

BƯỚC ĐẦU NGHIÊN CỨU GIÁ TRỊ CỦA MÔ HÌNH HỌC MÁY TRONG PHÁT HIỆN BẤT THƯỜNG TRUNG THẮT TRÊN X-QUANG NGỰC THẲNG

Preliminary Evaluation of a Machine Learning Model for Detecting Mediastinal Abnormalities on Frontal Chest Radiographs

Võ Tấn Đức, Trần Thị Mai Thùy*, Nguyễn Linh Thoại*,
Trần Thanh Vỹ**, Phan Công Chiến*, Lê Hữu Nhật Minh***,
Nguyễn Quang Hiên****, Nguyễn Thanh Huy*****,
Nguyễn Lương Thanh Hà******

SUMMARY

Objective: To preliminarily evaluate the feasibility and performance of a machine learning model in detecting mediastinal abnormalities on frontal chest radiographs.

Methods: A retrospective cross-sectional study was conducted at the University Medical Center Ho Chi Minh City from September 2018 to June 2025. The study dataset comprised 735 frontal chest radiographs, including 441 cases with mediastinal abnormalities (such as 260 cases of mediastinal tumors and 181 cases of thoracic aortic aneurysms), and 294 normal cases. All cases were correlated with chest computed tomography or magnetic resonance imaging to determine the presence or absence of mediastinal abnormalities. The chest radiographs were preprocessed to standardize image quality before being used to train a ResNet34 model using a supervised learning approach. After training, the model received frontal chest radiographs as input and classified them into two groups: normal and abnormal mediastinum. The diagnostic performance of the model was evaluated using five-fold cross-validation.

Results: In the mediastinal lesion group, males predominated at 59.2%, and the mean age was significantly higher than that of the non-lesion group (58.4 ± 19.1 versus 43.4 ± 14.2 years; $p < 0.05$). The difference in gender distribution between the two groups was statistically significant ($p < 0.05$). The two most common radiographic signs in the lesion group were abnormal mediastinal contour (66.7%) and mediastinal widening (30.4%). Regarding the diagnostic performance, the ResNet34 model achieved an area under the curve (AUC-ROC) of 0.95 (95% CI: 0.93-0.98) and an accuracy of 0.87 (95% CI: 0.82-0.92). Other performance metrics included an F1-score of 0.88 (95% CI: 0.82-0.94), sensitivity of 0.83 (95% CI: 0.72-0.94), specificity of 0.93 (95% CI: 0.88-0.98), positive predictive value of 0.95 (95% CI: 0.92-0.98), and negative predictive value of 0.79 (95% CI: 0.69-0.88).

* Bệnh viện Đại học Y Dược
Thành phố Hồ Chí Minh

* Đại học Y Dược
Thành phố Hồ Chí Minh

*** Đại học Y khoa Yale,
Connecticut, Hoa Kỳ

**** Bệnh viện Montefiore, Đại
học Y Albert Einstein, Bronx, New
York, Hoa Kỳ

***** Đại học Northwestern,
Chicago, Illinois, Hoa Kỳ

***** Đại học Missouri State,
Khoa Công nghệ thông tin,
Springfield, Missouri, Hoa Kỳ

SUMMARY

Conclusion: Preliminary results suggest that the ResNet34 model achieved promising performance in detecting mediastinal abnormalities, including mediastinal tumors and thoracic aortic aneurysms, on frontal chest radiographs.

Keywords: machine learning model, chest radiograph, mediastinal lesion.

I. ĐẶT VẤN ĐỀ

Trung thất chứa nhiều cấu trúc giải phẫu quan trọng và có phổ bệnh lý đa dạng. Trong nghiên cứu này, nhóm bất thường được giới hạn ở hai nhóm bệnh đã được xác nhận trên CLVT hoặc cộng hưởng từ: u trung thất, bao gồm các khối u hoặc khối hạch tạo hình ảnh choán chỗ trung thất, và phình động mạch chủ ngực. X-quang ngực thẳng là phương tiện thường được lựa chọn trong đánh giá ban đầu, nhưng độ nhạy bị hạn chế bởi bản chất hình ảnh hai chiều và sự chồng lấp của tim, mạch máu, xương và nhu mô phổi [1], [2]. Sự phát triển nhanh chóng của trí tuệ nhân tạo, đặc biệt là các mô hình học sâu, đã mở rộng tiềm năng ứng dụng trong chẩn đoán hình ảnh, góp phần nâng cao khả năng phát hiện và phân loại tổn thương. Một phân tích tổng quan gần đây cho thấy các mô hình học sâu trong phát hiện nốt phổi trên X-quang đạt độ nhạy từ 85–99% và độ đặc hiệu từ 83–98% [3]. Các bất thường có thể biểu hiện bằng trung thất rộng, biến dạng hoặc mất các đường bờ trung thất như bờ cạnh khí quản phải, cửa sổ chủ-phổi, đường cạnh động mạch chủ, ngách đơn-thực quản và các đường cạnh sống; tuy nhiên tổn thương nhỏ hoặc nằm sau các cấu trúc bình thường vẫn có thể bị bỏ sót. X-quang nghiêng có thể cung cấp thêm thông tin định khu, nhưng không phải lúc nào cũng được thực hiện trong quy trình chụp thường quy; vì vậy nghiên cứu này tập trung vào X-quang ngực thẳng, là tư thế có tính sẵn có cao hơn trong thực hành.

Mạng nơ-ron tích chập có khả năng tự động học các đặc trưng hình ảnh. ResNet34 sử dụng các kết nối tắt, giúp hạn chế suy giảm độ chệch và cho phép huấn luyện mạng sâu với độ phức tạp tính toán vừa phải. Kiến trúc này được lựa chọn vì phù hợp với bài toán phân loại nhị phân trên bộ dữ liệu có quy mô trung bình và đã được

sử dụng rộng rãi trong phân loại hình ảnh [4]. Từ thực tiễn trên, nghiên cứu nhằm bước đầu đánh giá hiệu năng của mô hình ResNet34 trong phát hiện u trung thất và phình động mạch chủ ngực trên X-quang ngực thẳng, với CLVT hoặc cộng hưởng từ lồng ngực làm chuẩn tham chiếu.

II. ĐỐI TƯỢNG, PHƯƠNG PHÁP NGHIÊN CỨU

Đối tượng nghiên cứu

Dân số nghiên cứu: Bệnh nhân có hình X-quang ngực thẳng được chụp tại Bệnh viện Đại học Y Dược Thành phố Hồ Chí Minh trong thời gian từ 09/2018 đến 06/2025.

Tiêu chuẩn chọn mẫu:

Bệnh nhân có X-quang ngực thẳng đủ chất lượng chẩn đoán bao gồm hình ảnh thấy được toàn bộ lồng ngực, không bị xoay, độ xuyên thấu tia X vừa đủ và có CLVT hoặc cộng hưởng từ lồng ngực đối chiếu được thực hiện trong khoảng không quá 1 tháng. Khoảng thời gian này được lựa chọn nhằm tăng tính tương đồng về tình trạng bệnh giữa hai lần khảo sát và giảm khả năng xuất hiện hoặc tiến triển đáng kể của tổn thương trong thời gian đối chiếu. Nghiên cứu chỉ sử dụng X-quang ngực thẳng; X-quang nghiêng không được phân tích do không được thực hiện đồng nhất ở toàn bộ mẫu.

2. Phương pháp nghiên cứu

Thiết kế nghiên cứu: Nghiên cứu cắt ngang hồi cứu.

Phương pháp thu thập số liệu: Tất cả X-quang ngực thẳng thỏa tiêu chuẩn chọn mẫu được đưa vào nghiên cứu. Do dữ liệu hồi cứu được thu nhận trong thời gian dài và từ các quy trình chụp thường quy, thông số chụp chi tiết (bao gồm kV, mAs và khoảng cách bóng phát tia - bộ

thu nhận ảnh) không đồng nhất và không được truy xuất đầy đủ cho mọi trường hợp. Các hình được xác nhận đủ chất lượng chẩn đoán được sử dụng.

Các biến số được thu thập gồm tuổi, giới tính, đường kính ngang lớn nhất trung thất, diện tích đường bờ trung thất bất thường. Tăng đường kính ngang lớn nhất trung thất được định nghĩa là đường kính ngang lớn nhất của trung thất trên 6 cm trên X-quang ngực tư thế sau – trước hoặc trên 8 cm trên X-quang ngực tư thế trước-sau. Đường kính ngang được đo ở vùng trung thất nằm trên vị trí tĩnh mạch chủ trên đổ vào nhĩ phải. Đường bờ trung thất bất thường được định nghĩa là sự hiện diện của đường bờ bất thường ngoài các đường trung thất thấy rõ trên hình X-quang ngực thẳng bao gồm bờ động mạch chủ lên, quai và động mạch chủ xuống, bờ tĩnh mạch chủ trên, bờ nhĩ phải, thân động mạch phổi và bờ thất trái. Hai dấu hiệu này được ghi nhận dưới dạng biến nhị phân.

Quy trình xây dựng mô hình và đánh giá hiệu quả của mô hình bao gồm các bước chính sau:

Bước 1. Thu thập và chuẩn bị dữ liệu hình ảnh: tất cả hình X-quang ngực thẳng được thu thập và kiểm tra trước khi đưa vào phân tích. Các trường hợp nghiên cứu được phân loại dựa trên kết quả đối chiếu với chụp cắt lớp vi tính hoặc cộng hưởng từ lồng ngực nhằm xác định nhóm bình thường và nhóm có bất thường trung thất.

Bước 2. Tiền xử lý hình ảnh X-quang: Dữ liệu DICOM được đọc bằng pydicom và xử lý trong môi trường Python 3, sử dụng các thư viện pydicom [5], NumPy [6], OpenCV [7], scikit-image [8], PyTorch [9], torchvision [10] và scikit-learn [11]. Quy trình gồm chuẩn hóa hướng ảnh và thang giá trị điểm ảnh, điều chỉnh cửa sổ hiển thị, giảm nhiễu, tăng tương phản cục bộ bằng CLAHE [12], chuẩn hóa cường độ và thay đổi kích thước ảnh phù hợp với đầu vào của mạng. Các bước được thực hiện tự động bằng NumPy, OpenCV và scikit-image nhằm hạn chế điều chỉnh chủ quan giữa các trường hợp. Bước này giúp các cấu trúc giải phẫu và dấu hiệu bất thường trên hình được thể hiện rõ hơn trước khi đưa vào mô hình học sâu.

Bước 3. Huấn luyện mô hình ResNet34: Sau tiền xử lý, hình ảnh được đưa vào mô hình ResNet34 và

huấn luyện theo phương pháp học có giám sát. Các vùng bất thường được người huấn luyện mô hình đánh dấu bằng hộp chữ nhật trên phần mềm LabelMe và gán nhãn phân loại là bất thường trung thất. Các dấu hiệu tăng đường kính ngang trung thất và diện tích đường bờ trung thất bất thường trên X-quang ngực thẳng được hai bác sĩ chẩn đoán hình ảnh có kinh nghiệm 5 năm và kinh nghiệm trên 10 năm đọc. Mô hình sẽ tự động học các đặc điểm được huấn luyện và thực hiện bài toán phân loại nhị phân: có bất thường trung thất và không có bất thường trung thất. Nhãn tham chiếu trung thất bình thường hay bất thường, bao gồm u trung thất hoặc phình động mạch chủ ngực, xác định từ kết quả chụp cắt lớp vi tính hoặc cộng hưởng từ được đọc bởi bác sĩ chẩn đoán hình ảnh có kinh nghiệm trên 5 năm.

Bước 4. Đánh giá hiệu quả mô hình: Hiệu quả chẩn đoán của mô hình được đánh giá bằng phương pháp kiểm định chéo 5 lần. Phương pháp này giúp kiểm tra khả năng hoạt động ổn định của mô hình trên nhiều tập dữ liệu khác nhau, hạn chế sai lệch khi chỉ sử dụng một lần chia dữ liệu duy nhất.

Xử lý số liệu: Dữ liệu được quản lý bằng phần mềm Microsoft Excel 365 và phân tích thống kê bằng phần mềm STATA, ngôn ngữ Python 3 với các thư viện Pandas, NumPy, SciPy và Scikit-learn.

3. Đạo đức trong nghiên cứu

Nghiên cứu được chấp thuận và cho phép bởi Hội đồng Đạo đức Bệnh viện Đại học Y Dược Thành phố Hồ Chí Minh, 06/GCN-HĐĐĐ ngày 29 tháng 01 năm 2026 và thực hiện theo đúng quy định của đơn vị. Tất cả các thông tin của người bệnh được bảo mật theo qui định của Hội đồng Đạo đức Bệnh viện.

III. KẾT QUẢ

1. Đặc điểm chung

Trong mẫu nghiên cứu, phân bố giới tính có sự khác biệt có ý nghĩa thống kê giữa hai nhóm ($p < 0,05$). Ở nhóm có tổn thương trung thất, nam giới chiếm tỷ lệ 59,2%, với 261 trường hợp, cao hơn so với nữ giới. Ở nhóm còn lại, nữ giới chiếm ưu thế hơn với tỷ lệ 53,1%.

Về tuổi, nhóm có tổn thương trung thất có tuổi trung bình $58,41 \pm 19,1$, cao hơn có ý nghĩa thống kê so với tuổi trung bình của nhóm không có tổn thương là $43,43 \pm 14,2$. Ở nhóm có tổn thương trung thất, các trường hợp tổn thương xuất hiện chủ yếu ở nhóm từ 60 tuổi trở lên, chiếm 55,8%. Ngược lại, các bệnh nhân ở nhóm không có tổn thương trung thất có phân bố tuổi tập trung chủ yếu ở hai nhóm 18–39 tuổi và 40–59 tuổi, với tỷ lệ lần lượt là 40,5% và 40,8%.

Dấu hiệu tăng đường kính ngang lớn nhất trung thất được ghi nhận ở 134 trường hợp trong nhóm có tổn thương trung thất, chiếm 30,4%. Dấu hiệu hiện diện đường bờ trung thất bất thường được ghi nhận ở 294 trường hợp, tương ứng tỷ lệ 66,7%.

2. Mô hình học máy

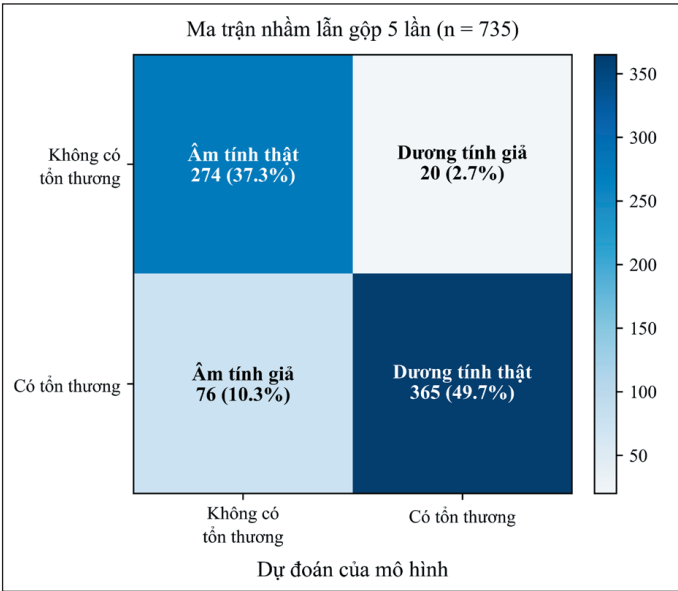
Hiệu suất của mô hình trong phát hiện tổn thương trung thất qua kiểm định chéo 5 lần được thể hiện ở Bảng 1.

Bảng 1. Hiệu suất của mô hình ResNet34 trong phát hiện tổn thương trung thất trên X-quang ngực qua kiểm định chéo 5 lần

Chỉ số	Lần 1	Lần 2	Lần 3	Lần 4	Lần 5	Trung bình	95% CI
Độ chính xác	0,912	0,851	0,863	0,905	0,816	$0,869 \pm 0,039$	0,820-0,918
Độ đặc hiệu	0,943	0,921	0,963	0,869	0,968	$0,933 \pm 0,040$	0,883-0,983
Độ nhạy	0,894	0,800	0,804	0,930	0,702	$0,826 \pm 0,089$	0,715-0,937
Giá trị tiên đoán dương	0,966	0,932	0,974	0,909	0,967	$0,949 \pm 0,028$	0,915-0,984
Giá trị tiên đoán âm	0,833	0,773	0,743	0,898	0,709	$0,791 \pm 0,075$	0,698-0,885
Điểm F1	0,928	0,861	0,881	0,920	0,814	$0,881 \pm 0,046$	0,823-0,938
ROC-AUC	0,967	0,945	0,976	0,958	0,927	$0,954 \pm 0,019$	0,931-0,978

Hiệu suất của mô hình tương đối ổn định giữa 5 lần, với độ chính xác dao động từ 0,816 (Lần 5) đến 0,912 (Lần 1), điểm F1 từ 0,814 (Lần 5) đến 0,928 (Lần 1) và diện tích dưới đường cong ROC từ 0,927 (Lần 5) đến

0,976 (Lần 3). Số ca tổn thương bị bỏ sót dao động từ 6 (Lần 4) đến 25 ca (Lần 5) mỗi lần, trong khi số ca dương tính giả chỉ từ 2 đến 8 ca mỗi lần.

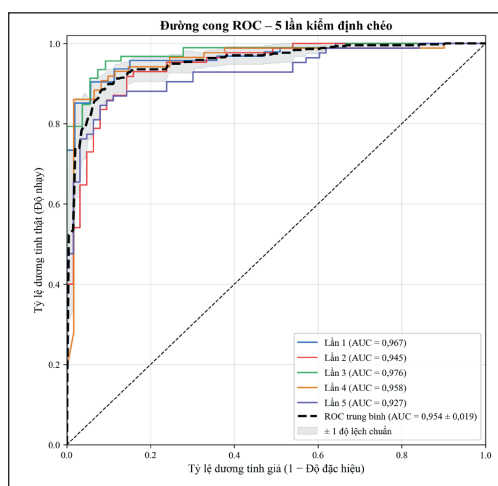


Hình 1. Ma trận nhầm lẫn gộp của mô hình ResNet34 trong phát hiện tổn thương trung thất qua kiểm định chéo 5 lần

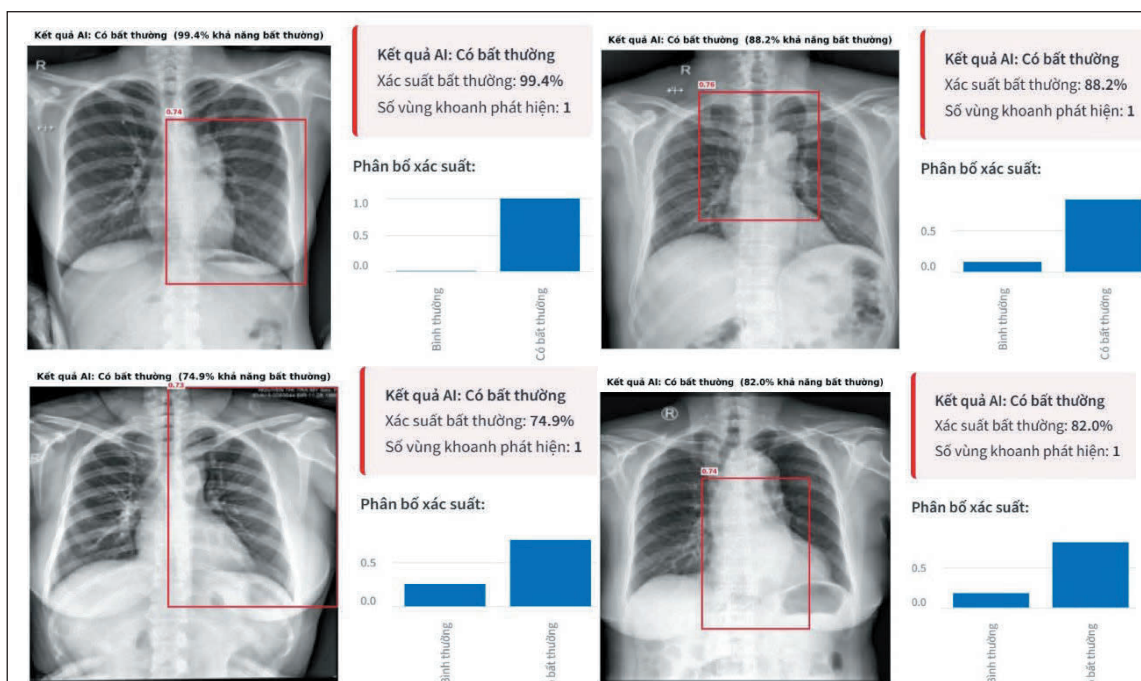
Khi gộp các dự đoán ngoài mẫu của 5 lần kiểm định chéo ($n = 735$), mô hình phát hiện đúng 365/441 trường hợp có bất thường trung thất và bỏ sót 76 trường hợp (17,2%); đồng thời xác định đúng 274/294 trường hợp không có bất thường và tạo 20 trường hợp dương tính giả (6,8%), như trình bày ở Hình 1. Hiệu năng gộp gồm độ nhạy 0,828, độ đặc hiệu 0,932, giá trị tiên đoán dương 0,948, giá trị tiên đoán âm 0,783 và độ chính xác 0,869; ROC-AUC trung bình của 5 lần là 0,954 (KTC 95%: 0,931-0,978). Tỷ số khả dĩ dương và âm lần lượt khoảng

12,2 và 0,18. Nghiên cứu chưa thực hiện phân tích sai số có hệ thống cho toàn bộ 76 âm tính giả và 20 dương tính giả; quan sát định tính cho thấy sai số có thể liên quan đến tổn thương nhỏ hoặc bị che lấp, biến dạng bờ trung thất kín đáo, các cấu trúc mạch máu/tim nổi bật và khác biệt về tư thế hoặc chất lượng ảnh.

Đường cong ROC của 5 lần kiểm định chéo và đường cong ROC trung bình được trình bày ở Hình 2, với diện tích dưới đường cong trung bình đạt 0,954 (95% CI 0,931 – 0,978).



Hình 2. Đường cong ROC của mô hình ResNet34 qua 5 lần kiểm định chéo và đường cong ROC trung bình



Hình 3. Một số ví dụ minh họa kết quả dự đoán của mô hình ResNet34 trên X-quang ngực

IV. BÀN LUẬN

Trong nghiên cứu của chúng tôi, tỷ lệ nam giới cao hơn so với nữ giới trong nhóm có tổn thương trung thất. Bên cạnh đó, nhóm có tổn thương trung thất cũng có tuổi trung bình cao hơn nhóm không có tổn thương, lần lượt là $58,41 \pm 19,10$ tuổi và $43,43 \pm 14,21$ tuổi. Sự khác biệt giới tính và tuổi có ý nghĩa thống kê ($p < 0,05$). Đặc điểm này có thể liên quan đến cấu trúc bệnh lý không đồng nhất của mẫu nghiên cứu, bao gồm các tổn thương tân sinh tại trung thất cũng như các trường hợp phình động mạch chủ ngực. Theo y văn, nam giới và tuổi cao là các yếu tố nguy cơ đã được ghi nhận của phình động mạch chủ ngực [13]. Bên cạnh đó, nghiên cứu của Mehrabi Nasab (2022) cũng cho thấy nam giới chiếm tỷ lệ cao hơn trong nhóm có kích thước động mạch chủ bất thường, đồng thời phần lớn người bệnh thuộc nhóm trên 60 tuổi [14]. Do đó, sự hiện diện của các trường hợp phình động mạch chủ ngực trong nhóm tổn thương trung thất có thể góp phần lý giải tỷ lệ nam giới và tuổi trung bình cao hơn ở nhóm có tổn thương so với nhóm còn lại.

Trong 441 trường hợp có bất thường trung thất, tăng đường kính ngang trung thất chỉ gặp ở 30,4%, trong khi ít nhất một đường bờ trung thất bất thường gặp ở 66,7%. Kết quả này cho thấy tăng đường kính ngang trung thất là dấu hiệu có độ bao phủ hạn chế nếu sử dụng đơn độc. Bất thường đường bờ xuất hiện thường xuyên hơn nhưng vẫn không hiện diện ở khoảng một phần ba trường hợp. Vì nghiên cứu chỉ ghi nhận hai dấu hiệu dưới dạng có/không và không phân loại theo từng đường bờ, chưa thể xác định đường bờ nào đóng góp nhiều nhất cho khả năng phát hiện. Việc không có X-quang nghiêng ở tất cả đối tượng cũng có thể làm giảm khả năng nhận diện hoặc định khu các tổn thương nằm sau tim, trước tim hoặc chồng lấp với các cấu trúc trung thất.

Về hiệu năng mô hình học máy, chúng tôi đánh giá hiệu suất của mô hình trên lớp dương tính, nhóm bệnh nhân có tổn thương trung thất, dựa trên kiểm định chéo 5 lần. Kết quả ghi nhận điểm F1 đạt $0,881 \pm 0,046$, độ nhạy $0,826 \pm 0,089$, độ đặc hiệu $0,933 \pm 0,040$, giá trị tiên đoán dương $0,949 \pm 0,028$, giá trị tiên đoán âm $0,791 \pm 0,075$, độ chính xác $0,869 \pm 0,039$ và diện tích dưới đường cong

ROC $0,954 \pm 0,019$. Hiệu suất dao động nhẹ giữa các lần kiểm định, phản ánh tính ổn định của mô hình. Sự dao động này có thể do phân bố không đều của các trường hợp khó giữa các lần. Mô hình đạt độ đặc hiệu cao và độ nhạy ở mức khá, cho thấy khó khăn trong việc giải quyết bài toán phát hiện tổn thương trung thất trên X-quang ngực thẳng, khi các cấu trúc xung quanh che khuất một số tổn thương kích thước nhỏ.

Cho đến nay, số lượng nghiên cứu tập trung phát triển các mô hình học máy chuyên biệt nhằm phát hiện bất thường trung thất trên X-quang ngực còn hạn chế. Hầu hết các nghiên cứu hiện nay tiếp cận bất thường trung thất gián tiếp thông qua các dấu hiệu cổ điển như tăng đường kính ngang trung thất hoặc bóng tim to. Khi so sánh với một số mô hình học máy trong đánh giá bất thường trên X-quang ngực nói chung, mô hình của chúng tôi mặc dù tập trung vào một bài toán hẹp hơn nhưng vẫn cho thấy hiệu năng khá tương đồng. Nghiên cứu gần đây của Kai (2025) sử dụng kết hợp Vision Transformer, EfficientNetV2 và Light Gradient Boosting Machine trên bộ dữ liệu từ hai cơ sở y tế Việt Nam, ghi nhận điểm F1 đạt 0,668, độ nhạy 0,596, độ đặc hiệu 0,931, độ chính xác 0,842 và diện tích dưới đường cong 0,897 [15]. Mô hình chúng tôi có điểm F1, độ nhạy, độ chính xác và diện tích dưới đường cong cao hơn, trong khi độ đặc hiệu tương đương. Sự khác biệt này có thể là do chúng tôi thiết kế mô hình đặc thù để đánh giá tổn thương trung thất, thay vì xây dựng mô hình bao quát nhiều bệnh lý lồng ngực như nghiên cứu này. Bên cạnh đó, hiệu suất mô hình của chúng tôi cũng cao hơn nghiên cứu của Nguyen (2022) [16]. Trong nghiên cứu này, tác giả ứng dụng hệ thống VinDr-CXR trên hơn 6000 hình X-quang ngực nhằm phân biệt hình X-quang bình thường và bất thường; trong đó, các dấu hiệu trung thất trên X-quang được đánh giá bao gồm động mạch chủ bụng rộng, bóng tim to và động mạch phổi lớn. Sự khác biệt này có thể giải thích là do nghiên cứu của tác giả được thực hiện trên bộ dữ liệu ngoài phân bố, trong khi mô hình của chúng tôi được đánh giá trong điều kiện kiểm tra chéo trên cùng dữ liệu huấn luyện. Bên cạnh đó, chúng tôi cũng tìm thấy một số khác biệt khi so sánh với các nghiên cứu tập trung phân tích tổn thương trung thất. Nghiên cứu của Ribeiro

(2024) xây dựng mô hình nhằm phát hiện giãn động mạch chủ trên X-quang ngực bằng mô hình EfficientNet và DenseNet [17], ghi nhận EfficientNet đạt hiệu suất tốt hơn DenseNet ở hầu hết các chỉ số, với điểm F1 0,825, độ đặc hiệu 0,894, độ nhạy 0,823, độ chính xác 0,867 và diện tích dưới đường cong 0,927. So sánh với EfficientNet trong nghiên cứu của Ribeiro, mô hình ResNet34 của chúng tôi cho kết quả tương đương về độ chính xác và độ nhạy, nhưng cao hơn về độ đặc hiệu, điểm F1 và diện tích dưới đường cong. Sự khác biệt này có thể được giải thích do nghiên cứu của Ribeiro sử dụng quy trình học chuyển giao hai giai đoạn với bộ dữ liệu huấn luyện lớn, trong khi mô hình của chúng tôi được huấn luyện chủ yếu trên dữ liệu bệnh viện riêng với cỡ mẫu có giới hạn hơn, điều có thể ảnh hưởng đến khả năng học đặc trưng phong phú ở lớp có tổn thương. Tuy nhiên, khi so sánh với nghiên cứu của van Beek (2023) [18], hiệu suất của mô hình trong nghiên cứu hiện tại nhìn chung có xu hướng thấp hơn, đặc biệt về độ đặc hiệu. Trong nghiên cứu của van Beek, tổn thương trung thất được đánh giá gián tiếp qua dấu hiệu bóng tim to và tăng đường kính ngang trung thất. Đối với dấu hiệu bóng tim to, độ nhạy của mô hình trong hai nhóm cấp cứu và chăm sóc ban đầu dao động từ 0,850 đến 0,880, trong khi độ đặc hiệu dao động từ 0,884 đến 0,962. Đối với dấu hiệu tăng đường kính ngang trung thất, một dấu hiệu liên quan trực tiếp đến phát hiện bất thường trung thất trên X-quang ngực, van Beek và cộng sự ghi nhận độ nhạy từ 0,800–1,000 và độ đặc hiệu từ 0,970–0,993. Trong nghiên cứu của chúng tôi, mô hình ResNet34 đạt độ nhạy $0,826 \pm 0,089$ và độ đặc hiệu $0,933 \pm 0,040$. Sự khác biệt này

có thể liên quan đến khác biệt về mục tiêu phân loại, đặc điểm bộ dữ liệu, định nghĩa tổn thương và phương pháp xây dựng nhãn tham chiếu giữa hai nghiên cứu.

Đây là nghiên cứu hồi cứu đơn trung tâm với cỡ mẫu còn hạn chế và chưa có tập kiểm định ngoài độc lập, do đó hiệu năng có thể bị thay đổi khi áp dụng ở cơ sở khác. Nghiên cứu bước đầu tập trung vào xây dựng mô hình học máy nên chưa đi sâu phân tích chi tiết từng kiểu đường bờ trung thất trên X-quang ngực. Các nghiên cứu tiếp theo cần sử dụng dữ liệu đa trung tâm, so sánh với kết quả đọc của bác sĩ chẩn đoán hình ảnh để thấy được giá trị của mô hình học máy trong phát hiện u trung thất hoặc phình động mạch chủ ngực trên X-quang ngực thẳng. Ngoài ra, tiêu chuẩn yêu cầu có hình cắt lớp vi tính hoặc cộng hưởng từ đối chiếu có thể gây sai lệch chọn mẫu và làm cho phổ bệnh trong nghiên cứu khác với quần thể X-quang ngực thường quy. Tỷ lệ bất thường cao trong bộ dữ liệu cũng có thể ảnh hưởng đến các giá trị tiên đoán, do đó các giá trị tiên đoán dương hoặc âm cần được diễn giải thận trọng khi áp dụng cho quần thể có tỷ lệ bệnh khác.

V. KẾT LUẬN

Kết quả từ bộ dữ liệu gồm 735 hình X-quang ngực thẳng cho thấy mô hình ResNet34 đạt hiệu quả tương đối khả quan trong phát hiện bất thường trung thất bao gồm u trung thất và phình động mạch chủ ngực. Tuy vậy, mô hình cần được kiểm định ngoài, phân tích sai số chi tiết và so sánh với kết quả của bác sĩ chẩn đoán hình ảnh đọc trước khi xem xét ứng dụng lâm sàng.

TÀI LIỆU THAM KHẢO

1. Von Kodolitsch Y, Nienaber CA, Dieckmann C, et al. Chest radiography for the diagnosis of acute aortic syndrome. *Am J Med.* 2004;116(2):73-7. doi:10.1016/j.amjmed.2003.08.030
2. Wernecke K, Vassallo P, Pötter R, Lückener HG, Peters PE. Mediastinal tumors: sensitivity of detection with sonography compared with CT and radiography. *Radiology.* 1990;175(1):137-43. doi:10.1148/radiology.175.1.2179987
3. Rehman A, Khan A, Fatima G, Naz S, Razzak I. Review on chest pathologies detection systems using deep learning techniques. *Artif Intell Rev.* 2023;1-47. doi:10.1007/s10462-023-10457-9
4. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016:770-778.
5. *Pydicom: An open-source DICOM library for Python.* Zenodo; 2024. <https://doi.org/10.5281/zenodo.13824606>
6. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P. Array programming with NumPy. *Nature.* 2020;585(7825):357-362. doi:10.1038/s41586-020-2649-2
7. Bradski G. The OpenCV Library. *Dr Dobb's Journal of Software Tools.* 2000;25(11):120-125.
8. Van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD. scikit-image: image processing in Python. *PeerJ.* 2014;2:e453. doi:10.7717/peerj.453
9. Paszke A, Gross S, Massa F, Lerer A, Bradbury J. PyTorch: An Imperative Style, High-Performance Deep Learning Library. presented at: Advances in Neural Information Processing Systems 32; 2019; <https://doi.org/10.48550/arXiv.1912.01703>
10. Marcel S, Rodriguez Y. Torchvision the machine-vision package of torch. presented at: Proceedings of the 18th ACM International Conference on Multimedia; 2010; <https://doi.org/10.1145/1873951.1874254>
11. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research.* 2011;12:2825-2830.
12. Zuiderveld K. Contrast Limited Adaptive Histogram Equalization. In: Heckbert PS, ed. *Graphics Gems IV.* Academic Press; 1994:474-485.
13. Faiza Z, Sharman T. Thoracic Aorta Aneurysm. *StatPearls [Internet].* StatPearls Publishing; 2026.
14. Mehrabi Nasab E, Athari SS. The prevalence of thoracic aorta aneurysm as an important cardiovascular disease in the general population. *J Cardiothorac Surg.* 2022;17(1):51. doi:10.1186/s13019-022-01767-0
15. Kai C, Kasai S, Teramoto R, et al. Classifying abnormalities in chest radiographs from Vietnam using deep learning for early detection of cardiopulmonary diseases. *Frontiers in Radiology.* 2025;doi:10.3389/fradi.2025.1703927
16. Nguyen NH, Nguyen HQ, Nguyen NT, et al. Deployment and validation of an AI system for detecting abnormal chest radiographs in clinical settings. *Front Digit Health.* 2022;4:890759. doi:10.3389/fdgth.2022.890759
17. Ribeiro E, Cardenas DAC, Dias FM, Krieger JE, Gutierrez MA. Explainable artificial intelligence in deep learning-based detection of aortic elongation on chest X-ray images. *Eur Heart J Digit Health.* 2024;5(5):524-534. doi:10.1093/ehjdh/ztae045
18. Van Beek EJR, Ahn JS, Kim MJ, Murchison JT. Validation study of machine-learning chest radiograph software in primary and emergency medicine. *Clin Radiol.* 2023;78(1):1-7. doi:10.1016/j.crad.2022.08.129

TÓM TẮT

Mục tiêu: Bước đầu khảo sát giá trị của mô hình học máy trong phát hiện bất thường trung thất trên hình X-quang ngực thẳng.

Đối tượng và phương pháp nghiên cứu: Nghiên cứu cắt ngang hồi cứu được thực hiện tại Bệnh viện Đại học Y Dược Thành phố Hồ Chí Minh từ tháng 9/2018 đến tháng 6/2025. Dữ liệu nghiên cứu bao gồm 735 hình X-quang ngực thẳng, trong đó có 441 trường hợp bất thường trung thất (bao gồm 260 ca u trung thất, 181 ca phình động mạch chủ ngực) và 294 trường hợp bình thường. Tất cả các trường hợp đều được đối chiếu với hình ảnh chụp cắt lớp vi tính hoặc cộng hưởng từ lồng ngực để xác định trung thất bình thường hay bất thường. Các hình X quang được tiền xử lý nhằm chuẩn hóa chất lượng trước khi đưa vào huấn luyện mô hình ResNet34 theo phương pháp học có giám sát. Mô hình sau khi huấn luyện sẽ nhận hình X-quang ngực và phân loại thành hai nhóm: trung thất bình thường và bất thường. Giá trị chẩn đoán của mô hình được đánh giá thông qua kỹ thuật kiểm định chéo 5 lần.

Kết quả: Trong nhóm có tổn thương trung thất, nam giới chiếm ưu thế với tỷ lệ 59,2% và độ tuổi trung bình cao hơn so với nhóm không có tổn thương ($58,4 \pm 19,1$ so với $43,4 \pm 14,2$ tuổi; $p < 0,05$). Sự khác biệt về phân bố giới tính giữa hai nhóm cũng có ý nghĩa thống kê ($p < 0,05$). Hai dấu hiệu X-quang thường gặp nhất ở nhóm có tổn thương là hiện diện bờ trung thất bất thường (66,7%) và tăng đường kính ngang trung thất (30,4%). Về giá trị chẩn đoán trong nhóm tổn thương, mô hình ResNet34 đạt diện tích dưới đường cong là 0,95 với khoảng tin cậy (KTC) 95% là 0,93-0,98 và độ chính xác là 0,87 (KTC 95% là 0,82-0,92). Các chỉ số giá trị khác bao gồm: điểm F1 đạt 0,88 (KTC 95% là 0,82-0,94), độ nhạy 0,83 (KTC 95% là 0,72-0,94), độ đặc hiệu 0,93 (KTC 95% là 0,88-0,98), giá trị tiên đoán dương 0,95 (KTC 95% là 0,92-0,98) và giá trị tiên đoán âm 0,79 (KTC 95% là 0,69-0,88).

Kết luận: Kết quả bước đầu cho thấy mô hình ResNet34 có hiệu năng khả quan trong phát hiện bất thường trung thất, bao gồm u trung thất và phình động mạch chủ ngực, trên X-quang ngực thẳng. .

Từ khóa: mô hình học máy, X-quang ngực thẳng, tổn thương trung thất.

Người liên hệ: Trần Thị Mai Thùy. Email: thuy.ttm@umc.edu.vn

Ngày nhận bài: 01/06/2026. Ngày nhận phản biện: 06/06/2026. Ngày chấp nhận đăng: 25/06/2026