scientific reports

Check for updates

OPEN

Diagnostic performance of an artificial intelligence software for diabetic retinopathy organised screening in a pilot study

Emma Altobelli[1]✉, Marco Baroni[2], Marco Ciancaglini[2], Paolo Carpineto[3] & Francesco Masedu[4]

Diabetic retinopathy (DR) is one of the most common complications of diabetes mellitus (DM) with a prevalence that varies from 10% to 61% in different countries. From healthcare to the precise prevention, diagnosis, and management of diseases, Artificial Intelligence (AI) is progressing rapidly in interdisciplinary fields, including ophthalmology. We aimed to explore the efficacy of artificial intelligence (AI)-based screening for diabetic retinopathy in type 2 DM patients. High-resolution colour fundus photographs were obtained for all patients using a non-mydriatic camera (Nidek AFC-330). DAIRET software uses a machine learning algorithm to identify the most important signs of DR. The software provides an output with a binary result: negative, in case of no referable disease or positive, when the presence of signs of DR is detected. The photographs were also manually graded negative or positive for DR by an expert ophthalmologist, masked to AI responses. Expert clinical opinion was used as the reference (gold standard), and the results were compared with those of the AI responses. A significant association was observed between DAIRET and expert diagnosis ($\chi^2 = 18.6$, $p < 0.001$). Disease prevalence, according to expert assessment, was 20.2% (33/163). Observed agreement between DAIRET and Expert diagnosis was 69.3%. Agreement metric was: Gwet's AC1 = 0.47 (95% CI: 0.33–0.61), indicating moderate agreement. Diagnostic performance of the DAIRET test indicates good sensitivity (73%), a specificity (68%) and high negative predictive value (NPV=90.8%), suggesting that DAIRET is more effective at excluding diabetic retinopathy than at confirming it. ROC analysis stratified by gender showed no difference in performance (p>0.05). Overall, these results suggest that DAIRET may be effectively integrated into screening workflows to rule out diabetic retinopathy, reduce specialist workload, and prioritize referrals. In particular, for daily practical use, DAIRET is managed within the MètaClinic electronic medical record. In-house data processing without going to an external cloud is a significant strength. Larger, prospective studies are needed to validate these findings, optimise thresholds, and assess performance across diverse populations and clinical settings.

Keywords Artificial intelligence, Diabetes, Retinopathy, Screening

According to data from the International Diabetes Federation, 589 million adults (20-79 years) are living with diabetes – 1 in 9. This number is predicted to rise to 853 million by 2050 (https://diabetesatlas.org/Latest access 13 January 2026)[1].

Diabetic retinopathy (DR) is one of the most common complications of diabetes mellitus (DM) and it is the primary eye disease that causes blindness in the population[2–4], with a prevalence that varies from 10% to 61% in different countries[5]. Moreover, the economic costs associated with DR and its complications are substantial[6]. It is important to underline that DR is often asymptomatic until an advanced stage, when it is less amenable

[1]Section of Epidemiology and Public Health Unit, Department of Life, Health and Environmental Sciences, University of L'Aquila, Piazzale Salvatore Tommasi 1, Coppito, 67100 L'Aquila, Italy. [2]Department of Life, Health and Environmental Sciences, Section of Clinical and Molecular Medicine, University of L'Aquila, 67100 L'Aquila, Italy. [3]Department of Medical, Oral and Biotechnological Sciences, University "G. D'Annunzio" of Chieti-Pescara, 66100 Chieti, Italy. [4]Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, Via Vetoio, Coppito, 67100 L'Aquila, Italy. ✉email: emma.altobelli@univaq.it

to treatment; therefore, screening is recommended to prevent sight loss. Diagnosis usually involves multiple consultations with retina specialists, often not available promptly, leading to treatment delays. Current guidelines by the American Academy of Ophthalmology recommend that patients diagnosed with diabetes undergo yearly screenings for the detection of DR[7].

From healthcare to the precise prevention, diagnosis, and management of diseases, Artificial Intelligence (AI) is progressing rapidly in various interdisciplinary fields[8,9], including ophthalmology. In particular, pivotal clinical trials have demonstrated that autonomous AI systems can achieve high diagnostic accuracy in primary care settings, establishing a solid foundation for DR screening[10]. AI for DR detection could be a clinically and economically viable alternative. Ophthalmology is a potential exemplar speciality in the application of medical AI, in particular with its use in the context of DR. In recent years, automated retinal photography reading platforms built on AI standards have been developed[11,12]. Most studies on DR screening using AI software have been performed in developed countries. In low- and middle-income countries, AI has been promoted in order to strengthen and improve clinical practice and to this end. It is important to underline that many barriers exist that could prevent the successful development and adoption of well-performing, context-specific AI tools, such as, for example, limited data availability[13].

In Germany, AI-based screening for diabetic retinopathy is viewed positively and is considered 'fundamentally suitable for future use' according to the National Care Guideline[14].

In the United States of America, the first automated screening system for DR was approved back in 2018[15]. Recently, in Italy, diagnostic-accuracy studies have been conducted, which have demonstrated that Artificial Intelligence (AI) for DR detection could be a clinical alternative[16,17].

Indeed, AI algorithms could improve the efficacy of screening and might be implemented for clinical use after thorough validation in a real-life setting[18].

We aimed to explore the efficacy of artificial intelligence (AI)-based screening for diabetic retinopathy (DR) in type 2 diabetes mellitus (T2DM) patients.

Materials and methods

Study design and objective

This observational pilot study aimed to evaluate the diagnostic performance of DAIRET (acronym for *Diabetes Artificial Intelligence for RETinopathy*), an AI–based software designed for automated screening of DR within an organised screening program. This study adhered to the tenets of the Declaration of Helsinki. The authorization protocol to process the data was obtained on 3 April 2024 (protocol number FFORIC24.01) by Institutional Review Board of Department of Life, Health and Environmental Sciences of the University of L'Aquila, Italy.

Inclusion and exclusion criteria are shown in Fig. 1.

High-resolution colour fundus photographs were obtained for all patients using a non-mydriatic camera (Nidek AFC-330). Usually, two 45° fields per eye were acquired: one centred on the optic disc and one on the macula (Fig. 2). The images were uploaded to the electronic medical record (EMR) platform MetaClinic (Meteda) for processing. DAIRET software uses a machine learning algorithm to identify the most important signs of DR, such as microaneurysms (MA), small red dots representing bulges in blood vessel walls, haemorrhages and exudates, which are signs of leaking fluid or blood. The software provides an output with a binary result: negative, in case of no referable disease or positive (Fig. 3), when the presence of signs of DR is detected (Fig. 3).

The photographs were also manually graded as negative or positive for DR by an expert ophthalmologist, masked to AI responses. Expert clinical opinion was used as the reference (gold standard), and the results were compared with those of the AI responses.

Sample size calculation

Sample size was calculated using a power analysis for a single proportion based on Cohen's arcsine transformation. Due to the absence of preliminary pilot data or established literature regarding the expected performance of the AI software, a medium effect size (Cohen's h = 0.3) was assumed. With a significance level of 0.01 and a statistical power of 90% in a two-sided test, the minimum required sample size was determined to be 166 patients (n = 165.33).

Study population

A total of 170 consecutive diabetic patients enrolled in the DR screening in the Eye Unit of San Salvatore Hospital, L'Aquila (Italy), between September 2024 and June 2025, were recruited. Patients not performant or whose tests were classified as ambiguous or inconclusive were excluded from the analysis. Specifically, 7 DAIRET tests (4.1%) were excluded, yielding a final analytic sample of 163 patients.

Clinical variables

Descriptive statistics (mean and standard deviation) were computed for demographic and clinical variables, including age, glycaemic control, lipid profile, renal function, liver enzymes, blood pressure, and diabetes duration (Table 1).

Diagnostic comparison and confusion matrix

Using expert opinion as the reference standard, a 2×2 confusion matrix was constructed to compare DAIRET test results with expert diagnosis. From this matrix, diagnostic performance measures were calculated and summarised in Table 2.

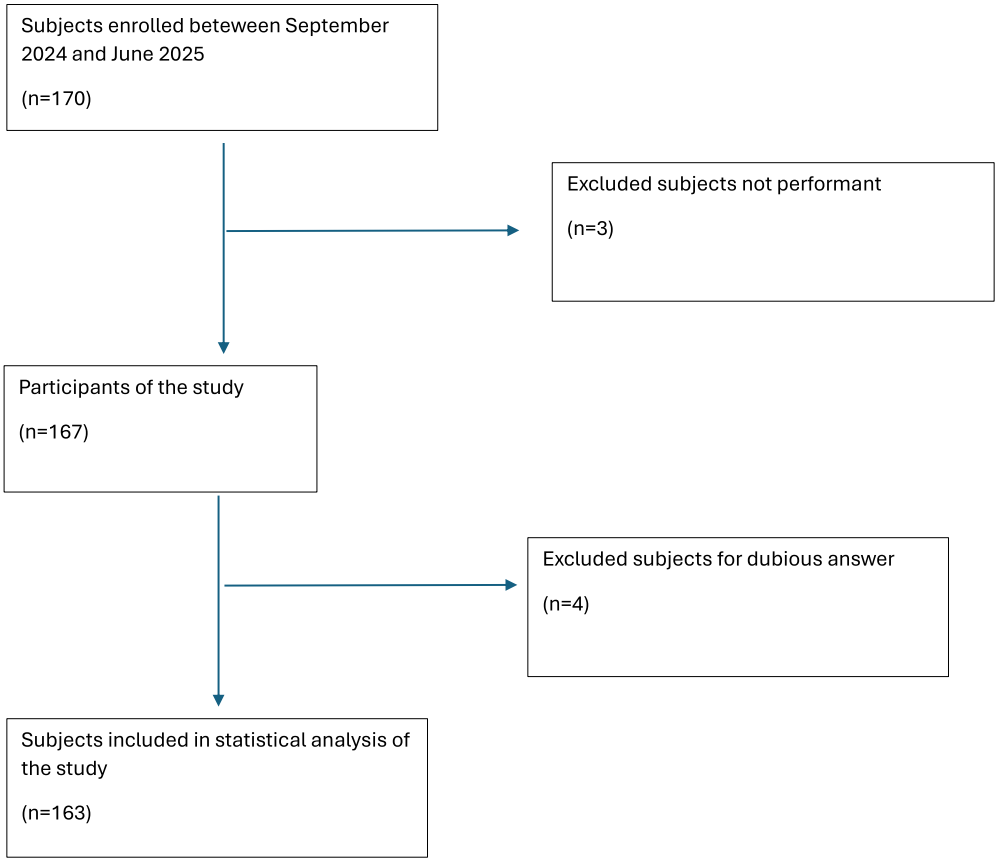

Fig. 1. Participant Flow Diagram. Selection process of the study population. A total of 170 patients with Type 2 Diabetes Mellitus were initially enrolled between September 2024 and June 2025. Out of these, 7 subjects were excluded: 3 due to ungradable fundus images (not performant) and 4 due to dubious clinical records, resulting in a final sample of 163 participants for AI-based statistical analysis.

Statistical analysis

A χ^2 test was used to assess the association between DAIRET and expert diagnosis using Cramér's V to quantify the strength of the statistical association.

The Inter-rater agreement was assessed using percent agreement, Cohen's kappa, and Gwet's AC1[19]. The proportion of patients with diabetic retinopathy (DR) (33/163 ≈ 20.2%), namely only 1 in 5 patients has the disease, is much lower than that of those without DR.

Given the imbalanced disease prevalence, Gwet's AC1 was considered the primary agreement metric.

In our data, disease prevalence according to the reference standard is 20.2%, resulting in unbalanced marginal distributions. The Inter-rater agreement was assessed using percent agreement, Cohen's kappa and Gwet's AC1. Cohen's kappa is well known to be sensitive to prevalence and marginal imbalance, often producing artificially low values despite substantial observed agreement (*"kappa paradox"*)[20,21].

Gwet's AC1 is robust to prevalence effects and provides a more stable estimate of agreement. The discrepancy between observed agreement (k=0.69, 95% CI=(0.16, 0.44)) and kappa (0.30), typically when kappa is affected by skewed prevalence, may underestimate the true level of agreement, describing a fair agreement. The Gwet'AC1 (AC1=0.47, 95% CI=(0.33, 0.61)).

Diagnostic accuracy measures (sensitivity, specificity, predictive values, accuracy, and likelihood ratios) were estimated with 95% confidence intervals (CIs). Proportion CIs were calculated using the exact binomial method. Likelihood ratio CIs were calculated on the log scale and back-transformed. Using expert opinion as the reference standard, a 2×2 confusion matrix was constructed to compare DAIRET test results with expert diagnosis. From this matrix, diagnostic performance measures were calculated and summarised in Table 2.

Subgroup analysis ROC analyses were additionally performed stratified by sex to assess potential gender differences in diagnostic performance. All statistical analyses were performed using Stata software, version 17 (StataCorp, College Station, TX, USA). Statistical significance was set at $p<0.05$ unless otherwise specified.

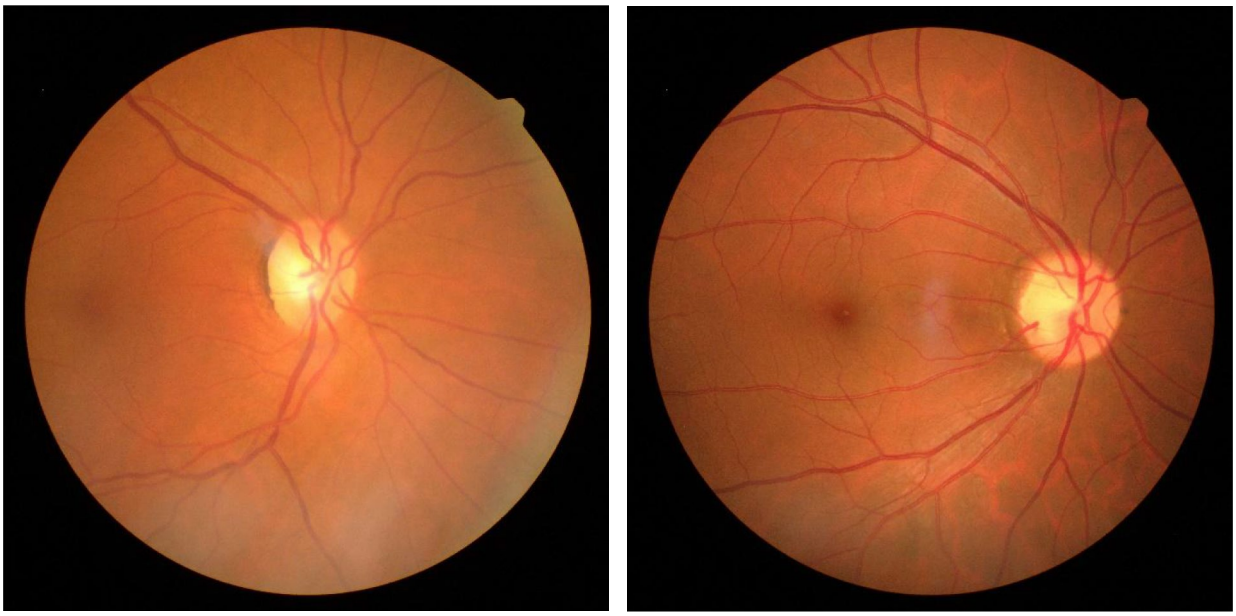

Fig. 2. Color fundus photographs centered on the optic disc (Left) and on the macula (Right). The images illustrate the standard fields captured during the screening process to allow for both automated AI analysis and clinical validation of the optic nerve head and macular region.

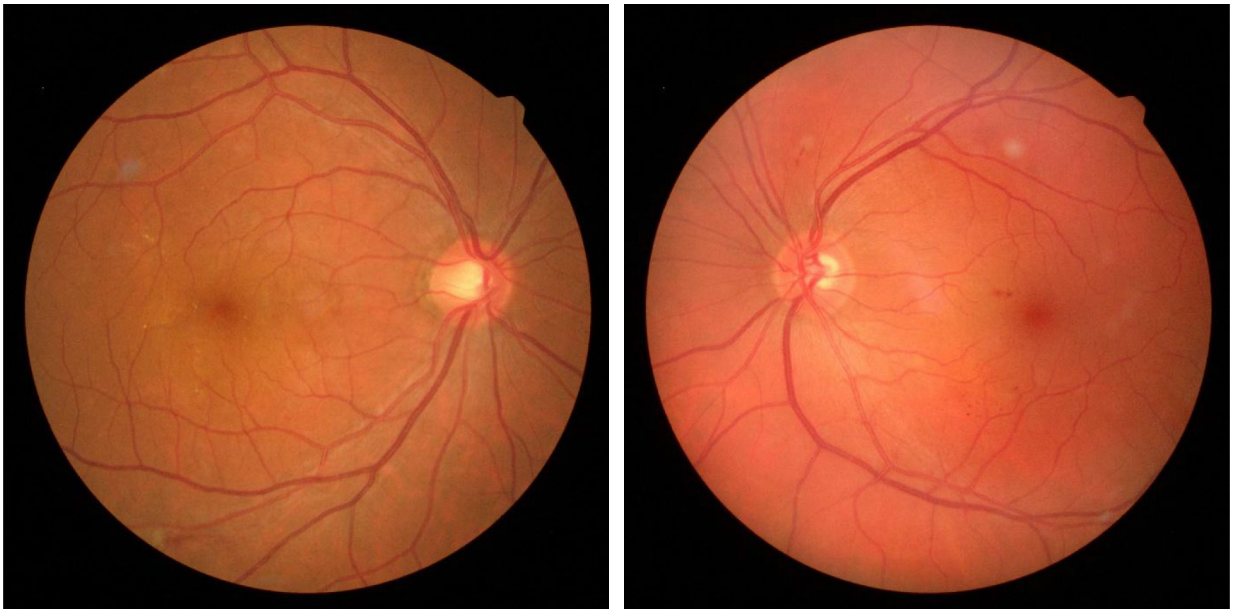

Fig. 3. Fundus colour photograph of a patient without any signs of DR (Left) and of a patient with Mild DR (Right). The right image demonstrates the presence of microaneurysms and small red macular dots, representing characteristic bulges in blood vessel walls identified by the DAIRET software during the screening phase.

Results

Patient characteristics

A total of 170 consecutive diabetic patients were initially enrolled in the study. Following the screening protocol, 7 subjects (4.1%) were excluded: 3 due to ungradable fundus images (media opacities) and 4 due to incomplete or dubious clinical records. This resulted in a final analytic sample of 163 participants. Mean age was 65.7 ± 12.6 years. Descriptive statistics for clinical variables are reported in Table 1.

Clinical features	Males			Females		
	Observed	Mean	SD	Observed	Mean	SD
Age	111	64.9	12.9	52	67.2	11.8
HbA1c (%)	110	7.2	1.2	52	7.0	1.0
Body mass index (kg/m^2)	108	28.0	4.7	52	28.3	5.6
Total cholesterol (mg/dL)	104	152.9	43.0	49	162.9	38.5
HDL cholesterol (mg/dL)	105	47.3	14.7	51	52.2	13.4
LDL cholesterol (mg/dL)	104	83.3	34.3	49	85.8	32.9
Triglycerides (mg/dL)	105	116.9	76.9	49	123.7	55.6
Serum creatinine (mg/dL)*	110	1.1	0.7	52	0.8	0.2
Glomerular filtration rate	110	86.1	25.3	52	83.1	20.2
AST (U/L)	97	22.5	8.2	45	21.6	6.1
ALT (U/L)	97	23.6	14.0	45	20.2	7.3
Systolic blood pressure (mmHg)	104	128.2	13.1	50	129.0	15.5
Diastolic blood pressure (mmHg)	104	73.1	8.0	50	74.6	9.1
*(eGFR, mL/min/1.73 m^2)						

Table 1.. Distribution of clinical characteristics according to gender.

DAIRET test parameters	Estimate	95% CI
Sensitivity (Se)	0.73	0.56 – 0.85
Specificity (Sp)	0.68	0.60 – 0.76
Positive Predictive Value (PPV)	0.37	0.26 – 0.49
Negative Predictive Value (NPV)	0.91	0.83 – 0.95
Accuracy	0.69	0.62 – 0.76
Positive Likelihood Ratio (LR+)	2.31	1.66 – 3.20
Negative Likelihood Ratio (LR−)	0.4	0.23 – 0.70

Table 2.. DAIRET performance.

Association between DAIRET and Expert diagnosis

A significant association was observed between DAIRET and expert diagnosis (χ^2 = 18.6, p < 0.001), with a Cramér's V of 0.34, indicating a moderate association.

Disease prevalence, according to expert assessment, was 20.2% (33/163).

Agreement analysis

Observed agreement between DAIRET and Expert diagnosis was 69.3%.

Agreement metrics were: Cohen's kappa = 0.30 (95% CI: 0.16–0.44), indicating *fair agreement*; and Gwet's AC1 = 0.47 (95% CI: 0.33–0.61), indicating *moderate agreement*.

Given the unbalanced prevalence, Gwet's AC1 was considered a more reliable measure of agreement than Cohen's kappa.

Diagnostic performance of the DAIRET test

Using expert opinion as the reference standard, the DAIRET test showed the following performance (Table 2):

These results indicate good sensitivity and high negative predictive value, suggesting that DAIRET is more effective at excluding diabetic retinopathy than at confirming it.

ROC analysis

ROC analysis yielded an AUC of 0.71, indicating acceptable overall diagnostic accuracy. The cut-off was 0.5, corresponding to a Sensitivity of 73% and a Specificity of 68%. Bootstrap validation confirmed the robustness of these estimates. ROC analysis stratified by gender (Fig. 4) showed no difference in performance. The AUC values have been: AUC_{Female} = 71.5%; AUC_{Male} = 69.3% (χ^2=0.04, p=0.84).

Discussion

The integration of Artificial Intelligence (AI) into medical screening represents a fundamental turning point in modern healthcare, acting as a powerful ally in improving the accuracy of early diagnosis. Its importance lies primarily in its ability to analyse huge amounts of data in real time. For example, in mammography screening, it improves breast cancer detection[8]. In recent years, several studies have been conducted using artificial intelligence (AI) to diagnose DR, demonstrating their efficiency, cost-effectiveness, and ease of use. Artificial intelligence (AI) is also revolutionizing diabetic retinopathy screening by rapidly analysing fundus images with

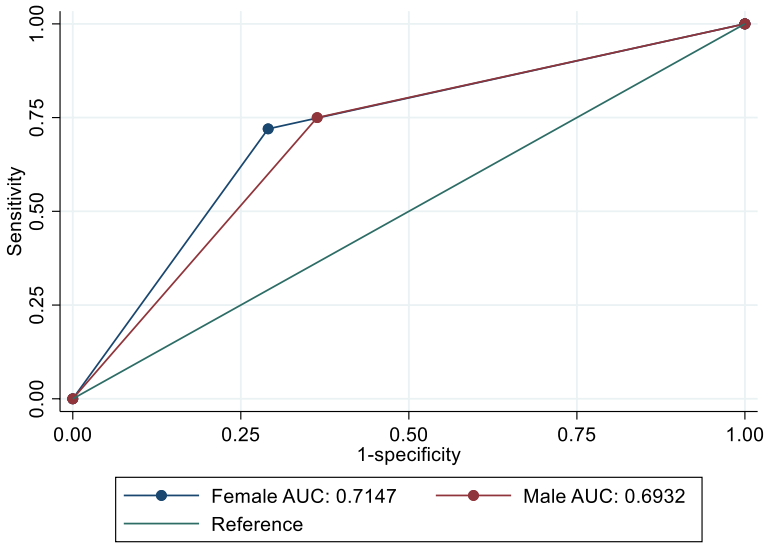

Fig. 4. Gender-Stratified ROC Curve Analysis. Comparison of AI diagnostic performance between Female participants and Male participants. The analysis highlights the software's consistency across genders, with an Area Under the Curve (AUC) of 71.5% for females and 69.3% for males (χ^2=0.04, p=0.84).

great accuracy, sensitivity, specificity for identifying diabetic retinopathy and related eye diseases and adequate levels of safety[15,22–24].

Artificial intelligence (AI) is revolutionizing diabetic retinopathy screening by rapidly analysing fundus images with high accuracy and safety. While this pilot study was conducted in a high-resource setting, the workflow efficiency observed supports the broader global application of such tools in resource-limited environments where specialist access is constrained. This is further supported by a recent scoping review highlighting the growing role of AI in diabetic eye care, particularly for screening populations at risk of sight loss in low-income and middle-income countries (LMICs)[25,26].

In our pilot study, the AI-based DAIRET test demonstrated moderate diagnostic accuracy for detecting diabetic retinopathy. The observed sensitivity (73%) and high negative predictive value (91%) indicate that the software performs acceptably in excluding disease, supporting its potential role as a preliminary first-line triage aid within organised diabetic retinopathy screening programmes. When placing these findings in the context of existing AI literature, large-scale clinical trials and validation studies—such as those by Gulshan et al.[18] and the pivotal trial by Abramoff et al.[15]—generally report higher diagnostic performance metrics, with sensitivities and specificities frequently exceeding 85–90%. Similarly, prospective real-world evaluations in national cohorts, such as the screening program in Thailand by Ruamviboonsuk et al.[27], and extensive diagnostic accuracy evaluations reviewed by Tufail et al.[28] and Xie et al.[29], demonstrate systems trained on larger database infrastructures that achieve high diagnostic metrics under heterogeneous field conditions.

The lower baseline sensitivity and specificity observed in our pilot study can be attributed to the preliminary sample size, real-world screening conditions, and the single-grader reference standard. However, the high NPV (90.8%) achieved by DAIRET matches the clinical requirements for entry-level public health screening workflows, where a primary objective is to safely identify low-risk individuals to reduce specialist burden and optimize referral pathways[28,29]. Conversely, the relatively low positive predictive value (37%) and modest positive likelihood ratio (LR+ = 2.31) represent a significant trade-off, underscoring that DAIRET cannot be utilized as a stand-alone confirmatory diagnostic test, but rather as an initial screening filter

The agreement analysis showed moderate concordance between DAIRET and expert diagnoses. Although Cohen's kappa suggests only fair agreement, Gwet's AC1 indicates a more robust level of agreement after accounting for the imbalanced disease prevalence (20.2%), highlighting the importance of selecting appropriate agreement metrics in diagnostic accuracy studies.

Subgroup analysis showed consistent diagnostic performance across genders, with an AUC of 71.5% for females and 69.3% for males. The lack of a statistically significant difference (p=0.84) suggests that the DAIRET algorithm maintains a stable diagnostic accuracy regardless of the patient's sex. This consistency is a relevant finding for the implementation of the tool in unselected populations within organized screening programs.

Overall, these results suggest that DAIRET may be effectively integrated into screening workflows to rule out diabetic retinopathy, reduce specialist workload, and prioritise referrals. In particular, for daily practical use, DAIRET is managed within the MètaClinic electronic medical record: DR classification data are available within a few minutes. Regarding data protection regulations, in-house data processing without use of an external cloud is a significant strength.

Undoubtedly, some limitations of this study should be noted. First, DAIRET can diagnose and grade DR

Through an AI algorithm, it is not suitable for some patients. For example, it is not possible to obtain fundus photographs from some DM patients due to small pupils or poor image quality caused by cataract opacity. Second,

larger, prospective studies are needed to validate these findings, optimise thresholds, and assess performance across diverse populations and clinical settings. This is especially true in population-based organised screening.

Third, the reference gold standard in this pilot study relied on the manual clinical evaluation of a single expert ophthalmologist. Although highly experienced, this approach introduces a potential limitation regarding subjective inter-observer variability, whereas a multi-reader consensus panel or central reading center would provide an even more robust reference standard for large-scale trials.

On the other hand, it should be emphasized that there are points highlighting the importance of AI alongside screening are: rapid and effective screening, especially distinguishing between healthy patients and those requiring specialist intervention, diagnostic support for ophthalmologists in managing large patient volumes, and improving access to care.

Conclusions

In conclusion, this pilot study demonstrates that the DAIRET software provides a reliable and consistent diagnostic output for diabetic retinopathy screening. The high Negative Predictive Value (90.8%) supports its use as an effective 'rule-out' tool, capable of identifying healthy individuals and reducing the workload for ophthalmologists. Furthermore, our findings indicate that the software's performance is stable across genders, with no statistically significant differences in diagnostic accuracy between male and female participants (p=0.84).

While further large-scale studies are needed to refine these results, this automated workflow represents a promising step toward more efficient and accessible screening programs, both in high-resource settings and in resource-constrained environments.

Data availability

The datasets generated and analysed during the current study are stored within the MètaClinic electronic medical record system at the University of L'Aquila. Due to privacy and data protection regulations regarding in-house processing, the data are not publicly available but can be made available from the corresponding author on reasonable request.

Received: 12 March 2026; Accepted: 11 June 2026
Published online: 16 June 2026

References

1. https://diabetesatlas.org/ (Latest access 13 January 2026).
2. Lin, K. Y., Guppy, F. H., Burgess, A. S. & Hanout, M. Systematic review of diabetes monitoring systems and their impact on patient outcomes. *Diabetes Spectr.* **29**(2), 92–8. https://doi.org/10.2337/diaspect.29.2.92 (2016).
3. Vujosevic, S. et al. Screening for diabetic retinopathy: new perspectives and challenges. *Lancet Diabetes Endocrinol.* **8**(4), 337–47. https://doi.org/10.1016/S2213-8587(19)30411-5 (2020).
4. Sivaprasad, S., Gupta, B., Crosby-Nwaobi, R. & Evans, J. Prevalence of diabetic retinopathy in various ethnic groups: a worldwide perspective. *Surv. Ophthalmol.* **57**(4), 347–70 (2012).
5. Ruta, L. M. et al. Prevalence of diabetic retinopathy in Type 2 diabetes in developing and developed countries. *Diabet Med.* **30**, 387–98 (2013).
6. Hex, Nick et al. Estimation of the direct health and indirect societal costs of diabetes in the UK using a cost of illness model. *Diabet Med.* **41**(9), e15326. https://doi.org/10.1111/dme.15326 (2024) (**Epub 2024 Jun 18**).
7. AAO PPP Retina/Vitreous Committee, Hoskins center for quality eye care. diabetic retinopathy PPP 2019. American Academy of Ophthalmology, 2019. (Latest access 7 December 2025).
8. Altobelli, E., Angeletti, P. M., Ciancaglini, M. & Petrocelli, R. The future of breast cancer organized screening program through artificial intelligence: a scoping review. *Healthc. Switz. Open Sour. Preview* **13**(4), 378 (2025).
9. Sheng, B. et al. An overview of artificial intelligence in diabetic retinopathy and other ocular diseases. *Front. Public Health* **10**, 971943. https://doi.org/10.3389/fpubh.2022.971943 (2022) (**eCollection 2022**).
10. Kubin, Anna-Maria. et al. Comparison of 21 artificial intelligence algorithms in automated diabetic retinopathy screening using handheld fundus camera. *Ann. Med.* **56**(1), 2352018. https://doi.org/10.1080/07853890.2024.2352018 (2024) (**Epub 2024 May 13**).
11. Sheng, B. et al. An overview of artificial intelligence in diabetic retinopathy and other ocular diseases. *Front. Public Health.* **28**(10), 971943. https://doi.org/10.3389/fpubh.2022.971943 (2022) (**eCollection 2022**).
12. Huang, X. et al. Artificial intelligence promotes the diagnosis and screening of diabetic retinopathy. *Front. Endocrinol.* **13**, 946915. https://doi.org/10.3389/fendo.2022.946915 (2022) (**eCollection 2022**).
13. Ciecierski-Holmes, Tadeusz, Singh, Ritvij, Axt, Miriam, Brenner, Stephan & Barteit, Sandra. Artificial intelligence for strengthening healthcare systems in low- and middle-income countries: a systematic scoping review. *NPJ Digit Med.* **5**(1), 162. https://doi.org/10.1038/s41746-022-00700-y (2022).
14. Bundesärztekammer, Kassenärztliche Bundesvereinigung, Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. Nationale VersorgungsLeitlinie Typ-2-Diabetes – Langfassung, Version 3.0. AWMF-Register-Nr. nvl-001. 2023. https://register.awmf.org/assets/guidelines/nvl-001l_S3_Typ-2-Diabetes_2024-12.pdf Stand: 15.06.2025.
15. Abràmoff, M. D. et al. Pivotal trial of an autonomous AIbased diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med.* **1**, 39. https://doi.org/10.1038/s41746-018-0040-6 (2018).
16. Piatti, A. et al. Feasibility and accuracy of the screening for diabetic retinopathy using a fundus camera and an artificial intelligence pre-evaluation application. *Acta Diabetol.* **61**(1), 63–68. https://doi.org/10.1007/s00592-023-02172-2 (2024) (**Epub 2023 Sep 7**).
17. Piatti, A. et al. Diabetic retinopathy screening with confocal fundus camera and artificial intelligence - assisted grading. *Eur. J. Ophthalmol.* **35**(2), 679–688. https://doi.org/10.1177/11206721241272229 (2025) (**Epub 2024 Aug 7**).
18. Gulshan, V. et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA.* **316**(22), 2402–10. https://doi.org/10.1001/jama.2016.17216 (2016).
19. Bellemo, V. et al. Artificial intelligence using deep learning to screen for referable and vision threatening diabetic retinopathy in Africa: a clinical validation study. *Lancet Digit Health.* **1**(1), e35-44 (2019).
20. Gwet, K. L. *Handbook of Inter-Rater Reliability* 4th ed. (Advanced Analytics, 2014).
21. Cicchetti, D. V. & Feinstein, A. R. High agreement but low kappa: II Resolving the paradoxes. *J. Clin. Epidemiol.* **43**(6), 551–8 (1990).
22. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics.* **33**(1), 159–74 (1977).

23. Ting, D. S. W. et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. **318**(22), 2211–23 (2017).
24. Pei, X. et al. Efficacy of artificial intelligence-based screening for diabetic retinopathy in type 2 diabetes mellitus patients. *Diabetes Res. Clin. Pract.* **184**, 109190. https://doi.org/10.1016/j.diabres.2022.109190 (2022).
25. Niemeijer, M., Abramoff, M. D. & van Ginneken, B. Fast detection of the optic disc and fovea in color fundus photographs. *Med. Image Annal.* **13**(6), 859–70. https://doi.org/10.1016/j.media.2009.08.003 (2009).
26. Cleland, C. R. et al. Artificial intelligence for diabetic retinopathy in low-income and middle-income countries: a scoping review. *BMJ Open Diabetes Res. Care.* **11**(4), e003424. https://doi.org/10.1136/bmjdrc-2023-003424 (2023).
27. Ruamviboonsuk, P. et al. Real-world performance of a deep-learning system for diabetic retinopathy screening in Thailand: a prospective cohort study. *Lancet Digit. Health.* **4**(6), e444–e453. https://doi.org/10.1016/S2589-7500(22)00070-4 (2022).
28. Tufail, A. et al. Automated diabetic retinopathy image assessment software: diagnostic accuracy and cost-effectiveness compared with human graders. *Ophthalmology.* **124**(3), 343–351. https://doi.org/10.1016/j.ophtha.2016.11.009 (2017).
29. Xie, Y. et al. Automated advanced technology for diabetic retinopathy screening: a systematic review and meta-analysis. *Lancet Digit. Health.* **2**(5), e254–e264. https://doi.org/10.1016/S2589-7500(20)30060-1 (2020).

Author contributions

EA designed the study, coordinated the research, and drafted the manuscript. MB and MC were responsible for clinical data collection and ophthalmological assessments. PC contributed to the clinical supervision. FM performed the statistical analysis, ROC curve modelling, and contributed to drafting the methodology. All authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Consent to participate

This study adhered to the tenets of the Declaration of Helsinki. The authorization protocol to process the data was obtained on 3 April 2024 (protocol number FFORIC24.01) by Institutional Review Board of Department of Life, Health and Environmental Sciences of the University of L'Aquila, Italy. In addition it was approved by official protocol for the implementation of tele-retinography for the early diagnosis of diabetic retinopathy in the Abruzzo Region, Italy. (https://bura.regione.abruzzo.it/sites/bura.regione.abruzzo.it/archivio_bura/2016/Speciale_63_1.html). All participants provided informed consent to participate in the screening program at the Eye Unit of San Salvatore Hospital, L'Aquila.

Consent for publication

The patients enrolled in the study provided consent for the use of their clinical data and fundus images for research and publication purposes. All data were anonymized prior to analysis to ensure patient privacy.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-58158-x.

Correspondence and requests for materials should be addressed to E.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.