

ỨNG DỤNG TRÍ TUỆ NHÂN TẠO TRONG CHẨN ĐOÁN BỆNH GLÔCÔM VÀ THEO DÕI NGẮN HẠN

Phan Nhã Uyên^{1,2,✉}, Đoàn Kim Thành³ và Phạm Trọng Văn¹

¹Trường Đại học Y Hà Nội

²Bệnh viện Trung ương Huế

³Trường Đại học Y khoa Phạm Ngọc Thạch

Nghiên cứu đánh giá độ chính xác chẩn đoán có bắt cặp nhằm xác định giá trị của Cybersight AI trong phát hiện glôcôm nguyên phát tại thời điểm ban đầu và sau 3 tháng. Phân tích được thực hiện trên 78 mắt của 42 người bệnh tại Trung tâm Mắt, Bệnh viện Trung ương Huế, gồm 42 mắt glôcôm nguyên phát và 36 mắt không glôcôm, có đủ kết quả trí tuệ nhân tạo ở cả hai thời điểm. Chẩn đoán của bác sĩ nhãn khoa dựa trên khám lâm sàng, soi góc, OCT lớp sợi thần kinh võng mạc và thị trường được sử dụng làm tiêu chuẩn tham chiếu. Tại thời điểm ban đầu, độ nhạy, độ đặc hiệu, độ chính xác và AUC lần lượt là 61,9%, 86,1%, 73,1% và 0,762. Sau 3 tháng, các giá trị tương ứng là 61,9%, 94,4%, 76,9% và 0,784. Chênh lệch AUC theo cặp không có ý nghĩa thống kê ($p = 0,356$), tương tự độ chính xác tổng thể ($p = 0,371$). Cybersight AI có khả năng phân biệt mức trung bình, phù hợp hơn với vai trò hỗ trợ phân luồng và quyết định lâm sàng hơn là xét nghiệm chẩn đoán độc lập.

Từ khóa: Glôcôm, trí tuệ nhân tạo, học sâu, độ chính xác chẩn đoán, độ nhạy và độ đặc hiệu.

I. ĐẶT VẤN ĐỀ

Glôcôm là nhóm bệnh lý thần kinh thị giác mạn tính, tiến triển và không hồi phục, được xem là một trong những nguyên nhân hàng đầu gây mù lòa trên toàn cầu. Theo Tổ chức Y tế Thế giới, số người mắc glôcôm có thể vượt quá 100 triệu vào năm 2040.¹ Do tổn thương thị thần kinh và mất thị trường không thể phục hồi, phát hiện sớm có vai trò quan trọng trong bảo tồn chức năng thị giác và làm chậm tiến triển bệnh. Tuy nhiên, chẩn đoán sớm glôcôm còn gặp nhiều thách thức do biểu hiện lâm sàng không đặc hiệu, triệu chứng tiến triển âm thầm, và sự phụ thuộc lớn vào chuyên môn của bác sĩ nhãn khoa, vốn còn thiếu hụt ở nhiều khu vực trên thế giới.² Trong bối cảnh đó, sự phát triển của

trí tuệ nhân tạo (artificial intelligence, AI), đặc biệt là các mô hình học sâu (deep learning), đã mở ra những cơ hội mới trong hỗ trợ chẩn đoán glôcôm thông qua phân tích hình ảnh võng mạc và đĩa thị.³

Mặc dù AI đã được ứng dụng trong nhiều lĩnh vực nhãn khoa khác nhau như phát hiện bệnh võng mạc đái tháo đường và thoái hóa hoàng điểm liên quan tuổi, việc ứng dụng AI trong phát hiện glôcôm vẫn đang tiếp tục được đánh giá và tối ưu hóa.^{4,5} Một trong những tiêu chí quan trọng để đánh giá hiệu quả của AI là khả năng phân biệt chính xác giữa người mắc bệnh và người không mắc bệnh, được phản ánh qua các chỉ số như độ nhạy, độ đặc hiệu và diện tích dưới đường cong đặc tính hoạt động của bộ nhận dạng (ROC) (AUC).⁶

Ước tính có tới 75% người mắc glôcôm tại các quốc gia có thu nhập thấp và trung bình (LMICs) không được chẩn đoán cho đến khi đã xuất hiện suy giảm thị lực đáng kể.⁷ Điều này chủ yếu là do thiếu các hệ thống sàng lọc hiệu

Tác giả liên hệ: Phan Nhã Uyên

Trường Đại học Y Hà Nội

Email: phannhauyen15@gmail.com

Ngày nhận: 04/05/2026

Ngày được chấp nhận: 08/06/2026

quả, sự hạn chế về nhân lực nhãn khoa, và cơ sở hạ tầng y tế chưa đầy đủ.⁸ Gánh nặng mắt thị lực do glôcôm thường có tương quan nghịch với mức thu nhập quốc gia, phản ánh sự bất bình đẳng đáng kể trong khả năng tiếp cận các dịch vụ chăm sóc mắt.⁹

Cybersight AI là nền tảng trí tuệ nhân tạo dựa trên điện toán đám mây, có khả năng phân tích ảnh đáy mắt và đưa ra điểm nguy cơ glôcôm. Tuy nhiên, hiệu quả của hệ thống này trong thực hành lâm sàng tại Việt Nam vẫn cần được làm rõ. Vì vậy, nghiên cứu này được thực hiện với 2 mục tiêu: 1) Xác định giá trị chẩn đoán của Cybersight AI trong phát hiện glôcôm nguyên phát và phân tích tỷ lệ phát hiện đúng theo đặc điểm lâm sàng; 2) So sánh sự thay đổi hiệu quả chẩn đoán của Cybersight AI giữa thời điểm ban đầu và sau 3 tháng theo dõi.

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP

1. Đối tượng

Các bệnh nhân trưởng thành (≥ 18 tuổi) đến khám hoặc điều trị nhãn khoa tại Trung tâm Mắt trong thời gian nghiên cứu được xem là đủ điều kiện tham gia nghiên cứu. Bằng phương pháp chọn mẫu thuận tiện, nghiên cứu thu nhận 78 mắt của 42 bệnh nhân, bao gồm 42 mắt glôcôm nguyên phát và 36 mắt thuộc nhóm chứng. Mắt được sử dụng làm đơn vị phân tích.

2. Phương pháp

Thiết kế nghiên cứu: Nghiên cứu tiến cứu đánh giá độ chính xác chẩn đoán, có đo lặp bất cặp trên cùng một mắt sau 3 tháng theo dõi.

Địa điểm và thời gian nghiên cứu: Nghiên cứu được thực hiện tại Trung tâm Mắt, Bệnh viện Trung ương Huế, Việt Nam, từ tháng 12/2024 đến tháng 9/2025.

Cỡ mẫu

Cỡ mẫu cho nghiên cứu xác định một phương pháp chẩn đoán được tính theo công thức ước lượng độ nhạy/độ đặc hiệu cho biến

nhị phân:

$$n = Z_{1-\alpha/2}^2 \frac{P(1-P)}{d^2 \times \text{Prev}}$$

Trong đó $Z_{1-\alpha/2} = 1,96$ với độ tin cậy 95%; P là độ nhạy (hoặc độ đặc hiệu) kỳ vọng dựa trên y văn ($P \approx 0,80$); d là sai số tuyệt đối cho phép ($d = 0,10$); và Prev là tỷ lệ hiện mắc dự kiến của glôcôm trong quần thể nghiên cứu ($\text{Prev} \approx 0,50$ đối với mẫu giàu ca bệnh tại tuyến chuyên khoa). Cỡ mẫu tối thiểu ước tính cho mỗi nhóm là 31 mắt; thực tế nghiên cứu thu nhận 42 mắt glôcôm và 36 mắt chứng, đáp ứng yêu cầu cỡ mẫu.

Khám lâm sàng và thu nhận hình ảnh

Các bệnh nhân được khám mắt toàn diện bởi bác sĩ nhãn khoa, bao gồm sinh hiển vi đèn khe, soi góc tiền phòng bằng kính Goldmann ba mặt gương, đo thị trường tự động Humphrey và chụp OCT lớp sợi thần kinh võng mạc bằng máy NIDEK RS-3000. Sau giãn đồng tử, ảnh đáy mắt màu được chụp bằng máy Zeiss VISUCAM 524. Với mỗi mắt, ít nhất một ảnh đạt chất lượng tốt, căn chỉnh phù hợp, bao gồm đĩa thị và hoàng điểm, được lựa chọn để phân tích bằng Cybersight AI.

Chẩn đoán theo tiêu chuẩn tham chiếu

Chẩn đoán glôcôm nguyên phát hoặc không glôcôm được xác lập độc lập bởi hai bác sĩ nhãn khoa có kinh nghiệm trên 5 năm trong lĩnh vực glôcôm, dựa trên toàn bộ hồ sơ lâm sàng, OCT và thị trường. Các trường hợp bất đồng được phân xử bởi chuyên gia glôcôm cao cấp thứ ba. Tất cả bác sĩ đánh giá được làm mù với kết quả Cybersight AI.

Tiêu chuẩn chẩn đoán

Mắt được chẩn đoán glôcôm nguyên phát khi có ít nhất 2 trong 3 tiêu chuẩn: tổn thương đĩa thị gợi ý glôcôm, gồm VCDR $\geq 0,7$ hoặc bất đối xứng VCDR giữa hai mắt $\geq 0,2$ sau khi đối chiếu với kích thước đĩa thị.^{11,12} Mỏng RNFL trên OCT, với chiều dày toàn bộ hoặc theo vùng thấp hơn bách phân vị thứ 5 của cơ sở dữ liệu

tham chiếu RS-3000 sau khi loại trừ ảnh kém chất lượng hoặc lỗi phân đoạn; và tổn thương thị trường phù hợp glôcôm trên thị trường tự động tiêu chuẩn.¹³ Tổn thương thị trường được xác định khi có ít nhất 2 trong 3 dấu hiệu: ≥ 3 điểm giảm nhảy cảm liên kề ở vị trí điển hình với $p < 0,05$, trong đó có ít nhất 1 điểm $p < 0,01$; PSD bất thường với $p < 0,05$; hoặc GHT ngoài giới hạn bình thường. Lỗm đĩa đơn thuần chỉ được xem là dấu hiệu hỗ trợ và không đủ để xác định chẩn đoán.¹⁴

Mắt được xếp vào nhóm chứng không glôcôm khi có thị trường bình thường, RNFL trên OCT trong giới hạn bình thường và hình thái đĩa thị không gợi ý glôcôm.

Các mắt có đục môi trường trong suốt ảnh hưởng chất lượng ảnh đáy mắt, bệnh toàn thân hạn chế hợp tác, tiền sử chấn thương mắt, glôcôm bẩm sinh hoặc glôcôm thứ phát được loại khỏi nghiên cứu.

Xét nghiệm chỉ số: Cybersight AI

Ảnh đáy mắt đủ điều kiện được tải lên nền tảng Cybersight AI để phân tích tự động. Hệ thống cung cấp điểm nguy cơ glôcôm liên tục từ 0% đến 100% và phân loại nhị phân dương tính hoặc âm tính theo ngưỡng mặc định. Kết quả AI được ghép nối theo từng mắt với chẩn đoán tham chiếu để phân tích.

Biến số nghiên cứu và phân tích thống kê

Độ chính xác chẩn đoán được đánh giá bằng độ nhạy, độ đặc hiệu, giá trị dự báo dương tính, giá trị dự báo âm tính, độ chính xác tổng thể và tỷ số khả dĩ dương/âm. Khả năng phân biệt được đánh giá bằng diện tích dưới đường cong ROC. Mức độ phù hợp giữa Cybersight AI và chẩn đoán lâm sàng được đánh giá bằng hệ số Cohen's κ . So sánh AUC giữa hai thời điểm được thực hiện bằng kiểm định DeLong; so sánh độ chính xác phân loại bất cặp được thực hiện bằng kiểm định McNemar. Ngưỡng ý nghĩa thống kê được xác định khi $p < 0,05$.



Hình 1. Các chỉ số do Cybersight AI cung cấp

3. Đạo đức nghiên cứu

Đề cương nghiên cứu đã được Hội đồng Khoa học Trường Đại học Y Hà Nội thông qua ngày 18 tháng 8 năm 2024 và được Hội đồng

Đạo đức phê duyệt ngày 25 tháng 2 năm 2025 (IRB-VN01.001/IRB00003121/FWA00004148). Nghiên cứu được tiến hành phù hợp với Tuyên bố Helsinki.

III. KẾT QUẢ

1. Đối tượng nghiên cứu và đặc điểm ban đầu

Nghiên cứu bao gồm 78 mắt của 42 bệnh nhân, trong đó có 42 mắt glôcôm nguyên phát và 36 mắt không glôcôm. Các đặc điểm ban đầu của quần thể nghiên cứu được trình bày ở Bảng 1. Tuổi trung bình của đối tượng nghiên cứu là $63,6 \pm 10,3$ tuổi, còn tuổi trung vị là 66

tuổi (IQR: 60 – 70). Phân bố theo bên mắt là tương đương nhau, với 39 mắt phải và 39 mắt trái (mỗi bên chiếm 50,0%). Tổng thể, 42/78 mắt (53,8%) được xếp vào nhóm glôcôm và 36/78 mắt (46,2%) vào nhóm chứng. Tỷ lệ nam giới ở nhóm glôcôm cao hơn có ý nghĩa thống kê so với nhóm chứng (64,3% so với 30,6%; χ^2 , $p = 0,006$). Ngược lại, không có sự khác biệt có ý nghĩa thống kê giữa hai nhóm về tuổi (Mann–Whitney, $p = 0,085$) và bên mắt ($p = 1,000$).

Bảng 1. Đặc điểm nhân khẩu học và lâm sàng ban đầu của mẫu nghiên cứu (n = 78 mắt)

Đặc điểm	Tổng	Glôcôm (n = 42)	Không Glôcôm (n = 36)	Giá trị p
Tuổi (trung bình $\pm$ SD)	$63,6 \pm 10,3$	$65,1 \pm 11,2$	$61,9 \pm 8,9$	0,085 [†]
Trung vị (IQR)	66 (60 – 70)	66 (61 – 74)	66 (55 – 67)	
<i>Giới tính, n (%)</i>				0,006 [§]
Nam	38 (48,7%)	27 (64,3%)	11 (30,6%)	
Nữ	40 (51,3%)	15 (35,7%)	25 (69,4%)	
<i>Mắt, n (%)</i>				1,000 [§]
Phải (MP)	39 (50,0%)	21 (50,0%)	18 (50,0%)	
Trái (MT)	39 (50,0%)	21 (50,0%)	18 (50,0%)	

*Khác biệt có ý nghĩa thống kê ($p < 0,05$). SD: độ lệch chuẩn; IQR: khoảng tứ phân vị; [†]Kiểm định Mann-Whitney; [§] Kiểm định Chi-square

**Tuổi được tính theo người bệnh; các đặc điểm nhóm bệnh và bên mắt được trình bày theo đơn vị mắt

2. Độ chính xác chẩn đoán của AI tại thời điểm khám ban đầu

Tại thời điểm ban đầu, Cybersight AI đạt độ nhạy 61,9% và độ đặc hiệu 86,1% trong phát hiện glôcôm. Khi đối chiếu với chẩn đoán lâm sàng, Cybersight AI ghi nhận 26 dương tính thật, 31 âm tính thật, 5 dương tính giả và 16 âm tính giả. Giá trị dự báo dương tính là 83,9%, giá trị dự báo âm tính 66,0% và độ chính xác tổng thể 73,1%. Mô hình có khả năng phân biệt ở mức chấp nhận được, với AUC 0,762; LR+ 4,46 và LR– 0,442. Mức độ phù hợp với chẩn đoán lâm sàng ở mức trung bình, với chỉ số

Youden 0,480 và Cohen's κ 0,470. Ngưỡng tối ưu theo Youden là 30%.

3. Tỷ lệ Cybersight AI phát hiện đúng theo tổ hợp đặc điểm lâm sàng trong nhóm glôcôm

Trong 42 mắt glôcôm nguyên phát, nhóm có tổn thương thị trường kèm tổn thương đầu thị thần kinh/RNFL nhưng không tăng nhãn áp chiếm tỷ lệ cao nhất, với 24 mắt (57,1%). Cybersight AI phát hiện đúng 19 mắt ở thời điểm ban đầu (79,2%) và 20 mắt sau 3 tháng (83,3%). Nhóm có đủ cả ba yếu tố gồm nhãn áp tăng, tổn thương thị trường và tổn thương đầu thị thần kinh/RNFL có 4 mắt; AI phát hiện đúng

4 mắt ban đầu và 3 mắt sau 3 tháng. Ngược lại, 2 mắt chỉ có nhãn áp tăng kèm tổn thương thị trường đều không được AI phát hiện ở cả hai

thời điểm. Nhìn chung, AI phát hiện tốt hơn ở các mắt có biểu hiện tổn thương đầu thị thần kinh/RNFL rõ ràng.

Bảng 2. Phân bố tổ hợp đặc điểm lâm sàng và tỷ lệ Cybersight AI phát hiện đúng trong nhóm glôcôm nguyên phát

Tổ hợp đặc điểm lâm sàng trong nhóm glôcôm	Số mắt, n (%)	AI dương tính ban đầu, n (%)	AI dương tính sau 3 tháng, n (%)
Nhãn áp tăng + tổn thương thị trường	2 (4,8)	0 (0,0)	0 (0,0)
Nhãn áp tăng + tổn thương đầu thị thần kinh/RNFL	1 (2,4)	1 (100,0)	1 (100,0)
Tổn thương thị trường + tổn thương đầu thị thần kinh/RNFL	24 (57,1)	19 (79,2)	20 (83,3)
Có đồng thời cả 3 yếu tố	4 (9,5)	4 (100,0)	3 (75,0)
Tổ hợp khác hoặc không đủ 2 yếu tố trên	11 (26,2)	2 (18,2)	2 (18,2)
Tổng	42 (100,0)	26 (61,9)	26 (61,9)

*Nhãn áp tăng được xác định khi IOP > 21 mmHg. Tổn thương đầu thị thần kinh/RNFL bao gồm bất thường hình thái đĩa thị gợi ý glôcôm và/hoặc mỏng RNFL phù hợp glôcôm. Các nhóm được phân loại không chồng lấp

4. Độ chính xác chẩn đoán của AI tại thời điểm theo dõi sau 3 tháng

Sau 3 tháng, độ nhạy của Cybersight AI không đổi ở mức 61,9%, trong khi độ đặc hiệu tăng lên 94,4%. Số dương tính giả giảm từ 5

xuống 2, làm PPV tăng lên 92,9%. AUC cải thiện nhẹ lên 0,784; LR+ tăng từ 4,46 lên 11,14, trong khi LR- chỉ giảm nhẹ còn 0,403. Cohen's k tăng lên 0,548. Ngưỡng tối ưu theo Youden là 20%.

Bảng 3. Giá trị chẩn đoán của AI trong phát hiện glôcôm tại thời điểm ban đầu và sau 3 tháng theo dõi (n = 78 mắt)

Thông số	Ban đầu	KTC 95%	Sau 3 tháng	KTC 95%
Độ nhạy	61,9%	46,8 – 75,0	61,9%	46,8 – 75,0
Độ đặc hiệu	86,1%	71,3 – 93,9	94,4%	81,9 – 98,5
Giá trị dự báo dương tính	83,9%	67,4 – 92,9	92,9%	77,4 – 98,0
Giá trị dự báo âm tính	66,0%	51,7 – 77,8	68,0%	54,2 – 79,2
Độ chính xác	73,1%	62,3 – 81,7	76,9%	66,4 – 84,9
AUC	0,762	0,645 – 0,862	0,784	0,673 – 0,879
LR+	4,46	—	11,14	—
LR-	0,442	—	0,403	—

Thông số	Ban đầu	KTC 95%	Sau 3 tháng	KTC 95%
Chỉ số Youden	0,480	—	0,564	—
Cohen κ	0,470	—	0,548	—
Ngưỡng cắt tối ưu	30%	—	20%	—

Bảng 4. Ma trận nhầm lẫn của hệ thống AI trong phân loại glôcôm tại thời điểm ban đầu và sau 3 tháng theo dõi

Ban đầu	AI (+)	AI (-)	Tổng	Sau 3 tháng	AI (+)	AI (-)	Tổng
Glôcôm (+)	TP = 26	FN = 16	42	Glôcôm (+)	TP = 26	FN = 16	42
Glôcôm (-)	FP = 5	TN = 31	36	Glôcôm (-)	FP = 2	TN = 34	36

TP: dương tính thật; FP: dương tính giả; FN: âm tính giả; TN: âm tính thật

5. So sánh hiệu quả của AI giữa các thời điểm

So sánh bắt cặp cho thấy AUC tăng nhẹ sau 3 tháng, từ 0,762 lên 0,784, nhưng không có ý nghĩa thống kê ($\Delta AUC = +0,023$; $p = 0,356$).

Độ chính xác cũng tăng từ 73,1% lên 76,9%, song khác biệt không có ý nghĩa ($\chi^2 = 0,800$; $p = 0,371$). Trong 78 mắt, 56 mắt được phân loại đúng ở cả hai thời điểm, 17 mắt sai ở cả hai thời điểm và 5 mắt có kết quả không tương hợp.

Bảng 5. So sánh thống kê hiệu quả chẩn đoán của AI tại thời điểm ban đầu và sau 3 tháng theo dõi

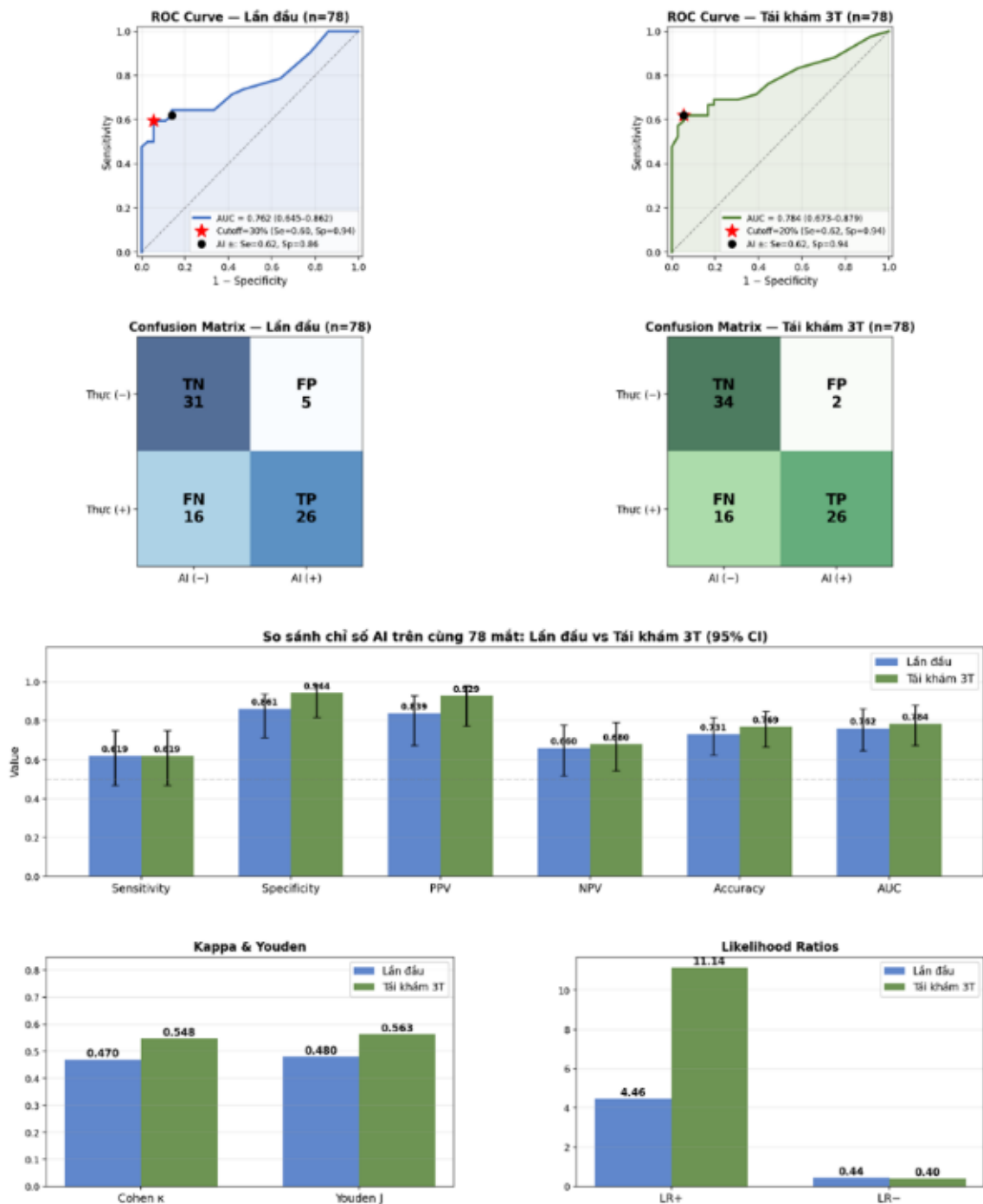
Phép kiểm	Thống kê	Giá trị p	Diễn giải
Kiểm định DeLong (ΔAUC)	+0,023	0,356	Không có ý nghĩa thống kê
Kiểm định McNemar (độ chính xác)	$\chi^2 = 0,800$	0,371	Không có ý nghĩa thống kê

AUC: diện tích dưới đường cong đặc tính hoạt động của bộ nhận dạng (receiver operating characteristic curve). Ngưỡng ý nghĩa thống kê được xác định tại $p < 0,05$

6. Phân tích đường cong ROC và minh họa trực quan

Đường cong ROC của hệ thống AI trong phát hiện glôcôm tại thời điểm ban đầu và sau 3 tháng theo dõi được trình bày ở Biểu đồ 1. Đánh giá trực quan cho thấy khả năng phân biệt tổng thể của mô hình giữa hai thời điểm

nhìn chung tương đương, mặc dù đường cong tại thời điểm 3 tháng có xu hướng cải thiện nhẹ. Các điểm vận hành tương ứng với kết quả phân loại nhị phân của AI (dương tính/âm tính) cũng được biểu diễn, cùng với ngưỡng cắt tối ưu xác định theo nguyên tắc tối đa hóa chỉ số Youden.



Biểu đồ 1. Hiệu quả chẩn đoán của AI trong phát hiện glôcôm: đường cong ROC, đối chiếu với chẩn đoán lâm sàng và so sánh các chỉ số hiệu quả tại thời điểm ban đầu và sau 3 tháng theo dõi (n = 78 mắt)

IV. BÀN LUẬN

Trong nghiên cứu đánh giá độ chính xác chẩn đoán bắt cặp trên 78 mắt, Cybersight AI cho thấy khả năng phân biệt glôcôm ở mức trung bình. Tại thời điểm ban đầu, AUC đạt 0,762, độ nhạy 61,9%, độ đặc hiệu 86,1% và độ chính xác tổng thể 73,1%. Sau 3 tháng, AUC tăng nhẹ lên 0,784, độ đặc hiệu tăng lên 94,4%, trong khi độ nhạy không thay đổi ở mức 61,9%. Kết quả này cho thấy Cybersight AI có giá trị hỗ trợ phát hiện glôcôm nguyên phát, nhưng hiệu quả chưa đủ để sử dụng như một xét nghiệm chẩn đoán độc lập.

Về mục tiêu đánh giá giá trị chẩn đoán, độ đặc hiệu tương đối cao cho thấy kết quả AI dương tính có ý nghĩa hỗ trợ xác nhận glôcôm, đặc biệt trong nhóm người bệnh đã có nghi ngờ lâm sàng. Điều này phù hợp với LR+ tăng từ 4,46 tại thời điểm ban đầu lên 11,14 sau 3 tháng. Ngược lại, LR- chỉ đạt 0,442 ban đầu và 0,403 sau 3 tháng, cho thấy kết quả AI âm tính chỉ làm giảm xác suất mắc bệnh ở mức hạn chế. Do đó, AI âm tính không đủ để loại trừ glôcôm, nhất là ở các trường hợp còn nghi ngờ trên lâm sàng, OCT hoặc thị trường.¹⁷

Độ nhạy còn khiêm tốn có thể được giải thích qua đặc điểm của các trường hợp âm tính giả. Trong 16 mắt glôcôm bị AI bỏ sót, phần lớn là các trường hợp giai đoạn sớm, có tỷ lệ lõm đĩa theo chiều dọc chưa rõ, VCDR < 0,7, và tổn thương thị trường còn khu trú. Ở giai đoạn này, biến đổi hình thái đĩa thị trên ảnh đáy mắt hai chiều thường kín đáo. Trong khi đó, khám lâm sàng bằng kính 90D cho phép đánh giá đĩa thị theo không gian lập thể, nhận diện tốt hơn độ sâu lõm đĩa, hình thái vành thần kinh võng mạc và các tổn thương khu trú. Ngoài ra, tiêu chuẩn tham chiếu trong nghiên cứu dựa trên tổng hợp khám lâm sàng, soi góc, OCT lớp sợi thần kinh võng mạc và thị trường, trong khi Cybersight AI chủ yếu phân tích ảnh đáy mắt màu. Sự khác

biệt về lượng thông tin đầu vào này có thể làm giảm khả năng phát hiện glôcôm sớm của hệ thống.

Phân tích theo tổ hợp đặc điểm lâm sàng cũng cho thấy Cybersight AI phát hiện tốt hơn ở các mắt đã có tổn thương cấu trúc rõ trên đầu thị thần kinh hoặc RNFL. Ngược lại, các mắt chưa có biến đổi cấu trúc điển hình dễ bị AI bỏ sót, dù có nhãn áp tăng hoặc tổn thương thị trường. Nhận xét này nhấn mạnh rằng hiệu quả của AI dựa trên ảnh đáy mắt phụ thuộc nhiều vào mức độ biểu hiện hình thái của tổn thương gai thị và lớp sợi thần kinh võng mạc.

Các trường hợp dương tính giả ở nhóm chứng chủ yếu gặp ở những mắt có lõm đĩa sinh lý lớn, VCDR khoảng 0,4 – 0,6, hoặc có đặc điểm đĩa thị không điển hình như đĩa thị nghiêng, teo quanh đĩa rộng và cận thị trục. Những đặc điểm này có thể làm tăng điểm nguy cơ glôcôm trên ảnh đáy mắt, mặc dù OCT và thị trường không ủng hộ chẩn đoán. Vì vậy, VCDR đơn thuần không nên được diễn giải tách rời khỏi kích thước đĩa thị, hình thái vành thần kinh võng mạc, OCT và dữ liệu chức năng.

Về mục tiêu so sánh hiệu quả sau 3 tháng, nghiên cứu ghi nhận AUC tăng nhẹ từ 0,762 lên 0,784 và độ chính xác tổng thể tăng từ 73,1% lên 76,9%; tuy nhiên, các khác biệt này không có ý nghĩa thống kê. Số dương tính giả giảm từ 5 mắt xuống còn 2 mắt, trong khi số âm tính giả không thay đổi. Sự cải thiện này nhiều khả năng liên quan đến chất lượng ảnh đầu vào tốt hơn, kỹ thuật viên quen hơn với quy trình chụp, giãn đồng tử ổn định hơn và lựa chọn ảnh phù hợp hơn, hơn là do thay đổi đặc tính nội tại của mô hình AI.

Chất lượng ảnh là yếu tố quan trọng ảnh hưởng đến kết quả của Cybersight AI. Trong 5 mắt dương tính giả tại thời điểm ban đầu, có 3 ảnh có chất lượng giới hạn về độ nét và độ tương phản. Ảnh kém nét, chiếu sáng không

đồng đều, căn chỉnh chưa tối ưu hoặc đục môi trường trong suốt nhẹ có thể làm sai lệch nhận diện bờ đĩa thị, vùng lõm đĩa và tổn thương quanh gai. Do đó, để tối ưu hiệu quả của AI, ảnh đáy mắt đầu vào cần bảo đảm đĩa thị được thể hiện rõ, đủ nét, đủ tương phản, chiếu sáng đồng đều, không bị che khuất và nên được chụp sau giãn đồng tử đạt yêu cầu.

Việc ngưỡng cắt tối ưu thay đổi từ 30% tại thời điểm ban đầu xuống 20% sau 3 tháng cũng gợi ý có sự dao động trong hiệu chuẩn xác suất giữa các lần đánh giá. Vì vậy, điểm nguy cơ do AI cung cấp không nên được áp dụng cứng nhắc trong mọi bối cảnh. Trong thực hành, cần diễn giải kết quả AI dựa trên chất lượng ảnh, đặc điểm đĩa thị, nguy cơ lâm sàng, OCT và thị trường. So với các mô hình học sâu đa phương thức, Cybersight AI có thể có giới hạn do chủ yếu dựa trên ảnh đáy mắt màu, trong khi chẩn đoán glôcôm cần tích hợp cả dữ liệu cấu trúc và chức năng.

Nghiên cứu có ưu điểm là thiết kế bắt cặp, tiêu chuẩn tham chiếu dựa trên đánh giá toàn diện của bác sĩ nhãn khoa và báo cáo đầy đủ các chỉ số vận hành như độ nhạy, độ đặc hiệu, giá trị dự báo, AUC, tỷ số khả dĩ, chỉ số Youden và Cohen's κ . Tuy nhiên, nghiên cứu còn hạn chế bởi cỡ mẫu nhỏ, thời gian theo dõi ngắn, phân tích theo mắt và chưa phân tầng đầy đủ theo giai đoạn glôcôm, chất lượng ảnh, cận thị hoặc kiểu hình đĩa thị đặc biệt. Các nghiên cứu tiếp theo nên có thiết kế tiền cứu đa trung tâm, cỡ mẫu lớn hơn, chuẩn hóa quy trình chụp ảnh và phân tích sâu các yếu tố ảnh hưởng đến sai số phân loại của AI.¹⁸

V. KẾT LUẬN

Trong nghiên cứu bắt cặp trên 78 mắt của 42 người bệnh, gồm 42 mắt glôcôm nguyên phát và 36 mắt không glôcôm, Cybersight AI cho thấy khả năng phân biệt ở mức trung bình, độ nhạy ổn định sau 3 tháng và xu hướng tăng độ

đặc hiệu nhưng chưa có ý nghĩa thống kê. Kết quả gợi ý AI nên được sử dụng như công cụ hỗ trợ phân luồng và quyết định lâm sàng, không thay thế sàng lọc độc lập. Cần nghiên cứu tiền cứu đa trung tâm, cỡ mẫu lớn và chuẩn hóa quy trình ghi ảnh trước khi triển khai rộng rãi.

Khuyến nghị về chất lượng ảnh

Để đạt hiệu quả chẩn đoán tối ưu khi sử dụng Cybersight AI, ảnh đáy mắt cần được chụp sau khi giãn đồng tử đầy đủ; bảo đảm độ phân giải, độ tương phản và độ sáng đồng đều; căn chỉnh đúng tâm với trường nhìn bao gồm đĩa thị và hoàng điểm; tránh các ảnh có ảo ảnh do mi mắt, đục môi trường trong suốt hoặc nhòe do chuyển động. Cần có quy trình kiểm soát chất lượng ảnh chuẩn hóa và kỹ thuật viên được đào tạo thống nhất; các ảnh không đạt tiêu chuẩn nên được chụp lại trước khi đưa vào phân tích bằng AI để tránh kết quả dương tính giả hoặc âm tính giả do chất lượng ảnh.

TÀI LIỆU THAM KHẢO

1. Tienan Qi HL, Frühn L, Löw K, Cursiefen C, Prokosch V. Understanding Glaucoma: Why it Remains a Leading Cause of Blindness Worldwide. *Klin Monbl Augenheilkd*. 2025 Jul;242(7):712-717.
2. Tonti E, Tonti S, Mancini F, et al. Artificial Intelligence and Advanced Technology in Glaucoma: A Review. *Journal of personalized medicine*. 2024;14(10).
3. Ling XC, Chen HS, Yeh PH, et al. Deep Learning in Glaucoma Detection and Progression Prediction: A Systematic Review and Meta-Analysis. *Biomedicine*. 2025;13(2).
4. Anton N DB, Curteanu S, Catălin L, Ilie OD, Târcoveanu F, Bogdănici CM. Comprehensive Review on the Use of Artificial Intelligence in Ophthalmology and Future Research Directions. *Diagnostics (Basel)*. 2022 13:100.

5. Shahriari MH, Asadi F, Moghaddasi H, Roshanpour A, Sharifipour F, Khorrami Z. Applications of machine learning in glaucoma diagnosis based on tabular data: a systematic review. *BMC Biomedical Engineering*. 2025;7(1):9.
6. Ashtari-Majlan M, Dehshibi MM, Masip D. Glaucoma diagnosis in the era of deep learning: A survey. *Expert Systems with Applications*. 2024;256:124888.
7. SS. S. Glaucoma blindness-A rapidly emerging non-communicable ocular disease in India: Addressing the issue with advocacy. *J Family Med Prim Care*. 2020 31(5):2200-2206.
8. Wolvaardt EH, V. H. Update on glaucoma. *Community Eye Health Journal*. 2021;34(112):31.
9. Wu H WY, Li F, Liu Z, Shi F. The national, regional, and global impact of glaucoma as reported in the 2019 Global Burden of Disease Study. *Arch Med Sci*. 2023 19(6):1913-1919.
10. Jaccard N. AI for Glaucoma Care. *Glaucoma Today*. 2022.
11. Foster P, Buhrmann R, Quigley H, Johnson G. The definition and classification of glaucoma in prevalence surveys. *The British journal of ophthalmology*. 2002;86:238-242.
12. Crowston JG, Hopley CR, Healey PR, Lee A, Mitchell P, Blue Mountains Eye S. The effect of optic disc diameter on vertical cup to disc ratio percentiles in a population based cohort: the Blue Mountains Eye Study. *Br J Ophthalmol*. 2004;88(6):766-770.
13. Myers JGaH, David S. and Schuman, Joel S. Is It Glaucoma or Not? Watch for RNFL OCT Artifacts. *EyeNet Magazine, American Academy of Ophthalmology*. 2024.
14. Ruia S, Tripathy K. Humphrey Visual Field. In: *StatPearls*. Treasure Island (FL) ineligible companies. Disclosure: Koushik Tripathy declares no relevant financial relationships with ineligible companies.: StatPearls Publishing 2025.
15. La Bruna S, Rai A, Mao G, et al. The OCT RNFL Probability Map and Artifacts Resembling Glaucomatous Damage. *Transl Vis Sci Technol*. 2022;11(3):18.
16. Tello A, Vazquez E, Villamizar SJ, et al. Diagnostic accuracy of the non-compliance of the ISNT rule in glaucoma: Systematic review and meta-analysis. *European journal of ophthalmology*. 2025:11206721251392629.
17. Tham YC, Li X, Wong TY, Quigley HA, Aung T, Cheng CY. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121(11):2081-2090.
18. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ open*. 2016;6(11):e012799.
19. Huang X, Sun J, Gupta K, et al. Detecting glaucoma from multi-modal data using probabilistic deep learning. *Front Med (Lausanne)*. 2022;9:923096.
20. Crowston JG, Hopley CR, Healey PR, Lee A, Mitchell P. The effect of optic disc diameter on vertical cup to disc ratio percentiles in a population based cohort: the Blue Mountains Eye Study. *Br J Ophthalmol*. 2004;88(6):766-770.

Summary

APPLICATION OF ARTIFICIAL INTELLIGENCE IN THE DIAGNOSIS OF GLAUCOMA AND SHORT-TERM FOLLOW-UP

This paired diagnostic accuracy study evaluated Cybersight AI for detecting primary glaucoma at baseline and 3 months. The analysis included 78 eyes from 42 patients at Hue Central Hospital Eye Center, including 42 primary glaucoma eyes and 36 non-glaucoma eyes with complete paired AI outputs. Comprehensive ophthalmologist diagnosis based on clinical examination, gonioscopy, retinal nerve fiber layer OCT and visual fields served as the reference standard. At baseline, sensitivity, specificity, accuracy and AUC were 61.9%, 86.1%, 73.1% and 0.762, respectively. At 3 months, the corresponding estimates were 61.9%, 94.4%, 76.9% and 0.784. The paired AUC difference was not statistically significant ($p = 0.356$), and overall accuracy did not differ significantly ($p = 0.371$). Cybersight AI showed moderate discrimination with stable sensitivity and a non-significant trend toward higher specificity, supporting its role as an adjunct for triage and clinical decision support rather than a stand-alone diagnostic test.

Keywords: Glaucoma, artificial intelligence, fundus photography, diagnostic accuracy, ROC-AUC, triage.