Dovepress
Taylor & Francis Group

ORIGINAL RESEARCH

Comparison of the Aireen System with Telemedicine Evaluation by an Ophthalmologist – A Real-World Study

Martin Šín [ID][1], Renata Ženíšková[1], Martin Slíva[2], Kamila Dvořák [ID][2,3], Jozefína Vaľková[2], Jan Bayer[2], Barbora Karasová[2], Jan Tesař[1], Dana Fillová[4], Martin Prázný[5,6]

[1]Department of Ophthalmology, Military University Hospital Prague, 1st Faculty of Medicine, Charles University, Prague, Czech Republic; [2]Aireen a.s., Prague, Czech Republic; [3]Department of Natural Sciences, Faculty of Biomedical Engineering, Czech Technical University, Prague, Czech Republic; [4]Eye Centre Prague a.s., Prague, Czech Republic; [5]3rd Department of Internal Medicine, General University Hospital in Prague, Prague, Czech Republic; [6]3rd Department of Medicine - Department of Endocrinology and Metabolism, 1st Faculty of Medicine, Charles University, Prague, Czech Republic

Correspondence: Martin Šín, Department of Ophthalmology, Military University Hospital Prague, 1st Faculty of Medicine, Charles University, U Vojenské nemocnice 1200, 169 02, Praha 6, Prague, Czech Republic, Email sin.martin@uvn.cz

Purpose: This study aimed to compare general ophthalmologists, retina specialists, and Aireen AI screening system with the clinical reference standard of a three-member high-level expert committee for diabetic retinopathy (DR) in the evaluation of fundus images for DR.

Patients and Methods: The study was designed as a diagnostic, multicenter, cross-sectional, non-randomized diagnostic study. The cohort included in the clinical investigation consisted of 1274 patients with diabetes mellitus (DM) type I or II. Each patient underwent one-field fundus photography using a non-mydriatic camera to assess findings of DR. One hundred and nineteen subjects (9.3%) were excluded from the clinical investigation based on Aireen system assessment. In the clinical investigation, all images were assessed at three independent levels of evaluation: 1) general ophthalmologists (GO) – without subspecialty training in the retina; 2) retina specialists (RS); and 3) system Aireen. In cases where there may be disagreements amongst groups, the image is referred for assessment by the Diabetic Retinopathy Board (DRB).

Results: The overall prevalence of any DR was 31.9% (368 cases out of 1154 DM), according to the DRB. Overall concordance between AI system Aireen and GO and RS assessments in the detection of DR from fundus photography occurred in 734 cases (63.6%). The number of disagreements between Aireen system, GO and RS evaluation occurred in 420 (36.4%) cases. Sensitivity for GO was 87.0% (95% CI: 83.6; 90.4), for RS was 82.9% (95% CI: 79.1; 86.7), and for AI system Aireen was 92.1% (95% CI: 89.3; 94.9). Specificity was 76.5% (95% CI: 73.5; 79.5), 81.2% (95% CI: 78.5; 83.9), and 90.7% (95% CI: 88.7; 92.7) for GO, RS and AI system Aireen, respectively.

Conclusion: This real-world study illustrates the potential use of AI system Aireen in screening for DR. It exhibits higher sensitivity and specificity compared to telemedicine evaluation of one field fundus image.

Keywords: diabetic retinopathy, artificial intelligence, screening, fundus image

Introduction

Diabetic retinopathy (DR) is the most common and serious ocular complication in diabetic patients. It remains the leading cause of vision loss in many developed countries.[1] Of the 246 million people worldwide with diabetes, about a third have signs of DR, and a third of these might have vision-threatening retinopathy, defined as proliferative retinopathy or macular edema.[2]

Between 1980 and 2008, the occurrence of advanced DR and significant visual loss in individuals with diabetes decreased in populations with better diabetes management.[3] However, the overall prevalence of visual impairment and blindness resulting from DR rose considerably between 1990 and 2015, as indicated in the most recent report from the Vision Loss Expert Group of the Global Burden of Disease Study.[4] This increase was primarily due to the growing

prevalence of type 2 diabetes in less developed countries. Detecting cases requiring timely ophthalmic examination and treatment to prevent permanent visual loss is vital for screening for DR, as early treatment leads to improved outcomes.[5] Nonetheless, numerous countries lack adequate resources for nationwide screening programs.

Artificial intelligence (AI) systems using digital fundus photography instruments have been developed for DR screening to partially address the increased demand for screening related to a burgeoning population with diabetes in the world. The advantages of AI screening systems include the convenience of point-of-care access and the potentially lower operating cost, as a result of automatic interpretation of the images and appropriate referral to an eye care specialist. Another innovative approach represents the use of telemedicine and portable imaging devices, which are changing the screening strategies and are also improving the cost-effectiveness of diabetic screening.[6]

The purpose of this study was to compare general ophthalmologists, retina specialists, and the Aireen AI screening system with the clinical reference standard of three-member high-level expert committee for DR screening in the evaluation of fundus photographs for detection of DR.

Materials and Methods

A diagnostic, multicenter cross-sectional, non-randomized, diagnostic study was preregistered (SUKL no. sukls65076/2022). The recruitment of the patients started on 27 May 2022. The end of the recruitment period for this study was 29 July 2022.

Data analysis was performed from February to July 2023. The protocol was approved by the Alpha Institutional Review Board (registration no. EK/36/2022) and site-specific institutional review boards, where required (registration no. 22/21). All participants provided written informed consent. The study was conducted in accordance with the International Conference on Harmonization Good Clinical Practice and the Declaration of Helsinki.

Based on the clinical study design, no indeterminate index test or reference standard was expected, and the study was not focused on analyzing variability. The investigation was not focused on the grade of DR. No adverse events were reported during the clinical study. The full study protocol can be accessed upon request from Aireen a.s.

Study Population

Four Czech primary diabetes-care centers participated in the study recruitment. The intended sample size of the study was at least 1070 subjects. This number corresponds to the recommended minimum sample size for a diagnostic study of a medical device with 10% prevalence of DR in the population, an intercepting of 10% change in the values of sensitivity and specificity of the diagnostic test, a statistical power of the test at least 80% and a significance level of 95% (alpha = 0 0.05).[7]

Patients with scheduled regular visits were sequentially assessed for eligibility by medical records review before being invited to participate in the study. No further patient subselection was performed. Participants were aged 18 years or older, and all had diabetes. Exclusion criteria included contraindication to fundus photography and unwillingness to participate in the study. The flow of the participants is displayed in Figure 1.

One thousand two hundred and seventy-four patients with DM type I or II were included in the clinical investigation. One hundred and nineteen subjects (9.3%) were excluded from the clinical investigation based on assessment of the Aireen system due to inadequate quality of images and one subject did not meet the recruitment criteria of age (18 and older).

In total, 1154 patients underwent final evaluation in the clinical investigation. Demographic data and the history of DM are summarized in Table 1.

Photographic Procedure

Study participants underwent non-mydriatic imaging of both eyes using a digital fundus camera with a resolution of 1.69 megapixels (Canon CR-2 AF). All fundus examinations were performed without pharmacological dilation. This approach was chosen to reflect real-world screening conditions and may have contributed to the percentage of non-assessable images. For each eye, 1 image centered on the macula was taken by a healthcare worker at diabetes clinics who has no experience in taking images with a fundus camera. The fundus photographs were then uploaded to cloud storage, where the Aireen AI system determined whether DR was present or not. If the image quality precluded an interpretation of the level of DR by the AI system, the images were considered positive (to avoid false negative errors).
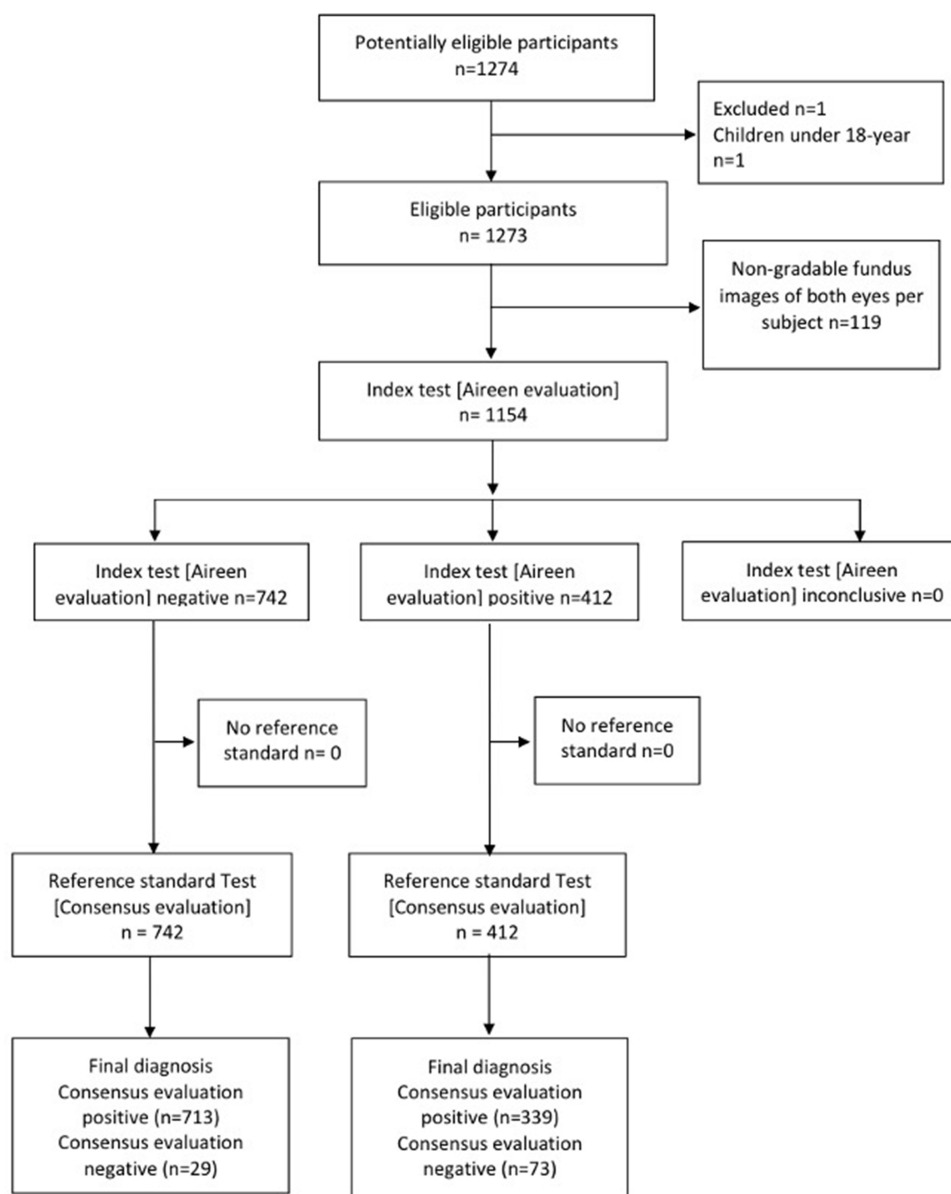
Figure I Standard for Reporting Diagnostic Accuracy (STARD) Flow Chart.

The AI System Aireen

The Aireen system represents a commercially available AI system composed of multiple independent neural networks.

Before the classification into DR or non-DR class, each image is pre-screened for sufficient quality to prevent erroneous results and potential harm to the patient. Only images with verified quality entered the classification process.

The quality assessment is performed using several separate neural networks that independently confirm whether the image depicts the retina, is adequately sharp, contains the macula and optic disk and is captured by an optical camera. Additionally, the system automatically identifies whether the image is from the right or left eye. All these evaluations occur within seconds after image upload, providing instant feedback to the camera operator. In cases where the images are not of adequate quality, the operator can promptly capture new ones and upload them to the Aireen system.

The DR classifier itself consists of three sequential steps. The first step aims to standardize the input from various image formats for subsequent network processing. During this stage, the image is cropped and resized to predefined dimensions, ensuring the removal of unnecessary edges.

Table I Demographic Data

		N = 1154
Age (years)	N/missing data	1154 / 0
	Mean (SD)	59.7 (14.2)
	Minimum/Median/Maximum	18/62/90
Duration of DM (years)	N/missing data	1151/3
	Mean (SD)	10.7 (8.9)
	Minimum/Median/Maximum	1/9/55
Gender	Male	594 (51.5%)
	Female	560 (48.5%)

The second step involves a neural network focused on image classification, built upon the well-established EfficientNet V2 architecture. The output of this network is an eight-dimensional vector, in which the network assesses the probability of identifying DR on the evaluated image, as well as the likelihood of identifying the presence of Microaneurysms and/or Hemorrhages, Hard Exudates, Cotton Wool Spots, Laser Scars, Intraretinal Microvascular Abnormalities, New Vessels, and/or Fibrous Proliferation.

This eight-dimensional probability vector is then input into the final post-processing step, which determines whether the image can be classified as DR or non-DR.

Evaluation Process

The output of the image evaluation is the result of the analysis of the image of the retina of both the left and the right eye. The output and its parameters are as follows:

- DR symptoms are present – 1 (positive),
- No symptoms of DR – 0 (negative),
- Cannot be evaluated – 1 (positive).

In the clinical investigation, all images were assessed at three independent levels of evaluation:

- General Ophthalmologists – without subspecialty training in retina,
- Retina Specialists,
- The AI system Aireen.

Agreement between the three levels of assessment was determined based on consensus in the DR classification. In cases where there may be disagreement amongst the three groups, the image is referred for assessment by the Diabetic Retinopathy Board. The Diabetic Retinopathy Board represents a commission of three highly specialized (10+ years of practice in DR) members. The DRB did not know the results of either evaluator; its knowledge was limited to the presence of a disparity between the evaluations. Since the DRB was a three-member board, each slide was voted on, and the result was a majority opinion (ie 2 or 3 members). The commission's decisions were considered as a gold standard and all evaluations in a clinical investigation are compared with their opinions.

From the number of potential detections of DR findings (Positive Findings) and no findings (Negative Findings), an estimate of Sensitivity, Specificity, and Positive Predictive Value (PPV) with respect to the reference (Positive Labels and Negative Labels) was subsequently derived as a basic measure of the reliability of the diagnosis. The reliability is presented in Table 2. Equations for Sensitivity, Specificity and PPV are displayed in Equations (1), (2), (3), respectively.

Table 2 True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) Parameters are Used in the Evaluation of Specificity, Sensitivity and PPV. TP is the Number of Perfectly Identified DR Images. TN is the Number of Perfectly Detected Non-DR Images

	Positive Labels (Reference)	Negative Labels (Reference)
Positive Findings (Aireen)	TP	FP
Negative Findings (Aireen)	FN	TN

The test positivity cut-off was pre-specified before this study. The index test results have only two output values (positive/negative). The cut-off value was selected according to the sensitivity/specificity ratio.

$$Sensitivity = \frac{TP}{TP+FN} \tag{1}$$

$$Specificity = \frac{TN}{TN+FP} \tag{2}$$

$$PPV = \frac{TP}{TP+FP} \tag{3}$$

Results

Of the 1273 recruited patients that met the requirements, 1154 (90.6%) patients had readable fundus image of at least one eye, and 119 (9.4%) patients did not have acceptable photographs in either eye. Based on the evaluation by the DRB, the prevalence of DR eyes in our cohort was 31.9% (368/1154). Overall concordance between the AI system Aireen and the GO and the RS assessments in the detection of diabetic retinopathy from fundus photography occurred in 734 cases (63.6%). The number of disagreements between the Aireen system and the GO and the RS evaluation occurred in 420 (36.4%) cases. More detailed results can be seen in Table 3. With respect to evaluations by GO vs RS, the number of disagreements occurred in 320 (27.7%) cases. Additional details can be found in Table 4. As mentioned above, a total of 420 cases had been evaluated by DRB. In 100 (8.7%) cases, there was a disagreement between system Aireen and the concordant agreement of human evaluation. In 320 (27.7%) cases, there were disagreement between two different humans' evaluation (GO vs RS). More comprehensive findings are presented in Table 5. Table 6 presents calculated Sensitivity, Specificity and Positive Predictive Value.

Table 3 Numbers of Concordances/Disagreements in the Detection of DR by the System Aireen Vs GO and/or RS

	N = 1154
Disagreements Aireen (1) vs GO and/or RS (0)	143 (12.4%)
Disagreements Aireen (0) vs GO and/or RS (1)	277 (24.0%)
All concordances Aireen vs GO and/or RS	734 (63.6%)
All disagreements Aireen vs GO and/or RS	420 (36.4%)

Table 4 Number of Disagreements in the Detection of Diabetic Retinopathy by GO Vs RS

	N = 1154
Disagreements GO (0) vs RS (1)	134 (11.6%)
Disagreements GO (1) vs RS (0)	186 (16.1%)
All disagreements between the GO and RS	320 (27.7%)

Table 5 Evaluation of the DRB in Case of Disagreement Aireen Vs GO and/or RS

	N = 420
Concordances Aireen (1) vs DRB (1)	70 (16.7%)
Concordances Aireen (0) vs DRB (0)	248 (59.0%)
Disagreements Aireen (1) vs DRB (0)	73 (17.4%)
Disagreements Aireen (0) vs DRB (1)	29 (6.9%)
All concordances Aireen vs DRB	318 (75.7%)
All disagreements Aireen vs DRB	102 (24.3%)

Table 6 Sensitivity, Specificity and Positive Predictive Values

		Aireen	**GO**	**RS**
N = 1154	Sensitivity	92.1% (95% CI: 89.3; 94.9)	87.0% (95% CI: 83.6; 90.4)	82.9% (95% CI: 79.1; 86.7)
	Specificity	90.7% (95% CI: 88.7; 92.7)	76.5% (95% CI: 73.5; 79.5)	81.2% (95% CI: 78.5; 83.9)
	PPV	82.3% (95% CI: 78.6; 85.9)	63.4% (95% CI: 59.2; 67.6)	67.3% (95% CI: 63.1; 71.7)

Discussion

Our current study recruited participants in a real-world setting in a general diabetology office on the intention-to-screen principle. Slightly more than 90% of images were of readable quality, which is in good agreement with other similar studies.[8,9]

The gold standard for the detection and classification of diabetic retinopathy is stereoscopic color fundus photographs in 7 standard fields, as defined by the Early Treatment Diabetic Retinopathy Study (ETDRS) group (ETDRS group 1991). Although this technique is accurate and reproducible, it is labor intensive and requires skilled photographers, which makes it less than ideal for mass screening programs.

A single 45° fundus image was deemed sufficient for DR screening and patient referral by the American Academy of Ophthalmology (AAO).[10] Ultimately, the implementation of undemanding protocols would increase patients' adherence to DR screening programs, contributing to their success.

In this study, the overall prevalence of any DR was 31.9% (368 cases out of 1154 DM) according to evaluation by the DRB. This represents almost the same prevalence of 31.2% (304 cases out of 973 DM) that was reported by the study in a similar setting.[11] However, the prevalence of DR is variable in different studies from countries with similar socio-economic status, reporting the prevalence of any DR was 15.5% in patients with Type 2 DM.[12] In Type 1 DM patients, the prevalence of any DR was 29.2%. In the Swedish study (population-based), the prevalence of any DR was 27.9% and 41.8% in Type 2 and Type 1 DM patients, whereas, in the Danish study, it was 21.2% and 54.3% in Type 2 and Type 1, respectively.[13,14] These diverse results indicate a multifactorial influence on the incidence of DR, the most important of which include the duration of DM, socio-economic status and lifestyle.

The results of this study show that the AI system in a primary care setting robustly exceeded the sensitivity and specificity of 92.1% (95% CI: 89.3; 94.9) and 90.7% (95% CI: 88.7; 92.7), respectively, in detecting DR presence. The results are in good agreement with the specificity and sensitivity reported in other studies.[8,15,16] In this study, human evaluators reached lower values of sensitivity and specificity. Counterintuitively, the RS group reached lower sensitivity as compared to GO, however in the specificity there was an opposite trend. This suggests that GO had a cautious approach to image evaluation and leaned more towards an increase in the number of positive cases. Sensitivity is

a patient safety criterion, because its primary role is to identify patients with diabetes, who are more likely to have DR requiring further evaluation.

On the other hand, the RS group reached higher specificity values, but both groups were approximately 10% lower than the AI system. The values of specificity and sensitivity are in accordance with previous findings in a study focusing on comparison of telemedicine DR screening, with different number of fundus photo fields. In this study, sensitivity varies between 69% and 78% and specificity between 85% and 99%, utilizing a different fundus photography device.[17]

This real-world study illustrates the potential use of the AI system Aireen in screening for DR. It exhibits higher sensitivity and specificity as compared to telemedicine evaluation of 1 field fundus photograph. We are aware that having no clinical ophthalmoscopy as a control is our study's main weakness. On the other hand, our primary aim was to simulate real-world conditions as best we could, and hence we consider this as the main strength of the study. Moreover, data variability can significantly impact AI performance, as differences in patient demographics may limit model generalizability and reduce diagnostic accuracy across diverse populations.

Given the current low rate of compliance with the recommendation for an annual diabetic retina examination, this AI system can be considered a useful adjunct in the detection of DR and seems to be more accurate than telemedicine for routine retinal screening.

Conclusion

This study underscores the significant potential of the Aireen AI screening system in the detection of DR by comparing the performance of general ophthalmologists, retina specialists, and the Aireen AI system against a clinical reference standard provided by a DRB.

The findings from this prospective, multicenter, cross-sectional study suggest that the Aireen AI system can effectively enhance DR screening processes, offering a reliable and efficient alternative to traditional methods. Its integration into clinical practice could lead to improved screening accuracy, potentially benefiting a larger population of patients with DM through earlier and more precise detection of diabetic retinopathy. This study supports the continued development and implementation of AI technologies in ophthalmology to advance patient care and screening efficiency.

Abbreviations

AI, artificial intelligence; DM, diabetes mellitus; DR, diabetic retinopathy; DRB, diabetic retinopathy board; ETDRS, early treatment diabetic retinopathy study; FN, false negative; FP, false positive; GO, general ophthalmologists; PPV, positive predictive value; RS, retina specialists; STARD, standard for reporting diagnostic accuracy; TN, true negative; TP, true positive.

Acknowledgments

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Disclosure

Mr Martin Slíva, Ms Kamila Dvořák, Ms Jozefína Valková and Mr Jan Bayer report grants from the Ministry of Industry and Trade of the Czech Republic, during the conduct of the study. Professor Martin Prázný reports personal fees from Aireen, during the conduct of the study. The authors report no other conflicts of interest in this work.

References

1. Fong DS, Aiello L, Gardner TW, et al. Retinopathy in diabetes. *Diabetes Care*. 2003;27(Supplement 1):S84–S87. doi:10.2337/diacare.27.2007.s84
2. Saaddine JB. Projection of diabetic retinopathy and other major eye diseases among people with diabetes mellitus. *Arch Ophthalmol*. 2008;126 (12):1740. doi:10.1001/archopht.126.12.1740
3. Sabanayagam C, Banu R, Chee ML, et al. Incidence and progression of diabetic retinopathy: a systematic review. *Lancet Diabetes Endocrinol*. 2019;7(2):140–149. doi:10.1016/S2213-8587(18)30128-1
4. Chew EY, Davis MD, Danis RP, et al. The effects of medical management on the progression of diabetic retinopathy in persons with type 2 diabetes. *Ophthalmology*. 2014;121(12):2443–2451. doi:10.1016/j.ophtha.2014.07.019
5. Flaxel CJ, Adelman RA, Bailey ST, et al. Diabetic retinopathy preferred practice pattern®. *Ophthalmology*. 2020;127(1):P66–P145. doi:10.1016/j. ophtha.2019.09.025
6. Vujosevic S, Aldington SJ, Silva P, et al. Screening for diabetic retinopathy: new perspectives and challenges. *Lancet Diabetes Endocrinol*. 2020;8 (4):337–347. doi:10.1016/S2213-8587(19)30411-5
7. Bujang MA, Adnan TH. Requirements for minimum sample size for sensitivity and specificity analysis. *J Clin Diagn Res*. 2016;10(10):YE01–YE06. doi:10.7860/jcdr/2016/18129.8744
8. Ming S, Xie K, Lei X, et al. Evaluation of a novel artificial intelligence-based screening system for diabetic retinopathy in community of China: a real-world study. *Intl Ophthalmol*. 2021;41(4):1291–1299. doi:10.1007/s10792-020-01685-x
9. Ipp E, Liljenquist D, Bode B, et al. Pivotal evaluation of an artificial intelligence system for autonomous detection of referrable and vision-threatening diabetic retinopathy. *JAMA Network Open*. 2021;4(11):e2134254–e2134254. doi:10.1001/jamanetworkopen.2021.34254
10. Williams GA, Scott IU, Haller JA, Maguire AM, Marcus D, McDonald HR. Single-field fundus photography for diabetic retinopathy screening: a report by the American Academy of Ophthalmology. *Ophthalmology*. 2004;111(5):1055–1062. doi:10.1016/j.ophtha.2004.02.004
11. Lim JI, Regillo CD, Sadda SR, et al. Artificial intelligence detection of diabetic retinopathy: subgroup comparison of the eyeart system with ophthalmologists' dilated examinations. *Ophthalmol Sci*. 2023;3(1):100228. doi:10.1016/j.xops.2022.100228
12. Ondrejkova M, Jackuliak P, Martinka E, et al. Prevalence and epidemiological characteristics of patients with diabetic retinopathy in Slovakia: 12-month results from the DIARET SK study. *PLoS One*. 2019;14(12):e0223788. doi:10.1371/journal.pone.0223788
13. Bøgelund Larsen M, Erik Henriksen J, Grauslund J, Pető T. Prevalence and risk factors for diabetic retinopathy in 17 152 patients from the island of Funen, Denmark. *Acta Ophthalmologica*. 2017;95(8):778–786. doi:10.1111/aos.13449
14. Heintz E, Wiréhn AB, Peebo BB, Rosenqvist U, Levin LÅ. Prevalence and healthcare costs of diabetic retinopathy: a population-based register study in Sweden. *Diabetologia*. 2010;53(10):2147–2154. doi:10.1007/s00125-010-1836-3
15. He J, Cao T, Xu F, et al. Artificial intelligence-based screening for diabetic retinopathy at community hospital. *Eye*. 2019. doi:10.1038/s41433-019-0562-4
16. 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. doi:10.1001/jamaophthalmol.2019.2923
17. Salongcay RP, Martin C, Michael C, et al. One-field, two-field and five-field handheld retinal imaging compared with standard seven-field Early Treatment Diabetic Retinopathy Study photography for diabetic retinopathy screening. *Br J Ophthalmol*. 2023:bjophthalmol–321849. doi:10.1136/bjo-2022-321849