Ophthalmic
Research

Research Article

Ophthalmic Res 2023;66:1286–1292
DOI: 10.1159/000534098

Received: February 17, 2023
Accepted: September 8, 2023
Published online: September 27, 2023

# Diagnostic Accuracy of Automated Diabetic Retinopathy Image Assessment Softwares: IDx-DR and Medios Artificial Intelligence

Andrzej Grzybowski[a] Divya Parthasarathy Rao[b] Piotr Brona[c]
Kalpa Negiloni[d] Tomasz Krzywicki[e] Florian M Savoy[f]

[a]Institute for Research in Ophthalmology, Foundation for Ophthalmology Development, Poznan, Poland; [b]Department of AI R & D, Remidio Innovative Solutions Inc., Glen Allen, VA, USA; [c]Department of Ophthalmology, Poznan City Hospital, Poznan, Poland; [d]Department of Clinical Research, Remidio Innovative Solutions Pvt Ltd, Bangalore, India; [e]Department of Mathematical Methods of Informatics, University of Warmia and Mazury, Olsztyn, Poland; [f]Department of AI R & D, Medios Technologies, Remidio Innovative Solutions, Singapore, Singapore

**Keywords**

Diabetic retinopathy · Artificial intelligence · Medios · IDx-DR

**Abstract**

***Introduction:*** Numerous studies have demonstrated the use of artificial intelligence (AI) for early detection of referable diabetic retinopathy (RDR). A direct comparison of these multiple automated diabetic retinopathy (DR) image assessment softwares (ARIAs) is, however, challenging. We retrospectively compared the performance of two modern ARIAs, IDx-DR and Medios AI. ***Methods:*** In this retrospective-comparative study, retinal images with sufficient image quality were run on both ARIAs. They were captured in 811 consecutive patients with diabetes visiting diabetic clinics in Poland. For each patient, four non-mydriatic images, 45° field of view, i.e., two sets of one optic disc and one macula-centered image using Topcon NW400 were captured. Images were manually graded for severity of DR as no DR, any DR (mild non-proliferative diabetic retinopathy [NPDR] or more severe disease), RDR (moderate NPDR or more severe disease and/or clinically significant diabetic macular edema [CSDME]), or sight-threatening DR (severe NPDR or more severe disease and/or CSDME) by certified graders. The ARIA output was compared to manual consensus image grading (reference standard). ***Results:*** On 807 patients, based on consensus grading, there was no evidence of DR in 543 patients (67%). Any DR was seen in 264 (33%) patients, of which 174 (22%) were RDR and 41 (5%) were sight-threatening DR. The sensitivity of detecting RDR against reference standard grading was 95% (95% CI: 91, 98%) and the specificity was 80% (95% CI: 77, 83%) for Medios AI. They were 99% (95% CI: 96, 100%) and 68% (95% CI: 64, 72%) for IDx-DR, respectively. ***Conclusion:*** Both the ARIAs achieved satisfactory accuracy, with few false negatives. Although false-positive results generate additional costs and workload, missed cases raise the most concern whenever automated screening is debated.

## Introduction

Diabetes is a global epidemic and one of the world's fastest-growing diseases. The number of patients with diabetic retinopathy (DR) and sight-threatening DR (STDR)

karger@karger.com
www.karger.com/ore

Karger
OPEN ACCESS

Correspondence to:
Divya Parthasarathy Rao, drdivya@remidio.com

is also expected to rise. There are only a few established nationwide DR screening programs, and overall DR screening services remain inadequate in most of the developing world and even some developed countries [1]. This is further compounded by the increasing resources needed for the implementation and maintenance of comprehensive DR screening programs [2].

One of the proposed solutions to this global issue is the use of automated diabetic retinopathy image assessment software (ARIA) to grade fundus images instead of or alongside human graders. There are multiple ARIAs currently available with many more being developed worldwide [1]. Although there is an abundance of studies looking into the performance of a single ARIA, studies comparing multiple ARIAs are currently rare, as direct comparison is often difficult [3]. Based on previous studies, it is clear that the performance of even state-of-the-art algorithms may vary considerably [3, 4]. We set out to analyze the performance of two modern ARIAs, IDx-DR and Medios Artificial Intelligence (AI).

## Materials and Methods

### *Study Design*

In this retrospective-comparative study, the performance of two different ARIAs in screening for DR was compared to human graders (reference standard). The screening for DR was conducted and retinal images were obtained from diabetic clinics in Poznan, Poland, between March 2020 and April 2021. The Institutional Review Board of the Foundation for Ophthalmology Development, Poznan, Poland approved the project (Application No. 2/2022) and waived the need for IRB approval and written informed consent from the participants for this retrospective study. The study was in adherence to the tenets of the Declaration of Helsinki. All the extracted images were anonymized, and no change in the clinical pathway was anticipated.

The primary outcome of the study was to assess the sensitivity and specificity of ARIAS in detecting referable diabetic retinopathy (RDR). The secondary outcomes were to assess the positive (PPV) and negative predictive values (NPV) of ARIAS to detect RDR and to assess the sensitivity of ARIAS in detecting STDR.

### *Sample Size*

Using an alpha error of 0.05, a precision rate of 10% (two-sided), an estimated sensitivity of 85%, and an estimated incidence of RDR (International Clinical Diabetic Retinopathy [ICDR] – moderate non-proliferative DR [NPDR] and/or presence of clinically significant diabetic macular edema [CSDME]) to be 7%, the sample size calculated was 700 participants. Given these assumptions and expecting that 10% of subjects may be qualified as insufficient quality, a sample size of 800 subjects was chosen.

### *Inclusion and Exclusion Criteria*

The retinal images of subjects with established diabetes mellitus that were captured at the time of DR screening were included. Those that did not have at least one disc and one macula-centered image of sufficient quality were excluded from the study. Additionally, subjects who received treatment for DR (lasers or intraocular injections) were excluded.

### *Retinal Image Acquisition*

The screening process involved undilated fundus images captured using a Topcon camera NW400 by trained operators who followed a specific imaging protocol. For each patient, a total of four images (45° field of view each) were captured. They included one image centered on the optic disc and one centered on the macula for each eye. Additional images were taken to ensure sufficient quality. Retinal images were obtained from 811 consecutive patients with established diabetes mellitus who underwent screening for DR. Images deemed of sufficient quality graded by the IDx-DR AI software were selected. A total of 3,200 sufficient quality images from 811 patients were used for the study.

### *Reference Standard Grading*

The patients with images of sufficient quality were split into two sub-datasets. 362 patients were graded by three Polish retina specialists and 491 by three certified graders in India. All the graders are masked to the output of the AI and to each other's grading. Images were graded for severity of DR based on ICDR severity classification as no DR, mild NPDR, moderate NPDR, severe NPDR, and proliferative diabetic retinopathy (PDR). Macular edema was determined by the presence of surrogate markers like hard exudates. If hard exudates were found within 1 DD of the fovea, macular edema was determined as significant and labeled as CSDME present. Image grading was done on a per-eye basis. The final diagnosis for each patient was determined by the stage of DR of the more affected eye. Consensus image grading was regarded as the final reference standard based on Polish and Indian graders for the comparison of both AI systems. All the analysis was performed at the patient level.

### *Definitions*

RDR was defined as moderate NPDR and more severe disease (moderate NPDR, severe NPDR, PDR) and/or the presence of CSDME. STDR was defined as severe NPDR and more severe disease (severe NPDR, PDR), and/or the presence of CSDME.

### *AI Analysis Using Automated Grading Systems*

We used two different ARIAs, i.e., Medios AI for DR (Medios Technologies, Remidio Innovative Solutions, Singapore) and IDx-DR (Digital Diagnostics, IA, USA). The retinal images were run on both the ARIAs to screen for DR. Both the systems processed images deemed to be of sufficient quality by the IDx-DR system. IDx-DR results were recorded during live screening and all images captured for the patient were analyzed on a per-patient basis. Two images per eye that passed the AI quality check were submitted to the AI for DR analysis. For Medios AI analysis, anonymized images for each patient were securely transferred to a cloud platform, and the images were analyzed on an automated script version of the AI on a server instead of a manual analysis through the standard iPhone app deployment.

Both the AI systems are based on convolutional neural networks, with the Medios system being based on the Inception-V3 architecture. A detailed description of the model is provided in the

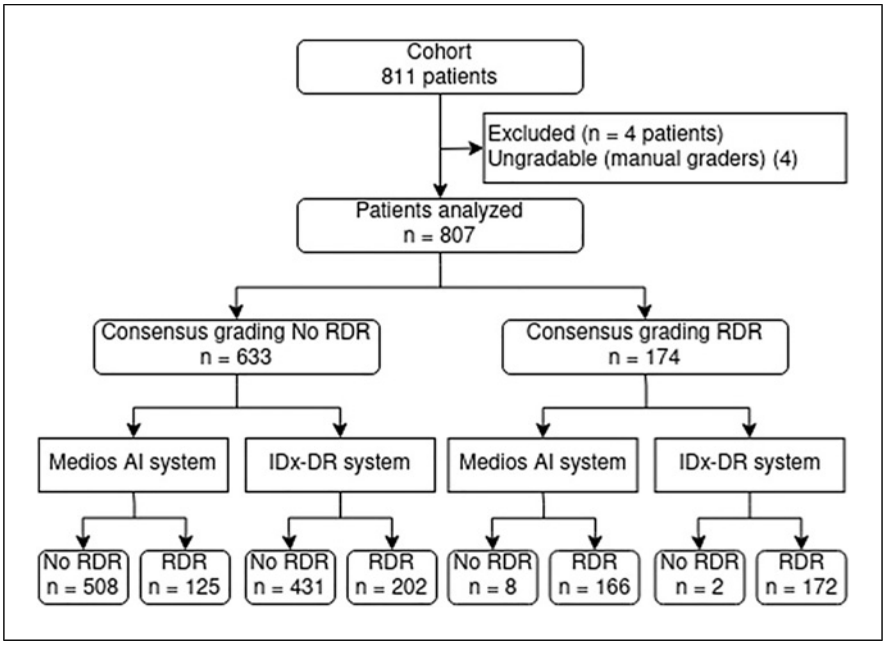

**Fig. 1.** STARD flow diagram showing the patient breakdown for the referable diabetic retinopathy (RDR) analysis.

literature [5]. In brief, the Medios AI algorithm evaluated two possible outputs: "no signs of DR detected" (non-RDR) and "signs of DR detected" (RDR). Report was generated on a per-patient basis. The IDx-DR system also has an image quality and a diagnostic algorithm. The IDx-DR system outputs the stage of DR and generates a per-patient report.

*Statistical Analysis*

All data were stored in Microsoft Excel sheets and Apache Parquet files and were analyzed using R and Python programming languages along with NumPy, Pandas, Scikit-learn, and SciPy libraries. The diagnosis of the AI using Medios and IDx-DR AI systems was tabulated against the consensus image diagnosis (reference standard) by constructing 2 × 2 tables. The sensitivity, specificity, PPV, and NPV with 95% CIs were calculated. Inter-rater agreement for Polish and Indian graders was measured by calculating the kappa statistic.

## Results

The study included the images of 811 patients. Image quality analysis was evaluated as part of the clinical workflow using the IDx-DR AI system. Four patients deemed ungradable by the graders were excluded. In total, 807 patients were included for further analysis. An additional 2 patients were removed from STDR analysis as they did not have a consensus for a STDR diagnosis despite being labeled as RDR by consensus. Figure 1 presents the STARD diagram of retinal image selection in the study.

Grades from the Polish and Indian graders were converted to no DR, any DR, RDR, and STDR before computing consensus. Based on consensus grading, there was no evidence of DR in 543 patients (67%). Any DR was seen in 264 (33%), of which 174 (22%) were RDR and 41 (5%) were STDR. The inter-rater agreement (Cohen's kappa) for Poland graders was 0.679 (ophthalmologist 1), 0.904 (ophthalmologist 2), and 0.848 (ophthalmologist 3). For the Indian graders, kappa was 0.632 (ophthalmologist 1), 0.916 (ophthalmologist 2), and 0.87 (ophthalmologist 3).

IDx-DR AI system gives an output at a stage level. 567 patients (567/807; 70.3%) were flagged positive for DR (mild, moderate, or threaten). 374 of them were also categorized as RDR (46.3%, moderate or threaten) and 189 as STDR (23.4%, threaten). Out of 174 patients with ground truth labeled as RDR, the IDx-DR system detected 172. This translates to a sensitivity of 99% (95% CI: 96, 100%). The Medios AI gives a binary output. It detected the presence of RDR in 291 patients (36.1%). It correctly identified 166 of the 174 patients with a ground-truth diagnosis of RDR. This translates to a sensitivity of 95% (95% CI: 91, 98%). The diagnostic abilities of both the AI systems including sensitivity, specificity, PPV, and NPV are tabulated in Tables 1–4.

## Discussion

With the rising burden of DR, the importance of early detection and screening cannot be overstated. To address this glaring need, advanced technologies like AI software

**Table 1.** Performance of Medios and IDx-DR with ground truth cutoff at any DR (*n* = 264)

| | ARIAs | | | | |
|---|---|---|---|---|---|
| | Medios AI | | IDx-DR | | |
| | positive | negative | positive | negative | total |
| Reference standard (consensus grading) | | | | | |
| Positive (mild NPDR and above and/or CSDME), *n* (%) | 235 (29) | 29 (4) | 262 (32) | 2 (0.2) | 264 |
| Negative (no DR), *n* (%) | 56 (7) | 487 (60) | 305 (38) | 238 (29) | 543 |
| Total | 291 | 516 | 567 | 240 | 807 |

NPDR, non-proliferative diabetic retinopathy; DR, diabetic retinopathy; CSDME, clinically significant diabetic macular edema.

**Table 2.** Performance of Medios and IDx-DR with ground truth cutoff at RDR (*n* = 174)

| | ARIAs | | | | |
|---|---|---|---|---|---|
| | Medios AI | | IDx-DR | | |
| | positive | negative | positive | negative | total |
| Reference standard (consensus grading) | | | | | |
| Positive (moderate NPDR and above and/or CSDME), *n* (%) | 166 (21) | 8 (1) | 172 (21) | 2 (0.2) | 174 |
| Negative (no DR and mild NPDR), *n* (%) | 125 (15) | 508 (63) | 202 (25) | 431 (53) | 633 |
| Total | 291 | 516 | 374 | 433 | 807 |

NPDR, non-proliferative diabetic retinopathy; DR, diabetic retinopathy; CSDME, clinically significant diabetic macular edema; RDR, referable diabetic retinopathy.

**Table 3.** Performance of Medios and IDx-DR with ground truth cutoff at STDR (*n* = 41)

| | ARIAs | | | | |
|---|---|---|---|---|---|
| | Medios AI | | IDx-DR | | |
| | positive | negative | positive | negative | total |
| Reference standard (consensus grading), *n* (%) | | | | | |
| Positive (severe, PDR, or CSDME) | 40 (5) | 1 (0.1) | 39 (5) | 2 (0.2) | 41 |

DR, diabetic retinopathy; PDR, proliferative diabetic retinopathy; CSDME, clinically significant diabetic macular edema; STDR, sight-threatening diabetic retinopathy.

**Table 4.** Performance analysis of AI system compared to reference standard

| Variables | For any DR | | For RDR | | For STDR | |
|---|---|---|---|---|---|---|
| | Medios AI | IDx-DR | Medios AI | IDx-DR AI | Medios AI | IDx-DR AI |
| Sensitivity | 89 (85, 93) | 99 (97, 100) | 95 (91, 98) | 99 (96, 100) | 98 (87, 100) | 95 (83, 99) |
| Specificity | 90 (87,92) | 44 (40, 48) | 80 (77, 83) | 68 (64, 72) | NA | 80 (77, 83) |
| PPV | 81 (76, 85) | 46 (42, 50) | 57 (51, 63) | 46 (41, 51) | NA | 21 (15, 27) |
| NPV | 94 (92, 96) | 99 (97, 100) | 98 (97, 99) | 100 (98, 100) | NA | 100 (99, 100) |

Values are % (95% CI).

have emerged as promising tools for DR screening. These AI systems are meticulously developed and optimized using diverse datasets. Before implementing such AI software in real-world scenarios, it is crucial to conduct comparisons among different solutions available. In our evaluation, we examined the performance of two ARIAs: Medios AI and IDx-DR. The results exhibited comparable performance in terms of sensitivity, with Medios AI achieving 95% and IDx-DR achieving 99% in identifying RDR, respectively. These findings underscore the potential of these software solutions in facilitating early detection and screening of DR.

Overall, the prevalence of DR in the sample analyzed was 33% for any DR and 22% for RDR, significantly higher than commonly reported in other studies. Scottish national diabetic retinopathy screening program reported rates of RDR between 4.3% and 7%, large primary care-based screening in California reported RDR rate of 8.2% and a hospital-based study in Ethiopia found any DR rate of 18.9% [6–8]. It is also much higher than previous estimates for DR prevalence in Poland [9]. This is likely a side effect of the original screening set-up. The screening is based around diabetic clinics and diabetes medical centers, therefore selecting for a higher risk population with other diabetic complications or difficult-to-control disease. A similarly high prevalence of DR was found in a study of 297 patients attending a tertiary center for diabetes care in India with DR prevalence of 40.8% [5].

Only patients who initially had images of sufficient quality for IDx-DR during the initial screening were included in this study. The IDx-DR image quality assistant process was used as part of the original screening program and is not evaluated herein. Out of 811 patients deemed gradable by IDx-DR, only 4 (less than 0.5%) were excluded by the manual graders indicating that overall, the images selected for this study have good image quality.

The accuracy measures for Medios AI are in line with previously published studies. Natarajan et al. [10] reported accuracy of the Medios AI offline, smartphone-based version, with sensitivity and specificity pairs of 100% and 88.4% for RDR and 85.2% and 92.0% for any DR. In the aforementioned study based in a tertiary diabetes center, Medios AI achieved 98.8% and 86.7% sensitivity and specificity for any DR [5]. In another India-based study of 900 prospectively included patients, Medios AI achieved 83.3%, 95.5% sensitivity and specificity for any DR, and 93% and 92.5%, respectively, for RDR [11].

Crucially, all of the abovementioned studies were done using images gathered with the Remidio FOP mobile smartphone-based camera in contrast to using a stationary, full-size automatic fundus camera for this study. Previous studies describing Medios AI were smartphone-based, with the algorithm app being run on a smartphone, which was also used to take the fundus pictures. This is the first study outside of India to investigate using Medios AI with images from a stationary fundus camera in a real-world screening scenario. Images captured with different cameras may differ in resolution, level of detail, contrast, noise, and other parameters that may influence the accuracy of an algorithm. It is unclear whether the software or human graders may benefit from higher resolution images and provide a more robust golden standard, and if so, how significant the difference is. For this study, the previous smartphone-based results obtained by Medios AI seem to translate into comparable accuracy when using dedicated stationary fundus camera. These results are similar to another study where Medios AI was evaluated on Topcon images in an Indian population. This demonstrates generalizability of the model's performance on a desktop system. This device agnostic approach is particularly useful in screening programs that have already invested in camera systems and would want to move toward an AI-based approach without having to replace expensive cameras. IDx-DR exceeded the sensitivity measures of Medios AI at the cost of lower specificity. We have previously reported sensitivity and specificity of IDx-DR of 94% and 95% when compared to a single reader [3, 12]. In this study, IDx-DR retained excellent sensitivity at 99%, with a significantly lower specificity. This was more pronounced for any DR, with IDx-DR specificity of only 44%. This may be explained in part by the fact that IDx-DR has been specifically marketed for the detection of more than mild DR, and the specificity for detection of RDR is much higher at 68% with a 99% sensitivity. For comparison, in the pivotal trial that led to IDx-DR receiving FDA approval, where IDx-DR was compared against a diagnosis based on a 7-field ETDRS study with stereoscopic images and OCT, it achieved 87% sensitivity and 90% specificity for more than mild DR [13]. Both systems over-referred mild cases (false positives included 55% mild, 45% no DR by Medios AI; 38% mild, 62% no DR cases by IDx-DR). Another possibility of lower specificity could be referral of patients with similar lesions and concurrent pathologies that were not evaluated as part of this grading. Overall, both systems achieved satisfactory accuracy, particularly when patient safety is concerned with excellent NPVs, meaning very few patients received a false-negative result. Although false-positive results generate additional costs and workload due to the increase in referrals, it is the patients with missed disease that raise the most concern whenever automated screening is debated.

Many of the studies regarding automated analysis of DR from fundus images are sponsored or even performed directly by the respective software's owner company, which raises questions regarding bias. As previously mentioned, Medios AI does not offer a dedicated desktop application at this point; therefore, it was necessary to submit the images to Remidio, owner of the Medios AI algorithm, for a remote analysis on their system. As the authors of this study collaborated remotely, we could not directly oversee or verify the Medios AI output on site. Upon reviewing the study methodology, we considered this to be a source of potential bias and asked Remidio for a way to independently verify some of the software's results. We submitted the subset of images assessed by Polish graders, through a dedicated API (application programming interface) provided by Remidio with live results. Images were anonymized without changing the image content. The results were in line with those previously submitted by Remidio for all but 3 patients.

Out of those 3 patients, for whom the initial Medios AI output differed from the verification, all three decisions changed from no RDR to RDR. For 2 of those patients, the new Medios AI result now matched the grader decision of RDR; for the remaining patient, the new Medios AI now disagreed with the grader consensus. All 3 of those patients had very subtle retinal signs. This is a study looking into Medios AI outside of the smartphone application which involves a custom-made deployment for the study. The discrepancies are likely due to challenges surrounding the implementation of the algorithm on different hardware or inconsistencies in image compression parameters between the version of the images submitted to Remidio for the first analysis and the version of the images sent for verification through the API.

Inter-grader and intra-grader variability is known among DR graders. Previous studies have shown that inter-grader kappa scores typically range from 0.40 to 0.65 in DR grading [13–19]. The kappa values for both Polish (0.68–0.90) and Indian graders (0.63–0.91) in the study showed similar variability but were well within the limits showing overall good agreement. This ensured reliability while having the data split and graded by both groups separately. The possible reasons for variability among graders could be identification and differentiation of subtle DR features (retinal hemorrhages, microaneurysms, hard exudates, new vessels, intraretinal microvascular abnormalities, neovascularization, and surrogate markers of macular edema), variation in image quality due to artifacts, brightness or contrast of images. This has been found in other studies and in other fields of medical imaging as well [13–15]. Gold standard grading in the current study was done based on a majority decision by the graders. As the individual grades were converted to a binary decision for each grader before computing consensus grading, there was a majority decision for any DR and RDR for each of the patients. The reliability of the human grading could be improved with an adjudication process for patients without a full consensus [14].

This study included only patients with good-quality, non-mydriatic images, which may not be representative of the whole screening cohort. Using a non-mydriatic protocol may underrepresent patients with smaller pupils or media opacities, particularly the elderly. In a previous study about DR grader reliability, based on the same screening program in Poland from which images for this study were taken, out of 495 patients only 335 were deemed to be of sufficient quality by IDx-DR and all three human graders [15]. How many of those low-quality screening encounters could we image and diagnose after mydriasis remains to be seen, as is the comparative performance of both systems in those patients.

In conclusion, our study compared the performance of two AI screening software, Medios AI and IDx-DR, in detecting RDR. Both software systems demonstrated robust performance, with high accuracy and sensitivity, highlighting their potential as reliable tools for screening DR in real-world settings. Continued research and validation in larger and diverse patient populations will be essential to strengthen the evidence base and ensure the widespread adoption of these AI screening tools. Our study underscores the promise of these AI systems for DR screening, facilitating early detection, and timely intervention for improved patient outcomes.

## Statement of Ethics

The Institutional Review Board of the Foundation for Ophthalmology Development, Poznan, Poland approved the project (Application No. 2/ 2022) and waived the need for IRB approval and written informed consent from the participants for this retrospective study. The study was in adherence to the tenets of the Declaration of Helsinki.

## Conflict of Interest Statement

A.G. has grants/contracts from Alcon, Bausch & Lomb, Zeiss, Hoya, Thea, Viatris, Teleon, J&J, Cooper Vision, Essilor, and Polpharma. A.G. has consulting fees/honoraria from Thea, Polpharma, and Viatris and stock with GoCheck Kids. D.R.P., F.M.S., and K.N. are employees of Remidio Innovative Solutions. Remidio Innovative Solutions, Inc., USA, and Medios Technologies are wholly owned subsidiaries of Remidio Innovative Solutions Pvt. Ltd, India. F.M.S. has patents (mentioned in ICMJE) and stock (ESOP and stock, Remidio Innovative Solutions Pvt. Ltd). Other authors declare no financial disclosures.

## Funding Sources

There was no funding for this study.

## Author Contribution

A.G.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, supervision, validation, visualization, roles/writing – original draft, and writing – review and editing. D.R.P.: conceptualization, formal analysis, investigation, methodology, project administration, resources, supervision, visualization, validation, and writing – review and editing. P.B.: conceptualization, data curation, formal analysis, investigation, methodology, roles/writing – original draft, and writing – review and editing. K.N.: formal analysis, writing – original draft, and writing – review and editing. T.K.: data curation, formal analysis, investigation, methodology, software validation, visualization, roles/writing – original draft, and writing – review and editing. F.M.S.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, resources, software, supervision, validation, visualization, and writing – review and editing.

## Data Availability Statement

The data that support the findings of this study are not publicly available due to ethical reasons and are available on request from Dr. Andrzej Grzybowski (email: ae.grzybowski@gmail.com).

## References

1 Grzybowski A, Brona P, Lim G, Ruamviboonsuk P, Tan GSW, Abramoff M, et al. Artificial intelligence for diabetic retinopathy screening: a review. Eye. 2020;34(3):451–60.

2 Ting DS, Cheung GC, Wong TY. Diabetic retinopathy: global prevalence, major risk factors, screening practices and public health challenges: a review. Clin Exp Ophthalmol. 2016;44(4):260–77.

3 Grzybowski A, Brona P. Analysis and comparison of two artificial intelligence diabetic retinopathy screening algorithms in a pilot study: IDx-DR and retinalyze. J Clin Med. 2021;10(11):2352.

4 Lee AY, Yanagihara RT, Lee CS, Blazes M, Jung HC, Chee YE, 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.

5 Sosale B, Sosale AR, Murthy H, Sengupta S, Naveenam M. Medios- an offline, smartphone-based artificial intelligence algorithm for the diagnosis of diabetic retinopathy. Indian J Ophthalmol. 2020;68(2):391–5.

6 Looker HC, Nyangoma SO, Cromie DT, Olson JA, Leese GP, Black MW, et al. Rates of referable eye disease in the scottish national diabetic retinopathy screening programme. Br J Ophthalmol. 2014;98(6):790–5.

7 Cuadros J, Bresnick G. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening. J Diabetes Sci Technol. 2009;3(3):509–16.

8 Tilahun M, Gobena T, Dereje D, Welde M, Yideg G. Prevalence of diabetic retinopathy and its associated factors among diabetic patients at debre markos referral hospital, northwest Ethiopia, 2019: hospital-based cross-sectional study. Diabetes Metab Syndr Obes. 2020;13:2179–87.

9 Kozioł M, Nowak MS, Udziela M, Piątkiewicz P, Grabska-Liberek I, Szaflik JP. First nationwide study of diabetic retinopathy in Poland in the years 2013–2017. Acta Diabetol. 2020;57(10):1255–64.

10 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.

11 Sosale B, Aravind SR, Murthy H, Narayana S, Sharma U, Gowda SGV, et al. Simple, mobile-based artificial intelligence algorithm in the detection of diabetic retinopathy (SMART) study. BMJ Open Diabetes Res Care. 2020;8(1):e000892.

12 Grzybowski A, Brona P. A pilot study of autonomous artificial intelligence-based diabetic retinopathy screening in Poland. Acta Ophthalmol. 2019;97(8):e1149–50.

13 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.

14 Krause J, Gulshan V, Rahimy E, Karth P, Widner K, Corrado GS, et al. Grader variability and the importance of reference standards for evaluating machine learning models for diabetic retinopathy. Ophthalmology. 2018;125(8):1264–72.

15 Grzybowski A, Brona P, Krzywicki T, Gaca-Wysocka M, Berlińska A, Święch A. Variability of grading DR screening images among non-trained retina specialists. J Clin Med. 2022;11(11):3125.

16 Guan MY, Gulshan V, Dai AM, Hinton GE. Who said what: modeling individual labelers improves classification. In: Thirty-second AAAI conference on artificial intelligence; 2018. arXiv:1703.08774v2.

17 Sedova A, Hajdu D, Datlinger F, Steiner I, Neschi M, Aschauer J, et al. Comparison of early diabetic retinopathy staging in asymptomatic patients between autonomous AI-based screening and human-graded ultra-widefield colour fundus images. Eye. 2022 Mar;36(3):510–6.

18 Gangaputra S, Lovato JF, Hubbard L, Davis MD, Esser BA, Ambrosius WT, et al. Comparison of standardized clinical classification with fundus photograph grading for the assessment of diabetic retinopathy and diabetic macular edema severity. Retina. 2013 Jul–Aug;33(7):1393–9.

19 Raumviboonsuk P, Krause J, Chotcomwongse P, Sayres R, Raman R, Widner K, et al. Deep learning vs. Human graders for classifying severity levels of diabetic retinopathy in a real-world nationwide screening program. arXiv; 2018 Oct 18. arXiv:1810.08290.