Vietnam Journal of Community Medicine, Vol. 67, No.11, 201-205

APPLICATION OF ARTIFICIAL INTELLIGENCE FOR BREAST CANCER RISK STRATIFICATION ON MAMMOGRAPHY AT BACH MAI HOSPITAL

Tran Thi Ngoc Mai[1,3]*, Vu Dang Luu[1,2], Luu Hong Nhung[1,2], Tang Van Dung[1], Tran Dang Khoa[3], Vu Trong Huy[3], Nguyen Huy Phuong[3], Nguyen Thi Kim Tien[3]

[1]*Hanoi Medical University - No. 1 Ton That Tung, Kim Lien, Hanoi*
[2]*Institute of Diagnostic and Interventional Radiology - Bach Mai Hospital - No. 78 Giai Phong Road, Kim Lien Ward, Hanoi, Vietnam*
[3]*Friendship Hospital - No. 1, Tran Khanh Du Street, Hai Ba Trung Ward, Hanoi*

Received: 29/07/2026
Reviced: 29/09/2026; Accepted: 19/11/2026

ABSTRACT

Objective: To describe the mammographic features of pathology-confirmed breast cancers and to evaluate the performance of artificial intelligence (AI) in supporting risk stratification of mammographically suspicious breast lesions.

Methods: This retrospective diagnostic-accuracy study included 601 women whose index mammograms were obtained at Bach Mai Hospital from January 2022 to April 2024: 277 pathology-confirmed cases and 324 cases with ≥2 years of negative follow-up (176 cancers, 425 non-cancers). Mammograms were analyzed using Lunit INSIGHT MMG version 1.1.2.7 (Lunit Inc., Seoul, Republic of Korea). Each case was assessed using three models: AI alone (positivity threshold ≥15%), independent radiologist assessment, and radiologist assessment combined with AI (BI-RADS ≥4A). AUCs were compared using DeLong's test, and accuracy was compared using McNemar's test.

Results: AI alone achieved an AUC of 0.828 (95% CI: 0.782-0.875), sensitivity 79.5%, specificity 90.1%, accuracy 87.0%. Independent radiologist assessment yielded an AUC of 0.936; radiologist assessment combined with AI achieved the highest AUC of 0.961, with sensitivity 88.1%, specificity 96.9%, and accuracy 94.3% ($p < 0.001$ versus both other models).

Conclusion: AI showed good performance in risk stratification on mammography. Combining AI with radiologist assessment achieved the highest classification performance, supporting the role of AI as an adjunctive reading tool.

Keywords: breast cancer, mammography, artificial intelligence, BI-RADS, ROC curve.

*Corresponding author
Email: maitn0602@gmail.com **Phone**: (+84) 946907366 **DOI:** 10.52163/yhc.v67i11.7156

ỨNG DỤNG TRÍ TUỆ NHÂN TẠO TRONG PHÂN LOẠI NGUY CƠ UNG THƯ VÚ TRÊN NHŨ ẢNH TẠI BỆNH VIỆN BẠCH MAI

Trần Thị Ngọc Mai[1,3*], Vũ Đăng Lưu[1,2], Lưu Hồng Nhung[1,2], Tăng Văn Dũng[1], Trần Đăng Khoa[3], Vũ Trọng Huy[3], Nguyễn Huy Phương[3], Nguyễn Thị Kim Tiền[3]

[1]Trường Đại học Y Hà Nội - Số 1 Tôn Thất Tùng, Kim Liên, Hà Nội
[2]Viện Điện quang chẩn đoán và can thiệp - Bệnh viện Bạch Mai - Số 78 Đường Giải Phóng, P. Kim Liên, TP Hà Nội, Việt Nam
[3]Bệnh viện Hữu Nghị - Số 1, phố Trần Khánh Dư, phường Hai Bà Trưng, TP. Hà Nội

Ngày nhận bài: 29/07/2026
Ngày chỉnh sửa: 29/09/2026; Ngày duyệt đăng: 19/11/2026

TÓM TẮT

Mục tiêu: Mô tả đặc điểm nhũ ảnh của ung thư vú được xác định bằng giải phẫu bệnh và đánh giá hiệu quả của trí tuệ nhân tạo (AI) trong hỗ trợ phân loại nguy cơ tổn thương nghi ngờ ung thư vú trên nhũ ảnh.

Phương pháp: Nghiên cứu đánh giá độ chính xác chẩn đoán, hồi cứu trên 601 bệnh nhân có phim nhũ ảnh ban đầu được thực hiện tại Bệnh viện Bạch Mai từ 01/2022 đến 04/2024, gồm 277 trường hợp có kết quả giải phẫu bệnh và 324 trường hợp theo dõi âm tính ≥2 năm (176 ung thư, 425 không ung thư). Hình ảnh được phân tích bằng Lunit INSIGHT MMG, phiên bản 1.1.2.7 (Lunit Inc., Seoul, Hàn Quốc). Mỗi trường hợp được đánh giá theo ba mô hình: AI đơn độc (ngưỡng dương tính ≥15%), bác sĩ đọc độc lập và bác sĩ đọc kết hợp AI (dương tính khi BI-RADS ≥4A). AUC được so sánh bằng kiểm định DeLong; độ chính xác được so sánh bằng kiểm định McNemar.

Kết quả: AI đơn độc đạt AUC 0,828 (95% CI: 0,782-0,875), độ nhạy 79,5%, độ đặc hiệu 90,1%, độ chính xác 87,0%. Bác sĩ đọc độc lập đạt AUC 0,936; bác sĩ đọc kết hợp AI đạt AUC cao nhất, 0,961, với độ nhạy 88,1%, độ đặc hiệu 96,9% và độ chính xác 94,3% ($p < 0,001$ so với hai mô hình còn lại).

Kết luận: AI có khả năng phân tầng nguy cơ tốt trên nhũ ảnh. Kết hợp AI với bác sĩ đọc phim cho hiệu quả phân loại cao nhất, ủng hộ vai trò của AI như một công cụ hỗ trợ đọc phim.

Từ khóa: ung thư vú, nhũ ảnh, trí tuệ nhân tạo, BI-RADS, đường cong ROC.

1. ĐẶT VẤN ĐỀ

Ung thư vú là ung thư thường gặp nhất ở phụ nữ trên toàn thế giới. Theo GLOBOCAN 2022, bệnh chiếm khoảng 2,3 triệu ca mắc mới, tương đương 11,6% tổng số ca ung thư mới trên toàn cầu[1]. Nhũ ảnh giữ vai trò chủ đạo trong sàng lọc, song hiệu quả diễn giải phụ thuộc vào kinh nghiệm người đọc, chất lượng hình ảnh và mật độ mô vú[2]. Đọc kép được sử dụng trong nhiều chương trình sàng lọc nhằm tăng độ an toàn chẩn đoán nhưng làm tăng khối lượng đọc; thử nghiệm MASAI cho thấy quy trình có hỗ trợ AI có thể giảm đáng kể khối lượng đọc so với đọc kép tiêu chuẩn[3].

Các hệ thống trí tuệ nhân tạo (AI) dựa trên học sâu có tiềm năng hỗ trợ phát hiện, định vị và đánh giá mức độ nghi ngờ của tổn thương trên nhũ ảnh, đóng vai trò như một người đọc thứ hai[4]. Hiệu quả của AI phụ thuộc vào đặc điểm quần thể, mật độ mô vú và ngưỡng điểm áp dụng, nên cần được đánh giá trên dữ liệu thực tế tại từng cơ sở. Việc triển khai lâm sàng còn phụ thuộc vào cách tích hợp AI vào quy trình, lựa chọn ngưỡng vận hành và giám sát hiệu quả[5]. Trong nghiên cứu này, hình ảnh được phân tích bằng Lunit INSIGHT MMG, phiên bản 1.1.2.7 (Lunit Inc., Seoul, Hàn Quốc). Phần mềm cung cấp điểm bất thường từ 0 đến 100 cho từng bên vú; điểm càng cao phản ánh mức độ tin cậy của thuật toán về sự hiện diện của tổn thương nghi ngờ, nhưng không được hiểu trực tiếp là xác suất mắc ung thư. Nghiên cứu được thực hiện nhằm: (1) mô tả đặc điểm nhũ ảnh của ung thư vú xác định bằng giải phẫu bệnh; (2) đánh giá hiệu quả của AI và vai trò hỗ trợ bác sĩ trong phân loại nguy cơ tổn thương nghi ngờ ung thư vú.

2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Thiết kế nghiên cứu

Nghiên cứu đánh giá độ chính xác chẩn đoán, hồi cứu.

2.2. Địa điểm và thời gian nghiên cứu

Trung tâm Điện quang, Bệnh viện Bạch Mai. Các phim nhũ ảnh ban đầu được thực hiện từ tháng 01/2022 đến tháng 04/2024; nhóm không sinh thiết được xác định bằng theo dõi âm tính tối thiểu 24 tháng.

2.3. Đối tượng nghiên cứu

Bệnh nhân được chụp nhũ ảnh đầy đủ hai tư thế CC và MLO, được phân tích bằng phần mềm Lunit INSIGHT MMG phiên bản 1.1.2.7, có kết quả giải phẫu bệnh hoặc được theo dõi tối thiểu 24 tháng. Loại trừ: thiếu dữ liệu; theo dõi không đủ thời gian; có chỉ định can thiệp nhưng không có mô bệnh học; ung thư đa ổ; hoặc đã điều trị trong thời gian theo dõi.

2.4. Cỡ mẫu, chọn mẫu

Chọn mẫu thuận tiện, gồm toàn bộ bệnh nhân đáp ứng tiêu chuẩn lựa chọn trong thời gian nghiên cứu (n = 601).

*Tác giả liên hệ
Email: maitn0602@gmail.com **Điện thoại**: (+84) 946907366 **DOI:** 10.52163/yhc.v67i11.7156

2.5. Biến số, chỉ số nghiên cứu

Tuổi, vị trí tổn thương, mật độ mô vú, các đặc điểm tổn thương trên nhũ ảnh (khối/vùng tăng đậm độ, vi vôi hóa, biến dạng cấu trúc, bất đối xứng), phân loại BI-RADS của bác sĩ đọc độc lập, điểm bất thường do AI cung cấp (thang điểm 0-100; sau đây gọi là điểm AI), kết quả giải phẫu bệnh hoặc tình trạng theo dõi, và các chỉ số hiệu quả chẩn đoán: AUC, độ nhạy, độ đặc hiệu, giá trị dự báo dương tính (PPV), giá trị dự báo âm tính (NPV) và độ chính xác.

2.6. Kỹ thuật, công cụ và quy trình thu thập số liệu

Nhũ ảnh được thực hiện trên hệ thống nhũ ảnh kỹ thuật số AMULET Innovality (FUJIFILM Nhật Bản). Hình ảnh được lưu trữ trên hệ thống PACS MINERVA và phân tích bằng Lunit INSIGHT MMG, phiên bản 1.1.2.7 (Lunit Inc., Seoul, Hàn Quốc). Phần mềm phân tích các ảnh CC và MLO, cung cấp điểm bất thường từ 0 đến 100 cho từng bên vú; điểm cao hơn biểu thị mức độ tin cậy lớn hơn của thuật toán về sự hiện diện của tổn thương nghi ngờ. Điểm này là đầu ra của thuật toán và không tương đương trực tiếp với xác suất ung thư. Với nhóm có giải phẫu bệnh, điểm AI phân tích là điểm của bên có tổn thương; với nhóm theo dõi âm tính, sử dụng điểm AI cao nhất giữa hai bên. Mỗi trường hợp được đánh giá theo ba mô hình: AI đơn độc, bác sĩ đọc độc lập và bác sĩ đọc kết hợp AI. Người đọc được làm mù thông tin lâm sàng, siêu âm và mô bệnh học. Giải phẫu bệnh là tiêu chuẩn tham chiếu cho nhóm sinh thiết; theo dõi âm tính tối thiểu 24 tháng xác định nhóm không ung thư. AI được coi là dương tính khi điểm ≥15%; hai mô hình bác sĩ được coi là dương tính khi BI-RADS ≥4A.

2.7. Xử lý và phân tích số liệu

Số liệu được quản lý và phân tích bằng IBM SPSS 20.00. Biến định lượng được trình bày bằng trung bình ± độ lệch chuẩn; biến định tính bằng tần số và tỷ lệ %. Hiệu quả chẩn đoán được đánh giá qua AUC, độ nhạy, độ đặc hiệu, PPV, NPV và độ chính xác. AUC của AI được tính từ điểm AI liên tục 0-100; AUC của hai mô hình bác sĩ được tính từ toàn bộ thang BI-RADS có thứ bậc. Các chỉ số độ nhạy, độ đặc hiệu, PPV, NPV và độ chính xác được tính tại ngưỡng AI ≥15% và BI-RADS ≥4A. So sánh AUC sử dụng kiểm định DeLong; so sánh độ chính xác sử dụng kiểm định McNemar. Ngưỡng ý nghĩa thống kê $p < 0,05$.

2.8. Đạo đức nghiên cứu

Nghiên cứu được Hội đồng Đạo đức trong nghiên cứu y sinh học của Bệnh viện Bạch Mai thông qua theo quyết định số 410/BM-HĐĐĐ, ngày 13/07/2026. Thông tin đối tượng nghiên cứu được bảo mật, chỉ phục vụ mục đích nghiên cứu; số liệu được thu thập và phân tích trung thực, khách quan.

3. KẾT QUẢ

3.1. Đặc điểm chung của nhóm nghiên cứu

Bảng 1. Đặc điểm chung của nhóm nghiên cứu

Đặc điểm	n	Tỷ lệ / Ghi chú
Tổng số bệnh nhân	601	100%
Nhóm có giải phẫu bệnh	277	46,1%
Ung thư trong nhóm giải phẫu bệnh	176	63,5% nhóm GPB
Không ung thư trong nhóm giải phẫu bệnh	101	36,5% nhóm GPB
Nhóm theo dõi 2 năm âm tính	324	53,9%
Tổng số ung thư	176	29,3% toàn mẫu
Tổng số không ung thư	425	70,7% toàn mẫu
Tuổi trung bình toàn bộ mẫu (năm)	46,6 ± 9,5	Trung vị 46; khoảng 14-84

Tuổi trung bình toàn mẫu là 46,6 ± 9,5 tuổi (trung vị 46; khoảng 14-84). Nhóm ung thư có tuổi trung bình cao hơn nhóm không ung thư (51,2 so với 44,8 tuổi; $p < 0,001$).

3.2. Đặc điểm nhũ ảnh của các trường hợp có giải phẫu bệnh

Bảng 2. Một số đặc điểm nhũ ảnh theo kết quả giải phẫu bệnh

Đặc điểm	GPB ung thư (n=176)	GPB không ung thư (n=101)	p
Tuổi trung bình (năm)	51,2 ± 12,0	40,1 ± 11,4	<0,001
Bên phải	90 (51,1%)	49 (48,5%)	0,673
Bên trái	86 (48,9%)	52 (51,5%)	0,673
Mật độ mô vú C/D theo bác sĩ	117 (66,5%)	67 (66,3%)	1,000
BI-RADS 4C-5 theo bác sĩ đọc	50 (28,4%)	0 (0,0%)	<0,001
Khối/vùng tăng đậm độ	114 (64,8%)	36 (35,6%)	<0,001
Vi vôi hóa	40 (22,7%)	0 (0,0%)	<0,001
Biến dạng cấu trúc	50 (28,4%)	1 (1,0%)	<0,001
Bất đối xứng/khác	5 (2,8%)	6 (5,9%)	0,217

Nhóm ung thư có tuổi trung bình cao hơn rõ rệt so với nhóm không ung thư (51,2 ± 12,0 so với 40,1 ± 11,4 tuổi; $p < 0,001$). Không ghi nhận sự khác biệt có ý nghĩa thống kê về bên tổn thương và tỷ lệ mô vú mật độ C/D giữa hai nhóm. Các dấu hiệu BI-RADS 4C–5, khối hoặc vùng tăng đậm độ, vi vôi hóa và biến dạng cấu trúc gặp nhiều hơn có ý nghĩa ở nhóm ung thư ($p < 0,001$). Trong đó, BI-RADS 4C–5 và vi vôi hóa chỉ được ghi nhận ở nhóm ung thư. Tỷ lệ bất đối xứng hoặc các dấu hiệu khác không khác biệt có ý nghĩa giữa hai nhóm ($p = 0,217$).

3.3. Điểm AI theo tình trạng tham chiếu và phân loại BI-RADS

Bảng 3. Điểm AI theo tình trạng tham chiếu và phân loại BI-RADS

Nhóm phân tích	n	Điểm AI TB ± SD	Trung vị (IQR)	p
Ung thư (GPB)	176	48,7 ± 36,4	42,0 (16,0-87,0)	<0,001
Không ung thư	425	6,0 ± 13,2	2,0 (1,0-6,0)	
BI-RADS 1-3 (bác sĩ độc lập)	466	7,3 ± 13,2	2,0 (1,0-7,0)	<0,001
BI-RADS ≥4A (bác sĩ độc lập)	135	57,3 ± 37,7	67,0 (21,0-96,0)	

Điểm AI trung bình cao hơn rõ rệt ở nhóm ung thư so với nhóm không ung thư (48,7 so với 6,0; $p < 0,001$) và ở nhóm BI-RADS ≥4A so với nhóm BI-RADS 1-3 (57,3 so với 7,3; $p < 0,001$). Xét theo từng mức BI-RADS cụ thể, điểm AI trung bình tăng dần từ 3,8 ở BI-RADS 1 lên 15,3 ở BI-RADS 3, 45,1 ở BI-RADS 4A và đạt 66,0 ở BI-RADS 5 ($p < 0,001$ giữa các nhóm).

3.4. Lựa chọn ngưỡng AI tối ưu trong phân loại nguy cơ

Bảng 4. Hiệu quả phân loại nguy cơ của AI theo các ngưỡng điểm

Ngưỡng AI (%)	Độ nhạy (%)	Độ đặc hiệu (%)	PPV (%)	NPV (%)	Độ chính xác (%)	Youden
10	79,5	85,9	70,0	91,0	84,0	0,654
15	79,5	90,1	76,9	91,4	87,0	0,697
20	72,2	93,2	81,4	89,0	87,0	0,653
30	61,9	95,1	83,8	85,8	85,4	0,570

Ngưỡng AI 15% cho chỉ số Youden cao nhất (0,697), với độ nhạy 79,5%, độ đặc hiệu 90,1% và độ chính xác 87,0%, và được lựa chọn làm điểm cắt cân bằng nhất trong bộ số liệu này.

3.5. Hiệu quả và so sánh giữa các mô hình phân loại nguy cơ

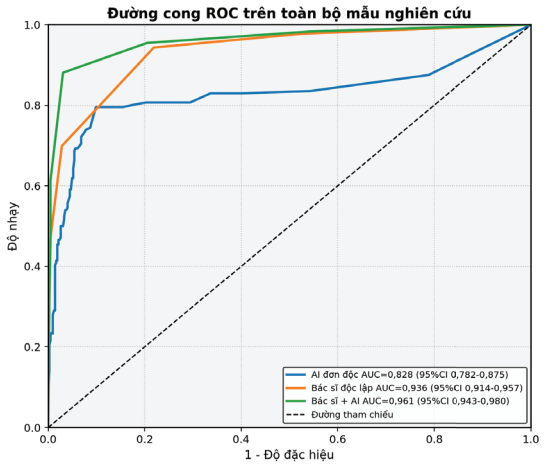
Bảng 5. Hiệu quả và so sánh giữa các mô hình trên toàn bộ mẫu nghiên cứu

Mô hình	AUC (95% CI)	Độ nhạy (%)	Độ đặc hiệu (%)	PPV (%)	NPV (%)	Độ chính xác (%)
AI đơn độc	0,828 (0,782-0,875)	79,5	90,1	76,9	91,4	87,0
Bác sĩ đọc độc lập	0,936 (0,914-0,957)	69,9	97,2	91,1	88,6	89,2
Bác sĩ + AI	0,961 (0,943-0,980)	88,1	96,9	92,3	95,2	94,3

So sánh cặp giữa các mô hình:

So sánh mô hình	Chênh lệch AUC	p (DeLong)	p (McNemar)
Bác sĩ đọc độc lập so với AI đơn độc	+0,108	<0,001	0,246
Bác sĩ + AI so với AI đơn độc	+0,133	<0,001	<0,001
Bác sĩ + AI so với bác sĩ đọc độc lập	+0,025	<0,001	<0,001

AI đơn độc đạt AUC 0,828; bác sĩ đọc độc lập đạt AUC 0,936. Mô hình bác sĩ đọc kết hợp AI đạt AUC cao nhất (0,961), đồng thời có độ nhạy và độ chính xác cao nhất trong ba mô hình. Bác sĩ đọc độc lập có AUC cao hơn AI đơn độc có ý nghĩa thống kê, nhưng khác biệt về độ chính xác chưa có ý nghĩa (p = 0,246). Mô hình bác sĩ kết hợp AI có AUC và độ chính xác cao hơn có ý nghĩa so với cả hai mô hình còn lại (p < 0,001).



Hình 1. Đường cong ROC của các mô hình trên toàn bộ mẫu nghiên cứu

Đường cong ROC cho thấy mô hình bác sĩ kết hợp AI nằm cao hơn bác sĩ đọc độc lập và AI đơn độc trên phần lớn dải giá trị 1 - độ đặc hiệu, phù hợp với kết quả AUC.

4. BÀN LUẬN

4.1. Đặc điểm chung và đặc điểm nhũ ảnh

Nhóm ung thư trong nghiên cứu này có tuổi trung bình cao hơn nhóm không ung thư, tương đồng với nghiên cứu của Yang và cộng sự, trong đó tuổi trung bình của nhóm ác tính là 51,4 tuổi, cao hơn nhóm lành tính 45,0 tuổi (p < 0,001)[6]. Sự khác biệt về tuổi này phù hợp với đặc điểm dịch tễ học đã biết của ung thư vú, khi nguy cơ mắc bệnh tăng dần theo tuổi.

Khối/vùng tăng đậm độ, biến dạng cấu trúc và vi vôi hóa đều gặp nhiều hơn có ý nghĩa ở nhóm ung thư. Biến dạng cấu trúc gần như đặc hiệu cho nhóm ung thư trong bộ số liệu này (28,4% so với 1,0% ở nhóm không ung thư), phù hợp với nhận định của Zhang và cộng sự, khi 228/346 tổn thương biến dạng cấu trúc trong nghiên cứu của họ được xác định là ác tính, và nguy cơ ác tính còn cao hơn khi tổn thương đi kèm các bất thường khác trên nhũ ảnh (73,6% so với 59,4%; p = 0,005)[7]. BI-RADS 4C-5 chỉ gặp ở nhóm ung thư, phù hợp với xu hướng tăng dần tỷ lệ ác tính theo mức độ BI-RADS mà Yang và cộng sự đã ghi nhận, từ 18,1% ở BI-RADS 4A lên 59,4% ở BI-RADS 4B và 90,0% ở BI-RADS 4C[6]. Tỷ lệ mật độ mô vú C/D không khác biệt giữa hai nhóm; tuy nhiên, mô vú đặc vẫn có thể làm giảm độ nhạy của nhũ ảnh quy ước, như Payne và cộng sự ghi nhận độ nhạy giảm từ 75,0% ở nhóm mật độ A xuống 51,3% ở nhóm mật độ D[8].

4.2. Điểm AI và ngưỡng phân loại nguy cơ

Điểm AI trung bình cao hơn rõ rệt ở nhóm ung thư và tăng dần theo mức độ nghi ngờ trên BI-RADS của bác sĩ đọc độc lập, phù hợp với xu hướng của Han và cộng sự[9] và Lee và cộng sự[10], dù giá trị điểm số tuyệt đối có khác biệt do đối tượng nghiên cứu không hoàn toàn tương đồng giữa các nghiên cứu. Sự chồng lấp điểm AI giữa các nhóm BI-RADS liền kề cho thấy AI có thể giúp bác sĩ nhận biết những trường hợp cần xem xét kỹ hơn, nhưng không thay thế hoàn toàn việc đánh giá hình thái tổn thương.

Trong các ngưỡng được khảo sát, ngưỡng 15% cho chỉ số Youden cao nhất, tạo được sự cân bằng tương đối giữa độ nhạy và độ đặc hiệu. Ngưỡng này thấp hơn ngưỡng 30,44% được Seker và cộng sự lựa chọn khi cùng sử dụng chỉ số Youden trên một phần mềm AI tương tự[11], trong khi Han và cộng sự ghi nhận ngưỡng phù hợp thay đổi giữa các nhóm BI-RADS tùy theo mục tiêu ưu tiên độ nhạy hay độ đặc hiệu của bối cảnh lâm sàng[9]. Sự khác biệt này cho thấy ngưỡng điểm AI tối ưu chịu ảnh hưởng đáng kể của đặc điểm quần thể, tỷ lệ hiện mắc ung thư trong mẫu và mục đích sử dụng; do đó ngưỡng 15% chỉ nên được xem là điểm cut off mô tả phù hợp với bộ số liệu hiện tại, cần được kiểm định trên một quần thể độc lập trước khi áp dụng rộng rãi.

4.3. Hiệu quả của AI so với bác sĩ đọc phim

AI đơn độc đạt AUC 0,828, thể hiện khả năng phân biệt tương đối tốt giữa nhóm ung thư và không ung thư, song thấp hơn có ý nghĩa so với bác sĩ đọc độc lập. So với bác sĩ, AI có độ nhạy cao hơn nhưng độ đặc hiệu thấp hơn rõ rệt, cho thấy AI có xu hướng nhạy với các bất thường hình

ảnh nhưng chưa phân biệt tốt bằng bác sĩ giữa các bất thường lành tính và ác tính, dẫn đến tỷ lệ dương tính giả cao hơn nếu sử dụng AI như công cụ đọc phim độc lập.

Khi bác sĩ sử dụng thêm kết quả AI, độ nhạy tăng đáng kể mà độ đặc hiệu chỉ giảm nhẹ; mô hình kết hợp đạt AUC và độ chính xác cao nhất. Kết quả này tương đồng với nghiên cứu của Yilmaz và cộng sự trên 1.063 phụ nữ, trong đó việc bác sĩ xem xét lại các trường hợp được AI đánh dấu giúp cải thiện kết quả đọc phim[12]. Trong nghiên cứu AI-STREAM năm 2025 trên 24.543 phụ nữ, bác sĩ có hỗ trợ AI phát hiện thêm 13,8% trường hợp ung thư mà không làm tăng đáng kể tỷ lệ gọi lại[13]. Các kết quả nhất quán này củng cố bằng chứng ủng hộ việc tích hợp AI như công cụ hỗ trợ đọc phim thường quy, thay vì công cụ chẩn đoán độc lập thay thế bác sĩ.

Đáng chú ý, độ chính xác giữa bác sĩ đọc độc lập và AI đơn độc không khác biệt rõ rệt, dù AUC của hai mô hình khác biệt có ý nghĩa. Hai kết quả này không mâu thuẫn, vì AUC phản ánh khả năng phân biệt trên toàn bộ dải điểm số, trong khi độ chính xác chỉ được tính tại một ngưỡng cụ thể đã chọn; điều này cho thấy tầm quan trọng của việc báo cáo đồng thời cả AUC và các chỉ số tại ngưỡng khi đánh giá một công cụ chẩn đoán mới.

4.4. Hạn chế của nghiên cứu và hướng khắc phục

Nghiên cứu có một số hạn chế: thiết kế hồi cứu, đơn trung tâm, cỡ mẫu khiêm tốn ở một số phân nhóm BI-RADS; nhóm không ung thư một phần dựa trên theo dõi âm tính 24 tháng thay vì mô bệnh học toàn bộ, có thể gây sai lệch phân loại nếu có ung thư tiến triển chậm; nghiên cứu chỉ thực hiện trên một hệ thống nhũ ảnh và một phiên bản phần mềm AI duy nhất (Lunit INSIGHT MMG 1.1.2.7) tại một cơ sở; việc đọc phim trong điều kiện hồi cứu cũng có thể khác với thực hành lâm sàng. Cần thực hiện các nghiên cứu tiến cứu, đa trung tâm, cỡ mẫu lớn hơn, đối chiếu mô bệnh học đầy đủ hơn, kiểm định ngưỡng AI trên quần thể độc lập, và đánh giá thêm theo mật độ mô vú cũng như trong điều kiện đọc phim thực tế.

5. KẾT LUẬN

AI có khả năng phân tầng nguy cơ tổn thương nghi ngờ ung thư vú ở mức tốt trên nhũ ảnh, nhưng hiệu quả phân loại khi sử dụng độc lập còn thấp hơn bác sĩ đọc phim. Mô hình bác sĩ đọc kết hợp AI đạt hiệu quả phân loại cao nhất trong ba mô hình được so sánh, cho thấy giá trị của AI chủ yếu nằm ở vai trò công cụ hỗ trợ đọc phim, giúp bác sĩ nhận diện thêm tổn thương nghi ngờ mà không thay thế vai trò đánh giá độc lập của bác sĩ. Ngưỡng điểm AI được xác định trong nghiên cứu phù hợp với bộ số liệu hiện tại nhưng cần được kiểm định trên các quần thể độc lập trước khi áp dụng rộng rãi vào thực hành sàng lọc và chẩn đoán ung thư vú.

6. TUYÊN BỐ VỀ ĐÓNG GÓP CỦA CÁC TÁC GIẢ

Trần Thị Ngọc Mai, Vũ Đăng Lưu, Lưu Hồng Nhung, Tăng Văn Dũng: xây dựng ý tưởng, thiết kế nghiên cứu, thu thập và phân tích số liệu, viết bản thảo. Trần Đăng Khoa, Vũ Trọng Huy, Nguyễn Huy Phương và Nguyễn Thị Kim Tiền: góp ý và chỉnh sửa bản thảo. Tất cả các tác giả đã đọc và đồng thuận với nội dung bản thảo cuối cùng.

7. TUYÊN BỐ VỀ XUNG ĐỘT LỢI ÍCH

Các tác giả cam kết không có xung đột lợi ích liên quan đến nghiên cứu này.

8. LỜI CẢM ƠN

Nhóm nghiên cứu xin trân trọng cảm ơn Bệnh viện Bạch Mai đã tạo điều kiện cho việc thu thập số liệu phục vụ nghiên cứu này.

9. TÀI LIỆU THAM KHẢO

[1] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-263. doi: 10.3322/caac.21834.

[2] Wu Y, Alagoz O, Ayvaci MU, et al. A comprehensive methodology for determining the most informative mammographic features. J Digit Imaging. 2013;26(5):941-947. doi: 10.1007/s10278-013-9588-5.

[3] Lang K, Josefsson V, Larsson AM, et al. Artificial intelligence-supported screen reading versus standard double reading in the Mammography Screening with Artificial Intelligence trial (MASAI): a clinical safety analysis of a randomised, controlled, non-inferiority, single-blinded, screening accuracy study. Lancet Oncol. 2023;24(8):936-944. doi: 10.1016/S1470-2045(23)00298-X.

[4] McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577(7788):89-94. doi: 10.1038/s41586-019-1799-6.

[5] Bahl M. Artificial intelligence in clinical practice: implementation considerations and barriers. J Breast Imaging. 2022;4(6):632-639. doi: 10.1093/jbi/wbac065.

[6] Yang Y, Liao T, Lin X, et al. Reducing unnecessary biopsies of BI-RADS 4 lesions based on a deep learning model for mammography. Front Oncol. 2025;15:1543553. doi: 10.3389/fonc.2025.1543553.

[7] Zhang S, Shao Z, Yi H, et al. A comprehensive analysis of imaging features and clinical characteristics to differentiate malignant from non-malignant mammographic architectural distortion. Gland Surg. 2024;13(5):669-683. doi: 10.21037/gs-24-110.

[8] Payne NR, Hickman SE, Black R, Priest AN, Hudson S, Gilbert FJ. Breast density effect on the sensitivity of digital screening mammography in a UK cohort. Eur Radiol. 2025;35(1):177-187. doi: 10.1007/s00330-024-10951-w.

[9] Han T, Yun H, Sur YK, Park H. Optimizing artificial intelligence thresholds for mammographic lesion detection: a retrospective study on diagnostic performance and radiologist-artificial intelligence discordance. Diagnostics (Basel). 2025;15(11):1368. doi: 10.3390/diagnostics15111368.

[10] Lee SE, Hong H, Kim E-K. Positive predictive values of abnormality scores from a commercial artificial intelligence-based computer-aided diagnosis for mammography. Korean J Radiol. 2024;25(4):343-350. doi: 10.3348/kjr.2023.0907.

[11] Seker ME, Koyluoglu YO, Ozaydin AN, et al. Diagnostic capabilities of artificial intelligence as an additional reader in a breast cancer screening program. Eur Radiol. 2024;34(9):6145-6157. doi: 10.1007/s00330-024-10661-3.

[12] Yilmaz E, Seker ME, Guldogan N, et al. Clinical application of AI in mammography: insights from a prospective study. Acad Radiol. 2025;32(9):5016-5027. doi: 10.1016/j.acra.2025.05.025.

[13] Chang Y-W, Ryu JK, An JK, et al. Artificial intelligence for breast cancer screening in mammography (AI-STREAM): preliminary analysis of a prospective multicenter cohort study. Nat Commun. 2025;16:2248. doi: 10.1038/s41467-025-57469-3.