

XÂY DỰNG HỆ HỖ TRỢ CHẨN ĐOÁN X-QUANG UNG THƯ VÚ BẰNG TRÍ TUỆ NHÂN TẠO

Bùi Mỹ Hạnh*, Lê Tuấn Linh*, Nguyễn Ngọc Cương*, Nguyễn Thanh Bình*, Lưu Tiến Đoàn*, Lê Duy Chung*, Giao Tất Vũ*, Ngô Thị Ly Ly*, Hoàng Thị Hồng Xuyên*, Nguyễn Đức Thắng**, Nguyễn Tú Anh**, Nguyễn Đức Dân**, Nguyễn Việt Dũng***, Trần Vĩnh Đức***, Nguyễn Hồng Quang***, Anh Nguyen****, Nguyễn Hoàng Phương**

Nhận bài: 20/07/2021; Chấp nhận đăng: 10/08/2021

© 2021 Trường Đại học Thăng Long.

Tóm tắt

Hiện nay, ung thư vú là loại ung thư phổ biến nhất ở phụ nữ nhiều nước trên thế giới nói chung và Việt Nam nói riêng. Trong các phương pháp sàng lọc phát hiện sớm ung thư vú thì chụp ảnh X-quang ung thư vú được coi là một phương pháp quan trọng và có giá trị cao. Việc đánh giá mức độ tổn thương trên ảnh X-quang ung thư vú được thực hiện qua hệ thống phân loại BI-RADS. Nghiên cứu này đề xuất phương pháp phân loại một ảnh mammogram của bệnh nhân theo ba loại: biểu hiện cho tuyến vú bình thường, không có tổn thương (BI-RADS 1), tuyến vú có tổn thương với khả năng lành tính cao (tương ứng với BI-RADS 23) và không đánh giá được tổn thương trên X-quang ung thư vú hoặc khả năng ung thư cao, cần dựa vào các thăm dò khác (BI-RADS 045). Hệ thống của chúng tôi được xây dựng dựa trên mạng nơ ron tích chập CNN với kiến trúc nền tảng là ResNet50. Hệ thống được huấn luyện trên bộ dữ liệu ảnh X-quang ung thư vú được xây dựng bởi các bác sĩ chẩn đoán hình ảnh Bệnh viện Đại học Y Hà Nội với 7.912 ảnh mammograms. Hệ thống này cho kết quả trên tập kiểm tra với độ đo độ tin cậy của hệ thống là macAUC = 0.75. Ngoài ra, chúng tôi cũng làm nghiên cứu so sánh độ tin cậy Acc (Accuracy), độ nhạy Sn (Sensitivity) và độ đặc hiệu Sp (Specificity) của các bác sĩ chẩn đoán hình ảnh X-quang ung thư vú của Bệnh viện Đại học Y Hà Nội và của hệ thống với gần 500 ảnh “chuẩn vàng” lấy ngẫu nhiên từ tập kiểm tra, kết quả cho thấy độ tin cậy chẩn đoán của hệ thống tương đương với “hội đồng” bác sĩ chẩn đoán hình ảnh có kinh nghiệm và cao hơn một bác sĩ chẩn đoán hình ảnh được chọn ngẫu nhiên. Như vậy hệ hỗ trợ có thể được coi như “bác sĩ chẩn đoán hình ảnh thứ hai” hỗ trợ các bác sĩ chẩn đoán hình ảnh trong việc tầm soát ung thư vú ở Việt Nam.

Từ khóa: Ảnh X-quang ung thư vú; Loại BI-RADS; Mạng nơ ron tích chập ResNet; Học sâu

* Bệnh viện Đại học Y Hà Nội

** Trường Đại học Thăng Long

*** Trường Đại học Bách Khoa Hà Nội

**** Auburn University, Auburn, Alabama, USA

1. Giới thiệu

1.1 Tình hình ung thư vú ở Việt Nam

Ung thư vú là loại ung thư phổ biến nhất ở phụ nữ nhiều nước trên Thế giới. Theo số liệu của cơ quan nghiên cứu ung thư quốc tế (GLOBOCAN) có khoảng 1.67 triệu trường hợp mới mắc (chiếm 25% tổng số bệnh ung thư nói chung) và hơn 522.000 phụ nữ chết vì ung thư vú năm 2012 [5]. “Theo thứ trưởng Bộ Y tế Trần Văn Thuấn, ung thư vú là bệnh ung thư hàng đầu ở phụ nữ cả ở các nước phát triển và đang phát triển. Tỷ lệ mắc bệnh ung thư vú đang gia tăng ở các nước đang phát triển do tuổi thọ tăng, tốc độ đô thị hóa tăng và thay đổi lối sống. Theo báo cáo ghi nhận ung thư toàn cầu Globocan 2020, tại Việt Nam, trong số ung thư ở nữ, số người mới mắc ung thư vú đứng hàng thứ nhất, với 21.555 người, chiếm 25,8%, tính theo cả hai giới đứng hàng thứ ba (sau ung thư gan và ung thư phổi). Tỷ lệ mới mắc chuẩn theo tuổi của ung thư vú ở nữ là 34.2 trên 100.000 dân. Ở cả hai giới, số tử vong do ung thư vú đứng hàng thứ tư (với 9.345 trường hợp) sau ung thư gan, ung thư phổi và ung thư dạ dày. Tỷ lệ tử vong chuẩn theo tuổi do ung thư vú là 13.8 trên 100.000 dân. Theo các chuyên gia, mặc dù có thể đạt được mục tiêu giảm một số nguy cơ gây ung thư vú thông qua dự phòng, nhưng chiến lược này không thể loại bỏ phần lớn ung thư vú đang phát triển ở các quốc gia có thu nhập thấp và trung bình, nơi ung thư vú được chẩn đoán ở giai đoạn rất muộn. Do đó, phát hiện sớm để cải thiện kết quả và sự sống còn của ung thư vú vẫn là nền tảng của việc kiểm soát ung thư vú, đồng thời tăng tiếp cận các liệu pháp điều trị phù hợp, tiên tiến là một giải pháp quan trọng trong kiểm soát ung thư vú” [12].

Một số khái niệm về các giai đoạn ung thư vú

được đề cập trong [10] như sau:

GIAI ĐOẠN 0 UNG THƯ VÚ

Bạn sẽ thường nghe nói về bệnh ung thư đã bắt đầu trong ống dẫn sữa hoặc tuyến sữa là “tại chỗ”, có nghĩa là nó đã ở “tại vị trí ban đầu”. Ung thư ở giai đoạn này được coi là không xâm lấn vì không có bằng chứng về các tế bào ung thư hoặc bất thường khác tồn tại ở nơi khác.

GIAI ĐOẠN I UNG THƯ VÚ

Ung thư ở giai đoạn này được coi là xâm lấn vì các tế bào khối u đã bắt đầu tấn công các mô vú khỏe mạnh. Tuy nhiên, nó được phân loại là “giai đoạn đầu”, do ung thư nằm trong một khu vực nhỏ.

GIAI ĐOẠN II UNG THƯ VÚ

Ở giai đoạn này, ung thư được coi là khu trú. Ung thư đã phát triển và / hoặc lan đến một vùng giới hạn của vú và các hạch bạch huyết.

GIAI ĐOẠN III UNG THƯ VÚ

Mặc dù ung thư đã di căn theo khu vực, nhưng nó vẫn chưa di căn đến xương hoặc các cơ quan xung quanh. Tuy nhiên, lúc này bệnh ung thư được coi là đã tiến triển và khó chữa hơn.

UNG THƯ VÚ GIAI ĐOẠN IV

Ở giai đoạn này, ung thư bây giờ được gọi là “ung thư vú di căn”, có nghĩa là nó đã di căn xa vú và vào xương, phổi, gan và / hoặc não. Hiện tại, vẫn chưa có phương pháp chữa trị nào được biết đến, nhưng việc điều trị có thể giúp giảm bớt các triệu chứng và kéo dài tuổi thọ.

1.2. Chụp X-quang vú

Các phương pháp sàng lọc phát hiện sớm ung thư vú bằng tự khám vú hằng ngày, khám sức khỏe định kỳ sau tuổi 40 và đặc biệt là chụp

X-quang tuyến vú (mammography) có vai trò rất quan trọng trong giảm tỷ lệ chết do ung thư vú, lên tới 20-22% [25]. Chính vì thế, chụp X-quang tuyến vú được coi là một phương pháp quan trọng và có giá trị cao trong chương trình sàng lọc phát hiện ung thư vú ở mỗi quốc gia. Thêm vào đó, chụp X-quang vú đối với những bệnh nhân nữ có triệu chứng lâm sàng, những người có tiền sử gia đình bị ung thư vú được khuyến cáo trong các hướng dẫn gần đây [22]. Theo đó, X-quang vú được chỉ định: để sàng lọc ung thư vú ở phụ nữ không có triệu chứng với độ tuổi sàng lọc từ 40 - 75 tuổi chụp X-quang vú 2 năm một lần và chụp 1 năm một lần với phụ nữ nguy cơ cao, để chẩn đoán ở những phụ nữ có triệu chứng tại vú như có dấu hiệu bất thường trên lâm sàng và siêu âm vú, định vị dẫn đường sinh thiết tổn thương vú.

Như vậy, nhu cầu chụp X-quang vú hiện nay là rất lớn. Chính vì số lượng chụp X-quang vú ngày càng tăng nên cần phải có nhiều bác sĩ chuyên về phân giải hình ảnh X-quang vú được gọi là bác sĩ chẩn đoán hình ảnh ung thư vú. Ở Việt Nam chưa có nghiên cứu về số lượng bác sĩ chẩn đoán hình ảnh ung thư vú đáp ứng với nhu cầu đang tăng lên của số lượng chụp X-quang tuyến vú. Tuy nhiên, theo một nghiên cứu dự báo ở Mỹ, số lượng bác sĩ về chẩn đoán hình ảnh X-quang tuyến vú sẽ thiếu hụt trầm trọng trong hai thập kỷ tới.

CÁC BƯỚC THỰC HIỆN KỸ THUẬT CHỤP X-QUANG VÚ [11]

- Giải thích cho bệnh nhân về quy trình chụp X-quang vú.
- Bệnh nhân cởi áo (áo ngoài và áo lót), có thể cho thay áo rộng, không cài cúc phía trước.
- Bệnh nhân đứng/ngồi quay mặt về máy chụp.

- Đặt vú lên bảng nhận ảnh, ép vú bởi tấm nhựa ép vú của máy. Ép đủ lực cho tuyến vú mỏng ra và dàn đều giúp quan sát rõ hơn tổn thương vú, giảm cường độ chiếu tia, ảnh không bị rung do di động (chú ý cảm giác đau của bệnh nhân).
- Chụp ít nhất ở hai tư thế: Tư thế thẳng trên-dưới (CC), tư thế chếch trong-ngoài (MLO). Có thể chụp tư thế nghiêng (ML) phối hợp xác định rõ hơn vị trí của tổn thương.
- Chụp ép phóng đại khu trú vùng nghi ngờ tổn thương hoặc tổn thương chưa được bộc lộ rõ trên các tư thế đã chụp.
- Hình ảnh chụp X-quang vú được xử lý và lưu trữ dưới dạng hình ảnh số hoặc được in phim.
- Siêu âm vú phối hợp (Hình ảnh X-quang vú luôn được đọc phối hợp với hình ảnh siêu âm vú).
- Phân tích hình ảnh, kết luận hình ảnh tuyến vú theo phân loại BIRADS.

1.3. Phân loại các tổn thương X-quang vú

Hiện nay, tại Việt Nam và Bệnh viện Đại học Y Hà Nội đang áp dụng hệ thống phân loại BI-RADS năm 2013 (Breast Imaging Reporting and Data System) để phân loại mức độ ác tính của các tổn thương trên X-quang tuyến vú [1]:

- BI-RADS 0: không đánh giá được tổn thương trên X-quang tuyến vú, cần dựa vào các thăm dò khác.
- BI-RADS 1: tuyến vú bình thường, không có tổn thương.
- BI-RADS 2: tuyến vú có tổn thương chắc chắn lành tính.
- BI-RADS 3: tuyến vú có tổn thương khả năng

lành tính cao, tỷ lệ ác tính <2%.

- BI-RADS 4: tuyến vú có tổn thương khả năng ác tính cao, tỷ lệ ác tính từ 2% đến 95%.
- BI-RADS 5: tuyến vú có tổn thương gần như chắc chắn ác tính, tỷ lệ ác tính > 95%.

Ở Việt Nam, ngoài sự thiếu hụt các bác sĩ chuyên về đọc phim X-quang tuyến vú còn có khoảng cách lớn về trình độ chuyên môn của các bác sĩ do sự phân tuyến trong y tế. Về số lượng bệnh nhân, các bệnh viện tuyến trung ương thường xuyên phải tiếp nhận số lượng lớn các bệnh nhân vượt tuyến đến trực tiếp, trong khi các bệnh viện tuyến cơ sở thì không có nhiều bệnh nhân đến khám. Qua đó, về trình độ các bác sĩ cũng không đồng đều ở các tuyến trong hệ thống y tế, điều này ảnh hưởng trực tiếp đến độ chính xác của chẩn đoán X-quang tuyến vú.

Để khắc phục được các vấn đề trên, một hệ thống hỗ trợ cho các bác sĩ đọc kết quả X-quang tuyến vú có thể giúp tăng độ chính xác của kết quả, giảm thời gian đọc kết quả ở tuyến trung ương và giảm sự chênh lệch về độ chính xác của các bác sĩ tại các tuyến cơ sở so với các tuyến trung ương cũng như giải quyết việc giảm sự thiếu hụt các bác sĩ chẩn đoán hình ảnh X-quang tuyến vú là một giải pháp hữu hiệu. Chính vì mục đích đó, Trường Đại học Thăng Long đã phối hợp với Bệnh viện Đại học Y Hà Nội phối hợp “nghiên cứu phát triển hệ hỗ trợ chẩn đoán X-quang ung thư vú bằng Trí tuệ nhân tạo”. Hệ hỗ trợ này nhằm mục đích phân loại mức độ BI-RADS trên ảnh X-quang vào các mức phân loại BIRADS 045, BI-RADS 1, BI-RADS 23.

Như chúng ta đã biết, trong khoảng 10 năm trở lại đây, Trí tuệ nhân tạo trong cách mạng công nghiệp 4.0 kết hợp với dữ liệu lớn, điện

toán đám mây và Internet vạn vật đang tạo ra những ứng dụng rộng rãi trong mọi hoạt động của đời sống xã hội, trong đó có lĩnh vực y học. Trí tuệ nhân tạo mới này với nền tảng là các kỹ thuật học sâu, đặc biệt là mạng nơ ron sâu như mạng nơ ron tích chập đã áp dụng rất hiệu quả trong phân tích ảnh y học bao gồm phân loại ảnh X-quang ung thư vú [7,17,20,21]. Chúng tôi áp dụng các kỹ thuật học sâu này vào nghiên cứu của chúng tôi. Bài báo gồm các mục sau đây: Mục 2 rà soát một số công trình liên quan đến các hệ tầm soát ung thư vú. Mục 3 đề xuất mô hình Trí tuệ nhân tạo hỗ trợ tầm soát ung thư vú. Mục 4 so sánh độ tin cậy của hệ phân loại ảnh X-quang ung thư vú với bác sĩ chẩn đoán hình ảnh. Mục 5 trình bày thí dụ về hoạt động của hệ hỗ trợ chẩn đoán X-quang ung thư vú bằng Trí tuệ nhân tạo. Cuối cùng là kết luận và công việc tiếp theo.

2. Một số công trình liên quan

Trong những năm gần đây, một số nhóm nghiên cứu đã áp dụng các phương pháp học sâu, nhất là áp dụng các mô hình mạng nơ ron tích chập để xây dựng hệ hỗ trợ chẩn đoán ung thư vú dựa trên ảnh X-quang vú. Bài báo [3] đã đề xuất mô hình mạng nơ ron dư sâu nhiều chiều (ResNet) để phân loại tự động ảnh ung thư vú vào các lớp u ác tính hoặc bình thường/u lành tính. Đặc biệt cách tiếp cận này bao gồm tập hợp của các mạng dư sâu (ResNet) với 6 ảnh đầu vào là ảnh chụp chụp thẳng (CC Image), ảnh cửa sổ chứa u chụp thẳng (Mass Mask of CC Image), ảnh cửa sổ chứa canxi chụp thẳng (Calcification Mask of CC Image), ảnh chụp xiên chéo (MLO Image), ảnh cửa sổ chứa u chụp xiên chéo (Mass Mask of MLO Image), ảnh cửa sổ chứa canxi chụp xiên chéo (Calcification Mask of MLO Image). Mô hình này sẽ sinh tự động cửa sổ phân đoạn

u hoặc vi canxi. Kết quả là hệ thống mRestNet phân loại ảnh ung thư vú vào một trong hai lớp “u ác tính” hoặc “bình thường/u lành tính” sử dụng 410 ảnh trong tập dữ liệu INBreast [19]. Trong một nghiên cứu khác [16], các tác giả sử dụng mạng nơ ron tích chập (CNN) để phân loại các u vú trước phân đoạn như các u lành tính hay ác tính bằng cách kết hợp học chuyển tiếp với tiền xử lý ảnh và tăng sự đa dạng của tập dữ liệu để đáp ứng việc học mô hình với dữ liệu hạn chế của tập dữ liệu DDSM [14]. Trong nghiên cứu này, các tác giả chỉ xem xét các ảnh ung thư vú có chứa các u, bao gồm 1.829 ảnh từ 997 bệnh nhân. Các dữ liệu ảnh được chia theo tỷ lệ các tập huấn luyện (80%), tập hiệu chỉnh mô hình (10%), tập kiểm tra (10%). Kết quả nghiên cứu cho thấy mô hình đề xuất cho kết quả tỷ lệ chẩn đoán bệnh dương tính đúng trên tổng chẩn đoán bệnh dương tính đúng cộng với chẩn đoán âm tính sai (gọi là recall hay độ nhạy cảm) là 0.934, trong khi kết quả chẩn đoán của bác sĩ chẩn đoán hình ảnh trong khoảng [0.745 – 0.923]. Trong một nghiên cứu khác [23], Li Shen đã xây dựng một thuật toán chẩn đoán ung thư vú dựa trên toàn ảnh mammogram. Cách tiếp cận này cho kết quả độ tin cậy của mô hình $AUC = 0,88$ với cơ sở dữ liệu ảnh DDSM và cho kết quả $AUC = 0,91$ với cơ sở dữ liệu ảnh INbreast. Trong một nghiên cứu mới đây [26], các tác giả đề xuất cách tiếp cận để phân loại và khoanh vùng canxi và các khối u trong ảnh ung thư vú. Các tác giả áp dụng các mạng nơ ron tích chập sâu để học đặc trưng một cách tự động và xây dựng bộ phân loại. Ở phương pháp này, bộ phân loại được huấn luyện trên các “mảnh cắt” (patches) vùng tổn thương được gán nhãn vùng canxi và các khối u trong ảnh ung thư vú, sau đó thích nghi

các “mảnh cắt” này trên ảnh mammogram đầy đủ để khoanh vùng bất thường. Thực nghiệm chỉ ra rằng, áp dụng mô hình VGGNet cho kết quả độ tin cậy đến 92,53%. Một nghiên cứu mới đây của các nhà khoa học trường Đại học New York, Hoa kỳ [6] về xây dựng hệ hỗ trợ tầm soát ung thư vú áp dụng mạng nơ ron tích chập sâu trên ảnh đa hướng nhằm phân loại các lớp ảnh theo phân loại quốc tế BIRADS thành 3 lớp BI-RADS 0 (không đầy đủ), BI-RADS 1 (bình thường) và BI-RADS 2 (xét nghiệm lành tính). Nhóm nghiên cứu áp dụng mạng nơ ron tích chập sâu huấn luyện trên 886.000 ảnh X-quang vú, mỗi bệnh nhân được chụp bốn ảnh ở các tư thế chụp L-CC (tư thế thẳng trên-dưới trái), L-MLO (tư thế chệch trong-ngoài trái), R-CC (tư thế thẳng trên-dưới phải), R-MLO (tư thế chệch trong-ngoài phải). Kết quả cho thấy kết quả của hệ thống MV-DCN đạt $macAUC = 0.688$ thấp hơn so với kết quả của bốn bác sĩ chẩn đoán hình ảnh X-quang ung thư vú có kinh nghiệm của trường Đại học New York, Hoa kỳ với $macAUC = 0,704$. Thí nghiệm chỉ ra rằng, khi các bác sĩ chẩn đoán hình ảnh X-quang ung thư vú tham khảo kết quả của hệ thống MV-DCN trước, sau đó mới đọc hình ảnh X-quang ung thư vú, thì kết quả cuối cùng của bác sĩ chẩn đoán hình ảnh và hệ thống MV-DCN nhận được là $macAUC = 0.735$. Trong một kết quả mới đây [28], khi số lượng ảnh dùng để huấn luyện mô hình đạt hơn một triệu ảnh thì kết quả của hệ thống MV-DCN đạt độ tin cậy $macAUC = 0.895$.

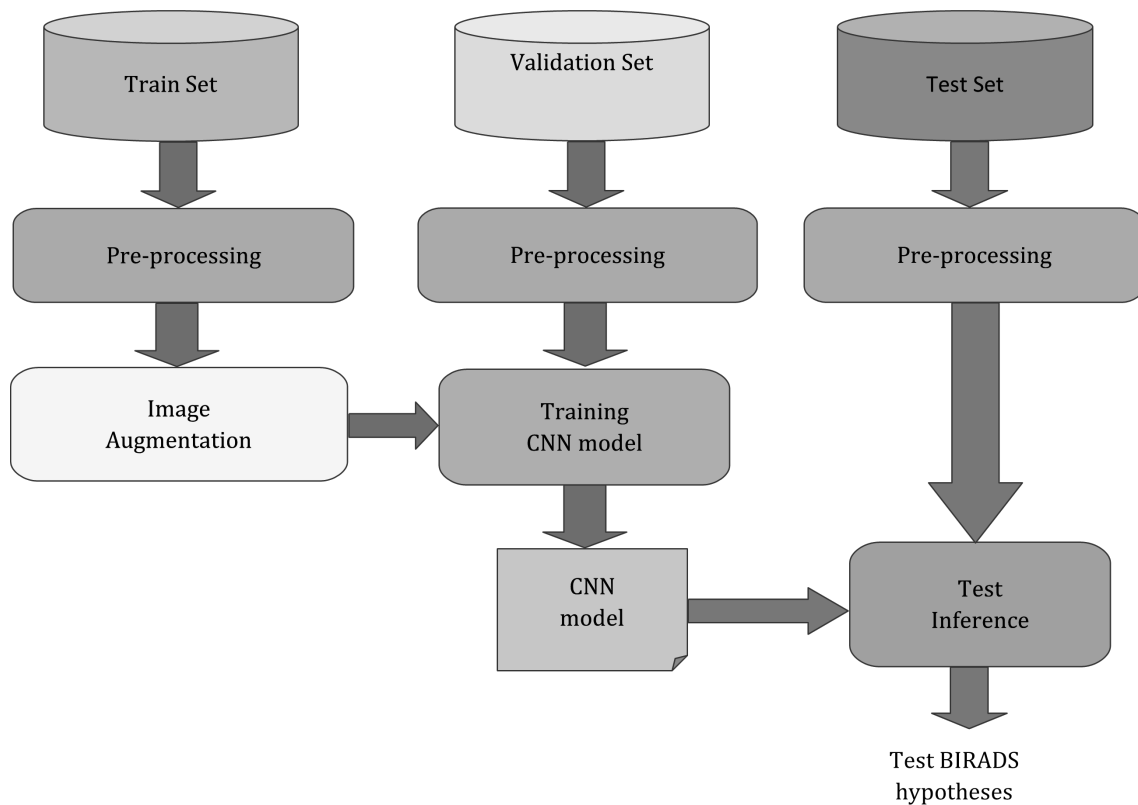
3. Mô hình Trí tuệ nhân tạo hỗ trợ tầm soát ung thư vú

3.1. Thiết kế mô hình

Sơ đồ tổng quát của hệ thống phân loại BI-RADS được mô tả ở Hình 1. Đầu tiên dữ liệu

ảnh trong file DICOM được chuyển về định dạng PNG 16 bits. Sau đó các ảnh này được thực hiện tiền xử lý. Tập dữ liệu huấn luyện (training set) được áp dụng một số kỹ thuật làm tăng cường dữ liệu ảnh, các dữ liệu mới này được sử dụng trong quá trình huấn luyện các mô hình mạng nơ ron tích chập training CNN. Tập điều chỉnh các

tham số (Validation set) được sử dụng để kiểm soát điều kiện dừng của quá trình huấn luyện mô hình. Mô hình mạng nơ ron tích chập CNN thu được sẽ được sử dụng để thử nghiệm với các ảnh trong tập dữ liệu kiểm tra. Kết quả thu được sẽ là xác suất của mỗi BI-RADS trên từng ảnh trong tập kiểm tra (TEST set).



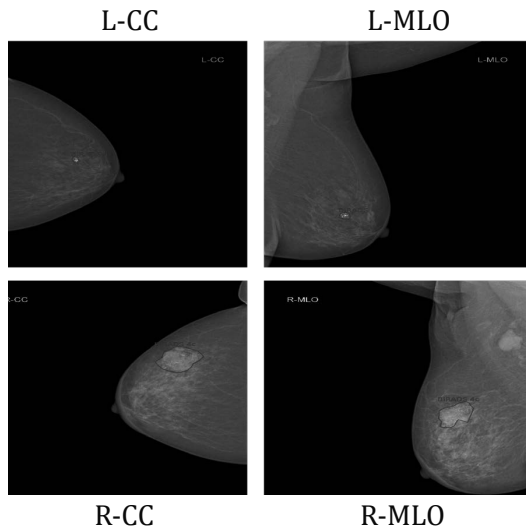
Hình 1. Sơ đồ kiến trúc tổng quát của hệ thống phân loại BIRADS

3.2. Lựa chọn dữ liệu và gán nhãn

Trong nghiên cứu này chúng tôi sử dụng tập dữ liệu ảnh X-quang vú của bệnh nhân Việt Nam được lựa chọn và thu thập bởi các bác sĩ chẩn đoán hình ảnh của Bệnh viện Đại học Y Hà Nội HMU-TL [20]. Mỗi bệnh nhân được chụp X-quang vú tư thế thẳng trên-dưới (CC), tư thế chếch trong-ngoài (MLO) ở vú trái (L) và vú phải

(R), thông thường nhận được bốn ảnh cho mỗi bệnh nhân: L-CC (tư thế thẳng trên-dưới trái), L-MLO (tư thế chếch trong-ngoài trái), R-CC (tư thế thẳng trên-dưới phải), R-MLO (tư thế chếch trong-ngoài phải). (Có thể có bệnh nhân không chụp đủ bốn ảnh).

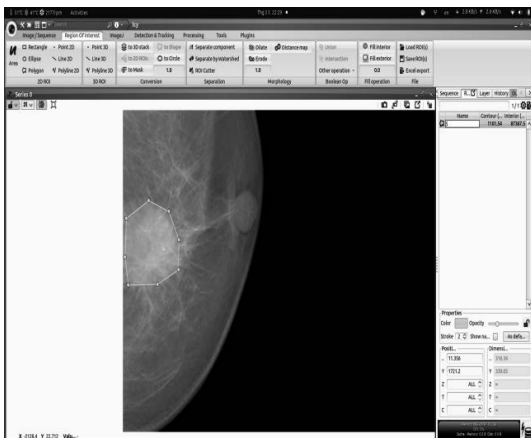
Thí dụ minh họa ảnh X-quang vú của một bệnh nhân, L-CC, L-MLO, R-CC, R-MLO ở Hình 2:



Hình 2. Thí dụ ảnh X-quang vú của 1 bệnh nhân

Với mỗi ảnh mammogram, các bác sĩ của Bệnh viện Đại học Y Hà Nội sẽ đưa ra 3 lựa chọn: thứ nhất là xác định đây là ảnh bình thường (ứng với mức BI-RADS 1), thứ hai là nếu có bất thường thì khoanh vùng khối u (có thể có nhiều khối u trong một ảnh mammogram) và chọn mức BI-RADS cho khối u (từ BI-RADS 2 đến BI-RADS 5), thứ ba là nếu ảnh không thể xác định rõ ràng thuộc BI-RADS nào thì sẽ xếp vào loại BI-RADS 0.

Thí dụ về việc khoanh vùng tổn thương và xác định ảnh thuộc BI-RADS nào được sử dụng phần mềm ICY dùng cho phân tích ảnh y - sinh [9]:



Hình 3. Khoanh vùng và gán nhãn vùng khối u

Cuối cùng, tệp có đuôi xml sẽ được tạo ra cùng với tên của tệp ảnh. Tệp này lưu các thông tin của từng điểm ảnh khoanh vùng khối u và thông tin về nhãn của ảnh.

Việc phân loại ảnh mammogram vào các BI-RADS do thực hiện bởi hai bác sĩ chẩn đoán hình ảnh X-quang ung thư vú với 10 năm kinh nghiệm. Nếu kết quả phân loại khác nhau thì bác sĩ chẩn đoán hình ảnh thứ ba với hơn 20 năm kinh nghiệm sẽ đọc ảnh X-quang ung thư vú này và sẽ đưa ra quyết định phân loại ảnh BI-RADS này.

Trong nghiên cứu của chúng tôi, các bác sĩ chẩn đoán hình ảnh X-quang vú của Bệnh viện Đại học Y Hà Nội cần phân loại các ảnh X-quang ung thư vú làm ba loại: 1) Bệnh nhân bình thường, không có tổn thương vú: kết thúc thăm khám. Ảnh mammogram của bệnh nhân được gán nhãn BI-RADS 1. 2) Bệnh nhân có u lành tính: cần tiếp tục kiểm tra theo định kỳ. Ảnh mammogram của bệnh nhân được gán nhãn BI-RADS 2 hoặc BI-RADS 3). 3) Bệnh nhân nghi ngờ u ác tính cao hoặc không đánh giá được tổn thương vú, cần kiểm tra và làm xét nghiệm thêm. Ảnh mammogram của bệnh nhân được gán nhãn BI-RADS 045 (nghĩa là ảnh của bệnh nhân thuộc BI-RADS 0 hoặc BI-RADS 4 hoặc BI-RADS 5).

Dữ liệu mà các bác sĩ chẩn đoán hình ảnh X-quang ung thư vú được chúng tôi lựa chọn, tiền xử lý và được chia thành 3 phần: tập huấn luyện chiếm 60%, tập hiệu chỉnh tham số chiếm 20% và tập kiểm tra chiếm 20% (Bảng 1). Tuy nhiên, trong các bài toán sử dụng kỹ thuật học sâu dữ liệu thường được chia theo tỷ lệ tập huấn luyện chiếm 80%, tập hiệu chỉnh tham số chiếm 10%

và tập kiểm tra chiếm 10%. Vì thực tế, ảnh BI-RADS 4 và BI-RADS 5 thu thập được ít hơn ảnh BI-RADS 0, do đó chúng tôi tăng số lượng của tập ảnh kiểm tra là 20% để có thể tăng gấp đôi số lượng ảnh BI-RADS 4 và BI-RADS 5 trong tập kiểm tra. Trong Bảng 1 có những dữ liệu không

đồng đều như số lượng ảnh L-CC không bằng với số lượng ảnh R-CC của tập huấn luyện. Điều này là do có một số bệnh nhân được yêu cầu chụp ảnh thêm một bên vú khi bác sĩ có những yêu cầu cần bổ sung thêm những ảnh đó để đánh giá thêm nguy cơ bệnh ở bên vú đó.

Bảng 1. Phân bố dữ liệu của các tập huấn luyện, tập hiệu chỉnh, tập kiểm tra của các ảnh mammograms theo BI-RADS và theo hướng chụp

Hướng chụp	Tập huấn luyện	Tập hiệu chỉnh	Tập kiểm tra	Tổng
BI-RAD 1	1925	460	444	2829
BI-RAD 1 L-CC	470	112	108	690
BI-RAD1 L-MLO	470	112	108	690
BI-RAD 1 R-CC	493	118	114	725
BI-RAD 1 R-MLO	492	118	114	724
BI-RAD 23	1407	346	374	2127
BI-RAD 23 L-CC	329	87	89	505
BI-RAD 23 L-MLO	331	87	89	507
BI-RAD 23 R-CC	374	86	98	558
BI-RAD 23 R-MLO	373	86	98	557
BI-RADS 045	2038	462	456	2956
BI-RADS 045 L-CC	523	119	114	756
BI-RADS 045 L-MLO	523	119	114	756
BI-RADS 045 R-CC	496	112	114	722
BI-RADS 045 R-MLO	496	112	114	722
BI-RADS 123045	5370	1268	1274	7912

3.3 Tiền xử lý và tăng cường dữ liệu ảnh

Khối tiền xử lý dữ liệu ảnh được mô tả ở Hình 4. Đầu tiên các ảnh R-CC và ảnh R-MLO được lật theo trục tung sao cho các ảnh vú của bệnh nhân luôn luôn ở cùng một mặt của ảnh. Sau đó, vùng ảnh vú được cắt (crop) bằng kỹ thuật xử lý ảnh đơn thuần: chúng tôi thực hiện xác định biên bằng cách bám theo đường viền trong ảnh. Sau đó ảnh này được rút gọn (resize) về kích thước

512 x 512 pixels. Cuối cùng từng ảnh được chuẩn hóa theo công thức 1.

$$X_{ij}norm = \frac{X_{ij} - \mu}{\delta} \dots\dots\dots (1)$$

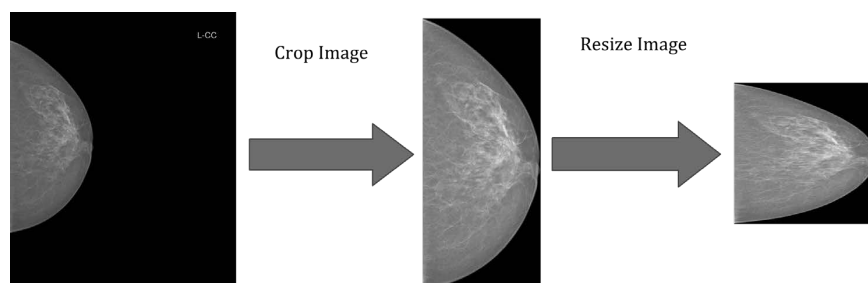
Trong công thức 1: Xij là một điểm ảnh trong ảnh, μ và δ là trung bình và độ lệch chuẩn được tính trên tất cả các pixels trong ảnh này.

Quá trình đa dạng hóa dữ liệu ảnh là quá trình tạo ra dữ liệu huấn luyện từ dữ liệu mà ta đang có được thực hiện bằng các phép biến đổi

hình học: chuyển đổi kích thước, dịch chuyển, xoay ảnh ... Hình 4 mô tả các ảnh thu được trong quá trình này.

Trong qua trình huấn luyện, đầu tiên các ảnh đầu với kích thước 512x512 pixels được biến đổi thành các ảnh mới bằng kỹ thuật biến đổi sau đây:

- **Scaling:** chuyển đổi kích thước các ảnh từ 90%-110% của kích thước của chúng, xoay theo trục. Giá trị của việc chuyển đổi kích thước các ảnh được lựa chọn ngẫu nhiên từ 90%-110%.
- **Translating:** dịch chuyển ảnh từ -10 đến +10 phần trăm theo trục. Điều đó nghĩa là các ảnh được dịch chuyển từ trái (giá trị âm), sang phải (giá trị dương), từ trên (giá trị âm), xuống dưới (giá trị dương). Tỷ lệ dịch chuyển này được chọn ngẫu nhiên trong khoảng [0,10%].
- **Rotating:** xoay ảnh từ -5 đến +5 độ. Điều đó nghĩa là các ảnh được xoay ngẫu nhiên từ 0 đến 5 độ. (giá trị âm: xoay sang trái, giá trị dương: xoay sang phải).



Hình 4. Thí dụ về tiền xử lý ảnh

Tất cả các bước biến đổi được thực hiện trên các ảnh vào để tạo ra các ảnh mới, sau đó các ảnh này được dùng trong quá trình huấn luyện ảnh. Quá trình này được thực hiện như sau:

Kiểm tra lần lượt từng ảnh X trong tập huấn luyện:

- Đưa ảnh X từ tập các ảnh vào bộ nhớ RAM của máy chủ.
- Biến đổi ảnh X vào ảnh X'. Các ảnh X' được lưu trong bộ nhớ RAM chứ không được lưu trong các ổ đĩa.
- Sử dụng các ảnh X' như dữ liệu vào của mô hình mạng nơ ron nhân tạo bằng thuật toán lan truyền ngược.

Chúng tôi muốn nhấn mạnh là việc đa dạng

hóa dữ liệu ảnh bằng cách thay thế hoặc biến đổi dữ liệu gốc được lưu vào bộ nhớ RAM của máy chủ. Việc này dẫn đến kết quả là mỗi epoch (epoch ở đây nghĩa là huấn luyện mạng nơ ron với tất cả dữ liệu huấn luyện cho một chu trình) sẽ có một tập dữ liệu huấn luyện mới. Cuối cùng, toàn bộ kích thước tập huấn luyện sau khi đa dạng hóa dữ liệu ảnh là số epoch nhân với kích thước tập dữ liệu gốc.

3.4. Huấn luyện mô hình phân loại BI-RADS

Kiến trúc mô hình mạng nơ ron tích chập cho phân loại BI-RADS được mô tả ở hình 5. Trong kiến trúc này, chúng ta thêm các tầng của mạng nơ ron tích chập với 32 bản đồ đặc trưng tiếp theo mô hình ResNet 50 để giảm tập đặc trưng được tạo ra từ các mô hình trước. Sau đó các

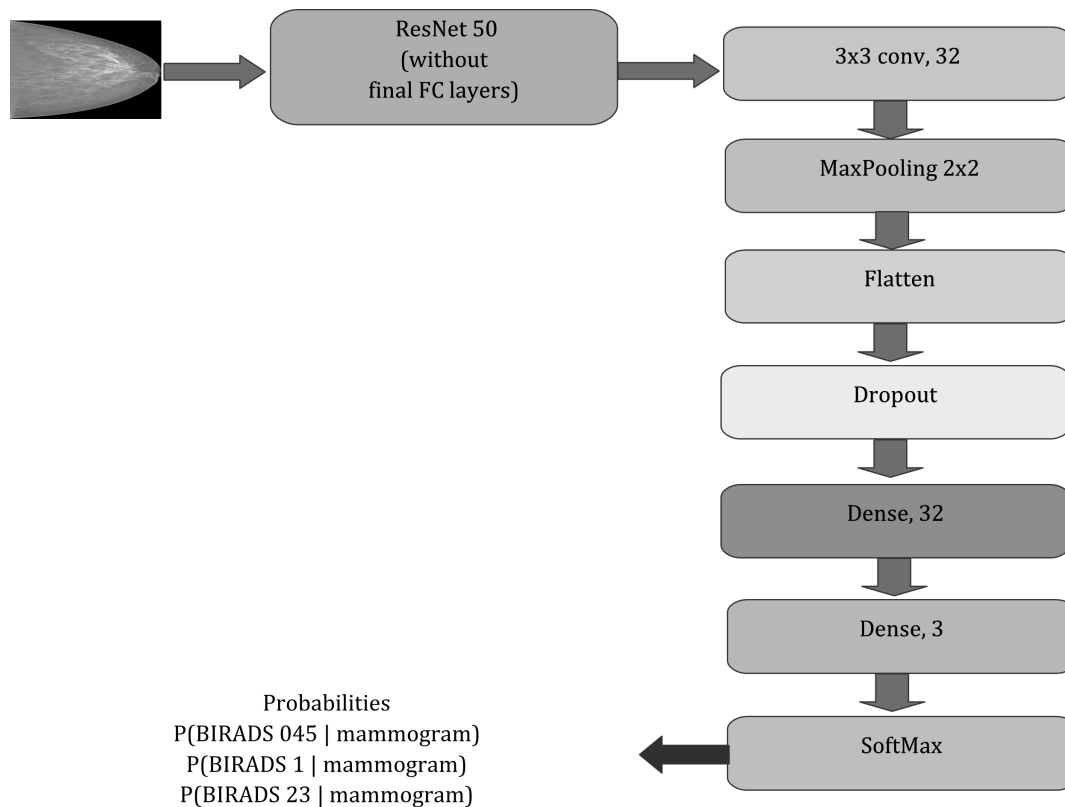
tầng Dense layers (hay Fully connected layers) cùng với tầng SoftMax layer được đưa vào mô hình dùng cho phân loại BI-RADS. Mạng nơ ron nhân tạo bao gồm 24 triệu tham số (backbone ResNet50).

Mỗi ảnh đầu vào được biểu diễn bởi tensor với số chiều là $512 \times 512 \times 3$. Kết quả đầu ra của tầng của mô hình ResNet50 được biểu diễn bởi tensor với số chiều là $16 \times 16 \times 2048$. Đầu ra của tầng Flatten là vector 2048 chiều ($8 \times 8 \times 32$). Kết quả đầu ra của mô hình là đầu ra của tầng SoftMax là vector ba chiều thể hiện xác suất của các BIRAD 045, BIRAD 1 và BIRAD 23 với một ảnh X-quang ung thư vú được đưa vào để mô hình phân loại.

Quá trình huấn luyện mô hình ResNet 50 (đã được huấn luyện trước trên tập ảnh ImageNet) được thực hiện trên máy chủ với các cấu hình sau đây đặt tại trường Đại học Thăng Long:

- CPU: Core i9-7900X (3.30 GHz)
- RAM: 128 GB
- 2 Hard disks: one SSD 500GB and one HDD 4TB
- 2 GPU NVIDIA RTX2080TI (11GB).

Máy chủ server được cài đặt hệ điều hành OS Linux Ubuntu 16.04.3 LTS với các thư viện hỗ trợ các thuật toán học sâu trên máy tính chuyên dụng GPU gồm Nvidia CUDA 8.0.61 và NVIDIA CUDA 6.0.21. Hệ thống học sâu được xây dựng trên Keras nền tảng Tensorflow back-end.



Hình 5. Kiến trúc mô hình mạng nơ ron tích chập cho phân loại BI-RADS (trên nền tảng ResNet50)

Chúng ta áp dụng các phương pháp Early Stopping và Reduce Learning Rate trên các kỹ thuật Plateau trong quá trình huấn luyện với các tham số sau đây:

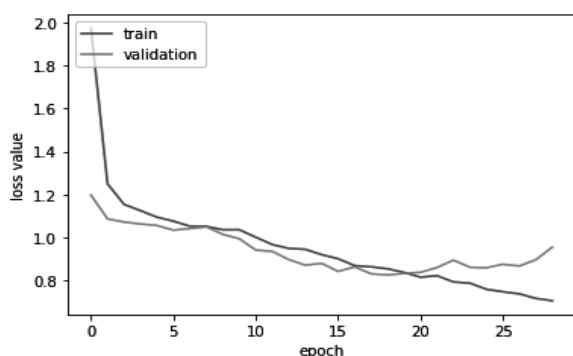
- Số epochs chịu đựng cho for Early Stopping :10.

(Đây là tham số của chức năng của Early Stopping của Tensorflow, là số epoch phải đợi trước khi dừng sớm nếu không có tiến bộ trên tập hiệu chỉnh tham số. Thông thường tham số này nằm trong khoảng từ 10 đến 100 và nhưng thường phụ thuộc vào tập dữ liệu và mô hình).

- Khởi đầu learning rate: 10^{-5} .

- Learning rate tối thiểu: 10^{-8} . (Giới hạn thấp nhất của learning rate)

(Tham số của chức năng ReduceLROnPlateau của Tensorflow).



Hình 6. Đồ thị giá trị mất mát của tập huấn luyện và tập điều chỉnh tham số tương ứng với các epochs trong quá trình huấn luyện mô hình

Hình 6 mô tả đồ thị biến thiên tập huấn luyện và tập điều chỉnh tham số trong quá trình huấn luyện mô hình. Qua đồ thị ta nhận thấy rằng, quá trình huấn luyện đạt giá trị tối ưu tại epoch 19, sau đó, vấn đề quá vừa dữ liệu (overfitting) bắt đầu. Cụ thể, giá trị mất mát của tập huấn luyện

(training set) tiếp tục tăng nhưng giá trị mất mát của tập hiệu chỉnh tham số (validation set) giảm. Do đó, mô hình đạt được tại epoch 19 được coi là tối ưu và được sử dụng cho việc thử nghiệm trên tập kiểm tra.

3.5. Kết quả thí nghiệm phân loại BI-RADS cho ảnh X-quang vú bệnh nhân Việt Nam

Trong nghiên cứu của chúng tôi, một số tiêu chí sau được áp dụng để đánh giá Hệ thống thông minh hỗ trợ chẩn đoán ung thư vú. Đối với chẩn đoán ung thư nói chung, các mẫu dương tính là những mẫu thuộc nhóm ác tính/bất thường và các mẫu âm tính là những mẫu thuộc nhóm lành tính/bình thường.

Các độ đo đánh giá chính và được sử dụng nhiều nhất áp dụng cho các hệ thống thông minh bao gồm độ tin cậy Acc (Accuracy), độ nhạy cảm Sn (Sensitivity), độ đặc hiệu Sp (Specificity) và AUC (Area Under Curve) – là đường cong biểu diễn xác suất và biểu diễn mức độ phân loại của mô hình [21]. Đây là các biện pháp hiệu quả được sử dụng nhiều nhất, được xác định như sau:

(1) Acc (Độ chính xác) biểu thị mức độ gần của lớp dự đoán so với lớp thực tế. Điều đó nói rằng chỉ số này chỉ ra tỉ lệ phần trăm các mẫu được phân loại đúng (bình thường và bất thường) trong tổng số mẫu.

(2) Sn (Độ nhạy) là tỷ lệ dương thực sự xác định tỷ lệ phần trăm của các mẫu bất thường được phân loại chính xác.

(3) Sp (Độ đặc hiệu) là tỷ lệ âm thực xác định tỷ lệ phần trăm của các mẫu bình thường được phân loại chính xác.

(4) AUC là một phương pháp đo phổ biến đại diện cho cách chọn mô hình tối ưu và bỏ qua các

mô hình tối ưu phụ. Đó là khu vực dưới đường cong ROC, là đường cong của tỷ lệ dương thực sự so với tỷ lệ dương tính giả. AUC nhận giá trị từ 0 đến 1. Một xét nghiệm chẩn đoán tốt có được khi AUC gần với 1. Các xét nghiệm có thể kiểm tra có AUC lớn hơn hoặc bằng 0,5 và nhỏ hơn 1.

Các công thức của Acc, Sn và Sp được đưa ra như sau:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sn = \frac{TP}{TP + FN}$$

$$Sp = \frac{TN}{TN + FP}$$

trong đó TP: dương tính thật, TN: âm tính thật, FP: dương tính giả, FN: âm tính giả.

Để lựa chọn độ đo đánh giá khi có hai lớp, thông thường ta chọn AUC là vùng dưới đường cong ROC (Receiver Operator Characteristic). ROC là độ đo đánh giá cho bài toán phân loại nhị phân. Tuy nhiên, trong bài toán của chúng ta, có ba lớp cần phân loại, do đó chúng ta không thể áp dụng trực tiếp độ đo này. Thay vào đó, chúng ta tính ba độ đo AUC, mỗi lần chúng ta xét một trong ba lớp như lớp dương và hai lớp còn lại là lớp âm. Chúng ta sử dụng trung bình macro của ba độ đo AUC, được ký hiệu là macAUC như độ đo chính cho hoạt động của hệ thống phân loại BI-RADS trong nghiên cứu của chúng tôi.

Chúng tôi tính AUC lần lượt từng trường hợp như sau:

Trường hợp 1. Tính BI-RAD 045 đối với các BI-RADS còn lại.

Với mỗi mẫu ảnh trong tập kiểm tra (1.274 mẫu ảnh), hệ thống trả về 3 xác suất: $P(X | \text{BI-RAD } 045)$, $P(X | \text{BI-RAD } 1)$, $P(X | \text{BI-RAD } 23)$. Tổng ba xác suất này bằng 1. Khi đó

chuyển đổi về đánh giá phân tách 2 lớp với xác suất $P(X | \text{BI-RAD } 045)$, $P(X | \text{các BI-RADS còn lại}) = P(X | \text{BI-RAD } 1) + P(X | \text{BI-RAD } 23)$.

(Tính False Positive Rate (FPR), True Positive Rate (TPR) theo từng ngưỡng thresholds: `fpr, tpr, thresholds = metrics.roc_curve(new_labels, new_probs, pos_label=0)`). Tính AUC dựa trên hai dãy số FPR và TPR: `auc = metrics.auc(fpr, tpr)`).

Trường hợp 2. BI-RAD 1 đối với các BI-RADS còn lại.

Trường hợp 3. BI-RAD 23 đối với các BI-RADS còn lại.

Với hai trường hợp 2 và 3, AUC được tính tương tự giống như trường hợp 1. Kết quả MacAUC là trung bình cộng của 3 AUC trên.

Bảng 2. Kết quả đánh giá của mô hình đề xuất

Giả thiết dự báo theo BIRADS	Kết quả của mô hình
AUC của BI-RADS 045 đối với hai lớp còn lại	0.724
AUC của BI-RADS 1 đối với hai lớp còn lại	0.801
AUC của BI-RADS 23 đối với hai lớp còn lại	0.724
macAUC	0.75

Bảng 2 chỉ ra rằng trung bình cộng của ba độ đo AUC như AUC của BI-RADS 045 đối với hai lớp còn lại, AUC của BI-RADS 1 đối với hai lớp còn lại, AUC của BI-RADS 23 đối với hai lớp còn lại và được gọi là macAUC nhận giá trị 0.75.

Bảng 3. Ma trận nhầm lẫn của tập dữ liệu kiểm tra với mô hình dựa trên nền tảng ResNet50

Giả thiết dự báo theo BI-RADS	Các tham chiếu			
	BI-RADS 045	BI-RADS 1	BI-RADS 23	Tổng
BI-RADS 045	279	146	31	456
BI-RADS 1	108	331	5	444
BI-RADS 23	129	118	127	374

Bảng 3 trình bày ma trận nhầm lẫn của tập dữ liệu kiểm tra với mô hình dựa trên nền tảng ResNet50. Nhìn vào kết quả trong Bảng 3, ta thấy:

- Trong số 456 ảnh của BI-RADS 045, hệ thống nhận dạng đúng 279 ảnh, chiếm 61%, và 146 ảnh bị nhận dạng nhầm thành ảnh của BI-RADS 1 và tỷ lệ này chiếm 32%. Ngoài ra, 31 ảnh cũng bị nhận dạng nhầm thành ảnh thuộc BI-RADS 23 và tỷ lệ này chiếm 6%.

- Trong số 444 ảnh của BI-RADS 1, hệ thống nhận dạng đúng 331 ảnh, chiếm 74%, và 108 ảnh bị nhận dạng nhầm thành ảnh của BI-RADS 045 và tỷ lệ này chiếm 24%. Ngoài ra, 5 ảnh cũng bị nhận dạng nhầm thành ảnh thuộc BI-RADS 23 và tỷ lệ này chiếm 1%.

- Trong số 374 ảnh của BI-RADS 23, hệ thống nhận dạng đúng 127 ảnh, chiếm 34%, và 129 ảnh bị nhận dạng nhầm thành ảnh của BI-RADS 045 và tỷ lệ này chiếm 34%. Ngoài ra, 118 ảnh cũng bị nhận dạng nhầm thành ảnh thuộc BI-RADS 1 và tỷ lệ này chiếm 31%.

Điều này chỉ ra rằng, hệ thống có khả năng phân loại tốt nhất các ảnh thuộc BI-RADS 1. Ngược lại, hệ thống phân loại ảnh thuộc BI-RADS 23 kém nhất. Trên thực tế, mẫu ảnh thuộc BI-RADS 1 dùng để huấn luyện mô hình nhiều

gấp đôi các mẫu học thuộc các lớp BI-RADS 23 và BI-RADS 045.

4. So sánh độ tin cậy của hệ phân loại ảnh X-quang ung thư vú với bác sĩ chẩn đoán hình ảnh

4.1 So sánh kết quả độ tin cậy của Hệ hỗ trợ với kết quả độ tin cậy của các bác sĩ chẩn đoán hình ảnh của của Bệnh viện Đại học Y Hà Nội.

Một trong những định nghĩa về hệ Trí tuệ nhân tạo được gọi là thông minh khi hệ thống có thể thực hiện các chức năng của mình bằng hoặc tốt hơn chuyên gia người có thể thực hiện. Ngoài ra, để hiểu rõ hơn những khả năng và hạn chế của hệ thống, chúng tôi tiến hành so sánh độ tin cậy của hệ trong việc phân loại ảnh vào các BI-RADS 045, BI-RADS 1 và BI-RADS 23 với các bác sĩ chẩn đoán hình ảnh X-quang ung thư vú của Khoa chẩn đoán hình ảnh, Bệnh viện Đại học Y Hà Nội.

(1) Các độ đo được áp dụng để đánh giá: Chúng tôi đánh giá theo độ đo của độ tin cậy Accuracy, độ nhạy cảm Sn (Sensitivity) và độ đặc hiệu Sp (Specificity). Trong đó, độ nhạy cảm Sn (Sensitivity) và độ đặc hiệu Sp (Specificity) cũng là độ đo mà chương trình trọng điểm quốc gia KC 4.0 của Bộ Khoa học và Công nghệ đưa ra đối với các sản phẩm tương tự của chương trình. Đối với bài toán tương tự cho phân tích ảnh X-quang ung thư vú, yêu cầu đạt được của độ nhạy cảm Sn (Sensitivity) và độ đặc hiệu Sp (Specificity) là 0.7 [13].

(2) Các ảnh Mammograms X-quang ung thư vú theo chuẩn DICOM được phân loại vào: 3 lớp BI-RADS 045, BI-RADS 1, BI-RADS 23.

(3) Nguyên tắc đánh giá: Hệ hỗ trợ được coi

là có ích, hỗ trợ cho bác sĩ khi độ tin cậy chẩn đoán của hệ so với tiêu chuẩn vàng bằng và cao hơn độ tin cậy chẩn đoán của bác sĩ (hoặc hội đồng bác sĩ) chẩn đoán hình ảnh X-quang ung thư vú.

(4) Ảnh được coi là tiêu chuẩn vàng dùng để đánh giá: Là các ảnh X-quang vú chuẩn DICOM do các bác sĩ chẩn đoán hình ảnh, Bệnh viện Đại học Y Hà Nội cung cấp. Các ảnh này không dùng để huấn luyện mô hình phân loại ảnh. Các ảnh này là các ảnh hồi cứu của bệnh án mà kết luận gán nhãn ảnh dựa trên kết luận đọc ảnh của các bác sĩ X-quang ung thư vú, kết hợp với chẩn đoán xác định trong bệnh án, và có thể kết hợp với kết quả sinh thiết vùng tổn thương vú của bệnh nhân.

(5) Các bước xây dựng ảnh “tiêu chuẩn vàng” và kết quả dán nhãn của các bác sĩ chẩn đoán hình ảnh: Các bác sĩ chẩn đoán hình ảnh Bệnh viện Đại học Y Hà Nội đã cung cấp cho nhóm CNTT trường Đại học Thăng Long 7.912 ảnh X-quang ung thư vú theo chuẩn DICOM theo phân loại quốc tế BI-RADS 0, BI-RADS 1, BI-RADS 2, BI-RADS 3, BI-RADS 4, BI-RADS 5. Để huấn luyện mô hình, chúng tôi nhóm các BI-RADS trên vào ba lớp BI-RADS 045 (gồm các ảnh thuộc BI-RADS 0, BI-RADS 4, BI-RADS5), BI-RADS 1, BI-RADS 23 (gồm các ảnh thuộc BI-RADS 2, BI-RADS 3), chúng tôi chia tập dữ liệu trên theo tỷ lệ 80 (tập huấn luyện) : 20 (tập hiệu chỉnh tham số) : 20 (tập kiểm tra). Chúng tôi có 1.570 ảnh tập trong tập kiểm tra để đánh giá hệ theo các độ đo AUC, macAUC. Ngoài ra, chúng tôi lấy ngẫu nhiên 500 ảnh trong tập kiểm tra dùng làm “chuẩn vàng” làm tiêu chuẩn để đánh giá theo các độ đo độ tin cậy Accuracy, độ nhạy cảm Sensitivity và độ đặc hiệu Specificity. Tuy nhiên, khi gửi 500 ảnh cho bác sĩ chẩn đoán hình ảnh của Bệnh viện Đại học

Y Hà Nội để đọc ảnh, thì có 32 ảnh các bác sĩ chẩn đoán hình ảnh thông báo máy không đọc được (có thể do lỗi máy đọc hoặc thao tác vì ở máy của chúng tôi vẫn đọc được), do đó chỉ có 468 ảnh được các bác sĩ chẩn đoán hình ảnh dùng làm “chuẩn vàng” dùng trong nghiên cứu đánh giá này.

- Ảnh “tiêu chuẩn vàng”: là 468 ảnh chọn lựa ngẫu nhiên trong tập kiểm tra được các bác sĩ chẩn đoán hình ảnh, Bệnh viện Đại học Y Hà Nội cung cấp, gọi là tiêu chuẩn vàng.

- Kết quả so sánh với một bác sĩ chẩn đoán hình ảnh: là kết quả gán nhãn của 1 bác sĩ chẩn đoán hình ảnh X-quang ung thư vú Bệnh viện Đại học Y Hà Nội đọc và gán nhãn cho một số ảnh trong số 468 ảnh. Có nhiều bác sĩ tham gia đọc ảnh, nhưng mỗi ảnh chỉ do 1 bác sĩ đọc.

- Kết quả so sánh với “hội đồng” bác sĩ chẩn đoán hình ảnh: là kết quả của hai bác sĩ chẩn đoán hình ảnh X-quang ung thư vú Bệnh viện Đại học Y Hà Nội có nhiều kinh nghiệm đọc và gán nhãn cho 468 ảnh theo cách sau:

-Nếu nhãn ảnh X-quang ung thư vú trùng với nhãn của ảnh của “hội đồng” bác sĩ chẩn đoán hình ảnh thì nhãn của ảnh này được xếp vào nhãn của Hội đồng bác sĩ.

-Nếu nhãn ảnh của của Hội đồng bác sĩ và nhãn của một bác sĩ chẩn đoán hình ảnh không trùng nhau, thì lấy nhãn ảnh của của Hội đồng bác sĩ là nhãn cuối cùng của ảnh được xem xét.

Bảng 4 thể hiện độ đánh giá độ tin cậy Accuracy, độ nhạy cảm Sensitivity và độ đặc hiệu Specificity của bác sĩ chẩn đoán hình ảnh và của hệ hỗ trợ chẩn đoán ung thư vú so với “tiêu chuẩn vàng”.

Bảng 4. Kết quả tính độ tin cậy theo accuracy, sensitivity, specificity

Giả thiết dự báo theo BIRADS	Một bác sĩ	"Hội đồng" bác sĩ	Hệ hỗ trợ
Độ nhạy cảm dự báo BI-RADS 045	0.168605	0.151163	0.604651
Độ đặc hiệu dự báo BI-RADS 045	0.942568	0.983108	0.736486
Độ nhạy cảm dự báo BI-RADS 1	0.781818	0.927273	0.830303
Độ đặc hiệu dự báo BI-RADS 1	0.475248	0.544554	0.544554
Độ nhạy cảm dự báo BI-RADS 23	0.625954	0.887240	0.312977
Độ đặc hiệu dự báo BI-RADS 23	0.845697	0.887240	0.970326
Độ tin cậy	0.512821	0.613248	0.602564
Độ nhạy cảm dự báo trung bình	0.525459	0.634288	0.582644
Độ đặc hiệu dự báo trung bình	0.754504	0.804968	0.794460

Nhận xét: Độ tin cậy của các phép đo Độ tin cậy (Accuracy), Độ nhạy cảm (Sensitivity) và Độ đặc hiệu (Specificity) của hệ hỗ trợ cao hơn hẳn độ tin cậy do 1 bác sĩ chẩn đoán hình ảnh ngẫu nhiên đọc ảnh và xấp xỉ độ tin cậy của Hội đồng bác sĩ trong đó có 1 bác sĩ nhiều kinh nghiệm.

Như vậy hệ hỗ trợ có độ tin cậy bằng hoặc cao hơn chuyên gia người, là bác sĩ chẩn đoán hình ảnh X-quang ung thư vú, nên hệ hoàn toàn có thể coi như "bác sĩ chẩn đoán hình ảnh" thứ hai hỗ trợ bác sĩ khi cần thiết trong việc đọc ảnh X-quang ung thư vú tại Bệnh viện Đại học Hà Nội và các bệnh viện khác ở Việt Nam.

4.2. So sánh kết quả độ tin cậy của Hệ hỗ trợ với kết quả độ tin cậy của các bác sĩ chẩn đoán hình ảnh của Đại học New York [6].

Các bác sĩ chẩn đoán hình ảnh của Đại học New York lấy ngẫu nhiên 500 bệnh nhân (mỗi bệnh nhân gồm 4 ảnh ở các tư thế chụp L-CC (tư thế thẳng trên-dưới trái), L-MLO (tư thế chéo trong-ngoài trái), R-CC (tư thế thẳng trên-dưới phải), R-MLO (tư thế chéo trong-ngoài phải).

Ảnh từ tập kiểm tra làm "chuẩn vàng" để đánh giá độ tin cậy chẩn đoán của mình. Bốn chuyên gia là các bác sĩ có kinh nghiệm về chẩn đoán hình ảnh X-quang ung thư vú của Đại học New York cùng đọc các ảnh X-quang ung thư vú của bệnh nhân và hội chẩn để thống nhất kết quả cuối cùng phân loại ảnh vào 3 lớp BI-RADS 0, BI-RADS 1, BI-RADS 2. Kết quả độ đo độ tin cậy của "hội đồng bác sĩ chẩn đoán hình ảnh" đạt $macAUC = 0.704$.

Trong khi đó, hệ hỗ trợ của chúng ta phân loại ảnh vào 3 lớp BI-RADS 045, BI-RADS 1, BI-RADS 23 với 1.274 ảnh trong tập kiểm tra đạt độ tin cậy $macAUC = 0,75$. (Lưu ý là cách tiếp cận của Đại học New York trong [6] là bác sĩ đọc ảnh X-quang ung thư vú để chẩn đoán theo từng bệnh nhân (tổng hợp của 4 ảnh của 1 bệnh nhân), còn hệ của chúng ta xây dựng chẩn đoán phân loại theo từng ảnh).

Nhận xét:

- Qua kết quả trên, chúng ta thấy để phân loại BIRADS thông qua đọc ảnh X-quang ung

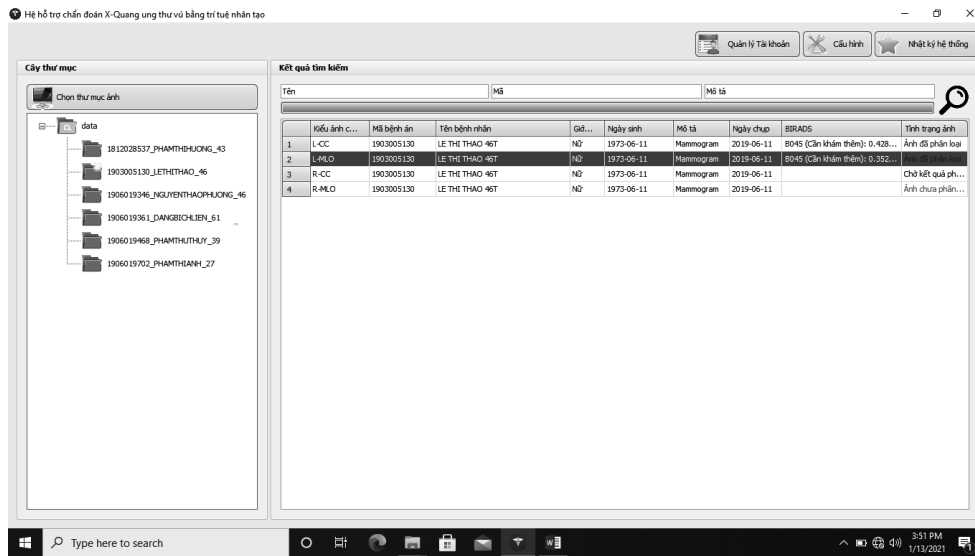
thư vú khi không có thông tin về bệnh nhân là một nhiệm vụ rất khó khăn thậm chí đối với các chuyên gia chẩn đoán hình ảnh có kinh nghiệm của Đại học New York, Hoa kỳ.

- Kết quả đánh giá độ tin cậy theo AUC, hệ hỗ trợ chẩn đoán ung thư vú của chúng ta cho độ tin cậy khi đánh giá trong tập kiểm tra đạt $macAUC = 0.75$ khi phân loại BI-RADS 045, BI-RADS 1, BI-RADS 23 cao hơn độ tin cậy $macAUC = 0.704$ của “hội đồng các bác sĩ chẩn đoán hình ảnh” của Đại học New York khi đọc 500 ảnh X-quang ung thư vú lấy ngẫu nhiên trong tập kiểm tra để phân loại BI-RADS 0, BI-RADS 1, BI-RADS 2. Cần nhấn mạnh là việc so sánh này mang tính tương đối và để tham khảo bởi vì “hội đồng các bác sĩ chẩn đoán hình ảnh” của Đại học New York và hệ của chúng ta không chẩn đoán (phân loại) cùng một tập ảnh.

5. Thí dụ về hoạt động của hệ hỗ trợ chẩn đoán X-quang ung thư vú bằng Trí tuệ nhân tạo

Hệ hỗ trợ này bao gồm các chức năng chính sau: (1) Lựa chọn ảnh X-quang ung thư vú để phân loại; (2) Phân loại ảnh X-quang ung thư vú; (3) Báo cáo kết quả Phân loại ảnh X-quang ung thư vú.

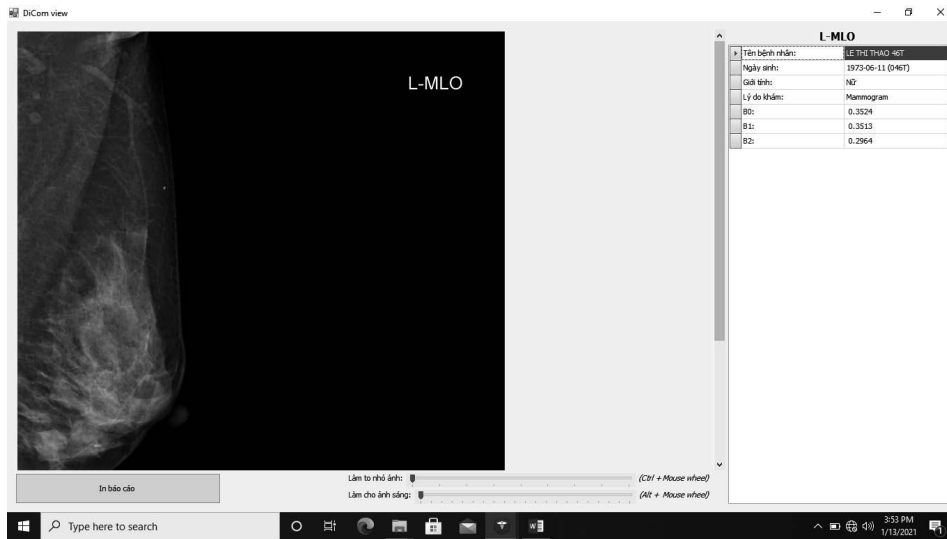
Trong hệ giao diện người dùng với bác sĩ của hệ hỗ trợ chẩn đoán X-quang ung thư vú bằng Trí tuệ nhân tạo, sau khi đăng nhập hệ thống thông qua chức năng “Tên đăng nhập”, người dùng đưa tên của mình vào và trong nút Mật khẩu, người dùng gõ mật khẩu của mình để có thể nhập vào hệ thống. Sau đó, người dùng có thể vào chức năng lựa chọn ảnh của bệnh nhân để phân loại ở Hình 7.



Hình 7. Lựa chọn ảnh X-quang ung thư vú để phân loại

Trong Hình 7, trong mục “Chọn thư mục ảnh”, bác sĩ chẩn đoán hình ảnh X-quang ung thư vú có thể chọn thư mục ảnh phù hợp để chọn ảnh cần phân loại. Ví dụ, trong hình 7, bác sĩ chẩn đoán hình ảnh X-quang ung thư vú chọn thư

mục “1903005130_Lxxxxxxx-46”, trong thư mục này chứa 4 ảnh X-quang ung thư như L-CC, L-MLO, R-CC, R-MLO. Trong hình 8, bác sĩ chẩn đoán hình ảnh đã chọn ảnh L-MLO với mã số “1903005130” của bệnh nhân Lxxxxxxx.



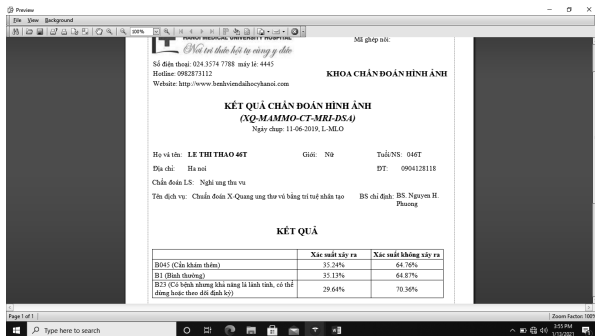
Hình 8. Kết quả phân loại ảnh X-quang ung thư vú

Trong Hình 8, kết quả phân loại ảnh L-MLO của mã bệnh án bệnh nhân “1903005130” nhận 3 giá trị xác suất phân loại BI-RADS 045 = 0.3524, BIRADS 1 = 0.3513, BI-RADS 23 = 0.2964. (Lưu ý, trong Hình 8, B0 ký hiệu cho BI-RADS 045, B1 ký hiệu cho BI-RADS 1, B2 ký hiệu cho BI-RADS 23).

ung thư vú của bệnh nhân này được minh họa tại Bảng 5.

Bảng 5. Kết quả phân loại ảnh X-quang ung thư vú

	Xác suất xảy ra	Xác suất không xảy ra
B045 (cần khám thêm)	35.24%	64.76%
B1 (bình thường)	35.13%	64.87%
B23 (có bệnh, nhiều khả năng là lành tính, có thể dừng hoặc theo dõi định kỳ)	29.64%	70.36%



Hình 9. In báo cáo kết quả chẩn đoán

Hình 9 minh họa một báo cáo kết quả của hệ hỗ trợ chẩn đoán ảnh X-quang ung thư vú. Đây là báo cáo kết quả phân loại ảnh X-quang ung thư vú của bệnh nhân Lxxxxxxx, 46 tuổi. Ảnh này được hệ thống PACS của Bệnh viện Đại học Y Hà Nội cung cấp. Kết quả phân loại ảnh X-quang

Tương ứng với báo cáo này, bệnh nhân Lxxxxxxx có xác suất cao nhất 35.24% - là xác suất ảnh thuộc BI-RADS 045 (còn xác suất ảnh thuộc BI-RADS 1 là 35.13% và xác suất ảnh thuộc BI-RADS 23 là 29.64%) nghĩa là bệnh nhân có thể thuộc một trong 3 mức phân loại ảnh

BI-RADS 0 hoặc BI-RADS 4 hoặc BI-RADS 5 nghĩa là kết luận còn không rõ ràng và cần phải tiếp tục khám và làm xét nghiệm thêm. Tất nhiên, kết quả này được tạo ra bởi hệ hỗ trợ chẩn đoán ảnh X-quang ung thư vú và chỉ mang tính tham khảo hỗ trợ cho bác sĩ khi cần. Bác sĩ chẩn đoán hình ảnh X-quang ung thư vú sẽ đưa ra kết luận cuối cùng dựa trên kết quả và những kinh nghiệm của mình.

6. Kết luận và công việc tiếp theo

Bài báo này trình bày việc xây dựng hệ hỗ trợ chẩn đoán chẩn đoán ung thư vú bằng Trí tuệ nhân tạo dựa trên ảnh X-quang ung thư vú nằm trong khuôn khổ hợp tác nghiên cứu giữa Trường Đại học Thăng Long và Bệnh viện Đại học Y Hà Nội. Đầu tiên chúng tôi đã xây dựng cơ sở dữ liệu ảnh X-quang ung thư vú, các ảnh được lựa chọn, gán nhãn theo hệ thống phân loại BI-RADS và khoanh vùng u. Sau đó các thao tác tiền xử, đa dạng dữ liệu được áp dụng trên các ảnh. Để thực hiện phân loại ảnh vào các nhóm BI-RADS, chúng tôi đã thiết kế một mạng nơ ron tích chập CNN với nền tảng ResNet50. Sau đó mạng nơ ron tích chập ResNet50 được huấn luyện với tập dữ liệu ảnh X-quang ung thư vú “chuẩn vàng” được cung cấp bởi các bác sĩ khoa chẩn đoán hình ảnh Bệnh viện Đại học Y Hà Nội để phân loại ảnh X-quang ung thư vú chuẩn DICOM vào ba lớp BI-RADS 045, BI-RADS 1, BI-RADS 23.

Kết quả đạt được độ tin cậy của hệ thống với độ đo $macAUC = 0.75$. Ngoài ra, chúng tôi xây dựng giao diện giữa bác sĩ và hệ hỗ trợ chẩn đoán X-quang ung thư vú để các bác sĩ dễ dàng sử dụng hệ.

Để đánh giá và so sánh hệ hỗ trợ chẩn đoán ung thư vú với các bác sĩ chẩn đoán hình ảnh

X-quang ung thư vú, chúng tôi lấy ngẫu nhiên gần 500 ảnh trong tập kiểm tra dùng làm “chuẩn vàng” làm tiêu chuẩn để đánh giá theo các độ đo độ tin cậy Accuracy, độ nhạy cảm Sensitivity và độ đặc hiệu Specificity. Kết quả so sánh cho thấy hệ hỗ trợ chẩn đoán ung thư vú bằng Trí tuệ nhân tạo có độ tin cậy xấp xỉ hoặc cao hơn “hội đồng” hoặc một bác sĩ chẩn đoán hình ảnh X-quang ung thư vú, nên hệ hoàn toàn có thể coi như “bác sĩ chẩn đoán hình ảnh” thứ hai hỗ trợ bác sĩ khi cần thiết trong việc đọc ảnh X-quang ung thư vú ở Việt Nam.

Theo nghiên cứu tại công trình [6], khi bác sĩ chẩn đoán hình ảnh X-quang ung thư vú tham khảo kết quả phân loại của hệ hỗ trợ chẩn đoán hình ảnh X-quang ung thư vú MV-DNN trước khi đọc để phân loại cùng một ảnh X-quang ung thư vú thì kết quả độ tin cậy “kết hợp bác sĩ và MV-DNN” luôn cao hơn độ tin cậy chỉ của hệ hỗ trợ hoặc chỉ của bác sĩ.

Trong thời gian tới chúng tôi sẽ tiếp tục thí điểm áp dụng hệ hỗ trợ chẩn đoán hình ảnh X-quang ung thư vú tại khoa chẩn đoán hình ảnh, Bệnh Viện Đại học Y Hà Nội để tiếp tục đánh giá và cải tiến hệ hỗ trợ này.

Về mặt đạo đức nghiên cứu

Nghiên cứu này đã được thông qua bởi Hội đồng đạo đức của Bệnh viện Đại học Y Hà Nội (Số Công văn chấp thuận là HMUIRB111). Trong nghiên cứu này, tất cả các bệnh nhân đồng ý bằng lời nói và/hoặc bằng văn bản ưng thuận cho sử dụng các dữ liệu và thông tin y tế của họ để phục vụ cho việc đào tạo và nghiên cứu của bệnh viện. Tất cả các chi tiết về thông tin bệnh nhân được giữ kín khi các dữ liệu được thu thập và phân tích bởi nhiều người.

Ghi nhận

Công trình này được hỗ trợ bởi Trường Đại học Thăng Long (thông qua đề tài nghiên cứu cơ sở của Trường Đại học Thăng Long do PGS.TS. Nguyễn Hoàng Phương làm chủ nhiệm đề tài) và Bệnh viện trường Đại học Y Hà Nội, Việt Nam thông qua dự án hợp tác nghiên cứu giữa hai đơn vị với tên là “Nghiên cứu phát triển hệ hỗ trợ chẩn đoán X-quang ung thư vú bằng Trí tuệ nhân tạo”, 11/2018. Trong dự án hợp tác này, PGS.TS. Nguyễn Hoàng Phương và PGS.TS. Bùi Mỹ Hạnh chịu trách nhiệm chính về mặt chuyên môn và phối hợp các hoạt động nghiên cứu của hai đơn vị thực hiện dự án. BS. Lê Tuấn Linh chịu trách nhiệm tổ chức thu thập và gán nhãn ảnh cũng như đánh giá hiệu quả phân loại ảnh X-quang ung thư vú của hệ hỗ trợ chẩn đoán ung thư vú được xây dựng.

Tài liệu tham khảo

- [1] (2013), ACR BI-RADS Atlas - Reporting System.
- [2] Curated Breast Imaging Subset of DDSM, <https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM>. (truy nhập: 19/07/2021).
- [3] Dhungel, N., Carneiro, G., and Bradley, A., (2017), Fully automated classification of mammograms using deep residual neural networks, Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), IEEE, United States of America, 310-314.
- [4] Expert Panel on Breast, I., L. Moy, S.L. Heller, et al., (2017), ACR Appropriateness Criteria® Palpable Breast Masses, J Am Coll Radiol, 14(5S): S203-S224.
- [5] Ferlay, J., Soerjomataram, I., Dikshit, R. et al., (2015), Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer, 136(5): E359-86.
- [6] Geras, K.J. et al., (2018), High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Networks. arXiv: 1703.07047v3 [cs.CV] 28 Jun 2018. (truy nhập: 19/07/2021).
- [7] Goodfellow, I., Bengio, Y., and Courville, A., (2016), Deep Learning, MIT Press.
- [8] He, K., Zhang, X., Ren, S., Sun, J., (2016), Deep residual learning for image recognition, In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770-778.
- [9] <http://icy.bioimageanalysis.org/download/> (truy nhập: 19/07/2021).
- [10] <https://www.californiaprotons.com/vi/breast-cancer/stages-grades/> (truy nhập: 19/07/2021).
- [11] <https://radiologyhanoi.com/khong-phan-loai/1812-chup-xquang-vu-quy-trinh-chup-xquang-vu-phan-do-nghi-ngo-ac-tinh-hinh-anh-xquang-vu-theo-birads.html>. (truy nhập: 19/07/2021).
- [12] <https://suckhoedoisong.vn/hon-21500-phu-nu-viet-mac-ung-thu-vu-nam-2020-n185530.html>. (truy nhập: 19/07/2021).
- [13] <https://www.most.gov.vn/vn/tag/1517/chuong-trinh-kc-4-0.aspx>. (truy nhập: 19/07/2021).
- [14] <http://www.eng.usf.edu/cvprg/Mammography/Database.html>. (DDSM: Digital Database for Screening Mammography).
- [15] <http://icy.bioimageanalysis.org/download/> (truy nhập: 19/07/2021).
- [16] Levy, D., Jain A., (2016), Breast Mass Classification

- from Mammograms using Deep Convolutional Neural Networks, arXiv:1612.00542v1[cs.CV] 2 Dec 2016. (truy nhập: 19/07/2021).
- [17] Litjens, G. et al., (2017), A Survey on Deep Learning in Medical Image Analysis. arXiv:1702.0574v1 [cs.CV] 19 Feb 2017.
- [18] Mammographic image analysis homepage, <http://www.mammoimage.org/databases/> (truy nhập: 19/07/2021).
- [19] Moreira, I.C., Amaral, I., Domingues, I. et al., (2012), INbreast: toward a full-field digital mammographic database, *Academic Radiology*, 19(2), 236-248.
- [20] Nguyen Duc Thang, Nguyen Viet Dung, Tran Vinh Duc, Anh Nguyen, Quang H. Nguyen, Nguyen Tu Anh, Nguyen Ngoc Cuong, Le Tuan Linh, Bui My Hanh, Phan Huy Phu, Nguyen Hoang Phuong, (2020), Building a X-ray Database for Mammography on Vietnamese patients and automatic Detecting ROI Using Mask-RCNN, In Book: V. Kreinovich, Nguyen Hoang Phuong (Eds.), *Soft Computing for Biomedical and Related Topics, Studies in Computational Intelligence* 899, Springer, 2020. 315-329.
- [21] Nguyễn Hoàng Phương, (2020), Tin học và Học sâu trong Y học, Nhà xuất bản Đại học quốc gia Hà Nội.
- [22] Sardanelli, F., Fallenberg, E.M., Clauser, P., et al., (2017), Mammography: an update of the EUSOBI recommendations on information for women, *Insights Imaging*, 8(1): 11-18.
- [23] Shen, L., (2017), End-to-end Training for Whole Image Breast Cancer Diagnosis using An All Convolutional Design, Poster at NIPS 2017 Workshop on Machine Learning for Health.
- [24] Shimoda, Y., Osanai, T., Nakayama, N. et al., (2016), De novo arteriovenous malformation in a patient with hereditary hemorrhagic telangiectasia, *J Neurosurg Pediatr.* 17(3): 330-5.
- [25] Tabar, L., Yen, A.M., Wu, W.Y., et al., (2015), Insights from the breast cancer screening trials: how screening affects the natural history of breast cancer and implications for evaluating service screening programs, *Breast J.* 21(1): 13-20.
- [26] Xi, P., Shu, C., Goubran, R., (2018), Abnormality Detection in Mammography using Deep Convolutional Neural Networks, ArXiv: 1803.01906v1 [cs.CV] 5 Mar 2018. (truy nhập: 19/07/2021).
- [27] Wing, P. and Langelier, M.H., (2009), Workforce shortages in breast imaging: impact on mammography utilization, *AJR Am J Roentgenol*, 192(2): 370-8.
- [28] Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., Jastrzebski, S., Fevry, T., Katsnelson, J., Kim, E., Wolfson, S., Parikh, U., Gaddam, S., Lin, L.L.Y., Ho, K., Weinstein, J.D., Reig, B., Gao, Y., Toth, H., Pysarenko, K., Lewin, A., Lee, J., Airola, K., Mema, E., Gao, Y., Toth, H., Pysarenko, K., Lewin, A., Lee, J., Airola, K., Mema, E., Chung, S., Hwang, E., Samreen, N., Kim, S.G., Heacock, L., Moy, L., Cho, K., and Geras, K.J., (2020), Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening, *IEEE Transactions on Medical Imaging*, Vol. 39, No. 4, April 2020. Pp. 1184-1194.