Artificial intelligence-based glaucoma screening in primary care: a cross-sectional study and economic viability analysis of an ongoing trial

Afonso Lima-Cabrita, Zahi Wehbi, Joana Pargana, Rafael Whitfield, Vasco Lobo, Bernardo Monteiro, Joana Ferreira, Quirina Ferreira, Federico Felizzi, Matilde Ourique, Sanna Leinonen, Marta Pazos, Ana Miguel, Panayiota Founti, Rafael Barão, Sophie Lemmens, Anja Tuulonen, Evelien Vandewalle, Sofia Theriaga, Rodrigo Marques, Ruben Hemelings, Carlos Marques-Neves, Mun Faria, Ingeborg Stalmans, Luís Abegão Pinto

Summary

Background Glaucoma is a leading cause of blindness and is often undiagnosed. Owing to novel artificial intelligence (AI) technology, population-wide screening could now be possible. Our aim was to evaluate whether AI-based glaucoma screening is clinically and economically feasible within publicly funded primary care.

Methods This cross-sectional study and economic analysis was part of the ongoing Glaucoma Portugal Screening Trial (Glaucoma POST; NCT05875090). In this substudy, we used individual participant baseline data from a single primary care screening facility in Lisbon, Portugal. Participants aged 55–65 years, with and without diabetes, were randomly selected from primary care screening registries and invited for glaucoma screening (ie, fundus photography and intraocular pressure) through existing diabetic retinopathy infrastructure. We excluded people who had comorbidities precluding attendance (eg, bilateral blindness or severe mobility limitations). Fundus photographs were analysed with an AI algorithm (MONA-GLC) to generate a glaucoma risk score. Referral to hospital for specialist evaluation was triggered by either a positive AI result or an intraocular pressure of 24 mm Hg or higher. Images were also independently graded by six glaucoma experts. Referred participants underwent visual field testing; glaucoma was diagnosed with adapted Thessaloniki Eye Study criteria. We assessed the feasibility, diagnostic performance, and cost-effectiveness over a 10-year horizon of integrating AI-based glaucoma screening into an existing diabetic retinopathy screening programme within a real-world primary care setting. We compared AI-based screening with standard care (ie, patients referred from primary care to hospital-based general ophthalmologists, with onward referral to glaucoma specialists as appropriate).

Findings Screening and recruitment was done between March 1 and Dec 31, 2023. Of 1038 invited individuals, 671 (65%) attended screening. After excluding 42 participants who missed visual field testing, 629 were included in the performance and prevalence analyses. Glaucoma was diagnosed in 40 of these 629 participants (6% [95% CI 5–9%]). From the 671 people screened, the AI algorithm referred 66 participants (10%), compared with 118 referrals (18%) through the adjudicated expert assessment. AI-algorithm sensitivity was 78% (95% CI 62–89%) and specificity 95% (93–97%). In a setting in which minimal staff training was required, AI-enhanced glaucoma screening achieved an incremental cost-effectiveness ratio of €1725 per quality-adjusted life-year (76·3% probability of cost-effectiveness at a threshold of €20 000 per quality-adjusted life-year) at 1% prevalence, becoming cost-saving at prevalences of 2% or more.

Interpretation Our results suggest that AI-enabled screening in primary care can be cost-effective and can reduce unnecessary expert referrals, even though a one-off screening round might not capture every case. AI could enhance efficiency and detection, enable earlier treatment, and prevent avoidable blindness.

Funding None.

Lancet Prim Care 2026

Published Online
https://doi.org/10.1016/j.lanprc.2026.100107

Department of Ophthalmology, Unidade Local de Saúde Santa Maria, Lisbon, Portugal (A Lima-Cabrita MD, J Pargana MD, R Whitfield MD, V Lobo MD, B Monteiro MD, J Ferreira MD PhD, Q Ferreira PhD, R Barão MD, C Marques-Neves MD PhD, M Faria MD PhD, L Abegão Pinto MD PhD); **School of Medicine, University of Lisbon, Lisbon Academic Medical Centre, Lisbon, Portugal** (A Lima-Cabrita, J Pargana, R Whitfield, V Lobo, B Monteiro, J Ferreira, Q Ferreira, R Barão, C Marques-Neves, M Faria, L Abegão Pinto); **Visual Sciences Study Centre, Faculty of Medicine, University of Lisbon, Lisbon, Portugal** (A Lima-Cabrita, J Pargana, R Whitfield, V Lobo, B Monteiro, J Ferreira, Q Ferreira, R Barão, C Marques-Neves, M Faria, L Abegão Pinto); **Research Group Ophthalmology, Department of Neurosciences, KU Leuven, Leuven, Belgium** (Z Wehbi MD, S Lemmens MD PhD, E Vandewalle MD PhD, I Stalmans MD PhD); **Italian Society for Artificial Intelligence in Medicine, Rome, Italy** (F Felizzi PhD, R Marques MD); **Unidade Saúde Publica Mafra, Unidade Local de Saude Santa Maria, Lisbon, Portugal** (M Ourique MD); **Tays Eye Centre, Tays, Tampere, Finland** (S Leinonen MD PhD, A Tuulonen MD PhD); **Department of Ophthalmology, Hospital Clinic of Barcelona, Barcelona, Spain** (M Pazos MD PhD); **Department of Ophthalmology, Private Hospital de la Baie, Avranches, France** (A Miguel MD PhD); **Moorfields Eye Hospital NHS**

Introduction

Glaucoma is a chronic disease and among the leading causes of irreversible blindness worldwide, affecting an estimated 76 million individuals.[1] The major public health effect of glaucoma stems from a long asymptomatic phase, resulting in substantial underdiagnosis and delayed treatment.[2] Even in high-income countries, up to 50% of individuals with glaucoma remain undiagnosed, frequently presenting with moderate to advanced disease at first detection.

Despite its substantial disease burden, glaucoma detection currently relies largely on opportunistic case finding.[3] Modelling studies suggest that screening could substantially reduce glaucoma-related visual impairment and blindness; however, important barriers have restricted implementation of population-wide programmes.[4] These barriers include the need for specialised diagnostic equipment and trained personnel, particularly in low-income and middle-income countries,[5,6] and the intrinsically low

Foundation Trust, London, UK (P Founti MD PhD); **Institute of Ophthalmology, University College London, London, UK** (P Founti); **Department of Ophthalmology, UZ Leuven, Leuven, Belgium** (S Lemmens, E Vandewalle, I Stalmans); **Population-Based Screening Management, National Health Service, Lisbon, Portugal** (S Theriaga MBA); **Singapore Eye Research Institute, Singapore National Eye Centre, Singapore** (R Hemelings PhD)

Correspondence to: Prof Luís Abegão Pinto, Department of Ophthalmology, Unidade Local de Saúde Santa Maria, Lisbon 1649-028, Portugal **abegao.pinto@ulssm.min-saude.pt**

Research in context

Evidence before this study

We searched PubMed, Scopus, and Google Scholar from database inception to Jan 1, 2023, for studies on glaucoma screening, artificial intelligence, fundus imaging, and cost-effectiveness, including articles and abstracts published in English, Portuguese, and French. We identified no systematic reviews or meta-analyses addressing the real-world implementation of artificial intelligence (AI)-based glaucoma screening. There were no previous studies prospectively testing AI glaucoma screening in primary care. Most existing evidence consists of retrospective analyses of curated fundus image datasets, prone to multiple sources of bias and poor generalisability.

Added value of this study

This study extends existing evidence by evaluating AI-enabled glaucoma screening within a taxpayer-funded and universal national health service, supporting generalisability beyond retrospective curated datasets. Implementation in services operated by trained non-physician personnel (eg, certified orthoptists and general practice clerks) provides valuable information for ongoing debates on task shifting and scalability in routine primary care. Benchmarking against a criteria-based glaucoma diagnosis established by expert glaucoma ophthalmologists clarifies AI's role as a standardised triage tool within real-world referral pathways and strengthens the evidence base for health system decision making in integrated screening programmes. By incorporating a cost-effectiveness evaluation, our study also informs on whether adoption of AI-enabled screening represents efficient use of health-system resources under real-world delivery conditions.

Implications of all the available evidence

Taken together, existing evidence and our findings suggest that AI-enabled glaucoma screening could be implemented within public primary care services as standard practice, shifting case finding closer to patients while preserving specialist capacity for confirmatory diagnosis and treatment. For policy, integration into established public screening pathways (eg, diabetic retinopathy programmes) with trained non-physician personnel could improve referral efficiency, but should be accompanied by clear governance, quality assurance, and equity-focused access planning. Future research should assess multicentre scale-up in primary care, long-term outcomes and service effects, patient-reported outcomes, performance across subgroups, and post-deployment performance surveillance.

positive predictive value of screening tests in a low-prevalence disease, which challenges cost-effectiveness.[7] As a result, most public health authorities do not currently endorse universal population-based glaucoma screening.[8] Instead, targeted or opportunistic screening of groups at high risk of glaucoma has been proposed as a more pragmatic strategy.[4]

Advances in artificial intelligence (AI) have renewed interest in scalable glaucoma screening approaches. AI systems based on optic disc-centred colour fundus photographs, acquired with relatively low-cost equipment and operated by trained non-physicians, have shown promising diagnostic performance in retrospective studies.[9–13] Such approaches could help overcome key barriers to implementation by reducing reliance on specialist expertise while enabling closer-to-patient risk stratification at scale, particularly in a primary care setting.

Informed by established population screening frameworks, including WHO screening criteria, glaucoma meets key requirements for screening, such as high disease burden, a long preclinical phase, availability of non-invasive tests, and effective treatments that slow progression when initiated early.[14] Nevertheless, implementation has been hindered by the absence of scalable, primary-care-ready diagnostic pathways and little real-world evidence on feasibility and cost-effectiveness, particularly in settings where specialist access is scarce.[12]

Many health-care systems have already established screening programmes for diabetic retinopathy, supported by existing infrastructure such as fundus cameras and trained personnel.[15,16] The use of this infrastructure for opportunistic glaucoma screening could represent an efficient and cost-effective implementation strategy, enabling integrated screening for the two most common causes of irreversible blindness.[17,18]

In this study, we aim to assess the feasibility, diagnostic performance, and cost-effectiveness of integrating AI-based glaucoma screening into an existing diabetic retinopathy screening programme within a real-world primary care setting.

Methods

Study design and population

In this cross-sectional study of an ongoing glaucoma screening trial (Glaucoma POST; NCT05875090), we evaluated diagnostic test accuracy through baseline participant data and did an economic analysis. Patient screening registries from primary care centres in the Santa Maria Local Health Unit area (Lisbon, Portugal) were reviewed to identify individuals aged 55–65 years. We chose this age range based on evidence from the Early Manifest Glaucoma Trial,[19] which identified individuals aged 60–65 years as an optimum target group for glaucoma screening sensitivity during the preclinical detectable phase. The lower age limit was extended to 55 years to enable assessment across a 10-year continuum and inform future refinement of age-based screening criteria.[19] Two cohorts were included: participants with diabetes undergoing routine diabetic retinopathy screening, in whom glaucoma screening was added, and participants without diabetes invited specifically

for glaucoma screening. Exclusion criteria included inability to provide informed consent or comorbidities precluding attendance (eg, bilateral blindness or severe mobility limitations). Eligible participants were randomly selected and invited by telephone by a general practice clerk, who explained the study objectives. Written informed consent was obtained before testing and explained the use of AI-based image assessment, possible screening outcomes, and referral pathways (appendix pp 2–4).

This study adhered to the Declaration of Helsinki and was approved by the local ethics committee (Comissão de ética do CAML number 271/22). Our multidisciplinary research team encompasses expertise ranging from the oversight and management of primary care screening programmes in the Lisbon area to a panel of experienced glaucoma clinician-researchers. The study was done within the Portuguese National Health Service, a tax-funded, universal health-care system in which primary care serves as the main entry point and coordinates access to specialist ophthalmology services.

The main trial was registered on ClinicalTrials.gov (NCT05875090) and is still recruiting. Participants will be invited for repeat screening, including intraocular pressure, face-to-face evaluation, and AI-derived glaucoma risk score, at intervals of 1–5 years based on their baseline risk assessment.

Procedures

First, a general practice clerk administered a demographic and socioeconomic questionnaire (appendix p 5) and participants self-reported both their sex (male or female) and ethnicity. This questionnaire enabled characterisation of the study population to support evaluation of model performance and factors associated with screening acceptability. Screening visits were done at a single designated primary care screening centre and included optic disc-centred colour fundus photography and intraocular pressure measurement.

Fundus imaging used a non-mydriatic fundus camera (Canon CR-2 Plus, Japan). Images were analysed with an AI-based glaucoma risk model (MONA-GLC, Belgium), which generates a continuous glaucoma risk score (range 0–1) based solely on image-derived features and independent of intraocular pressure. Images deemed of insufficient quality for automated analysis were flagged by the algorithm and recorded. Intraocular pressure was measured through rebound tonometry (iCare IC200, Finland) for all participants, with a mean of five measurements recorded. Screening procedures (photography and tonometry) were done by a certified orthoptist with previous experience in diabetic retinopathy screening, with support from a general practice clerk for schedule management and patient contact. There was a single day of study-specific training, covering fundus image acquisition, tonometry, AI score registration, and basic image quality assessment. Measurement tools were chosen for their availability and feasibility in primary care and for their ability to clearly inform referral decisions with non-invasive imaging. Participants were referred for hospital-based evaluation if the AI risk score exceeded the prespecified threshold of 0·73 or if intraocular pressure was 24 mm Hg or greater. The referral threshold was defined a priori based on external validation datasets and was not recalibrated with data from the present study.

Referred participants underwent specialist evaluation including slit-lamp examination, standard automated perimetry (SITA-Fast 24-2 protocol Humphrey, Zeiss, Germany), and optic disc and macular optical coherence tomography (OCT; SPECTRALIS Glaucoma Module, Heidelberg, Germany). Treatment decisions were left to the discretion of the treating ophthalmologist.

See Online for appendix

Glaucoma diagnosis was established with an adapted Thessaloniki Eye Study[20] criteria by two ophthalmologists and glaucoma specialists who jointly reviewed fundus images and visual fields while masked to referral origin (by AI and intraocular pressure screening or the human expert validation pathway). We applied both diagnostic criteria as used in the original Thessaloniki Eye Study to diagnose glaucoma (panel). Glaucoma severity was classified according to Ocular Hypertension Treatment Study criteria based on visual field defect severity.[21] OCT findings were not used as defining diagnostic criteria but were collected to document structural features.

We did an independent validation exercise with an expert panel of international ophthalmologists who are glaucoma specialists and they assessed fundus photographs from the entire screened cohort. After a calibration phase with a pilot set of images, six experts with the highest interobserver agreement for vertical cup-to-disc ratio estimation (intraclass correlation coefficient >0·8) were selected and organised into pairs. The dataset was divided into three subsets, with both eyes from each participant assessed together. Experts independently reviewed fundus photographs while masked to all clinical data, including intraocular pressure measurements and AI outputs. For each eye, graders estimated the probability of glaucoma (as a percentage), expected visual field status, and anticipated OCT findings based on optic nerve head appearance. Referral decisions were binary (yes or no; appendix p 6). Disagreements were resolved by adjudication from a third expert. Participants flagged by experts but not by the AI-based pathway were recalled for clinical assessment.

For AI–expert performance comparisons, analyses were restricted to referrals generated by the AI algorithm, which is based exclusively on fundus image-derived risk scores and does not incorporate intraocular pressure. Both AI and expert referrals were evaluated against the same reference standard.

The AI-based glaucoma risk model (MONA-GLC) used in this study is derived from the G-Risk framework described by Hemelings and colleagues.[12] The model was developed through supervised deep learning trained on large, annotated datasets of optic disc-centred colour fundus photographs and was designed to estimate the probability

Panel: **Adapted Thessaloniki Eye Study criteria**

Definition 1

The individual has at least one eye with both:

Glaucomatous optic disk: either thinning or notching of the optic disk rim, or asymmetry between the two eyes of more than 0·2 cup-to-disk ratio

Glaucomatous visual field defect: three contiguous points more than 5 dB depressed, or one point more than 10 dB depressed and either a glaucoma hemifield test index outside the typical limits or visual field findings only with typical characteristics of glaucoma damage (ie, nasal step or arcuate defect)

Definition 2

The individual has at least one eye with either:

- Visual field findings only with typical characteristics of glaucoma damage (ie, nasal step or arcuate defect).
- Optic nerve damage only: thinning or notching of the optic disk rim combined with matching asymmetry between the two eyes of more than 0·2 cup-to-disk ratio.
- Optic nerve damage and raised intraocular pressure: thinning or notching of the optic disk rim combined with matching intraocular pressure of 24 mm Hg or more.

These definitions were considered when the patient met reliability criteria (false-positive errors were <33%, false-negative errors were <33%, and fixation losses were <33%). A label of non-glaucoma visual field loss was applied when visual field defect was considered not secondary to glaucoma but could be explainable by other ophthalmic diseases or by other abnormalities such as lid ptosis and media opacities. When participants presented with incomplete glaucoma criteria, such as unilateral thinning without matching asymmetry, they were labelled as having incomplete glaucoma criteria. If no anomalous findings were found, participants were labelled as not having glaucoma.

The adaptations from the original Thessaloniki Eye Study criteria[20] were:

- Allowing visual field findings only with typical characteristics of glaucoma damage (to represent a "Glaucomatous visual field defect" instead of "CPSD in the 30–2 full-threshold test or PSD in the SITA-standard C-30 test with a p value index <0·05".
- Excluded the definition "incomplete data set but clinically believed to be glaucoma (home visits, unable or unreliable visual field test secondary to low vision or other medical impairment, unable to perform biomicroscopy because of severe media opacities prohibiting fundus visualisation".
- Excluded the definition "thinning or notching of the optic disk rim or asymmetry between the two eyes of more than 0.2 cup-to-disk ratio combined with: [...] b. History of high IOP currently controlled under medical or surgical treatment".

CPSD=corrected pattern standard deviation. IOP=intraocular pressure. PSD=pattern standard deviation. SITA=Swedish interactive threshold algorithm.

of glaucomatous optic neuropathy based solely on image-derived features. Before the present study, the model underwent extensive external validation across multiple independent population-based and screening-oriented datasets, encompassing a wide range of imaging conditions, disease prevalence, and demographic characteristics. These evaluations were used to characterise model discrimination and to define operating points relevant to screening applications, including referral thresholds.

One of the datasets used for external benchmarking was the Artificial Intelligence for Robust Glaucoma Screening (AIROGS) challenge dataset, which consists of colour fundus photographs collected within real-world screening programmes.[22] Within this benchmarking framework, the model showed diagnostic discrimination within the range considered suitable for glaucoma screening, consistent with screening-relevant performance envelopes reported across multiple AIROGS submissions. Importantly, these benchmarking results were used exclusively to contextualise threshold selection and anticipated screening performance rather than to optimise or retrain the model for the present cohort.[12] The referral threshold applied in this study (risk score ≥0·73) was fixed before study initiation, based on external validation datasets, including AIROGS[22] and the Gutenberg Health Study.[23] No model retraining, fine-tuning, or threshold adjustment was done with data from the present screening population.

A random 13% sample of participants who had both glaucoma and diabetic retinopathy screening was selected, and diabetic retinopathy screening results were additionally reported. Retinopathy was graded from fundus photographs acquired with the same fundus camera as for glaucoma screening.

Outcomes

In this cross-sectional baseline report, the main outcome was the glaucoma referral rate in the intention-to-treat screened population (all participants who attended the primary care screening visit and had fundus photography and intraocular pressure measurement), reported separately for the AI-based pathway (irrespective of intraocular pressure) and for the adjudicated expert grading pathway

(binary referral based on masked fundus photograph review with adjudication). Additional outcomes for this work were AI diagnostic performance; glaucoma and ocular hypertension prevalence; subgroup analyses by sex, ethnicity, and diabetes status; multivariable predictors of glaucoma diagnosis; expert-panel referral outcomes and diagnostic performance against the same reference standard; attendance patterns; pathway timing and circuit times; exploratory diabetic retinopathy findings in a random subsample; and cost-effectiveness over a 10-year horizon, including prespecified scenario and sensitivity analyses (including a societal perspective). All analyses that required glaucoma diagnosis were done per protocol, because glaucoma status could not be ascertained in participants who did not attend visual field testing. Beyond these outcomes, analyses from the main ongoing trial will have diagnostic agreement between referring decision and reading centre decision as the primary outcome and will include patient-reported outcome metrics, such as the Patient Satisfaction Questionnaire Short Form[24] and staff-reported outcomes (Intervention Appropriateness Measure).[25]

Statistical analysis

Based on previously published external validation of the AI model (expected sensitivity 90% and specificity 95%), assuming a glaucoma prevalence of 5%, 95% CIs, and a 10% margin of error, the required sample size for this analysis was estimated at 673 participants. The sample size was designed to ensure adequate precision around screening-level specificity and referral rates rather than to detect small differences in sensitivity.

We developed a state-transition (Markov) cohort model to compare AI-based glaucoma screening with standard care with multiple time horizons (5–30 years). Analyses were done from a health-care sector perspective in the base case, with a societal perspective explored in sensitivity analyses. Outcomes were costs, quality-adjusted life-years (QALYs), and incremental cost-effectiveness ratios (ICERs). Costs were expressed in 2024 euros and represent means across five European countries (Portugal, Germany, Belgium, Austria, and Italy). Economic robustness was assessed with Portuguese cost variables and utility values derived from EQ-5D-3L population norms. Operational costs were derived from equipment depreciation and personnel salaries distributed across annual screening volumes. These cost inputs reflect implementation costs associated with AI-assisted image analysis rather than proprietary valuation of the algorithm itself. The use of a free research licence for the present study did not influence cost assumptions, which reflect anticipated real-world implementation, rather than study-specific, pricing arrangements. Discounting assumptions and full model details are provided in the appendix (pp 7–38).

All analyses were done at the participant level unless otherwise specified. Continuous variables are reported as means (SD) or medians (IQR) and categorical variables as counts and percentages. Subgroup comparisons were done through Fisher's exact test or Wilcoxon rank-sum tests, as appropriate. AI diagnostic performance was assessed through sensitivity, specificity, Cohen's κ, receiver operating characteristic curves, and area under the curve (AUC), with comparisons done with DeLong's test and was benchmarked against Thessaloniki Eye Study criteria. A confusion matrix between AI referrals and glaucoma diagnosis was drawn. Multivariable logistic regression was used to identify factors associated with glaucoma diagnosis. Clinically relevant predictors were prespecified (age, sex, intraocular pressure, AI risk, diabetes diagnosis, and cup-to-disk ratio), and variables meeting $p<0{\cdot}10$ in univariable analyses were entered into the multivariable model. We accounted for within-participant correlation with cluster-robust standard errors. Participants missing visual field testing were excluded from primary analyses.

Statistical analyses were done in R (version 4.4.0).

Role of the funding source

There was no funding source for this study.

Although two authors (IS and RH) are associated with development of the AI model and provided technical advisory input, they had no role in study design, participant recruitment, data collection, outcome adjudication, statistical analysis, interpretation of results, or manuscript drafting. All analyses were done by investigators without commercial ties to the AI model, and all clinical diagnoses were established independently of the AI outputs.

Results

We recruited and screened participants from March 1 to December 31, 2023. The figure shows the number of individuals contacted, those who agreed to participate, and those ultimately included in the primary analysis. Table 1 summarises the baseline characteristics of the 671 screened participants. The appendix (p 39) illustrates the distributions of intraocular pressure and AI-derived glaucoma risk scores.

We assessed referral outcomes generated by the screening protocol. With the combined AI-based risk score and intraocular pressure screening protocol, 82 (12%) of 671 participants were referred for face-to-face evaluation. In referred participants, the mean intraocular pressure was 16·9 mm Hg (SD 6.1) and the mean AI-derived risk score was 0·74 (0·10). Of these 82 referrals, 62 (76%) were based solely on an elevated AI risk score (≥0·73), 16 (20%) on elevated intraocular pressure alone (≥24 mm Hg), and 4 (5%) on both criteria. Six participants reported a previous diagnosis of glaucoma at the time of screening. Three of these people met referral criteria under the screening protocol. Of all those that self-reported glaucoma, only one was ultimately diagnosed with glaucoma by our screening study. A total of 589 participants were neither referred by the AI-based pathway nor by intraocular pressure criteria. The AI algorithm flagged one eye as ungradable in eight participants; no participant had both eyes rejected, and therefore no participants were excluded due to image quality. Subsequent expert review of these cases did not

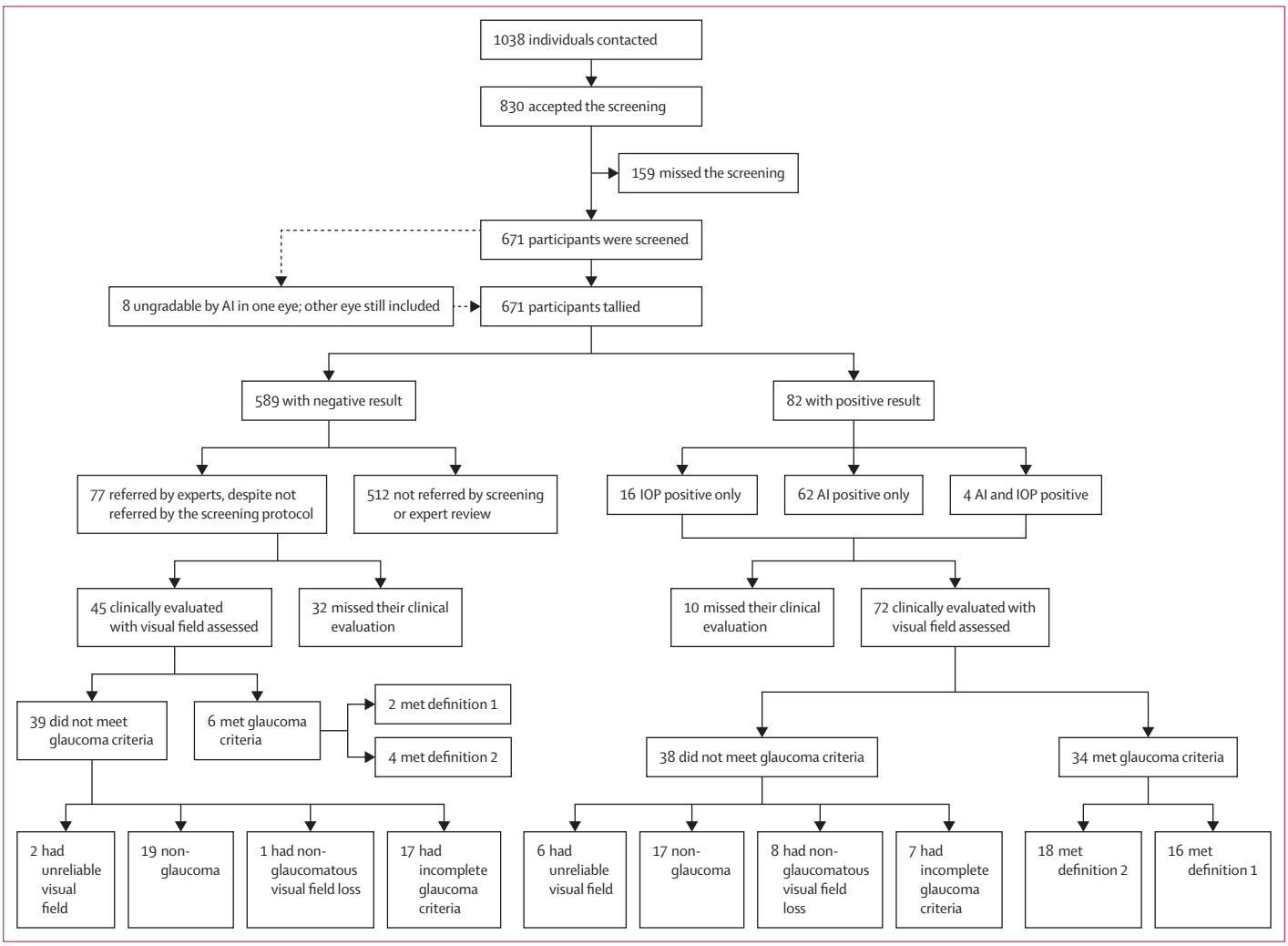

Figure: **Study participant flow diagram**
The diagram shows both the full cohort of people invited, accepted, screened, and with visual field assessment and those who were referred to face-to-face further assessment with visual fields. Glaucoma definition 1 is more stringent, requiring both characteristic optic nerve head changes and visual field, whereas definition 2 is more lenient, allowing for either optic nerve head or visual field if they are suggestive enough. These definitions are further detailed in the panel. Due to 42 participants (10 AI or IOP referred and 32 expert referred) not attending their visual field assessment, 629 of the 671 participants screened are used for analysis dependent on glaucoma diagnosis and prevalence. AI=artificial intelligence. IOP=intraocular pressure.

identify any additional glaucoma referrals. For the primary outcome (referral rate in the screened population), the AI pathway referred 66 (10%) of 671 participants. Of these, 58 (88%) attended visual field testing, and 31 (53%) of these 58 people met diagnostic criteria for glaucoma. Among the 16 participants referred due to elevated intraocular pressure alone, 14 (88%) underwent visual field testing, of whom three (21%) were diagnosed with glaucoma.

Human expert review identified 77 additional participants who were not referred by either the AI-based pathway or intraocular pressure criteria. Of these, 45 (58%) completed visual field testing, leading to the identification of six (13%) additional patients with glaucoma. Label distributions across referral pathways are detailed in the appendix (p 40). A flow diagram of positive screening outcomes is shown in the figure.

Glaucoma and ocular hypertension prevalence were assessed in the per-protocol cohort of participants who completed the reference standard clinical evaluation including visual field testing (n=629). This cohort includes participants recalled for evaluation after the expert-panel photograph review (ie, expert-flagged participants who were not referred by the AI and intraocular pressure pathway but underwent visual field testing). The prevalence of glaucoma was 6% (40/629; 95% CI 5–9%), and the prevalence of ocular hypertension was 3% (20/629; 95% CI 2–5%). In participants with diabetes, glaucoma prevalence was 7% (35/532; 95% CI 5–9%) and ocular hypertension prevalence was 3% (14/532; 95% CI 1–4%). In participants without diabetes, glaucoma prevalence was 5% (5/97; 95% CI 2–12%) and ocular hypertension prevalence was 4% (4/97; 95% CI 1–10%). In most participants for whom

	Cohort, n=671
Age, years	61·3 (58·7–63·9)
Self-reported sex	
Male	370 (55%)
Female	301 (45%)
Self-reported ethnicity	
White	641 (96%)
Black	28 (4%)
Asian	1 (<1%)
Native American	1 (<1%)
Reported previous glaucoma diagnosis	
Yes	6 (1%)
No	665 (99%)
Reported previous diabetes diagnosis	
Yes	564 (84%)
No	107 (16%)
Intraocular pressure, mm Hg	15·0 (3·9)
AI algorithm glaucoma risk value	0·57 (0·11)

Data are n (%), median (IQR), or mean (SD). AI=artificial intelligence.

Table 1: **Demographic characteristics of screened cohort**

glaucoma was detected, it was classified as mild (30/40; 75%). Five (13%) participants had moderate glaucoma, and five (13%) had advanced glaucoma in at least one eye. Structural and functional characteristics of subgroups are summarised in table 2.

An exploratory diabetic retinopathy assessment was done in a randomly selected subset of 76 participants. Two (3%) participants showed referable diabetic retinopathy (both graded as R2). One participant had background diabetic retinopathy (R1), two were unclassifiable, and the remaining 71 (93%) participants showed no evidence of diabetic retinopathy.

We did exploratory analyses of factors associated with screening attendance in a sample of 720 (69·4%) of the total 1038 invited individuals from a single primary care centre, of whom 343 (47·6%) attended screening. Attendance was not associated with sex (no missingness; males: 180 [47%] of 381 attended; females: 176 [52%] of 339 attended; p=0·82), age (no missingness; attended: median age 61 years [IQR 59–63]; did not attend: median age 61 years [58–63]; p=0·78), or family history of glaucoma (missingness: 335 [47%] of 720; family history of glaucoma: 34 [85%] of 40 attended; no family history of glaucoma: 282 [82%] of 345 attended; p=0·61). Attendance was significantly higher in participants with diabetes than those without diabetes (missingness: 301 [42%] of 720; patients with diabetes: 232 [86%] of 271; patients without diabetes: 110 [74%] of 148; p=0·0044)). Across professional groups (missingness: 344 [48%] 720), there was a statistically significant difference in attendance (p=0.0063; appendix p 41). Of note, the highest non-attendance rate (9 [47%] of 19) was observed in the managers group, while the armed forces group had no refusals (6 [100%] of 6). Participants with more than 12 years of education were more likely to miss their screening appointment than those with 5–12 years of education (missingness: 333 [46%] of 720; >12 years of education: 97 [77%] of 126; 5–12 years of education: 162 [87%] of 187; p=0·033; appendix p 41). The Appendix (p 41) reports the attendance of those with less than 5 years of education. The most frequently reported reason for declining participation was ongoing follow-up with an eye-care provider outside the public health-care system (appendix p 41). Of note, the source of missingness stems from participants who were contacted and did not answer, did not accept the screening invitation, or gave incomplete information. The median time from invitation to screening was 1 day (IQR 1–2). For those screened positive, the median time from screening to visual field testing was 45 days (IQR 34–60). In participants who did not attend the hospital-based evaluation after referral, no significant differences in baseline characteristics were observed compared with those who did attend, aside from the referral modality itself (AI-based, intraocular pressure-based, or expert-based; appendix p 42).

We assessed visit duration and operational timing of the screening protocol. Median screening duration was 5·1 mins (IQR 4·0–6·2) in participants without diabetes. In participants with diabetes, adding glaucoma screening increased the median visit duration by 2·0 mins (1·8–2·4). No additional staffing was required, and no major operational barriers were identified beyond non-attendance.

As a prespecified secondary outcome, we assessed the diagnostic performance of the AI algorithm against the reference standard in the per-protocol cohort. We evaluated the diagnostic performance of the AI algorithm against the reference standard. Across all referral pathways (AI-based, intraocular pressure-based, and expert-based), 159 participants were referred for further face-to-face evaluation. Fundus photographs from all 671 screened participants were analysed by the AI-based system. Of these 159 referred participants, 42 (26%) did not attend hospital-based evaluation and were excluded from the primary diagnostic performance analysis, leaving 629 participants (figure). According to the reference standard diagnosis, the AI-based pathway would have referred 31 (78%) of 40 participants diagnosed with glaucoma while missing nine patients (23%) with glaucoma. A confusion matrix of AI performance is shown in the appendix (p 43), and summary performance metrics are provided in table 3 and the appendix (p 44). Subgroup analyses showed a higher receiver operating characteristic AUC in participants with diabetes (0·913; 95% CI 0·810–0·956) than in those without diabetes (0·852; 95% CI 0·688–1·000; p=0·49), and in males (0·943; 95% CI 0·889–0·996) than in females (0·864; 95% CI 0·780–0·948; p=0·12), although these differences were not statistically significant. Due to little ethnic diversity (641 [96%] of 671 were White), performance metrics by ethnicity could not be assessed. When a less stringent reference definition was applied (incomplete glaucoma criteria; panel), discriminatory performance remained similar (AUC 0·903; 95% CI 0·865–0·939), although the

	Met glaucoma criteria		Did not meet glaucoma criteria				
	AI algorithm referred, n=31	Not AI algorithm referred, n=9	Incomplete glaucoma criteria, n=24	Non-glaucoma visual field loss, n=9	Unreliable visual field, n=8	Non-glaucomatous and expert referred, n=19	Non-glaucomatous and AI algorithm or intraocular pressure referred, n=17
Affected eyes	42	12	32	14	10	38	34
Age, years	61 (59 to64)	61 (59 to 64)	62 (59 to 64)	62 (60 to 63)	61 (58 to 62)	61 (58 to 63)	61 (59 to 64)
Diabetes	28 (90%)	7 (78%)	21 (88%)	8 (89%)	7 (88%)	13 (68%)	9 (53%)
p value	0·0091	0·40	0·029	0·098	0·18	0·50	..
Bilateral glaucoma	11 (35%)	3 (33%)	..	..	..	..	..
Intraocular pressure, mm Hg	14·7 (4·7)	17·1 (5·3)	15·4 (3·3)	15·6 (6·0)	19·0 (6·4)	13·0 (3·2)	15·8 (7·2)
p value	0·64	0·22	0·87	0·95	0·22	0·38	..
AI glaucoma risk	0·79 (0·04)	0·62 (0·08)	0·69 (0·08)	0·69 (0·11)	0·62 (0·15)	0·58 (0·08)	0·62 (0·14)
p value	<0·0001	0·27	0·063	0·16	0·81	0·081	..
Central corneal thickness, µm	539 (35)	549 (34)	548 (45)	564 (41)	550 (56)	558 (61)	545 (37)
p value	0·32	0·91	0·74	0·18	0·23	0·58	..
Spherical equivalent, DS	–0·60 (2·47)	0·52 (1·65)	–0·99 (2·09)	0·13 (1·68)	–0·03 (2·05)	1·57 (0·72)	0·35 (3·20)
p value	0·010	0·42	0·0053	0·25	0·17	0·40	..
Visual acuity, logMAR	0·10 (0·00 to0·10)	0·00 (0·00 to 0·10)	0·10 (0·00 to 0·20)	0·10 (0.00 to 0·30)	0·05 (0·00 to 0·10)	0·10 (0·00 to 0·20)	0·10 (0·00 to 0·10)
p value	0·93	0·11	0·60	0·54	0·90	0·94	..
Mean defect, dB	–3·3 (–8·5 to –2·0)	–2·4 (–3·9 to –0·7)	–2·6 (–5·4 to –1·6)	–15·4 (–20·6 to –7·5)	–5·9 (–14·9 to –1·6)	–2·4 (–6·6 to –1·2)	–1·6 (–3·0 to –0·8)
p value	0·0009	0·29	0·020	<0·0001	0·025	0·048	..
ppRNFL, µm	90 (14)	94 (10)	85 (17)	86 (26)	96 (12)	82 (11)	95 (11)
p value	0·12	0·74	0·023	0·37	0·81	0·036	..
MRW-BMO, µm	236 (35)	252 (34)	238 (30)	311 (70)	340 (120)	232 (26)	286 (50)
p value	0·0001	0·0086	0·0003	0·22	0·30	0·031	..
Outside typical limits of GHT	31 (74%)	9 (75%)	16 (47%)	12 (83%)	5 (55·6%)	25 (66%)	7 (21%)
p value	0·0022	0·0013	0·0007	0·0007	0·30	0·0001	..
Glaucoma severity[15]*							
Early glaucoma	26/42 (62%)	12 (100%)	..	..	..	..	..
Moderate glaucoma	9/42 (21%)	0	..	..	..	..	..
Advanced glaucoma	4/42 (10%)	0	..	..	..	..	..
Severe glaucoma	3/42 (7%)	0	..	..	..	..	..

Data are n, n (%), n/N (%), median (IQR), or mean (SD). This table includes participants who were sent for further testing through the original screening of AI-based fundus imaging and IOP and those who were sent for further testing through expert review of fundus imaging. Glaucoma severity follows the ocular hypertension study methodology.[15] AI=artificial intelligence. GHT=glaucoma hemifield test. LogMAR=logarithm of the Minimum Angle of Resolution. MRW-BMO=minimum rim width-Bruch's membrane opening thickness. ppRNFL=peripapillary retinal nerve fibre thickness. *Glaucoma severity is shown per eye rather than per person.

Table 2: **Function and structure of all referred participants and their eyes and their diagnostic label**

	AI algorithm	Human experts
Sensitivity	0·78 (0·62–0·89)	0·75 (0·59–0·87)
Specificity	0·95 (0·93–0·97)	0·91 (0·88–0·93)
Positive predictive value	0·53 (0·40–0·67)	0·36 (0·26–0·47)
Negative predictive value	0·98 (0·97–0·99)	0·98 (0·97–0·99)
Positive likelihood ratio	16·91 (11·28–25·34)	8·18 (5·99–11·16)
Negative likelihood ratio	0·24 (0·13–0·42)	0·28 (0·16–0·47)
Accuracy	0·94 (0·92–0·96)	0·90 (0·87–0·92)
Cohen's κ	0·60 (0·49–0·72)	0·44 (0·32–0·55)

Data are estimates (95% CI). AI=artificial intelligence.

Table 3: **Diagnostic performance for glaucoma**

optimum referral threshold was lower than when the standard adapted Thessaloniki Eye Study criteria are used (appendix p 44).

Referral outcomes from human expert screening were examined in relation to the established glaucoma diagnoses. After independent review of fundus photographs, experts unanimously agreed that 31 participants required referral and that 437 did not. Disagreement occurred in 203 cases. After adjudication, 87 of these were referred, resulting in a total of 118 expert-based referrals (18% of all screened participants). For comparison, single-expert assessment would have resulted in 234 referrals (35% of all screen participants). Using the same reference standard, paired expert assessment with adjudication detected 30 (75%) of 40 participants with glaucoma and missed 10 (25%) patients with glaucoma, while referring 52 more participants than the AI-based pathway. Comparative performance metrics are shown in table 3. The appendix (p 45) details referral pattern differences across AI-based and expert-based screening pathways.

Results from the univariable and multivariable regression analyses on glaucoma diagnosis, including plots and tables, are presented in the appendix (pp 46–48).

Cost-effectiveness of AI-based screening was evaluated across varying glaucoma prevalences over a 10-year horizon, with increasing prevalence associated with greater economic benefit. AI screening costs totalled €11·65 per participant (€8·00 for software licensing and €3·65 for operational costs). At 1% prevalence, AI screening resulted in incremental costs of €384 (95% CI –6217 to 6650; ICER €1725 per QALY [95% CI –114 000 to 124 000; probability of 76·3% cost-effectiveness at the €20 000 per QALY threshold). At 2% prevalence, the strategy became cost-saving (–€680 [95 CI% –7294 to 5579]; ICER –€3058 per QALY [–121 000 to 123 000]; probability of 81·0% cost-effectiveness at the €20 000 per QALY threshold), indicating dominance—meaning the intervention both improves health and saves money. At 5% prevalence (the base case), the cost reached –€1315 (95% CI –7939 to 4944); ICER –€5,916 per QALY (–127 559 to 123 218); and a probability of 83·3% cost-effectiveness at the €20 000 per QALY threshold, indicating dominance, increasing to –€1528 (–8154 to 4729) at 10% prevalence. QALY gains remained constant (0·222) across prevalence scenarios, whereas cost-savings increased by €1900 per participant from 1% to 10% prevalence. Populations at high risk achieved dominance with a more than 83% cost-effectiveness probability at conservative thresholds.

Economic robustness was assessed by varying infrastructure costs across plausible implementation scenarios. AI screening remained cost-effective across all scenarios. At 10 years, the low-cost scenario (€8·00 per screening, co-located with diabetic retinopathy infrastructure) cost –€1407 (95% CI –8022 to 4830; ICER –€6,411 per QALY [95% CI –140 000 to 122 000]; probability of 83·7% cost-effectiveness at the €20 000 per QALY threshold), indicating dominance, the base-case scenario (€11·65 per screening) cost –€1315 (95% CI –7939 to 4944; ICER –€5916 per QALY [95% CI –128 000 to 123 000]; probability of 83·3% cost-effectiveness), and the high-cost scenario (€18·95 per screening) cost –€1166 (95% CI –7787 to 5084; ICER –€5243 per QALY [95% CI –129 000 to 120 000]; probability of 82·6% of cost-effectiveness). At 30 years, savings ranged between €11 529 (95% CI –5·6 000 to 30 000) and €11 815 (–5·3 000 to 30 000) despite 2·4-fold cost variation. QALY gains remained stable (0·95 [95% CI –0·76 to 2·59] at 10 years; 0·96 [95% CI –0·77 to 2·58] at 30 years), with cost-effectiveness probabilities differing less than 1·5 percentage points, indicating little sensitivity to infrastructure costs. At the 10-year time horizon with 5% glaucoma prevalence, AI-enhanced screening showed cost-saving potential in the Portuguese context, with a mean incremental cost of –€1197 (95% CI –7518 to 4964) and cumulative QALY gains of 0·205 (–0·334 to 0·743). The ICER of –€5838 per QALY (–112 259 to 112 072) indicated dominance. The probability of cost-effectiveness increased substantially to 82·6% at the €20 000 per QALY threshold, reaching 87·1% at €50 000 per QALY, and 88·6% at €100 000 per QALY. This crossover to cost-savings occurs at a similar timepoint in both the European and Portuguese cost analyses, substantiating the robustness of the intervention's economic benefits across different utility assumptions and the proportionally scaled Portuguese population norms (0·868 for Portuguese values *vs* 0·985 for Dutch values), while maintaining the relative health state utility ratios.

Discussion

Our results show that AI-led glaucoma screening achieves clinically meaningful performance in a real-world primary care setting, with a sensitivity of 78%, specificity of 95%, and positive predictive value of 53%. In a slowly progressive and largely asymptomatic disease such as glaucoma, specificity is particularly important, as false-positive referrals are the main drivers of unnecessary follow-up, patient anxiety, clinician workload, and health-care costs.[26] The high specificity observed in this study therefore supports the suitability of this approach for targeted, primary-care-based screening strategies, particularly in settings where specialist access is low and efficient referral pathways are essential.

A key contribution of this study is the prospective evaluation of AI screening performance in routine care, benchmarked against expert human grading. Although overall diagnostic performance was similar, human graders referred nearly twice as many participants as the AI system (18% *vs* 10%), with substantially more false-positive referrals. These findings align with previous evidence showing greater interobserver variability and lower consistency in human graders than machine learning approaches for assessing fundus photographs.[27] Importantly, expert referral rates in this study reflect adjudicated decisions; individual human grading, as would occur in many real-world settings, would likely result in even higher referral rates, rendering such approaches inefficient at scale. The observed referral patterns are consistent with previous screening studies and likely reflect intrinsic diagnostic uncertainty rather than systematic bias in favour of AI.

Although a sensitivity of 78% implies that some glaucoma cases might be missed at a single screening round, the screening pathway incorporated two important safeguards: referral based on elevated intraocular pressure, irrespective of AI risk score, and repeated screening over time. Given the slow progression of screen-detected glaucoma, false-negative cases are unlikely to progress to clinically meaningful vision loss between screening rounds, particularly when longitudinal surveillance is applied. In contrast, false-positive referrals represent the dominant potential harm of screening, reinforcing the importance of maintaining high specificity in population-facing screening programmes.

Economic analyses showed that AI-based screening increased costs but also improved health outcomes, resulting in favourable cost-effectiveness over a 10-year horizon. At a willingness-to-pay threshold of €20 000 per QALY gained, the probability of cost-effectiveness exceeded 80% both in general population screening and when integrated into existing diabetic retinopathy infrastructure. These values compare favourably with other accepted ophthalmic interventions.[28] Incremental costs were primarily driven by screening and earlier treatment but were partly offset by reductions in visual impairment and associated productivity losses. The wide 95% CIs reflect substantial joint parameter uncertainty across model inputs, amplified through probabilistic sampling in the cost-effectiveness analysis.

Our cost analysis explicitly itemised AI-based screening costs into software licensing and operational components, resulting in a fully loaded base-case cost of €11·65 per screening event. When co-located within existing diabetic retinopathy screening programmes, the incremental cost of glaucoma screening was substantially lower than the base case, reflecting shared infrastructure and personnel. Although standalone implementation incurs higher initial costs, economic modelling found cost-effectiveness even under conservative assumptions, with convergence towards cost-savings by 10 years across all scenarios. These findings indicate that AI-enabled glaucoma screening can deliver value across a range of implementation contexts, although the magnitude of benefit depends on existing infrastructure.

Several limitations merit consideration. The study population included a high proportion of participants with diabetes, reflecting integration within a diabetic retinopathy screening pathway; therefore, the observed glaucoma prevalence should not be interpreted as representative of the general population aged 55–65 years. The association between diabetes and glaucoma remains controversial, and the European Glaucoma Society does not recognise diabetes as an established risk factor.[29] Consistent with this, AI performance did not differ significantly between participants with diabetes and those without diabetes, although the subgroup without diabetes was small. Participation and follow-up rates were similar to those reported in other community-based glaucoma screening studies. Although attrition could introduce selection bias, there is no clear reason to assume that non-attendance preferentially involved individuals with diagnosed glaucoma, particularly given the asymptomatic nature of glaucoma. The requirement for a separate hospital visit for confirmatory testing likely contributed to attrition and represents a structural limitation of the pathway rather than a limitation of the screening test itself. Applicability to other primary care screening settings might vary according to access to low-technology screening devices (fundus cameras and tonometers), trained personnel, and established referral pathways; training was straightforward in our programme because staff already had ophthalmic screening experience but might be more resource intensive in settings without existing eye-screening infrastructure.

To minimise potential bias related to model provenance, all performance evaluations, clinical adjudications, and economic analyses were done independently of the AI developers, with prespecified thresholds and blind reference standards.

A substantial proportion of referred participants had a typical intraocular pressure, underscoring the importance of optic disc assessment and supporting the relevance of structural screening approaches, particularly in normal tension glaucoma. Although OCT imaging was done in referred participants, it was not included in the diagnostic reference standard, as there are currently no guidelines that endorse OCT-based diagnostic criteria[29] and OCT-only diagnosis risks false positives. Although OCT inclusion could increase detection of very early disease, it would also increase referral burden. Given the slow progression of glaucoma, a longitudinal screening strategy with repeat assessments is likely to capture early disease over time; OCT data were therefore retained for future reanalysis as diagnostic standards evolve.

Finally, although AI-based screening was cost-effective across multiple scenarios, implementation should account for operational workload and sustainability, as cost-effectiveness alone does not guarantee feasibility. Future work will focus on patient-reported and staff-reported outcomes, sociodemographic determinants of participation,

and longitudinal changes in AI risk scores, which might allow dynamic refinement of referral thresholds, and patient perspectives on AI evaluation. Equity considerations will be central to scale-up, with targeted strategies to address differential uptake and non-attendance.

In summary, AI-based glaucoma screening can be effectively implemented in primary care settings and efficiently integrated within existing diabetic retinopathy screening programmes. The approach offers scalable performance similar to expert assessment, reduces unnecessary referrals, and is likely to be cost-effective across diverse health-system contexts. Continued follow-up will be essential to optimise implementation and maximise population-level benefits.

Contributors

Conceptualisation: LAP, RB, PF, SLei, SLem, AL-C, CM-N, AM, MP, AT, and EV. Methods: LAP, FF, JF, and AL-C. Data curation: AL-C, FF, VL, RM, BM, MO, JP, ST, and RW. Formal analysis: AL-C, FF, VL, BM, JP, and RW. Investigation: RB, FF, PF, SaL, SoL, AL-C, VL, AM, BM, MP, JP, AT, EV, and RW. Validation: LAP, RB, PF, SaL, SoL, AL-C, AM, MP, AT, and EV. Visualisation: AL-C and ZW. Writing of the original draft: LAP, FF, AL-C, and ZW. Review and editing: LAP, RB, FF, PF, SaL, SoL, AL-C, AM, MP, AT, EV, and ZW. Project administration: MF, JF, and QF. Resources: MF, JF, QF, RH, RM, CM-N, MO, IS, and ST. Software provision: IS and RH. Supervision: LAP, MF, JF, QF, and CM-N. Technical support and advisory input on AI model use: RH and IS. Access to raw data was limited to the academic investigators responsible for data curation and analysis (ALC, FF, VL, BM, JP, RW, and LAP), who had full and independent access, verified the data, and take responsibility for data integrity and analytical accuracy.

Declaration of interests

IS is cofounder, shareholder, and a consultant of Mona.Health, a KU Leuven and VITO spin-off to which the described AI model was transferred. IS reports grant or research support from Bausch + Lomb, Heidelberg, Santen, and Théa; participation in clinical trials sponsored by Omikron, Santen, and Théa; and honoraria or consulting fees from AbbVie, Bausch + Lomb, Omikron, Roche, Santen, and Théa. RH reports consulting fees and stock or stock options in Mona.Health. The investigators received an unrestricted, in-kind research licence to the MONA-GLC model for the conduct of this study. FF reports stock or stock options in Novartis and Bayer. MP reports grants from Carl Zeiss Meditec; consulting fees from AbbVie and Santen; honoraria for lectures or educational activities from AbbVie, Glaukos, Théa, Santen, and Alcon; travel support from AbbVie and Théa; and leadership or fiduciary roles with the European Glaucoma Society and the European Medicines Agency. AM reports grants from OphtaFrance, SamSara, and Bausch + Lomb; consulting fees from Glaukos, Alcon, Horus, AbbVie, Théa, Santen, CureCall, and Bausch + Lomb; and honoraria from Glaukos, Horus, AbbVie, Théa, iCare, Santen, Alcon, Bausch + Lomb, and MTS. SLem reports honoraria from Théa and Horus Pharma and participation on a data safety monitoring or advisory board for Chiesi. LAP reports consulting fees from Santen, NIDEK, Alcon, and Théa and participation on a data safety monitoring or advisory board for Santen. All other authors declare no competing interests.

Data sharing

De-identified individual participant data (including a data dictionary) cannot be made available to others. The curated fundus image dataset and AI algorithm output files cannot be made publicly available and cannot be shared. The statistical analysis code will be made available upon reasonable request to the corresponding author, beginning with publication and for the subsequent 5 years, for researchers who provide a methodologically sound proposal and agree to the required data and code use agreement. The economic model analysis (methods, parameter inputs, script, and results) is provided in the appendix (pp 7–38); any additional model materials are available on request under the same access criteria.

Acknowledgments

We thank MONA.Health for providing an unrestricted, in-kind research licence for the MONA-GLC algorithm and Stef Rommes for technical support regarding the MONA-GLC algorithm.

References

1 Stein JD, Khawaja AP, Weizer JS. Glaucoma in adults-screening, diagnosis, and management: a review. *JAMA* 2021; **325:** 164–74.
2 Chua J, Baskaran M, Ong PG, et al. Prevalence, risk factors, and visual features of undiagnosed glaucoma: the Singapore Epidemiology of Eye Diseases study. *JAMA Ophthalmol* 2015; **133:** 938–46.
3 Friedman D. Screening for glaucoma is challenging but should be actively pursued. *J Glaucoma* 2024; **33** (suppl 1)**:** S2.
4 Aspberg J, Heijl A, Bengtsson B. Screening for open-angle glaucoma and its effect on blindness. *Am J Ophthalmol* 2021; **228:** 106–16.
5 Hamid S, Desai P, Hysi P, Burr JM, Khawaja AP. Population screening for glaucoma in UK: current recommendations and future directions. *Eye* 2022; **36:** 504–09.
6 Kolomeyer NN, Katz LJ, Hark LA, et al. Lessons learned from 2 large community-based glaucoma screening studies. *J Glaucoma* 2021; **30:** 875–77.
7 Founti P, Stuart K, Nolan WP, Khawaja AP, Foster PJ. Screening strategies and methodologies. *J Glaucoma* 2024; **33** (suppl 1)**:** S15–20.
8 Gunzenhauser R, Coleman AL. Glaucoma screening guidelines worldwide. *J Glaucoma* 2024; **33** (suppl 1)**:** S9–12.
9 Hemelings R, Elen B, Barbosa-Breda J, et al. Accurate prediction of glaucoma from colour fundus images with a convolutional neural network that relies on active and transfer learning. *Acta Ophthalmol* 2020; **98:** e94–100.
10 Li Z, He Y, Keel S, Meng W, Chang RT, He M. Efficacy of a deep learning system for detecting glaucomatous optic neuropathy based on color fundus photographs. *Ophthalmology* 2018; **125:** 1199–206.
11 Chuter B, Huynh J, Hallaj S, et al. Evaluating a foundation artificial intelligence model for glaucoma detection using color fundus photographs. *Ophthalmol Sci* 2024; **5:** 100623.
12 Hemelings R, Elen B, Schuster AK, et al. A generalizable deep learning regression model for automated glaucoma screening from fundus images. *NPJ Digit Med* 2023; **6:** 112.
13 Buisson M, Navel V, Labbé A, et al. Deep learning versus ophthalmologists for screening for glaucoma on fundus examination: a systematic review and meta-analysis. *Clin Exp Ophthalmol* 2021; **49:** 1027–38.
14 WHO. Screening programmes: a short guide. Increase effectiveness, maximize benefits and minimize harm. Feb 6, 2020. https://www.who.int/europe/publications/i/item/9789289054782 (accessed Dec 12, 2025).
15 Peeters F, Rommes S, Elen B, et al. Artificial intelligence software for diabetic eye screening: diagnostic performance and impact of stratification. *J Clin Med* 2023; **12:** 1408.
16 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.
17 Treacy MP, O'Neill EC, Murphy M, et al. Opportunistic detection of glaucomatous optic discs within a diabetic retinopathy screening service. *Eur J Ophthalmol* 2016; **26:** 315–20.
18 Boverhof BJ, Corro Ramos I, Vermeer KA, et al. The cost-effectiveness of an artificial intelligence-based population-wide screening program for primary open-angle glaucoma in the Netherlands. *Value Health* 2025; **28:** 1317–26.
19 Leske MC, Heijl A, Hyman L, Bengtsson B, Dong L, Yang Z, and the EMGT Group. Predictors of long-term progression in the early manifest glaucoma trial. *Ophthalmology* 2007; **114:** 1965–72.
20 Topouzis F, Wilson MR, Harris A, et al. Prevalence of open-angle glaucoma in Greece: the Thessaloniki Eye Study. *Am J Ophthalmol* 2007; **144:** 511–19.

21 Gordon MO, Kass MA. The Ocular Hypertension Treatment Study: design and baseline description of the participants. *Arch Ophthalmol* 1999; **117:** 573–83.

22 de Vente C, Vermeer KA, Jaccard N, et al. AIROGS: Artificial Intelligence for Robust Glaucoma Screening challenge. *IEEE Trans Med Imaging* 2024; **43:** 542–57.

23 Wild PS, Zeller T, Beutel M, et al. Die Gutenberg Gesundheitsstudie [The Gutenberg Health Study]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2012; **55:** 824–29.

24 Marshall GN, Hays RD. The Patient Satisfaction Questionnaire Short Form (PSQ-18). 1994. https://www.rand.org/pubs/papers/P7865.html (accessed Jan 10, 2026).

25 Weiner BJ, Lewis CC, Stanick C, et al. Psychometric assessment of three newly developed implementation outcome measures. *Implement Sci* 2017; **12:** 108.

26 Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. *Front Public Health* 2017; **5:** 307.

27 Wu JH, Nishida T, Weinreb RN, Lin JW. Performances of machine learning in detecting glaucoma using fundus and retinal optical coherence tomography images: a meta-analysis. *Am J Ophthalmol* 2022; **237:** 1–12.

28 Azuara-Blanco A, Burr J, Ramsay C, et al, and the EAGLE study group. Effectiveness of early lens extraction for the treatment of primary angle-closure glaucoma (EAGLE): a randomised controlled trial. *Lancet* 2016; **388:** 1389–97.

29 Pazos M, Traverso CE, Viswanathan A, et al. European Glaucoma Society—terminology and guidelines for glaucoma, 6th edition. *Br J Ophthalmol* 2025; **109** (suppl 1): 1–212.