. They could show that the semi-automated DL system, which would function as a triage prior to secondary human assessment is equally effective as full human assessment, but less cost-intensive. Further assessments of the costs and requirements of transition processes from human-based screening to AI-assisted programs as well as from opportunistic to population-based programs are warranted.

Implications for the Future
Most patients with diabetes are not aware of their current status of DR. The National Health and Nutrition Examination Survey (NHANES) in the US, the Andhra Pradesh Eye Disease Study (APEDS) in India and the Singapore Epidemiology of Eye Disease (SEED) Study all found that more than 70% of the patients with diabetes assessed in these studies were not aware of their own DR status. Due to an increasing global population, the majority of patients with diabetes now live in low- and moderate-income countries, with potentially underfunded healthcare systems.

The International Diabetes Federation conducted the DR barometer study, with the aim to assess the awareness of DR and the access to care. This study of 2329 healthcare professionals from 41 countries reported several interesting insights: more than a third of diabetes specialists included in the survey reported that they would not discuss eye care with their patients with diabetes. Sufficient information on eye complications was only available of one in five primary care providers. The most substantial barriers to care according to the healthcare professionals were patients’ lack of knowledge or awareness of eye complications in 43%, the lack of importance given to eye examinations by patients in 33% and the high cost of care in 32%. Ophthalmologists included would complain about late screening in 66% and a lack of patient education material in 55%. These insights have to be interpreted critically due to the way this study was conducted, and may only give a snapshot of today’s situation.

There is undeniably a need for a global strategy to tackle the burden of DR. Patient education will be as important as raising awareness from the levels of primary care all the way through to tertiary care. Technical
advances, as well as lessons learnt from large screening services, may lead the way. Initiatives, such as the Diabetic Retinopathy Network (DR-NET), supported by the VISION 2020 LINKS Programme, provide opportunities for knowledge exchange on the construction of DR screening programmes, best practice on the development of national frameworks and procedures on governance and audit. Another consideration is the advances in technology, particularly in regards to the use of digital cameras as well as cloud-based and offline screening algorithms. It is estimated that 80% of the population of sub-Saharan Africa have a mobile device, and eHealth applications and screening programs could play an important role as healthcare devices.135

Deep-learning systems need to go through robust validation processes in multiethnic cohorts to avoid biases; the ownership of data on a global level is still not regulated.136 The European Union’s General Data Protection Regulation or California’s Consumer Privacy Act are examples of recent regulations to apply rules for data transfer and collection, and may be first steps in the right direction of data protection in a fast-growing online world.137,138 An interesting proposal about the ethics of sharing and using clinical imaging data for AI was put forward by Larson et al, stating that rather than discussing the ownership of data between patients and provider organisations, they propose that clinical data would not be owned at all in a traditional sense.139 Larson et al argue that everybody involved in dealing with the data has an obligation to ensure that the data would be used for the benefit of future patients and society. It is also important to bear in mind that by installing a screening service into an already stretched system with little resources for treatment, strategies from high resource countries will not work. Thus, for instance, criteria for referable disease could be adapted to focus on those being affected the most.140

Frequently asked questions about these DL systems almost always include the following three: will AI replace ophthalmologists, how can we open the black box, and can these algorithms be sued? The first question was addressed in a recent article by Korot et al, stating that although current AI systems may be good at dealing with high volume data related tasks; however, a clinician’s ability to interpret the complex and multivariate data-driven AI recommendations should be the next step for ophthalmologists.140 The black box problem refers to the interpretability of AI systems, and those end-to-end solutions will not provide insight to the clinicians regarding.141

As these DL systems are providing interesting results, clinicians as well as public health providers would be interested to be able to interpret, and scale the newly discovered insights, for example, individually stratifying certain risk factors. Current strategies of using again AI to analyse these black boxes by using reinforcement learning are providing a prosperous outlook, and hopefully, these black boxes can be turned into white boxes soon.142–144 So, what about the legal aspects of AI, can an algorithm be sued?145 The answer to this question will depend a lot on the regulators and the robust clinical evaluation prior to the approval.

As the current Covid19 pandemic causing global turmoil in 2020 caused by SarsCoV-2 is affecting all continents with an unclear outcome, telemedicine has arisen to the forefront of transforming healthcare delivery models.146,147 In Ophthalmology, the decoupling of clinical data gathering using virtual clinics and explanation of results using video consultations with treatments such as intravitreal injections for diabetic macular oedema will allow the sustained provision of eye care in the new era of social distancing. Not only have virtual clinics proven to be cost-effective and safe, but patients’ acceptance is also high.148,149 Once validated, simple machine learning techniques could support fully automated clinics supporting clinicians for instance by detecting non-glaucomatous visual field defects in virtual glaucoma clinics, hence enhancing patient safety.150 This model not only future proofs eye care provision but provides resilience should a second or third wave pandemic arise. Research in the patient acceptance, efficacy, safety, and cost-efficiency of these pathways will likely, in the coming years, dominate the implementation research fields of Ophthalmology.

Conclusions

The growth in global incident DM rates secondary to ageing populations and rising obesity levels are well recognised. While many parts of the developed world benefit from effective infrastructure for DR screening, there remains a scarcity of similar programmes in the developing world, where over 60% of the global DM burden is thought to exist. Fortunately, novel approaches in hardware and image analysis may afford the opportunities to address this unmet need. The latest imaging devices provide higher resolution and peripheral views previously restricted to clinical examination and the unprecedented technological environment has fostered a vibrant sphere of AI research in DR detection.
these efforts are beginning to translate into clinical benefit with the approval of some of the first autonomous AI-based DR detection tools. Hardware bears important financial considerations however the permeation of mobile smartphones in the developing world may provide a window of opportunity. Moreover, enhancing global advocacy through partnerships, such as that epitomised by the Vision 2020 LINKS programme, has helped establish robust frameworks for the implementation of DR screening in the developing world. By either engaging with clinicians through telemedicine platforms or cloud-based upload to an AI model, there are indications that we may yet be able to democratize the power of effective DR screening on a global scale.

Acknowledgments
The authors would like to thank Lauren Stafford of Big Picture Medical for the graphical design.

Disclosure
The authors report no conflict of interest in this work.

References


