

BREAST CANCER RISK PREDICTION MODELS: A SYSTEMATIC REVIEW

Tran Thi Thanh Huong^{1,2}, Nguyen Huong Giang^{1,2}, Do Vu Minh Ha^{1,2}, Bui Thi Oanh^{1,2}, Pham Tuong Van²

¹National Cancer Institute, K Hospital-No. 43 Quan Su, Cua Nam Ward, Hanoi City, Vietnam

²Institute of Preventive Medicine and Public Health, Hanoi Medical University-No. 1 Ton That Tung, Kim Lien Ward, Hanoi City, Vietnam

Received: 06/04/2026

Revised: 24/04/2026; Accepted: 25/06/2026

ABSTRACT

In Vietnam, breast cancer ranks as the leading malignancy among women, underscoring the importance of developing accurate risk prediction tools to facilitate screening and early diagnosis.

Objective: To review breast cancer risk prediction models developed worldwide and propose a suitable model for the Vietnamese context.

Methods: Models were evaluated based on AUC, sensitivity, specificity, and risk of bias assessed by the PROBAST tool. Meta-analysis using a random-effects approach was conducted to pool AUC estimates across studies.

Results: Eighteen models were identified, five of which had undergone external validation, using up to 19 predictors and sample sizes ranging from 222 to 1,455,493 participants. AUC values ranged from 0.46 to 0.80; traditional models (Gail, IBIS, BCSC, Rosner–Colditz) achieved AUCs of 0.55–0.70, while BRCAPRO and AI/deep learning models reached 0.70–0.80. The pooled AUC was 0.65 (95% CI: 0.63–0.68; $I^2 = 97.5\%$). Traditional models showed lower risk of bias than AI-based models per PROBAST.

Conclusion: The Gail model is a feasible initial option for Vietnam given its simplicity, but requires local adaptation and validation.

Keywords: risk prediction model, breast cancer.

*Corresponding author

Email: huongtranthanh@hmu.edu.vn **Phone:** (+84) 984568118 **DOI:** 10.52163/yhc.v67iCD8.5711

CÁC MÔ HÌNH DỰ BÁO NGUY CƠ UNG THƯ VÚ: NGHIÊN CỨU TỔNG QUAN HỆ THỐNG

Trần Thị Thanh Hương^{1,2}, Nguyễn Hương Giang^{1,2}, Đỗ Vũ Minh Hà^{1,2}, Bùi Thị Oanh^{1,2}, Phạm Tường Vân²

¹Viện Ung thư Quốc gia, Bệnh viện K-Số 43 Quán Sứ, Phường Cửa Nam, thành phố Hà Nội, Việt Nam

²Viện Đào tạo Y học dự phòng và Y tế công cộng, Trường Đại học Y Hà Nội-Số 1 Tôn Thất Tùng, phường Kim Liên, thành phố Hà Nội, Việt Nam

Ngày nhận bài: 06/04/2026

Ngày chỉnh sửa: 24/04/2026; Ngày duyệt đăng: 25/06/2026

TÓM TẮT

Ung thư vú là bệnh ung thư phổ biến nhất ở phụ nữ Việt Nam. Việc xây dựng mô hình dự báo nguy cơ ung thư vú đóng vai trò quan trọng trong sàng lọc và phát hiện sớm.

Mục tiêu: Tổng quan các mô hình dự báo nguy cơ ung thư vú trên thế giới nhằm đề xuất mô hình phù hợp bối cảnh Việt Nam.

Phương pháp: Nghiên cứu tổng quan hệ thống theo hướng dẫn PRISMA 2020, tổng hợp các nghiên cứu công bố đến ngày 31/12/2024. Các mô hình được đánh giá dựa trên độ nhạy, độ đặc hiệu, diện tích dưới đường cong AUC, và đánh giá nguy cơ sai lệch bằng công cụ PROBAST. Phân tích gộp được thực hiện đối với AUC bằng mô hình ảnh hưởng ngẫu nhiên.

Kết quả: Có 18 mô hình được phát triển, trong đó 5 mô hình đã được ngoại kiểm. Các mô hình sử dụng 19 biến số khác nhau, với kích thước mẫu từ 222 đến 1.455.493 người. Giá trị AUC của các mô hình dao động từ 0,46 đến 0,80, trong đó các mô hình truyền thống như Gail, IBIS, BCSC và Rosner-Colditz thường đạt AUC 0,55–0,70. Các mô hình dành cho nhóm nguy cơ cao (BRCAPRO) và các mô hình AI/deep learning cho kết quả cao hơn (AUC khoảng 0,70–0,80). Giá trị AUC gộp là 0,65 (95% CI: 0,63–0,68) với mức dị hợp cao ($I^2 = 97,5\%$). Đánh giá PROBAST cho thấy các mô hình truyền thống có nguy cơ sai lệch thấp hơn so với các mô hình AI.

Kết luận: Mô hình Gail có thể là lựa chọn phù hợp bước đầu tại Việt Nam nhờ tính khả thi, nhưng cần được hiệu chỉnh và kiểm định trên dữ liệu trong nước.

Từ khóa: mô hình dự báo nguy cơ, ung thư vú

1. ĐẶT VẤN ĐỀ

Ung thư vú là một trong những thách thức y tế lớn trên toàn cầu. Theo số liệu từ Cơ quan Nghiên cứu Ung thư Quốc tế (IARC) trong báo cáo GLOBOCAN 2022, ung thư vú xếp thứ 4 về nguyên nhân tử vong do ung thư, với tỷ suất mắc mới chuẩn hóa theo tuổi đạt 46,8/100.000 dân và tỷ suất tử vong chuẩn hóa theo tuổi là 12,7/100.000 người. Riêng tại Việt Nam, năm 2022 có 24.563 ca mới được ghi nhận, tương đương 13,6% tổng số ca ung thư, cùng với khoảng 10.000 ca tử vong ở phụ nữ (chiếm 8,3%). Tỷ suất mắc mới và tỷ suất tử vong chuẩn hóa theo tuổi lần lượt là 38/100.000 và 14,7/100.000 người [1].

Hiện nay, nhiều yếu tố nguy cơ ung thư vú đã được xác nhận qua nghiên cứu, bao gồm tuổi tác, các yếu tố nội tiết như tuổi bắt đầu có kinh nguyệt, các yếu tố sinh sản như tuổi sinh con đầu lòng, cũng như một số đặc điểm di truyền. Xuất phát từ cơ sở đó, các mô hình dự báo nguy cơ ung thư vú đã được xây dựng nhằm hỗ trợ công tác sàng lọc và phát hiện bệnh ở giai đoạn sớm.

Trên thế giới đã có nhiều mô hình dự báo nguy cơ ung thư vú được xây dựng và phát triển nhằm xác định tác động

kết hợp của các yếu tố nguy cơ, các chương trình sàng lọc và xét nghiệm đột biến gen, từ đó làm giảm gánh nặng của bệnh ung thư vú [2]. Sàng lọc dựa trên yếu tố nguy cơ có thể cải thiện hiệu quả chi phí và tối đa hóa lợi ích. Mô hình dự báo nguy cơ cá thể cho ung thư vú có thể được sử dụng trong thực hành lâm sàng nhằm hỗ trợ đưa ra quyết định về triển khai sàng lọc trên phạm vi rộng và đưa ra các chiến lược điều trị bệnh [3][4] [5] [6].

Tại Việt Nam hiện nay vẫn chưa có mô hình xác định nguy cơ ung thư vú, để có thể xây dựng được mô hình hiệu quả thích hợp với người dân Việt Nam, chúng tôi tiến hành nghiên cứu tổng quan các mô hình dự báo nguy cơ ung thư vú của các quốc gia khác nhau nhằm đưa ra gợi ý mô hình phù hợp cho người Việt Nam.

2. PHƯƠNG PHÁP

2.1. Đối tượng và địa điểm nghiên cứu

2.1.1. Đối tượng: Các bài báo trong nước và quốc tế liên quan tới nguy cơ, các mô hình dự báo nguy cơ ung thư vú.

*Tác giả liên hệ

Email: huongtranthanh@hmu.edu.vn Điện thoại: (+84) 984568118 DOI: 10.52163/yhc.v67iCD8.5711

2.1.2. Địa điểm nghiên cứu

Nghiên cứu sử dụng hai nhóm nguồn dữ liệu:

Các cơ sở dữ liệu điện tử chính thức:

Tên cơ sở dữ liệu điện tử	Liên kết truy cập
Google scholar	https://scholar.google.com.vn/
Pubmed	https://www.ncbi.nlm.nih.gov/pubmed/

Nguồn dữ liệu bổ sung:

Google scholar, ResearchGate

Tài liệu, công trình nghiên cứu, các bài báo cáo hội nghị trong và ngoài nước, các báo cáo tổng kết của Tổ chức Y tế Thế giới (WHO) và Bộ Y tế.

Tài liệu tham khảo của các nghiên cứu sẵn có. Các nguồn này chủ yếu được sử dụng nhằm hỗ trợ truy xuất tài liệu toàn văn và đối chiếu thông tin, không đóng vai trò là nguồn dữ liệu chính cho quá trình tổng hợp định lượng.

Cơ sở dữ liệu cập nhật đến ngày 31/12/2024.

2.2. Phương pháp nghiên cứu

Áp dụng phương pháp nghiên cứu tổng quan hệ thống.

2.2.1. Xác định câu hỏi nghiên cứu

Câu hỏi nghiên cứu “Các mô hình dự báo nguy cơ ung thư vú hiện nay có đặc điểm gì phương pháp, các biến số, độ chính xác và khả năng áp dụng thực tiễn?” được xây dựng theo khung dành cho nghiên cứu mô hình dự báo, bao gồm:

- Population (P): Phụ nữ có nguy cơ mắc ung thư vú (độ tuổi, tiền sử gia đình, mang gen BRCA1/BRCA2 hoặc có các yếu tố nguy cơ ung thư vú khác)
- Outcome (O) (S):
 - o Đánh giá hiệu quả các mô hình dự báo dựa trên các tiêu chí như độ đặc hiệu, độ nhạy, diện tích đường cong AUC, tỷ số dự báo/quan sát,
 - o Đánh giá tính khả thi và tác động của các mô hình đến các chiến lược dự phòng, phát hiện sớm ung thư vú.
- Predictors: Các yếu tố nhân khẩu học, sinh sản, nội tiết, tiền sử gia đình, yếu tố di truyền, hình ảnh học và các biến lâm sàng khác.
- Model type: Các mô hình dự báo nguy cơ ung thư vú truyền thống, mô hình thống kê đa biến, mô hình di truyền và mô hình machine learning/trí tuệ nhân tạo.
- Performance/Validation: Các chỉ số đánh giá hiệu năng mô hình như AUC, độ nhạy, độ đặc hiệu, calibration, cũng như thông tin về nội kiểm hoặc ngoại kiểm.

2.2.2. Xây dựng tiêu chuẩn lựa chọn/loại trừ

Dựa trên câu hỏi nghiên cứu, chúng tôi tiến hành lựa chọn các nghiên cứu thỏa mãn tiêu chí:

- Thỏa mãn các tiêu chí xây dựng câu hỏi nghiên cứu trong thời gian lựa chọn
- Nghiên cứu được viết bằng tiếng Anh hoặc tiếng Việt
- Các nghiên cứu sẽ bị loại trừ nếu:
- Không thỏa mãn các tiêu chí của câu hỏi nghiên cứu
- Nghiên cứu được công bố bằng các ngôn ngữ khác.

2.2.3. Phương pháp, chiến lược tìm kiếm và nguồn dữ liệu

- Các từ khóa tìm kiếm:

o Các từ khóa tiếng Việt được sử dụng trong việc tìm kiếm tài liệu là: ung thư vú, yếu tố nguy cơ ung thư vú, mô hình dự báo nguy cơ ung thư vú.

o Từ khóa tiếng Anh: breast cancer, breast cancer risk factors, breast cancer risk prediction model, breast cancer risk prediction model, breast cancer risk assessment, breast cancer risk prediction tool.

- Chuỗi tìm kiếm tài liệu trên Pubmed như sau: (“breast cancer”[Title/Abstract] OR “breast neoplasm”[Title/Abstract]) AND (“risk prediction model”[Title/Abstract] OR “risk assessment”[Title/Abstract] OR “prediction tool”[Title/Abstract]) AND (“validation”[Title/Abstract] OR “AUC”[Title/Abstract] OR “ROC”[Title/Abstract])

2.2.4. Lựa chọn nghiên cứu

- Sau khi tìm kiếm các nghiên cứu, nhóm nghiên cứu thực hiện sàng lọc nghiên cứu một cách độc lập, sau đó thống nhất để đưa đến kết quả các nghiên cứu được đưa vào nghiên cứu tổng quan.
- Các nghiên cứu được lựa chọn sau vòng sàng lọc được quản lý bằng phần mềm Zotero để lưu trữ tài liệu và loại bỏ các nghiên cứu trùng lặp.

2.2.5. Trích xuất dữ liệu

Dữ liệu từ các nghiên cứu đủ tiêu chuẩn được trích xuất theo biểu mẫu chuẩn hóa nhằm đảm bảo tính nhất quán trong quá trình tổng hợp và phân tích. Các thông tin được trích xuất bao gồm:

- Thông tin cơ bản: Tác giả (nhóm tác giả), thời gian tiến hành nghiên cứu, địa điểm (nơi thực hiện nghiên cứu), nhà xuất bản, tên tạp chí.
- Thông tin phương pháp: Thiết kế nghiên cứu, quy mô mẫu, phương pháp chọn mẫu và các biến số được đưa vào mô hình
- Kết quả nghiên cứu: Độ nhạy, độ đặc hiệu, diện tích đường cong AUC, tỷ lệ dự đoán/quan sát và tính khả thi, giá trị của mô hình dự báo nguy cơ trong sàng lọc và phát hiện sớm bệnh ung thư vú.

2.2.6. Đánh giá về chất lượng các bài báo

Với các bài báo được lựa chọn, chúng tôi tổng hợp và đánh giá chất lượng của mô hình dự báo bằng thang điểm PROBAST (Prediction Model Risk of Bias Assessment Tool) [7]. Đây là thang được sử dụng để đánh giá nguy cơ sai lệch với 4 nội dung và 20 tiêu chí về sự phù hợp và tính đại diện của quần thể nghiên cứu, khả năng đo lường và tính sẵn có của các biến dự báo, phương pháp đo lường kết quả đầu ra cũng như tính hiệu năng và khả năng mô hình hoá. Theo hướng dẫn của PROBAST, mỗi nội dung được phân loại thành: nguy cơ sai lệch thấp, nguy cơ sai cao và chưa rõ nguy cơ sai lệch.

Trong nghiên cứu này, để thuận tiện cho việc trực quan hóa kết quả bằng biểu đồ tổng hợp, các mức đánh giá được mã hóa nội bộ theo quy ước: 0 = sai lệch thấp; 1 = Không xác định; 2 = sai lệch cao. Việc mã hóa này chỉ phục vụ mục đích trình bày định lượng và không thay thế phương pháp đánh giá cơ bản của PROBAST.

2.2.7. Đánh giá mô hình

Phương pháp đo lường độ nhạy, độ đặc hiệu, diện tích dưới đường cong (AUC) cho các mô hình dự báo nguy cơ ung thư:

Để đánh giá hiệu năng các mô hình dự báo nguy cơ ung thư vú, diện tích dưới đường cong ROC (Area Under the Curve) được tính từ đường cong ROC, với giá trị dao động từ 0,5 (không có khả năng dự báo) đến 1,0 (dự báo hoàn toàn chính xác), trong đó $AUC \geq 0,7$ được xem là có khả năng dự báo ở mức chấp nhận được [8].

2.2.8. Phân tích tổng hợp

Bên cạnh tổng quan mô tả, nghiên cứu tiến hành phân tích tổng hợp (meta-analysis) đối với chỉ số hiệu năng của các mô hình, trong đó sử dụng diện tích dưới đường cong AUC. Từ mỗi nghiên cứu, nhóm nghiên cứu trích xuất giá trị AUC cùng khoảng tin cậy 95%.

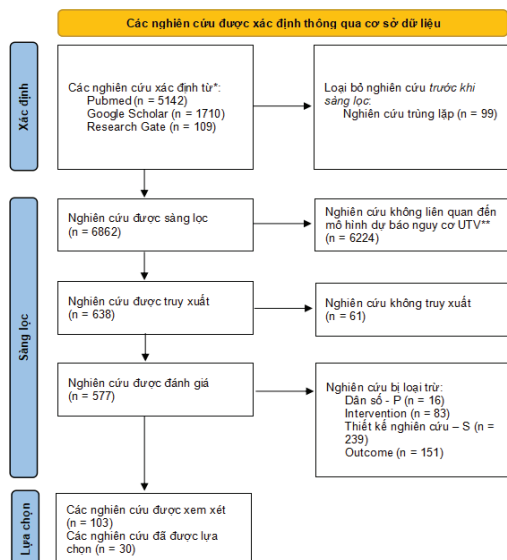
Các giá trị này được dùng làm cơ sở để xác định trọng số trong mô hình meta-analysis. Do sự khác biệt rõ rệt giữa dân số nghiên cứu, biến số và phương pháp mô hình hóa, tất cả phân tích tổng hợp được thực hiện bằng mô hình ảnh hưởng ngẫu nhiên.

Mức độ không đồng nhất giữa các nghiên cứu được lượng hóa thông qua chỉ số I^2 , kiểm định Q của Cochran và ước tính phương sai giữa các nghiên cứu (τ^2). I^2 cao được kỳ vọng trong bối cảnh tổng hợp nhiều mô hình dự báo khác nhau và nhiều quần thể khác nhau. Toàn bộ quá trình phân tích thống kê được thực hiện bằng ngôn ngữ lập trình R.

Do sự khác biệt đáng kể giữa các nghiên cứu về quần thể, nhóm nguy cơ, biến dự báo và phương pháp xây dựng mô hình (mô hình truyền thống, mô hình di truyền và mô hình AI/deep learning), kết quả phân tích gộp được diễn giải theo hướng mô tả xu hướng hiệu năng chung hơn là đại diện cho một giá trị hiệu năng chuẩn duy nhất.

3. KẾT QUẢ

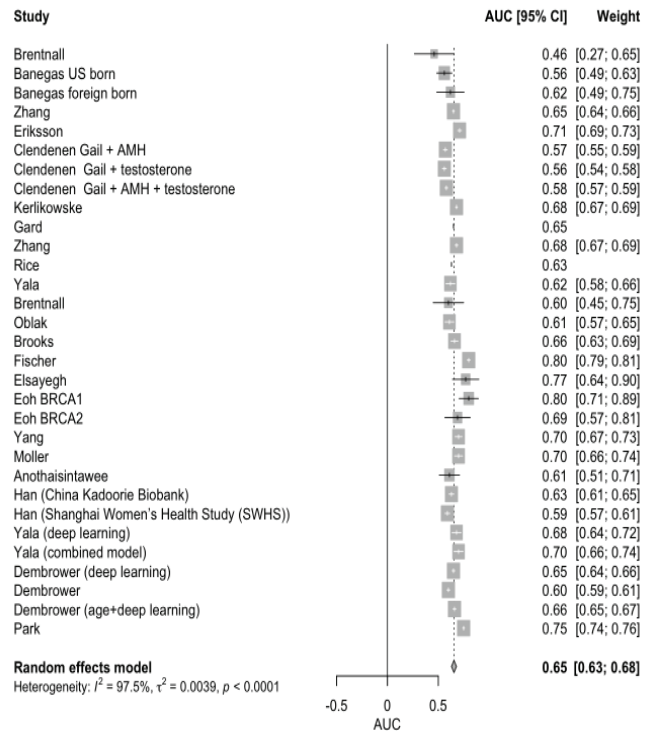
Quy trình tổng quan hệ thống được thực hiện theo hướng dẫn PRISMA2020. Tổng số 6.961 tài liệu được tìm kiếm từ ba cơ sở dữ liệu (PubMed, Google Scholar và ResearchGate). Sau khi loại bỏ trùng lặp và các bài không có toàn văn, 638 bài được sàng lọc chi tiết theo tiêu chí PRISMA và PICOS. Kết quả cuối cùng, lựa chọn được 30 nghiên cứu toàn văn về các mô hình dự báo nguy cơ ung thư vú. Trong đó, 18 mô hình dự báo nguy cơ UTV khác nhau được ghi nhận, bao gồm nghiên cứu phát triển mô hình mới, nghiên cứu hiệu chỉnh/cập nhật mô hình và nghiên cứu ngoại kiểm; 5 mô hình có dữ liệu ngoại kiểm độc lập.



Hình 1. Sơ đồ PRISMA

Các yếu tố được sử dụng trong các mô hình dự báo bao gồm: tuổi, dân tộc, các yếu tố sinh sản (tuổi có kinh, tuổi sinh con đầu lòng). Ngoài ra, các mô hình gần đây còn tích hợp thêm độ dày mô vú qua nhũ ảnh và yếu tố di truyền như đột biến BRCA1/BRCA2 hoặc tiền sử gia đình.

Biểu đồ 1. Phân tích tổng hợp giá trị AUC



Giá trị AUC giữa các nghiên cứu có sự dao động đáng kể, phản ánh sự khác biệt về thiết kế, quần thể nghiên cứu và phương pháp mô hình hóa.

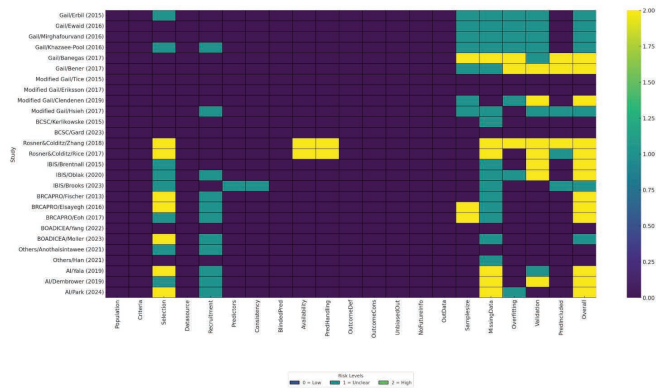
Các mô hình truyền thống (Gail, IBIS, BCSC, Rosner-Colditz) có AUC dao động từ 0,55–0,70, cho thấy khả năng phân biệt ở mức trung bình. Một số nghiên cứu ghi nhận AUC thấp (Brentnall: 0,46; Banegas: 0,56), cho thấy hạn chế khi áp dụng mô hình Gail cho các quần thể khác nếu không hiệu chỉnh.

Một số mô hình cải tiến (Eriksson, Yang/BOADICEA) đạt AUC khoảng 0,70, cao hơn mức trung bình của các mô hình cổ điển.

Nhóm mô hình dành cho người mang đột biến BRCA (BRCA1/BRCA2) có AUC cao hơn, khoảng 0,77–0,80, phù hợp với nhóm nguy cơ cao.

Các mô hình AI/deep learning (Yala, Dembrower, Park) đạt AUC khoảng 0,70–0,75, cho thấy tiềm năng cải thiện dự báo nhờ khai thác dữ liệu hình ảnh.

Kết quả phân tích gộp cho thấy giá trị AUC trung bình đạt 0,65 (95% CI: 0,63–0,68). Tuy nhiên, mức độ dị hợp là rất cao giữa các nghiên cứu ($I^2 = 97,5\%$). Sự dị hợp này có thể xuất phát từ khác biệt về quần thể nghiên cứu, loại biến dự báo được sử dụng, cũng như phương pháp xây dựng mô hình. Ngoài ra, phân tích gộp trong nghiên cứu bao gồm đồng thời các mô hình truyền thống, mô hình di truyền và mô hình AI/deep learning, cũng như các quần thể nguy cơ trung bình và nguy cơ cao. Do đó, giá trị AUC gộp chủ yếu phản ánh xu hướng hiệu năng tổng quát hơn là một giá trị đại diện có thể áp dụng trực tiếp trong thực hành lâm sàng.

Biểu đồ 2. Đánh giá sai lệch theo PROBAST

Biểu đồ 2 Kết quả đánh giá PROBAST được trình bày trong Biểu đồ 2, cho thấy sự phân hóa đáng kể về nguy cơ sai lệch giữa các mô hình được xem xét.

Các mô hình Gail (gốc và cải tiến) chủ yếu có nguy cơ sai lệch thấp, nhờ quy trình chuẩn hóa lâu dài; tuy nhiên, một số phiên bản mở rộng vẫn có tiêu chí chưa rõ hoặc nguy cơ cao ở biến dự báo.

Các mô hình BCSC và BOADICEA nhìn chung có nguy cơ sai lệch thấp, do được xây dựng trên dữ liệu lớn và được kiểm định tốt, dù một số tiêu chí phân tích còn chưa rõ.

Nhóm Rosner–Colditz và IBIS có mức sai lệch trung bình, với một số hạn chế liên quan đến cỡ mẫu và tính nhất quán của dữ liệu.

Ngược lại, các mô hình AI/deep learning có mức sai lệch biến thiên lớn hơn, đặc biệt ở các khía cạnh như xử lý dữ liệu thiếu, nguy cơ quá khớp và thiếu kiểm định độc lập.

Qua tổng hợp các nghiên cứu, các biến dự báo xuất hiện ổn định nhất giữa các mô hình bao gồm tuổi, tiền sử gia đình mắc ung thư vú, tiền sử sinh sản và các yếu tố nội tiết. Các biến này hiện diện trong phần lớn các mô hình truyền thống như Gail, IBIS, BCSC và Rosner–Colditz, cho thấy tính nhất quán tương đối cao giữa các nghiên cứu.

Các mô hình có external validation độc lập, đặc biệt BCSC và BOADICEA, nhìn chung cho thấy tính ổn định tốt hơn về hiệu năng và khả năng ngoại suy. Ngược lại, nhiều mô hình AI/deep learning mặc dù đạt AUC cao nhưng còn hạn chế về minh bạch thuật toán, xử lý dữ liệu thiếu và thiếu kiểm định trên quần thể độc lập.

4. BÀN LUẬN

Qua 30 nghiên cứu được tổng hợp về mô hình nguy cơ ung thư vú cho thấy các yếu tố nguy cơ được đưa vào mô hình bao gồm các yếu tố về tuổi; yếu tố liên quan tới nội tiết (tuổi có kinh lần đầu, tuổi mãn kinh); tiền sử sinh sản (tuổi có con lần đầu; có cho con bú hay không); yếu tố liên quan tới gen như tiền sử gia đình mắc ung thư vú hay sự có mặt của BRCA1; BRCA2...

Hiệu quả hay khả năng dự báo của mô hình cũng rất khác nhau tùy thuộc vào các yếu tố đầu vào đưa vào mô hình. Mỗi mô hình có những ưu điểm và hạn chế nhất định. Tuy nhiên, mô hình Gail và Gail cải tiến và được phát triển đầu tiên lại có tính ổn định cao và nguy cơ sai lệch không nhiều. Điều này có thể do các yếu tố nguy cơ được đưa vào mô hình là các yếu tố nguy cơ đã được khẳng định rõ rệt.

Phần lớn các nghiên cứu được tập trung vào nhóm quần thể người da trắng. Do vậy, khi áp dụng tại các quần thể châu Á thì cần hiệu chỉnh do các đặc điểm về gen cũng như lối sống khác nhau.

Rất ít các mô hình xây dựng bởi nghiên cứu được ngoại kiểm. Điều này có thể do việc tổ chức cũng như hạn chế về kinh phí

và nhân lực. Tuy nhiên, nếu không có việc kiểm định này thì cũng khó có thể khẳng định giá trị của mô hình đã xây dựng.

Mô hình có sử dụng các yếu tố về gen thì có giá trị cao hơn, nhưng không phải nơi nào cũng đầy đủ các yếu tố về nhân lực cũng như cơ sở vật chất để có thể thực hiện việc đánh giá nguy cơ này.

Tại Việt Nam, do chưa có đầy đủ và toàn diện các yếu tố nguy cơ liên quan tới nội tiết cũng như đặc điểm về gen, do vậy bước đầu có thể sử dụng mô hình Gail để đánh giá nguy cơ ung thư vú và trong tương lai có thể ứng dụng AI khi có đầy đủ các dữ liệu đầu vào. Điều này hoàn toàn khả thi khi triển khai mã định danh cá nhân và bệnh án điện tử trên phạm vi toàn quốc.

5. KẾT LUẬN

Có nhiều mô hình dự báo nguy cơ ung thư vú với mức độ hiệu năng và độ tin cậy khác nhau. Các mô hình truyền thống như Gail, BCSC và BOADICEA có ưu điểm về tính minh bạch, khả năng triển khai và đã được kiểm định trên nhiều quần thể, trong khi các mô hình AI/deep learning cho thấy tiềm năng cải thiện hiệu năng nhưng vẫn cần thêm bằng chứng về tính ổn định và khả năng ngoại suy.

Phần lớn các mô hình hiện nay được phát triển trên quần thể phương Tây, do đó cần được hiệu chỉnh và kiểm định trước khi áp dụng cho phụ nữ Việt Nam. Trong bối cảnh dữ liệu lâm sàng và di truyền còn hạn chế, mô hình Gail có thể được xem như một lựa chọn khả thi để cân nhắc trong giai đoạn đầu nhờ cấu trúc đơn giản và tính khả thi tương đối cao. Tuy nhiên, cần có thêm nghiên cứu đánh giá tính sẵn có của biến số, hiệu năng dự báo và khả năng hiệu chỉnh mô hình trên dữ liệu trong nước trước khi triển khai rộng rãi.

6. TÀI LIỆU THAM KHẢO

- [1] Cancer (IARC) T.I.A. for R. on Global Cancer Observatory. <https://gco.iarc.fr/>, accessed: 01/11/2022.
- [2] Zheng Y., Li J., Wu Z., et al. (2022). Risk prediction models for breast cancer: a systematic review. *BMJ Open*, 12(7), e055398. doi: 10.1136/bmjopen-2021-055398
- [3] Gail M.H., Brinton L.A., Byar D.P., et al. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst*, 81(24), 1879–1886. doi: 10.1093/jnci/81.24.1879
- [4] Colditz G.A. and Rosner B. (2000). Cumulative risk of breast cancer to age 70 years according to risk factor status: data from the Nurses' Health Study. *Am J Epidemiol*, 152(10), 950–964. doi: 10.1093/aje/152.10.950
- [5] Maas P., Barrdahl M., Joshi A.D., et al. (2016). Breast Cancer Risk From Modifiable and Nonmodifiable Risk Factors Among White Women in the United States. *JAMA Oncol*, 2(10), 1295–1302. doi: 10.1001/jamaoncol.2016.1025
- [6] Lee A., Mavaddat N., Wilcox A.N., et al. (2019). BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med*, 21(8), 1708–1718. doi: 10.1038/s41436-018-0406-9
- [7] Tice J.A., Bissell M.C.S., Miglioretti D.L., et al. (2019). Validation of the Breast Cancer Surveillance Consortium Model of Breast Cancer Risk. *Breast Cancer Res Treat*, 175(2), 519–523. doi: 10.1007/s10549-019-05167-4
- [8] Wolff RF, Moons KGM, Riley RD, et al. (2019). PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med*, 170(1), 51–58. doi: 10.7326/M18-1376