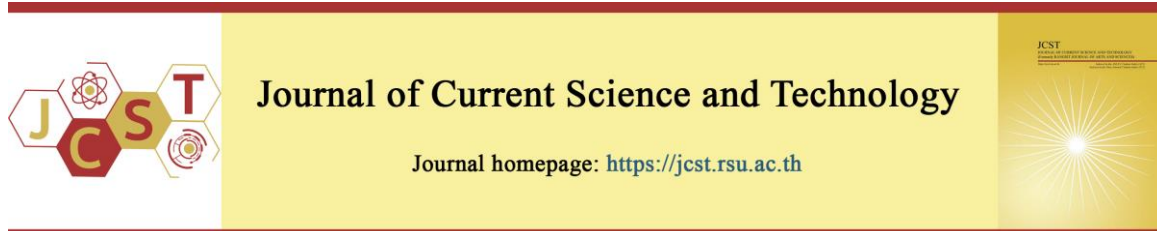


Cite this article: Phan, N. U., Pham, T. V., Doan, K. T., Hoang, L. P., Pham, N. V. T., Nguyen, M. T., Mai, Q. T., Jaccard, N., & Congdon, N. (2026). Diagnostic accuracy of cybersight AI for glaucoma detection: Sensitivity, specificity, and ROC analysis. *Journal of Current Science and Technology*, 16(4), Article 220. <https://doi.org/10.59796/jcst.V16N4.2026.220>



Diagnostic Accuracy of Cybersight AI for Glaucoma Detection: Sensitivity, Specificity, and ROC Analysis

Nha Uyen Phan^{1, 2, *} 
Trong Van Pham¹
Kim Thanh Doan³ 
Le Phuc Hoang⁴ 
Nhu Vinh Tuyen Pham² 
Minh Tuan Nguyen² 
Quoc Tung Mai¹ 
Nicolas Jaccard⁵ 
Nathan Congdon⁶ 

¹Department of Ophthalmology, Hanoi Medical University, Hanoi, Vietnam

²Eye Center, Hue Central Hospital, Hue, Vietnam

³Department of Ophthalmology, Pham Ngoc Thach University of Medicine, Ho Chi Minh City, Vietnam

⁴Children's Hospital 1, Ho Chi Minh City, Vietnam

⁵Visulytix Ltd, London, United Kingdom

⁶Centre for Public Health, Queen's University Belfast, Belfast, United Kingdom; Orbis International, New York, NY, USA

*Corresponding author; E-mail: phannhuyen15@gmail.com

Received 2 February 2026; Revised 22 May 2026; Accepted 16 June 2026; Published 15 September 2026

Abstract

Glaucoma is a chronic, progressive optic neuropathy and a leading cause of irreversible blindness worldwide. Early detection is crucial but is often limited by nonspecific clinical signs and the need for specialist expertise. Artificial intelligence (AI), particularly deep learning, has shown promise for glaucoma detection from ophthalmic images. This study independently validated an existing third-party AI tool, Cybersight AI, powered by the Visulytix Pegasus engine, for detecting primary glaucoma. No new algorithm was developed or proposed. A cross-sectional diagnostic accuracy study was conducted on 229 eyes from 121 patients, including 133 eyes with primary glaucoma and 96 controls, at Hue Central Hospital between December 2024 and September 2025. Fundus photographs were analyzed using the Cybersight AI platform, which estimated the vertical cup-to-disc ratio and generated a binary glaucoma referral recommendation. AI outputs were compared with diagnoses established by ophthalmologists using an expert multimodal clinical reference standard. Diagnostic performance was assessed using sensitivity, specificity, positive and negative predictive values, and the area under the receiver operating characteristic curve (AUC). At its default operating point, Cybersight AI achieved 51.9% sensitivity (95% CI 43.5–60.2) and 90.6% specificity (95% CI 83.1–95.0), with an AUC of 0.744 (95% CI 0.682–0.807). Positive and negative predictive values were 88.5% and 57.6%, respectively. The Youden-optimal cut-off was 10%, yielding 60.9% sensitivity and 85.4% specificity. Cybersight AI demonstrated moderate discrimination, high specificity, and modest sensitivity at its default binary setting in this Vietnamese tertiary-care cohort. It may serve as a screening or triage adjunct, although context-specific thresholding and further external validation are required before broad clinical deployment. Limitations include the single-center convenience sample, within-patient clustering, and comparison of a fundus-only index test against a multimodal reference standard. Future studies should prioritize multicenter validation in Vietnamese settings and comparison with expert fundus assessment alone.

Keywords: artificial intelligence, deep learning, diagnostic accuracy, glaucoma, sensitivity and specificity

1. Introduction

Glaucoma is a group of chronic, progressive, and irreversible optic neuropathies and is one of the leading causes of blindness worldwide (Jenchitr et al., 2024). According to the World Health Organization (WHO), an estimated 76 million people were affected by glaucoma in 2020, and this number is projected to exceed 100 million by 2040 (Qi et al., 2025). Early detection is a prerequisite for effective disease control and for preventing irreversible vision loss.

However, early diagnosis of glaucoma presents many challenges due to its non-specific clinical manifestations, the insidious progression of symptoms, and the heavy reliance on the expertise of ophthalmologists (Jin et al., 2024; Tonti et al., 2024), who are in short supply in many areas of the world. In this context, the development of artificial intelligence (AI), particularly deep learning models, has opened new opportunities for supporting glaucoma diagnosis through the analysis of retinal and optic disc images (Lan et al., 2025; Ling et al., 2025).

Although AI has been applied in various ophthalmic fields such as the detection of diabetic retinopathy and age-related macular degeneration (Anton et al., 2023), its application in glaucoma detection is still under evaluation and optimization (Shahriari et al., 2025). One of the key criteria for assessing the effectiveness of AI is its ability to accurately distinguish between affected and unaffected individuals, as reflected in metrics such as sensitivity, specificity, and the area under the receiver operating characteristic (ROC) curve (AUC) (Ashtari-Majlan et al., 2024; Chaurasia et al., 2022).

It is estimated that up to 75% of individuals with glaucoma in low- and middle-income countries (LMICs) remain undiagnosed until significant visual impairment has already developed (Senjam, 2020). This is largely due to the lack of effective screening systems, limited availability of ophthalmic personnel, and inadequate healthcare infrastructure (Wolvaardt & Hu, 2021). The burden of vision loss caused by glaucoma is often inversely correlated with national income levels, reflecting substantial inequities in access to eye care services (Wu et al., 2023).

Cybersight AI is a cloud-based artificial intelligence software developed by non-governmental organization (NGO) Orbis International. Cybersight is a non-profit telemedicine and online training platform aimed at enhancing the expertise, skills, and knowledge of ophthalmic professionals worldwide (Jaccard, 2022). However, the diagnostic accuracy of Cybersight AI has not been independently validated

in a Vietnamese population, whose optic-disc characteristics and glaucoma stage distribution may differ from those represented in the platform's training data. This validation gap limits confidence in its deployment within local healthcare systems.

2. Objectives

This study aimed to evaluate the diagnostic accuracy of Cybersight AI (Visulytix Pegasus) for detecting primary glaucoma in adults attending a Vietnamese tertiary eye center, using an expert multimodal clinical reference standard as the comparator, in a cross-sectional diagnostic accuracy design. The primary performance measures were sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curve.

3. Materials and Methods

This cross-sectional diagnostic accuracy study was conducted at the Eye Center, Hue Central Hospital, Vietnam, from December 2024 to September 2025. The study protocol was reviewed and approved by the Scientific Committee of Hanoi Medical University on August 18, 2024, and ethical approval was granted by the Ethics Committee of Hanoi Medical University on February 25, 2025 (Approval No.: IRB-VN01.001/IRB00003121/FWA00004148). The study was conducted in accordance with the Declaration of Helsinki.

Study Population and Design

Consecutive adult patients aged 18 years or older who attended the Eye Center for ophthalmic examination or treatment during the study period and met the eligibility criteria were enrolled using a convenience sampling approach. A total of 229 eyes from 121 patients were included, comprising 133 eyes with primary glaucoma and 96 eyes without glaucoma. Because classification was performed at the eye level, 16 patients contributed one glaucomatous eye and one non-glaucomatous eye and therefore appear in both groups. Eyes were used as the unit of analysis, and within-patient correlation was accounted for in the primary analysis (see Statistical analysis).

Clinical Examinations and Imaging Acquisition

All participants underwent comprehensive ophthalmic evaluation by trained ophthalmologists, including slit-lamp biomicroscopy, gonioscopy with a Goldmann three-mirror lens, standard automated perimetry using the Humphrey Visual Field Analyzer,

and optical coherence tomography (OCT) of the retinal nerve fiber layer (RNFL) using the NIDEK RS-3000 system (Wanichwecharungruang et al., 2011). Color fundus photographs were obtained after pharmacologic pupil dilation by an ophthalmic technician using a Zeiss VISUCAM 524 fundus camera. At least one well-centered, high-quality fundus image capturing the optic disc and macula was obtained for each eye. These images were used both for clinical documentation and for analysis by the Cybersight AI system.

Reference Standard Diagnosis

The reference standard diagnosis (primary glaucoma versus non-glaucoma) was established independently by two ophthalmologists with more than five years of experience in glaucoma care at the Eye Center. Reviewers had access to the complete clinical record, including findings from clinical examination, gonioscopy, OCT RNFL measurements, and visual field testing. Optic disc assessment, including evaluation of the vertical cup-to-disc ratio (VCDR) and inter-eye asymmetry, was performed on dilated slit-lamp examination using a 90D lens and corroborated by fundus photographs when required. OCT RNFL scans were assessed for quality and segmentation artifacts before interpretation, and visual fields were reviewed for reliability indices and glaucomatous defect patterns, including glaucoma hemifield test (GHT) and pattern standard deviation (PSD) results.

In cases of disagreement between the two primary reviewers regarding diagnostic classification or eligibility, a third senior ophthalmologist with specialized expertise in glaucoma adjudicated the final diagnosis. All clinicians involved in reference standard determination were masked to the Cybersight AI outputs.

Diagnostic Criteria

Eyes were classified as having primary glaucoma if at least two of the following three diagnostic domains were present. First, optic disc features suspicious for glaucomatous optic neuropathy on dilated examination and/or fundus photography, defined as a VCDR ≥ 0.7 or an inter-eye VCDR asymmetry ≥ 0.2 , interpreted in the context of optic disc size. Optic disc cupping findings were considered supportive rather than diagnostic in isolation (Crowston et al., 2004; Foster et al., 2002). Second, OCT evidence of glaucomatous RNFL thinning, defined as global and/or sectoral RNFL thickness values below the lower 5th percentile

of the RS-3000 normative database after exclusion of scans with poor quality or significant segmentation artifacts (Myers et al., 2024). Third, glaucomatous visual field defects on standard automated perimetry, defined by the presence of at least two of the following criteria: (i) a cluster of three or more contiguous non-edge points in a typical glaucomatous location with $p < 5\%$, including at least one point with $p < 1\%$ on the pattern deviation plot; (ii) abnormal PSD ($p < 5\%$); or (iii) GHT outside normal limits (Budenz et al., 2002; Ruia & Tripathy, 2025).

Eyes were classified as non-glaucomatous controls if they met all of the following criteria: normal visual field results with acceptable reliability indices and no glaucomatous defect pattern (Ruia & Tripathy, 2025); OCT RNFL measurements within normal limits without borderline or abnormal sectoral or global values on the RS-3000 normative database (La Bruna et al., 2022); and an optic disc appearance not suggestive of glaucoma, defined as a VCDR ≤ 0.4 , or a VCDR between 0.4 and 0.7 with a neuroretinal rim configuration consistent with the ISNT rule and no focal rim notching or disc hemorrhage on clinical examination or fundus photography (Tello et al., 2026).

Eyes were excluded if they had media opacities that compromised fundus image quality, including dense cataract, corneal scarring, or vitreous opacities; systemic conditions impairing cooperation during examinations; or a history of ocular trauma, congenital glaucoma, or secondary glaucoma.

Index Test: Cybersight AI Analysis

Eligible fundus photographs were analyzed using Cybersight AI, a cloud-based clinical decision-support tool maintained by Orbis International (powered by the Visulytix Pegasus engine, Visulytix Ltd., London, UK). The platform version/build number was not displayed to end users in the Cybersight interface during the study period. All analyses were performed using the system accessed between December 2024 and September 2025, and no user-visible version change was recorded during data collection. Before analysis, images were screened using the platform's automated image-quality assessment, and only images flagged as sufficient for analysis were included. For each image, the platform generated (1) a continuous glaucoma risk score expressed as a percentage from 0% to 100%, and (2) a binary glaucoma referral recommendation. In this dataset, the platform's default binary output corresponded to a 20% risk-score cut-off; the exact internal threshold underlying the binary output is not

disclosed to end users. The AI outputs were exported and matched at eye level with the corresponding reference standard diagnoses for analysis.

Outcome Measures and Statistical Analysis

The primary outcome measures were the sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve (AUC) of Cybersight AI for detecting primary glaucoma, using the clinical reference standard as the comparator (Figure 1). ROC curves and AUC values were derived from the continuous glaucoma risk scores. Sensitivity and specificity were calculated for the platform's default binary classification and for prespecified risk-score cut-off points at 10% intervals (0%, 10%, 20%, ..., 100%) to evaluate the trade-off between sensitivity and specificity. The Youden index ($J = \text{sensitivity} + \text{specificity} - 1$) was calculated for each cut-off point, and the threshold yielding the maximum J value was identified as the optimal operating point. Statistical analyses were performed using IBM SPSS Statistics version 20.0 (IBM Corp., Armonk, NY, USA). Ninety-five percent confidence intervals (CIs)

were calculated for all diagnostic metrics: the Wilson score method for sensitivity, specificity, and predictive values, and the DeLong method for the AUC. Because most patients contributed two eyes, the primary analysis accounted for within-patient clustering using a generalized estimating equations (GEE) model with an exchangeable working correlation structure and cluster-robust (sandwich) standard errors, with the patient as the clustering unit. The study is reported in accordance with the STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence.

4. Results

4.1 Patient Characteristics

A total of 229 eyes from 121 patients were included, comprising 133 glaucomatous and 96 non-glaucomatous eyes; 16 patients contributed one eye to each group. Mean age did not differ between groups (64.0 ± 11.4 years vs 61.1 ± 12.2 years; $p = 0.078$). Sex distribution was 48.9% male and 51.1% female in the glaucoma group, compared with 26.0% male and 74.0% female in the non-glaucoma group (Table 1).



Figure 1 Representative Cybersight AI output showing optic disc analysis and the continuous glaucoma risk score generated from a color fundus photograph

Table 1 Baseline Clinical Characteristics of Study Participants

Parameter	Primary Glaucoma (n=133)	Non-Glaucoma (n=96)	p-value
Mean Age (years)	64.0 ± 11.4	61.1 ± 12.2	$p = 0.078$
Gender (Male/Female)	Male 48.9%; Female 51.1%.	Male 26.0%; Female 74.0%	
Visual Acuity (BCVA)	0.51 ± 0.48 logMAR (Snellen 6/19)	0.35 ± 0.26 logMAR (Snellen 6/13)	$p < 0.001$
IOP (mmHg)	17.4 ± 8.3	14.1 ± 4.1	$p < 0.001$
AL (mm)	23.2 ± 0.88	23.2 ± 0.93	$p = 0.890$
CCT (μm)	523.13 ± 39.8	531.47 ± 29.8	$p = 0.071$

BCVA (converted to logMAR) was worse in the glaucoma group than in the non-glaucoma group (0.51 ± 0.48 vs 0.35 ± 0.26 logMAR; $p < 0.001$). IOP was higher in the glaucoma group (17.4 ± 8.3 mmHg vs 14.1 ± 4.1 mmHg; $p < 0.001$). Axial length did not differ between groups (23.2 ± 0.88 mm vs 23.2 ± 0.93 mm; $p = 0.890$), and central corneal thickness did not differ (523.1 ± 39.8 μ m vs 531.5 ± 29.8 μ m; $p = 0.071$).

When compared to the reference standard of clinical diagnosis by ophthalmologists, the Cybersight AI system yielded the following results (Table 2).

The ROC curve for the Cybersight AI continuous risk score yielded an AUC of 0.744 (95% CI 0.682–0.807) (Figure 2). Using the risk-score thresholds evaluated in Table 4, the Youden-optimal cut-off was 10% (maximum J). At the 10% cut-off,

sensitivity was 60.9% (95% CI 52.4–68.8) and specificity was 85.4% (95% CI 77.0–91.1) (Youden J = 0.463). The platform's default binary classification corresponded to a 20% risk-score cut-off, at which sensitivity was 51.9% (95% CI 43.5–60.2) and specificity was 90.6% (95% CI 83.1–95.0); positive and negative predictive values were 88.5% (95% CI 79.5–93.8) and 57.6% (95% CI 49.6–65.2), respectively (Table 3). Accounting for within-patient clustering (GEE), the corresponding estimates were sensitivity 51.9% (95% CI 41.9–61.7) and specificity 90.6% (95% CI 82.2–95.3).

From the contingency data shown in Table 2, the following diagnostic metrics were calculated and are presented in Table 3.

Table 2 Comparison between Cybersight AI and clinician diagnosis using a 20% risk-score threshold (AI+ if $\geq 20\%$)

	Clinician (+)	Clinician (–)
AI (+)	69 (True Positive)	9 (False Positive)
AI (–)	64 (False Negative)	87 (True Negative)

Table 3 Diagnostic performance of Cybersight AI at the 20% threshold versus clinician diagnosis

Diagnostic Metric	Value	95% CI
Sensitivity	51.9%	43.5–60.2
Specificity	90.6%	83.1–95.0
Area Under the ROC Curve (AUC)*	0.744	0.682–0.807
Youden-optimal cut-off from risk score	10% (see Table 4)	—
Positive predictive value (PPV)	88.5%	79.5–93.8
Negative predictive value (NPV)	57.6%	49.6–65.2

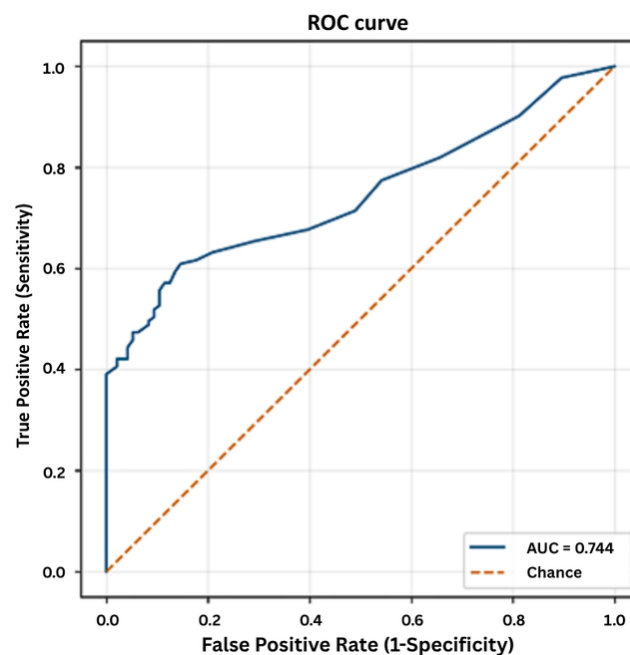


Figure 2 ROC Curve illustrating AI's ability to differentiate between glaucoma and non-Glaucoma

Table 4 Diagnostic performance of the AI tool across cut-off thresholds, including sensitivity, specificity, Youden index, and classification outcomes

Cut-off (%)	Se (%)	Sp (%)	Youden J	TP	FN	TN	FP
0	100.0	0.0	0.000	133	0	0	96
10	60.9	85.4	0.463	81	52	82	14
20	51.9	90.6	0.425	69	64	87	9
30	47.4	94.8	0.422	63	70	91	5
40	40.6	97.9	0.385	54	79	94	2
50	36.8	100.0	0.368	49	84	96	0
60	30.8	100.0	0.308	41	92	96	0
70	26.3	100.0	0.263	35	98	96	0
80	18.8	100.0	0.188	25	108	96	0
90	3.8	100.0	0.038	5	128	96	0
100	0.0	100.0	0.000	0	133	96	0

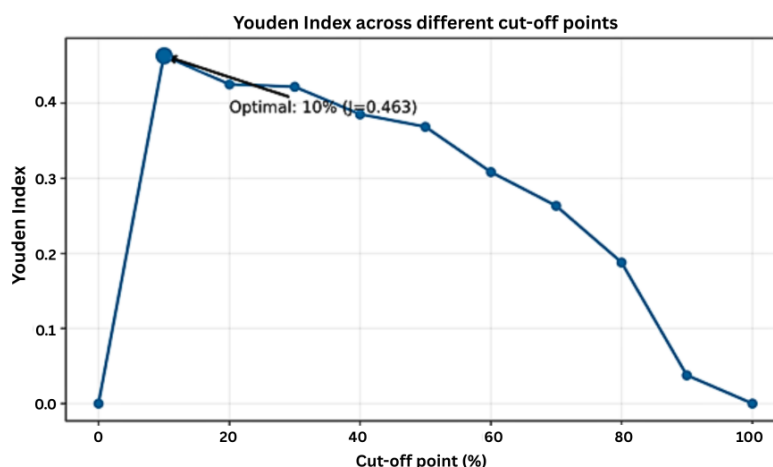


Figure 3 Youden Index across different cut-off points

The ROC analysis of the Cybersight AI continuous risk score yielded an AUC of 0.744 (95% CI 0.682–0.807). The Youden index was calculated across prespecified cut-off points (0–100%) (Figure 3). The maximum Youden index was observed at a cut-off of 10% ($J = 0.463$), corresponding to a sensitivity of 60.9% and a specificity of 85.4% (Table 4).

5. Discussion

In this single-center diagnostic accuracy study conducted at a tertiary eye hospital in Vietnam, the Cybersight AI system demonstrated high specificity (90.6%) but only moderate sensitivity (51.9%) at the platform’s default binary output, with an overall AUC of 0.744 (95% CI 0.682–0.807) when the continuous risk score was used to construct the ROC curve (Table 3). When risk-score thresholds were explored post hoc, the Youden-optimal cut-off was 10%, yielding a higher sensitivity (60.9%) at the cost of specificity (85.4%) (Table 4). These findings suggest that Cybersight AI may be more useful for reducing false-

positive referrals at the default setting (high specificity) than for maximizing case detection, unless the operating threshold is lowered.

A key methodological point is that ROC/AUC analyses require a continuous (or ordinal) test output, not a single binary classification. In this study, the ROC curve and AUC were derived from the continuous glaucoma risk score generated by Cybersight AI, whereas the sensitivity and specificity in Table 3 correspond to the platform’s default binary decision as returned by Cybersight AI (the numeric risk-score threshold underlying this binary output is not disclosed to users). Table 4 complements this by showing how sensitivity, specificity, and the Youden index change when alternative cut-offs are applied to the risk score, thereby identifying an “optimal” operating point under a specific criterion (maximum Youden J). This distinction resolves the apparent inconsistency: Table 3 reflects real-world “out-of-the-box” performance, while Table 4 reflects threshold optimization using the continuous score. Importantly,

AUC summarizes discrimination across all possible thresholds and is independent of any single cut-off.

Recent evidence indicates that deep learning systems for glaucoma detection can achieve higher discrimination (often AUC ~0.85–0.95) under optimized conditions, particularly when image quality is controlled and/or multimodal inputs (fundus ± OCT ± visual fields) are available. A recent systematic review and meta-analysis reported strong overall performance of deep learning for glaucoma diagnosis and progression prediction using fundus photography and OCT, although heterogeneity and study design differences were substantial (Ling et al., 2025). Similarly, a meta-analysis focusing on AI for glaucoma screening and clinical practice highlighted that performance varies widely by dataset, reference standard, and validation design (Chaurasia et al., 2022). In contrast to algorithm-development studies that report performance under controlled conditions and on curated datasets, for example, recent fundus-based deep-learning systems for glaucoma segmentation and classification (Shyamalee et al., 2024) the present work is an independent, real-world clinical validation of a deployed third-party platform against a multimodal expert reference standard, addressing a distinct and complementary evidence need. Our AUC of 0.744 is therefore lower than many “benchmark” reports, but it may better reflect real-world implementation in a Vietnamese tertiary-care workflow with heterogeneous imaging conditions and a reference standard based on comprehensive clinical testing.

Several factors may plausibly contribute to the comparatively modest sensitivity observed.

First, image quality and media opacity can substantially degrade fundus-based model performance. Recent work has shown that automated quality assessment and exclusion of poor-quality photographs can meaningfully improve glaucoma detection accuracy in independent populations (Chuter et al., 2024). In routine clinical practice, variable illumination, focus, and artefacts can reduce the visibility of subtle optic nerve head cues that fundus-only models rely upon.

Second, the AI index test in this study had access only to color fundus photographs, whereas the clinical reference standard incorporated multiple modalities (optic disc assessment, OCT RNFL, and Humphrey visual fields). This “information imbalance” can penalize the AI system when glaucoma is primarily manifest in structural OCT or functional VF changes with less obvious disc changes

on a single photograph. From an evaluation standpoint, this underscores the importance of clearly defining the intended use case: Cybersight AI is primarily designed as a screening/triage tool where OCT and VF may be unavailable, so a complementary analysis comparing AI to expert fundus examination alone would be a more like-for-like assessment of fundus-based screening.

Third, as a general consideration in fundus-based glaucoma screening, VCDR (or disc cupping) should be interpreted cautiously because physiologic cupping and disc size vary across individuals and populations. Therefore, fixed VCDR cut-offs are not universally applicable, and cupping findings are best treated as supportive signs rather than standalone diagnostic criteria, ideally interpreted alongside disc size and other corroborating information (e.g., RNFL/OCT or visual field when available).

Finally, dataset shift and limited representation of the local population in the algorithm’s training data may reduce generalizability. Contemporary work on fundus-based glaucoma AI emphasizes that external validation across diverse settings is essential, and performance may fall when applied outside the original data distribution.

The Youden index ($J = \text{sensitivity} + \text{specificity} - 1$) was used to identify a single operating point that maximized the sum of sensitivity and specificity, thereby implicitly assigning equal weight to false-positive and false-negative results (Youden, 1950). This transparent and commonly used approach is suitable for exploratory threshold selection, particularly when a balanced trade-off is desired. However, glaucoma screening programs often prioritize high sensitivity to minimize missed cases, whereas clinic-based decision-making may prioritize specificity to reduce unnecessary referrals and patient anxiety. Because the thresholds were explored post hoc using the same dataset employed for performance estimation, the Youden-optimal cut-off may be overly optimistic. Future studies should prespecify thresholds and validate the selected cut-off in an independent external cohort. Threshold selection could also be tailored using decision-analytic methods that account for local disease prevalence, referral capacity, and the relative harms of false-negative and false-positive results.

At the default binary setting, the low number of false positives suggests potential utility for reducing unnecessary referrals, which is valuable where specialist capacity is limited. Conversely, the false-negative burden indicates that Cybersight AI should

not be used as a stand-alone diagnostic tool. Its most appropriate role may be triage prompting comprehensive evaluation for eyes above a chosen risk threshold and supporting case-finding where resources are constrained. Recent prospective studies of offline AI systems embedded in smartphone-based fundus cameras aimed at resource-limited settings have reported higher sensitivities and specificities for referable glaucoma than observed here, underscoring both the promise of AI-enabled screening and the importance of local validation and workflow integration (Senthil et al., 2025; Soh et al., 2020).

Strengths of this study include (i) use of a multimodal clinical reference standard with expert adjudication and masking to AI outputs, (ii) reporting of both default binary performance and risk-score discrimination with ROC/AUC, and (iii) transparent exploration of operating thresholds (Table 4) aligned with contemporary reporting expectations for AI-centered diagnostic studies (Sounderajah et al., 2025).

This study has several limitations. First, the single-center design, convenience sampling, and absence of a priori sample-size calculation may limit generalizability. Second, eyes were treated as the unit of analysis, and 16 patients contributed one eye to each group. Although within-patient correlation was addressed using a GEE model as the primary analysis, residual clustering effects cannot be completely excluded. Third, the sex distribution differed between groups, with males comprising 48.9% of the glaucoma group and 26.0% of the control group. This imbalance may reflect referral and care-seeking patterns at a tertiary center, as well as possible sex-related differences in glaucoma risk, rather than a true population effect. Nevertheless, it may have confounded baseline comparisons and further limited generalizability. Fourth, glaucoma subtype, including open-angle and angle-closure glaucoma, was not systematically recorded. Consequently, subtype-specific analyses, which are particularly relevant in Southeast Asian populations, could not be performed and should be considered in future studies. Fifth, the AI index test relied solely on color fundus photographs, whereas the reference standard incorporated OCT and visual-field findings. Because independent fundus-only clinician assessments were not collected, a direct comparison between AI performance and expert evaluation of fundus photographs alone was not possible and should be prospectively prespecified. In addition, the reported PPV and NPV may not be applicable to screening settings because the prevalence in this enriched

sample differs from that in the general population. Finally, algorithm performance may change following software updates; therefore, explicit reporting of the AI version and operating conditions is essential for reproducibility (Guni et al., 2024; Sounderajah et al., 2025).

Future work should include (1) multi-center external validation across Vietnamese settings, (2) a prespecified evaluation comparing AI outputs with expert fundus examination alone to match the AI's information inputs, (3) incorporation of automated image quality scoring to quantify and control image-related failure modes (Chuter et al., 2024), and (4) exploration of models that integrate multimodal data (fundus + OCT + VF) or at least structured clinical features to improve sensitivity for early disease. Systematic reporting using STARD-AI and rigorous bias/applicability appraisal using emerging AI-focused tools would strengthen the evidence base for implementation (Guni et al., 2024; Sounderajah et al., 2025).

6. Conclusion

Cybersight AI demonstrated moderate discrimination (AUC 0.744, 95% CI 0.682–0.807) with high specificity (90.6%) but modest sensitivity (51.9%) at its default binary setting in a Vietnamese tertiary-care cohort. Threshold optimization using the continuous risk score identified a Youden-optimal cut-off of 10%, improving sensitivity at the expense of specificity. These findings support Cybersight AI as a potential screening/triage adjunct, while emphasizing the need for context-specific thresholding and further external validation prior to broad clinical deployment. These conclusions should be interpreted in light of the study's limitations, including its single-center convenience sample, the use of eyes as the unit of analysis with within-patient clustering, and a fundus-only index test compared against a multimodal reference standard.

7. Acknowledgements

The authors thank all patients who participated in this study and the clinical staff of the Eye Center, Hue Central Hospital, for their support with participant recruitment, clinical examinations, and study procedures. We are grateful to the ophthalmic technicians for pharmacologic pupil dilation and acquisition of high-quality color fundus photographs, which were used for clinical documentation and subsequently uploaded for Cybersight AI analysis.

Cybersight AI is a cloud-based platform developed by Orbis International; we acknowledge Cybersight/Orbis for enabling use of the platform for automated image analysis in this study. All individuals acknowledged have reviewed this statement and agreed to be named.

8. Abbreviations

Abbreviation	Full Term
AI	Artificial intelligence
AL	Axial length
AUC	Area under the receiver operating characteristic curve
BCVA	Best-corrected visual acuity
CCT	Central corneal thickness
CI	Confidence interval
FN	False negative
FP	False positive
FWA	Federalwide Assurance
GEE	Generalized estimating equations
GHT	Glaucoma hemifield test
IOP	Intraocular pressure
IRB	Institutional Review Board
ISNT	Inferior-Superior-Nasal-Temporal
LMICs	Low- and middle-income countries
logMAR	Logarithm of the minimum angle of resolution
NGO	Non-governmental organization
NPV	Negative predictive value
OCT	Optical coherence tomography
PPV	Positive predictive value
PSD	Pattern standard deviation
RNFL	Retinal nerve fiber layer
ROC	Receiver operating characteristic
SPSS	Statistical Package for the Social Sciences
STARD-AI	Standards for Reporting Diagnostic Accuracy Studies–Artificial Intelligence
TN	True negative
TP	True positive
VCDR	Vertical cup-to-disc ratio
VF	Visual field
WHO	World Health Organization

9. CRediT Statement

Nha Uyen Phan: Conceptualization, Methodology, Validation, Investigation, Resources, Data Curation, Visualization, Project Administration.

Trong Van Pham: Conceptualization, Project Administration.

Kim Thanh Doan: Conceptualization, Project Administration.

Le Phuc Hoang: Methodology, Validation.

Nhu Vinh Tuyen Pham: Validation, Investigation, Resources.

Minh Tuan Nguyen: Formal Analysis, Investigation, Data Curation, Visualization.

Quoc Tung Mai: Formal Analysis, Supervision, Funding Acquisition.

Nicolas Jaccard: Software, Validation.

Nathan Congdon: Methodology, Validation.

10. References

- Anton, N., Doroftei, B., Curteanu, S., Catălin, L., Ilie, O. D., Târcoveanu, F., & Bogdănici, C. M. (2023). Comprehensive review on the use of artificial intelligence in ophthalmology and future research directions. *Diagnostics*, 13(1), Article 100.
<https://doi.org/10.3390/diagnostics13010100>
- Ashtari-Majlan, M., Dehshibi, M. M., & Masip, D. (2024). Glaucoma diagnosis in the era of deep learning: A survey. *Expert Systems with Applications*, 256, Article 124888.
<https://doi.org/10.1016/j.eswa.2024.124888>
- Budenz, D. L., Rhee, P., Feuer, W. J., McSoley, J., Johnson, C. A., & Anderson, D. R. (2002). Comparison of glaucomatous visual field defects using standard full threshold and Swedish interactive threshold algorithms. *Archives of Ophthalmology*, 120(9), 1136-1141.
<https://doi.org/10.1001/archophth.120.9.1136>
- Chaurasia, A. K., Greatbatch, C. J., & Hewitt, A. W. (2022). Diagnostic accuracy of artificial intelligence in glaucoma screening and clinical practice. *Journal of Glaucoma*, 31(5), 285-299.
<https://doi.org/10.1097/IJG.0000000000002015>
- Chuter, B., Huynh, J., Bowd, C., Walker, E., Rezapour, J., Brye, N., ... & Christopher, M. (2024). Deep learning identifies high-quality fundus photographs and increases accuracy in automated primary open angle glaucoma detection. *Translational Vision Science & Technology*, 13(1), Article 23.
<https://doi.org/10.1167/tvst.13.1.23>
- Crowston, J. G., Hopley, C. R., Healey, P. R., Lee, A., & Mitchell, P. (2004). The effect of optic disc diameter on vertical cup to disc ratio percentiles in a population based cohort: The Blue Mountains Eye Study. *British Journal of Ophthalmology*, 88(6), 766-770.
<https://doi.org/10.1136/bjo.2003.028548>

- Foster, P. J., Buhrmann, R., Quigley, H. A., & Johnson, G. J. (2002). The definition and classification of glaucoma in prevalence surveys. *British Journal of Ophthalmology*, 86(2), 238-242.
<https://doi.org/10.1136/bjo.86.2.238>
- Guni, A., Sounderajah, V., Whiting, P., Bossuyt, P., Darzi, A., & Ashrafian, H. (2024). Revised tool for the quality assessment of diagnostic accuracy studies using AI (QUADAS-AI): protocol for a qualitative study. *JMIR Research Protocols*, 13(1), Article e58202.
<https://doi.org/10.2196/58202>
- Jaccard, N. (2022). *AI for glaucoma care*. Retrieved from <https://glaucomatoday.com/articles/2022-july-aug/ai-for-glaucoma-care>
- Jenchitr, W., Yokkampon, P., Ploysit, P., & Ausayakhun, S. (2024). The prevalence of visual impairment of the Elderly at University Eye Clinic. *Journal of Current Science and Technology*, 14(1), Article 9.
<https://doi.org/10.59796/jcst.V14N1.2024.9>
- Jin, Y., Liang, L., Li, J., Xu, K., Zhou, W., & Li, Y. (2024). Artificial intelligence and glaucoma: A lucid and comprehensive review. *Frontiers in Medicine*, 11, Article 1423813.
<https://doi.org/10.3389/fmed.2024.1423813>
- La Bruna, S., Rai, A., Mao, G., Kerr, J., Amin, H., Zemborain, Z. Z., ... & Hood, D. C. (2022). The OCT RNFL probability map and artifacts resembling glaucomatous damage. *Translational Vision Science & Technology*, 11(3), Article 18.
<https://doi.org/10.1167/tvst.11.3.18>
- Lan, C. H., Chiu, T. H., Yen, W. T., & Lu, D. W. (2025). Artificial intelligence in glaucoma: Advances in diagnosis, progression forecasting, and surgical outcome prediction. *International Journal of Molecular Sciences*, 26(10), Article 4473.
<https://doi.org/10.3390/ijms26104473>
- Ling, X. C., Chen, H. S. L., Yeh, P. H., Cheng, Y. C., Huang, C. Y., Shen, S. C., & Lee, Y. S. (2025). Deep learning in glaucoma detection and progression prediction: a systematic review and meta-analysis. *Biomedicines*, 13(2), Article 420.
<https://doi.org/10.3390/biomedicines13020420>
- Myers, J. G., Chen, T. C., Huang, D., & Schuman, J. S. (2024). Is it glaucoma or not? Watch for RNFL OCT artifacts. Retrieved from <https://www.aao.org/eyenet/article/is-it-glaucoma-or-not-watch-for-rnfl-oct-artifacts>
- Qi, T., Liu, H., Fröhn, L., Löw, K., Cursiefen, C., & Prokosch, V. (2025). Understanding glaucoma: Why it remains a leading cause of blindness worldwide. *Klinische Monatsblätter Für Augenheilkunde*, 242(07), 712-717.
<https://doi.org/10.1055/a-2617-1575>
- Ruia, S., & Tripathy, K. (2025). *Humphrey visual field*. StatPearls Publishing. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/36256759/>
- Senjam, S. S. (2020). Glaucoma blindness—A rapidly emerging non-communicable ocular disease in India: Addressing the issue with advocacy. *Journal of Family Medicine and Primary Care*, 9(5), Article 2200.
https://doi.org/10.4103/jfmpe.jfmpe_111_20
- Senthil, S., Rao, D. P., Savoy, F. M., Negiloni, K., Bhandary, S., Chary, R., & Chandrashekar, G. (2025). Evaluating real-world performance of an automated offline glaucoma AI on a smartphone fundus camera across glaucoma severity stages. *PLoS One*, 20(6), Article e0324883.
<https://doi.org/10.1371/journal.pone.0324883>
- Shahriari, M. H., Asadi, F., Moghaddasi, H., Roshanpour, A., Sharifipour, F., & Khorrami, Z. (2025). Applications of machine learning in glaucoma diagnosis based on tabular data: A systematic review. *BMC Biomedical Engineering*, 7(1), Article 9.
<https://doi.org/10.1186/s42490-025-00095-3>
- Shyamalee, T., Meedeniya, D., Lim, G., & Karunarathne, M. (2024). Automated tool support for glaucoma identification with explainability using fundus images. *IEEE Access*, 12, 17290-17307.
<https://doi.org/10.1109/ACCESS.2024.3359698>
- Soh, Z. D., Chee, M. L., Thakur, S., Tham, Y. C., Tao, Y., Lim, Z. W., ... & Cheng, C. Y. (2020). Asian-specific vertical cup-to-disc ratio cut-off for glaucoma screening: An evidence-based recommendation from a multi-ethnic Asian population. *Clinical & Experimental Ophthalmology*, 48(9), 1210-1218. <https://doi.org/10.1111/ceo.13836>
- Sounderajah, V., Guni, A., Liu, X., Collins, G. S., Karthikesalingam, A., Markar, S. R., ... & Ashrafian, H. (2025). The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nature*

- Medicine*, 31(10), 3283-3289.
<https://doi.org/10.1038/s41591-025-03953-8>
- Tello, A., Vazquez, E., Villamizar, S. J., Mejía-Salgado, G., Duarte-Bueno, L. M., Acuña, M. F., & Galvis, V. (2026). Diagnostic accuracy of the non-compliance of the ISNT rule in glaucoma: Systematic review and meta-analysis. *European Journal of Ophthalmology*, 36(2), 434–447.
<https://doi.org/10.1177/11206721251392629>
- Tonti, E., Tonti, S., Mancini, F., Bonini, C., Spadea, L., D'Esposito, F., ... & Zeppieri, M. (2024). Artificial intelligence and advanced technology in glaucoma: A review. *Journal of Personalized Medicine*, 14(10), Article 1062.
<https://doi.org/10.3390/jpm14101062>
- Wanichwecharungruang, B., Jaksataphorn, P., Yuttitham, K., Vanichvaranont, S., & Harncharoen, K. (2011). Optic disc area and diameter of the central retinal vein occlusion fellow eyes, determined by optical coherence tomography. *Journal of the Medical Association of Thailand*, 94(3), S76-S80.
- Wolvaardt, E., & Hu, V. H. (2021). Update on glaucoma. *Community Eye Health*, 34(112), Article 31.
- Wu, H., Wang, Y., Li, F., Liu, Z., & Shi, F. (2023). The national, regional, and global impact of glaucoma as reported in the 2019 Global Burden of Disease Study. *Archives of Medical Science: AMS*, 19(6), Article 1913.
<https://doi.org/10.5114/aoms/172929>
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32-35.
[https://doi.org/10.1002/1097-0142\(1950\)3:1%3C32::AID-CNCR2820030106%3E3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1%3C32::AID-CNCR2820030106%3E3.0.CO;2-3)