

DOI: 10.58490/ctump.2025i85.3398

KẾT QUẢ BƯỚC ĐẦU PHÂN LOẠI MẬT ĐỘ VÚ TRÊN PHIM NHŨ ẢNH BẰNG TRÍ TUỆ NHÂN TẠO TẠI BỆNH VIỆN THỐNG NHẤT

Võ Thành Toàn¹, Huỳnh Nguyên Thuận^{1*}, Vũ Trí Lộc²,
Nguyễn Thụy Nguyệt Quỳnh³, Đỗ Võ Công Nguyên¹

1. Bệnh viện Thống Nhất

2. Đại học Tân Tạo

3. Đại học Khoa học Tự nhiên, Đại học Quốc gia TP. Hồ Chí Minh

*Email: huynh@thuan.org

Ngày nhận bài: 04/3/2025

Ngày phản biện: 17/3/2025

Ngày duyệt đăng: 25/3/2025

TÓM TẮT

Đặt vấn đề: Chụp nhũ ảnh là phương pháp chính để sàng lọc và chẩn đoán sớm ung thư vú. Ứng dụng trí tuệ nhân tạo vào phân loại mật độ tuyến vú là cần thiết nhằm mang lại lợi ích cho người bệnh và giảm sự thiếu đồng nhất trong chẩn đoán. **Mục tiêu nghiên cứu:** Nghiên cứu sẽ xây dựng bộ dữ liệu tại Bệnh viện Thống Nhất và kết hợp với các bộ dữ liệu mở để tạo nên mô hình phân loại nhũ ảnh cho Bệnh viện Thống Nhất. **Đối tượng và phương pháp nghiên cứu:** Toàn bộ hình ảnh nhũ ảnh của phụ nữ tầm soát ung thư vú tại Bệnh viện Thống Nhất được lựa chọn và gán nhãn bởi 2 bác sĩ (mù đôi). Bộ dữ liệu kết hợp với bộ dữ liệu được công bố trước đây được dùng huấn luyện mô hình AI. **Kết quả** 13395 phim nhũ ảnh của 5506 phụ nữ trong đó có 698 bệnh nhân từ Bệnh viện Thống Nhất với tuổi trung bình là $48,7 \pm 12,6$ được đưa vào huấn luyện. Trên tập dữ liệu kiểm tra, mô hình đạt độ chính xác 76,8 % mức độ đồng thuận tốt với bác sĩ. Cụ thể trong đó độ nhạy tăng dần từ nhóm A với 31,8% tới nhóm D là 92,2%. Ngược lại độ đặc hiệu lại giảm từ nhóm A với 100% xuống nhóm D với 83,5%. **Kết luận:** Kết quả này chứng minh tiềm năng của việc ứng dụng trí tuệ nhân tạo để hỗ trợ đánh giá mật độ vú trong sàng lọc ung thư vú.

Từ khóa: trí tuệ nhân tạo, ung thư vú, mật độ vú.

ABSTRACT

PRELIMINARY RESULTS OF AUTOMATIC BREAST DENSITY CLASSIFICATION ON MAMMOGRAPHY BY ARTIFICIAL INTELLIGENCE AT THONG NHAT HOSPITAL

Vo Thanh Toan¹, Huynh Nguyen Thuan^{1*}, Vu Tri Loc²,
Nguyen Thuy Nguyệt Quỳnh³, Do Vo Cong Nguyen¹

1. Thong Nhat Hospital

2. Tan Tao University

3. University of Science - VNUHCM

Background: Mammography is the main method for screening and early diagnosis of breast cancer. Application of artificial intelligence in classification is necessary to bring benefits to patients and reduce inconsistency in diagnosis. **Objectives:** The research will build a Thong Nhat hospital dataset and combine it with open datasets to create a mammography classification model for Thong Nhat hospital. **Materials and methods:** All mammogram images of women screened for breast cancer at Thong Nhat Hospital were selected and labeled by 2 doctors (double blind). Their images combined with previously published dataset participate in model training. **Results:** Dataset included 13395 mammograms of 5506 women (698 patients from Thong Nhat hospital with an mean

age were 48.7 ± 12.6). On the test dataset, the model achieved an overall accuracy of 76.8 %, in good agreement with physicians. Specifically, the sensitivity gradually increases from group A with 31.8% to group D with 92.2%. On the contrary, the specificity decreased from group A with 100% to group D with 83.5%. **Conclusion:** These results demonstrate the potential of applying artificial intelligence to support breast density assessment in breast cancer screening.

Keywords: Artificial intelligence, breast cancer, breast density.

I. ĐẶT VẤN ĐỀ

Ung thư vú là bệnh ung thư phổ biến nhất ở phụ nữ trên toàn thế giới và là nguyên nhân gây tử vong thứ hai do ung thư ở nữ giới [1]. Việc phát hiện sớm ung thư vú đóng vai trò quan trọng trong việc cải thiện tiên lượng và tỷ lệ sống sót của bệnh nhân. Chụp nhũ ảnh là phương pháp chủ yếu để sàng lọc và chẩn đoán sớm ung thư vú [2].

Mật độ vú (MĐV) hay mật độ nhu mô tuyến vú là tỷ lệ mô tuyến và mô liên kết trên nhũ ảnh, là một yếu tố nguy cơ quan trọng đối với ung thư vú. Hiện nay, phương pháp phổ biến nhất để đánh giá MĐV là dựa trên đánh giá của bác sĩ Chẩn đoán hình ảnh theo thang 4 mức độ của Hệ thống Báo cáo và Dữ liệu Hình ảnh Vú (BI-RADS) của Hội Điện Quang Hoa Kỳ (ACR) [2]. Phụ nữ có MĐV cao có nguy cơ mắc ung thư vú cao gấp 4-6 lần so với phụ nữ có MĐV thấp, MĐV cao cũng làm giảm độ nhạy của nhũ ảnh trong việc phát hiện các tổn thương vú do hiệu ứng che khuất [3]. Việc đánh giá chính xác MĐV rất quan trọng trong việc xác định nguy cơ ung thư vú và đưa ra chiến lược sàng lọc phù hợp cho từng người. Tuy nhiên, đánh giá của bác sĩ Chẩn đoán hình ảnh thường mang tính chủ quan và thiếu sự nhất quán giữa các bác sĩ cũng như giữa các lần đánh giá khác nhau trên cùng một bác sĩ. Sự không thống nhất trong đánh giá MĐV có thể dẫn đến việc phân loại sai mức độ nguy cơ ung thư vú của phụ nữ và đưa ra khuyến cáo sàng lọc không phù hợp.

Trong bối cảnh đó, việc phát triển các phương pháp tự động hóa và chuẩn hóa việc đánh giá MĐV là rất cần thiết nhằm khắc phục các hạn chế của đánh giá bằng mắt thường và nâng cao chất lượng của sàng lọc ung thư vú. Do đó nghiên cứu này được tiến hành nhằm xây dựng bộ dữ liệu hình ảnh nhũ ảnh tuyến vú của Bệnh viện (BV) Thống Nhất kết hợp với các bộ dữ liệu lớn về nhũ ảnh tuyến vú đã được công bố để phục vụ quá trình xây dựng các mô hình học sâu. Chúng tôi huấn luyện mô hình học sâu để tự động phân loại mật độ vú trên nhũ ảnh kỹ thuật số theo 4 mức độ của thang phân loại BI-RADS (A, B, C, D).

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Đối tượng nghiên cứu

Toàn bộ bệnh nhân đến chụp nhũ ảnh sàng lọc tại BV Thống Nhất.

- **Thời gian, địa điểm:** Từ tháng 01/2022 đến tháng 06/2024, Khoa Chẩn đoán hình ảnh (CĐHA) – Bệnh viện Thống Nhất.

- **Tiêu chuẩn chọn mẫu:**

+ Các phim nhũ ảnh kỹ thuật số từ PACS của bệnh viện được chọn để huấn luyện và kiểm tra mô hình cần đạt tiêu chuẩn:

+ Phim nhũ ảnh có chất lượng tốt, không có các vấn đề như che khuất, chuyển động, hoặc lỗi kỹ thuật.

+ Phụ nữ không có tiền sử ung thư vú, phẫu thuật vú, hoặc đang mang thai.

- **Tiêu chuẩn loại trừ:** Phụ nữ có tiền sử ung thư vú, phẫu thuật vú, hoặc đang mang thai.

- **Phương pháp chọn mẫu:** Lấy mẫu thuận tiện, chọn các phim nhũ ảnh đạt tiêu chuẩn chọn mẫu.

2.2. Phương pháp nghiên cứu

- **Gắn nhãn dữ liệu:** Hai bác sĩ chuyên khoa chẩn đoán hình ảnh sẽ đọc phân loại mật độ vú và các hướng của tuyến vú một cách độc lập (mù đôi). Nếu kết quả đọc phim không sự đồng nhất giữa 2 bác sĩ, hội chẩn ý kiến trưởng khoa trước khi quyết định. Kết quả thu được sau khi đồng thuận sẽ được sử dụng để đưa vào huấn luyện mô hình.

- **Bộ dữ liệu:** Cùng với bộ dữ liệu thu thập tại BV Thống Nhất. Nghiên cứu sẽ thực hiện huấn luyện mô hình học máy với các bộ dữ liệu lớn được công bố trước đây như VinDr-Mammo [4], bộ dữ liệu của Hiệp hội điện quang Bắc Mỹ [5], và bộ dữ liệu Digital Database for Screening Mammography (DDSM) [6].

- **Huấn luyện mô hình:** Chúng tôi sử dụng thị giác máy tính (computer vision) và phương pháp học sâu để giải quyết 3 vấn đề: (1) phân loại hình ảnh chụp bên trái hoặc bên phải; (2) phân loại hình ảnh chụp góc MLO (Mediolateral oblique) hoặc góc CC (Craniocaudal); (3) phân loại đậm độ hình ảnh theo BI-RADS với 4 mức A, B, C, D của tuyến vú.

Cấu hình sử dụng trong huấn luyện mô hình chip Intel Core i9-10900K chạy với tốc độ 3.70GHz, 64GB RAM và bộ xử lý hình ảnh NVIDIA RTX A5000 (24GB GDDR6, 384-bit, 4x DisplayPort, 1x 8-pin).

- **Phân loại hình chụp bên phải hay trái:** Chúng tôi nhận thấy rằng tất cả các ảnh được chụp ở phía bên trái thường có nhiều pixel hơn với màu đen ở phía bên phải của ảnh. Trong trường hợp ảnh được chụp ở phía bên phải thì sẽ ngược lại. Do đó, chúng tôi chia đôi mỗi hình ảnh và so sánh số lượng pixel màu đen. Nếu có nhiều điểm ảnh màu đen hơn ở bên trái thì ảnh được chụp ở bên phải, ngược lại là ảnh được chụp ở bên trái.

- Xác định chế độ xem MLO hoặc CC

Bốn hình ảnh từ cùng một bệnh nhân được sử dụng. Đối với mỗi hình ảnh chúng tôi thực hiện các bước sau:

- Thuật toán Gaussian blur được áp dụng để giảm nhiễu và loại các chi tiết nhỏ.
- Thuật toán phát hiện cạnh Canny được sử dụng để xác định tất cả các đường viền trong ảnh.
- Khu vực với đường viền lớn nhất tương ứng với vùng ngực, được xác định dựa trên các đường viền được phát hiện trước đó.
- Cuối cùng hình chữ nhật bao quanh khu vực ngực được tìm thấy
- Chiều rộng và chiều cao của nó được sử dụng để tính tỷ lệ khung hình (chiều rộng chia cho chiều cao).

Khi có được tỷ lệ khung hình cho cả bốn hình ảnh, hai hình ảnh có tỷ lệ khung hình cao nhất được phân loại là chế độ xem MLO, trong khi hai hình ảnh còn lại được phân loại là chế độ xem CC. Phương pháp này hoạt động hiệu quả vì trên thực tế, vùng ngực trong ảnh MLO có chiều cao lớn hơn ảnh CC trong khi ảnh CC lại có chiều rộng lớn hơn.

- **Xác định đậm độ:** Chúng tôi sử dụng kiến trúc CNN với 5 khối, 4 khối đầu tiên thực hiện các tác vụ tương tự nhau với số lượng filter 3x3 lần lượt là 2.32, 64, 128 và 256. Mỗi khối được kèm theo một lớp chuẩn hóa theo batch, hàm kích hoạt ReLU và lớp tổng hợp (Max Pooling). Khối cuối cùng bao gồm 1 lớp kết nối đầu đủ (Full connected) với 256 đơn vị sử dụng hàm kích hoạt ReLU. Chúng tôi huấn luyện mô hình với hàm mất mát Crossentropy, 100 epochs, batch size 128, và thuật toán tối ưu Adam với learning rate 1e-4.

- **Đạo đức nghiên cứu:** Nghiên cứu hồi cứu không can thiệp đến quá trình khám và điều trị bệnh nhân, không được tài trợ bởi bất kỳ công ty nào và được Hội đồng Đạo đức trong nghiên cứu y sinh học Bệnh viện Thống Nhất chấp thuận thông qua ngày 25/7/2024.

III. KẾT QUẢ

3.1. Chuẩn bị dữ liệu cho huấn luyện

Trong thời gian từ tháng 01/2022 tới 06/2024, có 698 bệnh nhân thực hiện chụp nhũ ảnh tại BV Thống Nhất phù hợp tiêu chuẩn chọn của nghiên cứu với độ tuổi trung bình là $48,7 \pm 12,6$ năm. Sau khi kiểm tra kỹ lưỡng, chúng tôi lựa chọn được 2734 hình nhũ ảnh đạt chuẩn để thêm vào bộ dữ liệu.

Bảng 1. Gán nhãn dữ liệu

	A	B	C	D	
Bác sĩ 1	2,5%	15,8%	35,8%	45,9%	k = 0,71
Bác sĩ 2	3%	19,9%	43%	34,1%	
Sau khi hiệu chỉnh	3,3%	19,6%	43%	34,1%	

3.2. Bộ dữ liệu

Bảng 2. Các bộ dữ liệu

Tên bộ dữ liệu	RSNA	VinDr-Mammo	DDSM
Số bệnh nhân	8000	5000	4808*
Số hình nhũ ảnh	54706	20000	10661*
Đặc điểm dữ liệu	Bệnh nhân không nhất thiết chụp đủ 4 ảnh một lần Có cả các thông số của sinh thiết	Đủ 4 ảnh 1 lần Dân số Việt Nam Tỷ lệ hình của đậm độ A, B, C, D lần lượt là 0.50%, 9.55%, 76.45%, và 13.50%	Đủ toàn bộ các chủng tộc da trắng, da vàng, Mỹ gốc á, da đen ...

*bộ dữ liệu DDSM công bố lần đầu 1996 với 2890 bệnh nhân, nó được cập nhật liên tục bởi nhiều bệnh viện ở Hoa Kỳ, được duy trì bởi University of South Florida để cung cấp cho các nhà nghiên cứu, Dữ liệu chúng tôi thực hiện trong nghiên cứu là hình ảnh của hai bộ dữ liệu Mini-DDSM và CBIS-DDSM trích xuất từ DDSM [6].

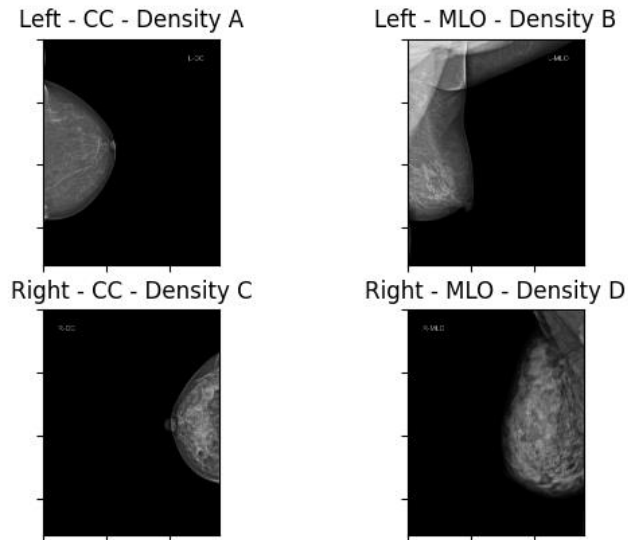
Nhận xét: Trước khi đưa vào huấn luyện, chúng tôi thực hiện huấn luyện trên từng bộ dữ liệu riêng lẻ. Bộ dữ liệu của RSNA cho kết quả dữ liệu rỗng (không gán nhãn mật độ vú) chiếm 46% nên không phù hợp gộp chung dữ liệu với bệnh nhân của BV Thống Nhất. Bộ dữ liệu VinDr-Mammo không cân đối giữa các nhóm đậm độ (A, B, C, D lần lượt là 0.50%, 9.55%, 76.45%, và 13.50%) dẫn tới huấn luyện đạt kết quả không tốt, chỉ đạt mức chính xác 50%. Cũng không phù hợp để gộp chung dữ liệu với bệnh nhân của BV Thống Nhất. Với bộ dữ liệu DDSM chúng tôi thực hiện lấy mẫu ngẫu nhiên 10661 hình ảnh của hai nghiên cứu trước đó Mini-DDSM và CBIS-DDSM từ 4808 bệnh nhân. Khi kết hợp với bộ dữ liệu gán nhãn tại BV Thống Nhất cho kết quả đạt mức tốt nhất với độ chính xác tới 76,8%.

Bảng 3. Phân bố bộ dữ liệu khi huấn luyện và kiểm tra

Phân bố bộ dữ liệu	Số hình ảnh	Tỷ lệ (%)
Bệnh viện Thống Nhất	2734	20,4
DDSM	10661	79,6
Tổng dữ liệu nghiên cứu	13395	100%

Nhận xét: Dữ liệu của BV Thống Nhất trong nghiên cứu này chiếm 20,4%, còn bộ dữ liệu DDSM chiếm 79,6% (bao gồm đầy đủ các chủng tộc tại Mỹ).

3.3. Hiệu suất của mô hình học sâu



Hình 1. Phân loại các góc chụp và đậm độ tuyến vú

Nhận xét: Kết quả nghiên cứu của chúng tôi cho thấy việc xác định vú bên trái, bên phải và góc chụp MLO hay CC có tỷ lệ chính xác 100%.

Bảng 4. Phân bố bộ dữ liệu huấn luyện

Phân bố bộ dữ liệu huấn luyện	Số hình ảnh	Tỷ lệ (%)
Dữ liệu huấn luyện (Training data)	8037	60
Dữ liệu kiểm chứng (Validating data)	3750	28
Kiểm tra (Test data)	1608	12
Tổng dữ liệu	13395	100%

Nhận xét: Từ bộ dữ liệu trên, chúng tôi thực hiện huấn luyện mô hình trên 8037 hình nhũ ảnh, 3750 hình ảnh sử dụng để kiểm chứng mô hình (validation) khi huấn luyện, và 1608 hình nhũ ảnh được sử dụng để kiểm tra độ chính xác của mô hình. Kết quả đạt độ chính xác tổng thể là 76,8%.

Bảng 5. Kết quả kiểm tra độ chính xác của mô hình ghi nhận bảng dưới đây

Dự đoán \ Thực tế	A	B	C	D
A	7	13	2	0
B	0	87	43	7
C	0	9	224	69
D	0	0	18	214

Bảng 6. Độ nhạy và độ đặc hiệu trong phân loại

	A	B	C	D
Độ nhạy	31,8	63,5	74,1	92,2
Độ đặc hiệu	100	96,04	83,9	83,5

Nhận xét: Bảng 5 và bảng 6 cho ta thấy sự chính xác trong phân loại từng nhóm khác nhau. Độ đặc hiệu cao nhất ở nhóm A với độ chính xác 100%.

IV. BÀN LUẬN

4.1. Bộ dữ liệu

Quan trọng nhất trong xây dựng một mô hình AI cần có dữ liệu phù hợp. Trong nghiên cứu của chúng tôi, để khắc phục việc dữ liệu mang tính cá nhân và không đa dạng, chúng tôi đã thử kết hợp với các bộ dữ liệu công khai trước đây. Trong đó, chúng tôi đã tham khảo và thử nghiệm với bộ dữ liệu VinDr-Mammo [4], mục đích chính của bộ dữ liệu là xây dựng công cụ nhận diện ung thư vú trên hình ảnh cho nên đa số hình ảnh nhũ ảnh thuộc mức C và D, nên chưa phù hợp với việc ứng dụng rộng rãi của mô hình chúng tôi muốn áp dụng. Ở mặt khác, bộ dữ liệu của RSNA[5] bao gồm các thông tin về thông số sinh thiết nên không phù hợp với chúng tôi, ngoài ra nó cũng chỉ gắn nhãn khoảng 46% dữ liệu về các đậm độ mô tuyến vú dẫn tới chúng tôi không thể huấn luyện. Chúng tôi quyết định sử dụng bộ dữ liệu của DDSM[6] vì đây là bộ dữ liệu được công bố với những dữ liệu hình ảnh đa dạng, đa chủng tộc từ nhiều trường đại học và bệnh viện lớn của Hoa Kỳ. Khi kết hợp bộ dữ liệu này tương thích với những dữ liệu của BV Thống Nhất.

4.2. Hiệu suất mô hình trí tuệ nhân tạo

Hiệu suất của mô hình học sâu đạt hiệu suất tương đối cao trong việc phân loại mật độ vú trên nhũ ảnh, với độ chính xác chung là 76,8%. Cụ thể trong đó độ nhạy tăng dần từ nhóm A với 31,8% tới nhóm D là 92,2%. Ngược lại độ đặc hiệu lại giảm từ nhóm A với 100% xuống nhóm D với 83,5%. Trong trường hợp của chúng tôi với các trường hợp nhóm C và D chúng tôi đạt được hiệu suất cao ở cả độ nhạy và đặc hiệu, điều này giúp bác sĩ chẩn đoán hình ảnh có thể cân nhắc các bước tiếp theo để theo dõi và can thiệp sớm các trường hợp nguy cơ cao. Kết quả nghiên cứu của chúng tôi có hiệu suất mô hình trong phân độ BI-RADS tương đương nhiều nghiên cứu trước đây sử dụng mô hình học sâu để phân loại mật độ vú. Cụ thể, nghiên cứu của Lehman và cộng sự[7] đạt độ chính xác mô hình là 77%; nghiên cứu của tác giả VeronicaMagni và cộng sự[8] ghi nhận độ chính xác của mô hình đạt 75% khi phân loại theo BI-RADS.

Hiệu suất tốt của mô hình nhờ vào một số yếu tố như: (1) quá trình lựa chọn và tiền xử lý và chuẩn hóa dữ liệu tối ưu, kết hợp dữ liệu thực tế tại khoa và bộ dữ liệu đóng góp từ đại học lớn trên thế giới (2) huấn luyện mô hình dựa trên nhân do nhiều bác sĩ gần (BV Thống Nhất, BV Hoa Kỳ), giúp giảm thiểu sự chủ quan trong đánh giá mật độ vú [6, 7]. Việc lựa chọn thận trọng các bộ dữ liệu như trên và sử dụng các tập dữ liệu với tiêu chí đánh giá phù hợp với BV Thống Nhất giúp cho mô hình của chúng tôi có thể ứng dụng rộng rãi.

Khi đánh giá hiệu suất của mô hình trên từng phân lớp mật độ, chúng tôi nhận thấy xu hướng tăng dần của độ nhạy từ tip A đến tip D và tăng dần độ đặc hiệu từ tip D về tip A. Điều này cho thấy mô hình gặp nhiều khó khăn hơn trong việc phân biệt các mức độ mật độ vú cao, đặc biệt là tip C và D. Kết quả này cũng được quan sát thấy trong nhiều nghiên cứu khác[7, 9] và có thể do sự khác biệt về hình thái học giữa mật độ vú cao là ít rõ ràng hơn so với sự khác biệt giữa vú mỡ và vú đặc.

So với đánh giá của bác sĩ, mô hình học sâu có một số ưu điểm như tính khách quan, nhất quán và khả năng xử lý nhanh trên quy mô lớn. Trong khi đánh giá của bác sĩ thường chịu ảnh hưởng của các yếu tố chủ quan như kinh nghiệm, sở thích cá nhân và điều kiện làm việc [9, 10] thì mô hình học sâu có thể đưa ra kết quả ổn định và lặp lại được với cùng một đầu vào. Trong một nghiên cứu đa trung tâm của Sprague và cộng sự, 83 bác sĩ Chẩn đoán hình ảnh đã đánh giá mật độ vú của cùng một nhóm phụ nữ. Kết quả cho thấy tỷ lệ

phụ nữ được phân loại vào nhóm vú đặc (mật độ vú mức C hoặc D theo BI-RADS) dao động từ 6% đến 85% tùy thuộc vào bác sĩ đọc phim. Tương tự, một nghiên cứu khác của Spayne và cộng sự trên 34 bác sĩ X quang cho thấy sự đồng thuận trong đánh giá mật độ vú giữa các lần đọc khác nhau của cùng một bác sĩ chỉ đạt từ 62% đến 87%. Sự không thống nhất trong đánh giá mật độ vú giữa các bác sĩ X quang có thể do sự khác biệt về kinh nghiệm, sở thích cá nhân, cách diễn giải tiêu chuẩn phân loại BI-RADS cũng như các yếu tố kỹ thuật như chất lượng hình ảnh, điều kiện quan sát [10]. Điều này dẫn đến việc phân loại sai và đưa ra khuyến cáo sàng lọc không phù hợp, gây lo lắng cho bệnh nhân hoặc bỏ sót các trường hợp cần theo dõi. Ngoài ra, đánh giá mật độ vú bằng mắt thường không cho phép định lượng chính xác tỷ lệ mô vú đặc. Đây là yếu tố quan trọng trong việc ước tính nguy cơ ung thư vú. Do đó, việc định lượng mật độ vú một cách tự động và chính xác có thể cung cấp thông tin chi tiết hơn về nguy cơ ung thư vú so với đánh giá định tính của bác sĩ. Bên cạnh đó ra, việc tích hợp công cụ tự động đánh giá mật độ vú vào quy trình đọc phim có thể giúp tiết kiệm đáng kể thời gian của bác sĩ.

Tuy nhiên, hiệu năng của mô hình học sâu phụ thuộc nhiều vào chất lượng và số lượng dữ liệu huấn luyện, trong khi kinh nghiệm và sự hiểu biết sâu rộng về giải phẫu và bệnh lý vú của bác sĩ là không thể thay thế hoàn toàn bằng các thuật toán [11]. Pháp lý hiện tại thì bác sĩ vẫn phải là người chịu trách nhiệm hoàn toàn cho các quyết định. Do đó, mô hình học sâu nên được xem như một công cụ hỗ trợ và bổ sung cho bác sĩ trong đánh giá mật độ vú, hơn là thay thế hoàn toàn vai trò của bác sĩ.

4.3. Ưu điểm và hạn chế

Việc sử dụng mô hình học sâu để hỗ trợ tự động hóa việc đánh giá mật độ vú cũng có nhiều ưu điểm. (1) Mô hình học sâu có thể hỗ trợ đưa ra kết quả nhanh chóng, nhất quán và khách quan, trong khi đánh giá của bác sĩ thường mất nhiều thời gian và chịu ảnh hưởng của các yếu tố chủ quan như kinh nghiệm, điều kiện làm việc. Với khả năng xử lý hàng nghìn phim chụp trong thời gian ngắn, mô hình học sâu có thể giúp tiết kiệm đáng kể nguồn lực và nâng cao hiệu suất của các chương trình sàng lọc. (2) Huấn luyện mô hình của chúng tôi dựa trên dữ liệu từ nhiều bác sĩ và nhiều trung tâm chụp khác nhau giúp tạo ra một tiêu chuẩn chung và giảm sự dao động trong đánh giá mật độ vú.

Tuy nhiên, phương pháp học sâu cũng có một số hạn chế cần lưu ý. Đầu tiên, hiệu suất của mô hình phụ thuộc rất nhiều vào chất lượng và số lượng dữ liệu huấn luyện. Để đạt được độ chính xác cao, mô hình cần được huấn luyện trên một tập dữ liệu lớn và đa dạng từ nhiều loại máy, nhiều cơ sở và nhiều quần thể khác nhau. Chúng tôi đã vượt qua điều này bằng cách sử dụng bộ dữ liệu lớn và thu thập kéo dài trên nhiều chủng tộc [6]. Hai là, việc triển khai mô hình học sâu vào lâm sàng đòi hỏi sự đầu tư về hạ tầng công nghệ thông tin, có thể gây khó khăn một số cơ sở y tế. Việc hợp tác giữa các cơ sở chuyên sâu về công nghệ thông tin như các viện nghiên cứu AI và bệnh viện là rất cần thiết. Việc tích hợp công cụ trí tuệ nhân tạo vào quy trình làm việc của bác sĩ đòi hỏi sự hợp tác chặt chẽ giữa nhiều phòng, ban, đơn vị. Mặc dù có tiềm năng lớn tuy nhiên hiệu quả lâm sàng thực tế của việc sử dụng mô hình học sâu trong đánh giá mật độ vú vẫn cần được chứng minh qua các thử nghiệm lâm sàng với cỡ mẫu lớn và thời gian theo dõi dài hơn.

4.4. Tiềm năng và ứng dụng trên lâm sàng

Về việc triển khai mô hình trên thực tế lâm sàng tại BV Thống Nhất sẽ được tiến hành theo các bước sau. (1) Trong nghiên cứu tiếp theo, chúng tôi xây dựng quy trình để ứng dụng mô hình vào quy trình làm việc, chúng tôi sẽ thiết kế nghiên cứu lâm sàng để đánh giá hiệu quả trong thực tế sử dụng với mục tiêu đầu tiên là đánh giá trong (internal validation) để đánh giá sự chính xác của AI và các bác sĩ của bệnh viện trong thực hành lâm sàng. (2) Đánh giá ngoài mô hình, việc này cần kết hợp với các bệnh viện khác để cho thấy được tính khách quan mô hình và sự thống nhất giữa các cơ sở y tế trong chẩn đoán. (3) Song song với đánh giá hiệu quả lâm sàng của mô hình chúng tôi cũng tiến hành đánh giá về hiệu quả của AI trong việc hỗ trợ các bác sĩ nhằm giảm thời gian cho việc đọc các phim nhũ ảnh. Đây là một chặng đường dài, chúng tôi sẽ tiếp tục cải thiện mô hình AI này với các thuật toán học tăng cường và gắn nhãn các trường hợp bất thường khác.

Tuy nhiên, để có thể đưa mô hình vào sử dụng rộng rãi, cần có những cải thiện và đánh giá bổ sung. Thứ nhất, cần liên tục cập nhật dữ liệu, mở rộng bộ dữ liệu huấn luyện cả về số lượng và tính đa dạng. Thứ hai, thực hiện các kiểm định đánh giá trong, đánh giá ngoài. Thứ ba, cần phát triển các giao diện người dùng trực quan hóa mô hình để tăng tính minh bạch và sự tin tưởng của bác sĩ. Thứ 4, cần tiến hành các nghiên cứu lâm sàng lớn hơn để đánh giá hiệu quả và tác động thực tế của việc áp dụng công cụ này đối với công tác chẩn đoán rộng rãi hơn. Cuối cùng, cần xây dựng cơ sở hạ tầng và quy trình phù hợp để tích hợp liền mạch kết quả của mô hình vào quy trình làm việc của bác sĩ, đồng thời cần xem xét các khía cạnh đạo đức và pháp lý liên quan đến việc sử dụng trí tuệ nhân tạo trong y tế.

V. KẾT LUẬN

Mô hình được huấn luyện và đánh giá trên một tập dữ liệu gồm 13395 phim nhũ ảnh của 5506 phụ nữ trong đó có 698 bệnh nhân từ BV Thống Nhất với tuổi trung bình là $48,7 \pm 12,6$. Trên tập dữ liệu kiểm tra, mô hình đạt độ chính xác tổng thể 76,8 % cho thấy hiệu suất cao và mức độ đồng thuận tốt với bác sĩ. Cụ thể trong đó độ nhạy tăng dần từ nhóm A với 31,8% tới nhóm D là 92,2%. Ngược lại độ đặc hiệu lại giảm từ nhóm A với 100% xuống nhóm D với 83,5%. Kết quả này chứng minh tiềm năng của việc ứng dụng trí tuệ nhân tạo để hỗ trợ đánh giá mật độ vú trong sàng lọc ung thư vú. Trong tương lai, chúng tôi sẽ tiếp tục cải thiện hiệu suất của mô hình bằng cách thử nghiệm các kiến trúc mạng nơ-ron tích chập khác nhau, tích hợp thêm các thông tin lâm sàng vào quá trình huấn luyện, đồng thời phát triển các mô hình bổ sung để phát hiện và định vị tổn thương trên nhũ ảnh.

TÀI LIỆU THAM KHẢO

1. Bray, F., et al., Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 2024. 74(3): p. 229-263, DOI: doi.org/10.3322/caac.21834.
2. Sickles, et al., Breast Imaging Reporting and Data System, in ACR BI-RADS® Atlas. 2013, American College of Radiology.
3. Freer, P.E., Mammographic breast density: impact on breast cancer risk and implications for screening. *Radiographics*, 2015. 35(2): p. 302-15.DOI: 10.1148/rg.352140106.
4. Nguyen, H.T., et al., VinDr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography. *Scientific Data*, 2023. 10(1): p. 277.DOI: 10.1038/s41597-023-02100-7.

5. Carr, C., et al. RSNA Screening Mammography Breast Cancer Detection. 2022; Available from: <https://kaggle.com/competitions/rsna-breast-cancer-detection>.
 6. Kopans, D., et al. Digital Database for Screening Mammography (DDSM). 2000 2000; <http://www.eng.usf.edu/cvprg/mammography/database.html>.
 7. Lehman, C.D., et al., Mammographic Breast Density Assessment Using Deep Learning: Clinical Implementation. *Radiology*, 2019. 290(1): p. 52-58.DOI: 10.1148/radiol.2018180694.
 8. Magni, V., et al., Development and Validation of an AI-driven Mammographic Breast Density Classification Tool Based on Radiologist Consensus. *Radiology: Artificial Intelligence*, 2022. 4(2): p. e210199.DOI: 10.1148/ryai.210199.
 9. Mohamed, A.A., et al., A deep learning method for classifying mammographic breast density categories. *Med Phys*, 2018. 45(1): p. 314-321.DOI: 10.1002/mp.12683.
 10. Sprague, B.L., et al., Variation in Mammographic Breast Density Assessments Among Radiologists in Clinical Practice: A Multicenter Observational Study. *Ann Intern Med*, 2016. 165(7): p. 457-464.DOI: 10.7326/m15-2934.
 11. Chan, H.P. and M.A. Helvie, Deep Learning for Mammographic Breast Density Assessment and Beyond. *Radiology*, 2019. 290(1): p. 59-60.DOI: 10.1148/radiol.2018182116.
-