

ỨNG DỤNG MÔ HÌNH HỌC SÂU ĐỂ PHÁT HIỆN VÀ PHÂN LOẠI GỠ CỔ XƯƠNG ĐÙI VÀ GỠ LIÊN MÁU CHUYỂN TRÊN X-QUANG

Phạm Tấn Hường¹, Nguyễn Vũ Đăng¹, Phạm Tấn Hạnh²
và Nguyễn Thành Tấn^{1,✉}

¹Trường Đại học Y Dược Cần Thơ

²Harvard Medical School

Gãy cổ xương đùi và gãy liên mấu chuyển là các tổn thương thường gặp trong chấn thương vùng háng. Việc phát hiện và phân loại chính xác có vai trò quan trọng trong lựa chọn phương pháp điều trị nhưng phụ thuộc nhiều vào kinh nghiệm bác sĩ. Nghiên cứu này sử dụng mô hình RF-DETR làm cơ sở để xây dựng mô hình phát hiện và phân loại gãy cổ xương đùi và gãy liên mấu chuyển trên phim X-quang. Mô hình RF-DETR được huấn luyện và kiểm định nội bộ trên 1.483 hình ảnh X-quang công khai. Đánh giá ngoài được thực hiện trên 93 phim X-quang tại bệnh viện, với tiêu chuẩn vàng là sự đồng thuận của hai bác sĩ chuyên khoa chấn thương chỉnh hình. Trên tập kiểm định nội bộ, mô hình đạt mAP@50 93,9%, precision 91,8%, recall 90,6%. Trên dữ liệu bệnh viện, mô hình đạt accuracy 95,7% (KTC 95%: 89,5 - 98,3%), cao hơn bác sĩ đa khoa 78,5% (KTC 95%: 69,1 - 85,6%), $p < 0,001$ (McNemar test), với hệ số kappa 0,936. Mô hình học sâu RF-DETR cho hiệu năng cao và có tiềm năng hỗ trợ chẩn đoán lâm sàng.

Từ khóa: Trí tuệ nhân tạo, gãy cổ xương đùi, gãy liên mấu chuyển, X-quang, RF-DETR.

I. ĐẶT VẤN ĐỀ

Gãy cổ xương đùi và gãy liên mấu chuyển là hai dạng gãy xương vùng háng thường gặp. Mỗi năm tại Hoa Kỳ có hơn 300.000 trường hợp gãy xương vùng háng, và con số này dự báo tiếp tục gia tăng theo quá trình già hóa dân số.¹ Hình ảnh X-quang là cận lâm sàng hỗ trợ bác sĩ chẩn đoán và đưa ra hướng điều trị. Tuy nhiên, độ chính xác phụ thuộc nhiều vào kinh nghiệm người đọc phim. Tỷ lệ bỏ sót gãy xương trên X-quang được ghi nhận dao động từ 2 - 10%, đặc biệt ở các trường hợp gãy không di lệch hoặc hình ảnh không điển hình.^{2,3} Hệ số đồng thuận liên bác sĩ trong phân loại Garden chỉ đạt kappa 0,03 - 0,56, phản ánh mức độ chủ quan

cao trong diễn giải hình ảnh.⁴

Sự phát triển của trí tuệ nhân tạo, đặc biệt là học sâu, đã mở ra hướng tiếp cận mới trong phân tích hình ảnh y khoa. Các mô hình học sâu đã được ứng dụng trong phát hiện và phân loại gãy xương vùng háng, với một số nghiên cứu ghi nhận độ chính xác lên tới 93,7% và AUC (Area Under the Curve) trên 0,90.^{5,6} Tuy nhiên, phần lớn các mô hình hiện có dựa trên kiến trúc CNN (Convolutional Neural Network) truyền thống hoạt động dựa trên đặc trưng cục bộ, phụ thuộc vào anchor box và Non-Maximum Suppression (NMS) hạn chế khả năng nắm bắt bối cảnh toàn cục của ảnh, điều đặc biệt quan trọng với các tổn thương có hình thái đa dạng và ranh giới không rõ.

RF-DETR (Roboflow Detection Transformer) là kiến trúc phát hiện đối tượng thế hệ mới dựa trên cơ chế Transformer, tích hợp backbone

Tác giả liên hệ: Nguyễn Thành Tấn

Trường Đại học Y Dược Cần Thơ

Email: nttan@ctump.edu.vn

Ngày nhận: 16/05/2026

Ngày được chấp nhận: 22/06/2026

DINOv2 với deformable attention, cho phép xử lý đầu cuối không cần anchor box hay NMS.⁷ Tuy nhiên, đến nay chưa có nghiên cứu nào đánh giá RF-DETR trong bài toán phát hiện và phân loại gãy cổ xương đùi và gãy liên mấu chuyển trên X-quang. Xuất phát từ thực tiễn đó, chúng tôi tiến hành nghiên cứu ứng dụng mô hình RF-DETR trong phát hiện và phân loại gãy cổ xương đùi và gãy liên mấu chuyển trên X-quang.

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP

1. Đối tượng

Tiêu chuẩn chọn mẫu

Tất cả các X-quang khung chậu thẳng đã được chẩn đoán xác định gãy cổ xương đùi hoặc gãy liên mấu chuyển.

Tất cả các trường hợp đều được 2 bác sĩ chuyên khoa chấn thương chỉnh hình có kinh nghiệm hơn 10 năm đọc một cách độc lập và kết quả trùng nhau dựa trên các tiêu chuẩn và phân loại đã được thống nhất.

Tiêu chuẩn loại trừ

X-quang có gãy xương vùng háng khác kèm theo như: gãy ổ cối, gãy chỏm xương đùi, gãy dưới liên mấu chuyển. Các trường hợp còn sụn tiếp hợp.

2. Phương pháp

Thiết kế nghiên cứu

Nghiên cứu mô tả cắt ngang.

Địa điểm và thời gian nghiên cứu

Nghiên cứu được thực hiện tại Bệnh viện Trường Đại học Y Dược Cần Thơ từ 2024 đến 2026.

Cỡ mẫu và phương pháp chọn mẫu

Chúng tôi sử dụng 1483 X-quang khung chậu được chia sẻ công khai trên nhiều nền tảng để xây dựng mô hình và 93 X-quang khung chậu thu thập theo phương pháp chọn mẫu thuận tiện từ bệnh viện để so sánh mô hình với bác sĩ đa khoa.

Nội dung nghiên cứu:

Quy trình xây dựng mô hình

*Tiêu chuẩn chọn bác sĩ đọc phim.

- Để đảm bảo chất lượng nhãn dán, tiêu chuẩn vàng được xây dựng bởi hai bác sĩ chuyên khoa Chấn thương chỉnh hình đáp ứng đầy đủ các tiêu chí sau:

- Có ít nhất 10 năm kinh nghiệm chẩn đoán và điều trị gãy cổ xương đùi và gãy liên mấu chuyển. Hai bác sĩ đọc phim độc lập với nhau.

- Đồng ý tham gia nghiên cứu và được thống nhất về quy ước các nhãn dán trước khi tiến hành.

- Khi hai bác sĩ đó đồng thuận, kết quả đọc phim X-quang đó được sử dụng làm tiêu chuẩn vàng. Khi có sự khác biệt, X-quang đó bị loại khỏi bộ dữ liệu.

* Quy ước phân loại 4 nhóm nhãn dán

- Dựa trên phân loại Garden cho gãy cổ xương đùi và phân loại AO cho gãy liên mấu chuyển, chúng tôi quy ước thành 4 nhóm thống nhất để các bác sĩ chỉnh hình đọc và chuẩn hóa bộ dữ liệu như sau:

+ Nhóm 1 - Gãy cổ xương đùi không di lệch (Garden I + II): đường gãy một phần hoặc toàn bộ cổ xương đùi nhưng không di lệch.

+ Nhóm 2 - Gãy cổ xương đùi di lệch (Garden III + IV): đường gãy rõ, di lệch một phần hoặc hoàn toàn.

+ Nhóm 3 - Gãy liên mấu chuyển vững (AO nhóm 1): gãy 2 mảnh, thành ngoài nguyên vẹn.

+ Nhóm 4 - Gãy liên mấu chuyển không vững (AO nhóm 2 + 3): gãy nhiều mảnh, có tổn thương thành ngoài hoặc thành sau trong, mất vững.

Gán nhãn dữ liệu

Sau khi đã có quy ước phân loại 4 nhóm thống nhất, việc gán nhãn cho bộ dữ liệu được thực hiện bằng công cụ gán nhãn chuyên biệt trên Roboflow, cho phép đánh dấu vùng gãy xương bằng khung giới hạn trực tiếp trên hình ảnh X-quang.

Mỗi ảnh được gán nhãn gồm hai thông tin: vị trí khung giới hạn xác định vùng tổn thương và nhãn lớp tương ứng (1 trong 4 nhóm). Việc xác định đúng cả vị trí và loại gãy là yêu cầu bắt buộc để mô hình học đúng đặc trưng.



Hình 1. Hình ảnh X-quang được dán nhãn theo nhóm đã được quy ước

Tiền xử lý và tăng cường dữ liệu

Toàn bộ ảnh X-quang sau khi gán nhãn trải qua quy trình tiền xử lý: Auto-Orient, Resize (640 × 640 pixel) và tăng cường dữ liệu: Lập ngang, xoay (-15° đến +15°), nghiêng (Biến dạng chéo ±10° theo chiều ngang, ±10° theo chiều dọc) nhằm đảm bảo tính đồng nhất đầu vào và nâng cao khả năng khái quát hóa của mô hình. Tăng cường dữ liệu tạo ra 3 phiên bản ngẫu nhiên từ mỗi ảnh gốc, tổng số ảnh huấn luyện tăng lên 3 lần so với bộ dữ liệu gốc.

Mô hình RF-DETR

Phát hiện gãy xương vùng háng trên X-quang là bài toán phát hiện vật thể trong y học, đòi hỏi mô hình vừa xác định được vị trí tổn thương vừa phân loại đúng nhóm gãy xương.

Trong nghiên cứu này, chúng tôi đã sử dụng mô hình Roboflow RF-DETR Object Detection một kiến trúc phát hiện vật thể dựa trên cơ chế Transformer (Detection Transformer - DETR)

được phát triển bởi Roboflow. Trong giai đoạn huấn luyện, mô hình được cung cấp nhiều hình ảnh đã được đánh dấu sẵn vị trí và loại tổn thương. Từ các dữ liệu này, mô hình học cách nhận biết những đặc điểm quan trọng của tổn thương như hình dạng, vị trí, đường viền và mối liên quan với các cấu trúc xung quanh. Thông qua quá trình học, mô hình dần điều chỉnh các tham số bên trong để dự đoán ngày càng gần với kết quả đã được gán nhãn. Sau khi được huấn luyện, mô hình có thể áp dụng trên những hình ảnh mới chưa được gán nhãn. Khi tiếp nhận ảnh đầu vào, RF-DETR sẽ phân tích toàn bộ hình ảnh, tập trung vào các vùng nghi ngờ, sau đó dự đoán vị trí tổn thương bằng khung giới hạn và phân loại tổn thương tương ứng.

Đánh giá giá trị của mô hình

Sau khi hoàn tất huấn luyện, mô hình được áp dụng trên bộ dữ liệu đánh giá độc lập. Kết quả dự đoán của mô hình được so sánh với kết quả đọc phim của bác sĩ đa khoa với tiêu chuẩn vàng đã được thống nhất.

Các chỉ số của mô hình:

+ Intersection over Union (IoU): đo mức độ chồng lấp giữa khung giới hạn dự đoán và khung giới hạn chuẩn. Một dự đoán được xác nhận là dương thực (True Positive - TP) khi $IoU \geq 0,5$.
 $IoU = \text{Diện tích giao nhau} / \text{Diện tích hợp nhất}$

+ Precision: tỷ lệ dự đoán đúng trong tổng số dự đoán dương tính:

$$\text{Precision} = TP / (TP + FP)$$

+ Recall: tỷ lệ vùng gãy được phát hiện trên tổng số vùng gãy thực sự:

$$\text{Recall} = TP / (TP + FN)$$

+ Mean Average Precision (mAP@50): chỉ số tổng hợp, tính bằng trung bình diện tích dưới đường cong Precision-Recall (AUC-PR) trên toàn bộ các lớp tại ngưỡng $IoU = 0,5$. Đây là chỉ số phổ biến nhất để đánh giá toàn diện hiệu năng của mô hình phát hiện vật thể.

Chỉ số đánh giá kết quả đọc phim của mô hình so với bác sĩ đa khoa: Bác sĩ đa khoa chỉ được cung cấp 93 hình ảnh X-quang khung chậu, không kèm theo bất kỳ thông tin gì thêm như: tên, tuổi, chẩn đoán.

+ Accuracy (Độ chính xác): tỷ lệ các trường hợp được phân loại đúng trên tổng số: $Accuracy = (TP + TN) / (TP + TN + FP + FN)$

+ Cohen's Kappa (κ): đánh giá mức độ đồng thuận giữa kết quả mô hình và tiêu chuẩn vàng, có hiệu chỉnh yếu tố ngẫu nhiên. Giá trị $\kappa > 0,80$ được xem là đồng thuận rất tốt; 0,61 - 0,80 là tốt; $\kappa = (Accuracy - pe) / (1 - pe)$

Phương pháp thu thập và xử lý số liệu

Các hình ảnh X-quang được thu thập và phân nhóm bởi hai bác sĩ chuyên khoa chấn thương chỉnh hình theo qui ước trước đó. Mô hình được huấn luyện trên Roboflow. Tất cả số liệu được thu thập và xử lý bằng phần mềm

SPSS 22. Sự khác biệt giữa mô hình và bác sĩ đa khoa được kiểm định bằng phép kiểm định McNemar trên cùng một mẫu. Mức ý nghĩa thống kê được xác định khi $p < 0,05$.

3. Đạo đức nghiên cứu

Nghiên cứu được chấp thuận của Hội đồng đạo đức trong nghiên cứu Y sinh học, Đại học Y Dược Cần Thơ, số 24.343.HV/PCT-HĐĐĐ ngày 28/06/2024. Thông tin cá nhân của người bệnh được bảo mật tuyệt đối, dữ liệu được mã hóa và chỉ sử dụng cho mục đích nghiên cứu.

III. KẾT QUẢ

1. Phân bố dữ liệu

Bộ dữ liệu xây dựng mô hình gồm 1,483 ảnh X-quang được thu thập và gán nhãn thành 4 nhóm theo tiêu chuẩn đã quy ước. Một bộ dữ liệu khác gồm 93 ảnh X-quang dùng để đánh giá mô hình.

Bảng 1. Phân bố dữ liệu

Loại tổn thương	Xây dựng mô hình	Đánh giá mô hình
Nhóm 1, n (%)	362 (24,4)	6 (6,5)
Nhóm 2, n (%)	426 (28,7)	45 (48,4)
Nhóm 3, n (%)	329 (22,2)	25 (26,9)
Nhóm 4, n (%)	366 (24,7)	17 (18,3)
Tổng	1483	93

Bộ dữ liệu huấn luyện gồm 1483 ảnh, trong đó nhóm 2 chiếm tỷ lệ cao nhất (28,7%), và nhóm 3 chiếm thấp nhất (22,2%). Tập đánh giá gồm 93 trường hợp, trong đó nhóm 1 chiếm tỷ

lệ cao nhất (48,4%), và nhóm 2 chiếm tỷ lệ thấp nhất (6,5%).

2. Kết quả các chỉ số của mô hình sau khi huấn luyện

Bảng 2. Kết quả các chỉ số của mô hình sau khi huấn luyện RF-DETR

Chỉ số	mAP@50	Precision	Recall
RF-DETR	93,9%	91,8%	90,6%

Mô hình RF-DETR sau huấn luyện đạt mAP@50 = 93,9%, Precision = 91,8% và Recall = 90,6% trên tập kiểm định. Kết quả cho thấy

mô hình có khả năng phát hiện và khoanh vùng gãy xương vùng háng với độ chính xác cao và tỷ lệ bỏ sót thấp, đủ điều kiện để tiến hành

đánh giá chẩn đoán trên tập dữ liệu độc lập. Trong đó: $mAP@50$: Mean Average Precision tại ngưỡng $IoU \geq 0,50$; $Precision = TP/(TP+FP)$; $Recall = TP/(TP+FN)$. Tập kiểm định độc lập

với tập huấn luyện.

3. Kết quả đánh giá giá trị chẩn đoán của mô hình

Bảng 3. Kết quả đọc phim của mô hình

Tiêu chuẩn vàng/AI	Nhóm 1	Nhóm 2	Nhóm 3	Nhóm 4
Nhóm 1	6	0	0	0
Nhóm 2	3	42	0	0
Nhóm 3	0	0	25	0
Nhóm 4	0	0	1	16

Trong 93 ca, mô hình phân loại đúng 89/93 trường hợp và 4 trường hợp phân loại sai. Có 3/45 trường hợp Nhóm 2 bị phân loại nhầm

thành Nhóm 1, và 1/17 trường hợp Nhóm 4 bị phân loại nhầm thành Nhóm.

Bảng 4. Kết quả so sánh chẩn đoán của mô hình và bác sĩ đa khoa

Phương pháp	Số ca đúng/ Tổng	Accuracy	Kappa(κ)
Mô hình AI	89/ 93	95,7%	0,936
Bác sĩ đa khoa	73/ 93	78,5%	0,669
Kiểm định McNemar ($b = 0, c = 16$)	$\chi^2 = 14,06$	$p = 0,0002$	

Kappa: Cohen's Kappa (< 0,40: kém; 0,41 - 0,60: vừa; 0,61-0,80: tốt; > 0,80: rất tốt). b: số ca AI sai - BSKĐ đúng; c: số ca AI đúng - BSKĐ sai. Mức ý nghĩa thống kê: $p < 0,05$.

Mô hình AI đạt Accuracy = 95,7% với Kappa = 0,936 (mức rất tốt), vượt trội so với bác sĩ đa khoa (Accuracy = 78,5%, Kappa = 0,669, mức tốt). Có 16 trường hợp mô hình AI phân loại đúng trong khi bác sĩ đa khoa phân loại sai ($c = 16$) và không có trường hợp nào theo chiều ngược lại ($b = 0$). Kiểm định McNemar cho $\chi^2 = 14,06$, $p = 0,0002$ ($p < 0,001$), cho thấy mô hình AI có khả năng phân loại gãy xương vùng háng tốt hơn bác sĩ đa khoa một cách có ý nghĩa thống kê.

IV. BÀN LUẬN

Phân bố dữ liệu

Bộ dữ liệu huấn luyện gồm 1.483 ảnh

X-quang, sự phân bố tương đối đồng đều giữa bốn nhóm là điều kiện thuận lợi để mô hình học tập cân bằng trên các lớp, giảm nguy cơ thiên lệch phân loại về nhóm đa số, một vấn đề phổ biến trong các bộ dữ liệu y tế. Để tăng tính đa dạng của tập huấn luyện, chúng tôi áp dụng tăng cường dữ liệu gồm lật ngang, xoay ngẫu nhiên $\pm 15^\circ$ và biến dạng chéo $\pm 10^\circ$, tạo ra 4.449 ảnh từ 1.483 ảnh gốc. Cách làm này phù hợp với thực hành chuẩn trong huấn luyện mô hình học sâu trên dữ liệu y tế có cỡ mẫu hạn chế, giúp cải thiện khả năng tổng quát hóa sang dữ liệu chưa thấy. Các nghiên cứu tương tự như González và cộng sự (2025) sử dụng YOLOv8 trên 986 ảnh được chọn từ hơn

10.000 ảnh cũng áp dụng tăng cường dữ liệu để cải thiện độ ổn định huấn luyện.⁸

Kết quả các chỉ số của mô hình sau khi huấn luyện

Trên tập kiểm định nội bộ, mô hình RF-DETR đạt $mAP@50 = 93,9\%$, Precision = 91,8% và Recall = 90,6%. Đây là những chỉ số phản ánh khả năng phát hiện và định vị vùng gãy xương với độ chính xác cao, tỷ lệ bỏ sót thấp và tỷ lệ dương tính giả được kiểm soát tốt. Kết quả này phù hợp với xu hướng chung của các nghiên cứu ứng dụng học sâu trong phát hiện gãy xương trên X-quang. So sánh trực tiếp, Beyaz và cộng sự (2020) sử dụng CNN chỉ đạt sensitivity 83% và specificity 73% trong phát hiện gãy cổ xương đùi, thấp hơn đáng kể so với recall 90,6% của mô hình chúng tôi.⁹

Sự lựa chọn kiến trúc RF-DETR có cơ sở vững chắc. RF-DETR tích hợp backbone DINOv2 với deformable attention, cho phép nắm bắt bối cảnh toàn cục của ảnh mà không phụ thuộc vào anchor box hay NMS.⁷ Điều này có ý nghĩa đặc biệt trong bài toán phát hiện gãy cổ xương đùi và gãy liên mấu chuyển, vì các tổn thương có hình thái đa dạng, ranh giới không rõ và đôi khi bị chồng lấp cấu trúc giải phẫu. Các nghiên cứu so sánh cho thấy mô hình Transformer-based hội tụ nhanh hơn và duy trì hiệu năng ổn định hơn so với CNN truyền thống trong các môi trường phát hiện đối tượng phức tạp.¹⁰ Ngoài ra, RF-DETR đã vượt qua ngưỡng 60% mAP trên benchmark COCO, cao nhất trong số các mô hình real-time tại thời điểm công bố cho thấy tiềm năng ứng dụng mạnh mẽ trong các bài toán đặc thù như phân tích ảnh y khoa.⁷

Một điểm mạnh của mô hình là khả năng xử lý đồng thời cả hai loại gãy trong một pipeline phát hiện đối tượng đầu cuối duy nhất, trong khi phần lớn các nghiên cứu trước đây chỉ tập trung vào một loại gãy. González và cộng sự (2025)

mặc dù phân loại cả ba vị trí gãy bằng YOLOv8 nhưng đạt accuracy phân loại chỉ tiết 86,2% thấp hơn so với 95,7% trên dữ liệu bệnh viện của chúng tôi.⁸ Điều này gợi ý rằng việc chuẩn hóa chặt chẽ quy trình gán nhãn bởi hai bác sĩ chuyên khoa chỉnh hình có kinh nghiệm trên 10 năm, với tiêu chuẩn loại bỏ ảnh khi có bất đồng, đã tạo ra bộ dữ liệu chất lượng cao, đóng góp quan trọng vào hiệu năng của mô hình.

Đánh giá kết quả đọc phim mô hình so với bác sĩ đa khoa

Mô hình RF-DETR phân loại đúng 89/93 trường hợp. Các trường hợp sai cho thấy mô hình không nhầm lẫn giữa tổn thương gãy cổ xương đùi và gãy liên mấu chuyển, mà chủ yếu sai lệch giữa các phân nhóm có đặc điểm hình ảnh gần nhau. Cụ thể, có 3 trường hợp gãy cổ xương đùi di lệch được phân loại thành gãy cổ xương đùi không di lệch và 1 trường hợp gãy liên mấu chuyển không vững được phân loại thành gãy liên mấu chuyển vững.

Các sai lệch này có thể liên quan đến tính chất khó phân biệt trên X-quang thường quy, đặc biệt ở các trường hợp đường gãy mờ, mức độ di lệch ít hoặc dấu hiệu mất vững không rõ ràng. Chất lượng phim, tư thế bệnh nhân khi chụp dẫn đến sự chồng lấp của các cấu trúc giải phẫu vùng háng cũng có thể ảnh hưởng đến khả năng nhận diện của mô hình. Ngoài ra, dữ liệu đầu vào để xây dựng mô hình chưa đủ lớn nên việc học của mô hình cũng còn hạn chế.

Trên tập 93 ảnh X-quang bệnh viện, mô hình RF-DETR đạt accuracy 95,7% (KTC 95%: 89,5 - 98,3%) với hệ số kappa = 0,936, vượt trội có ý nghĩa thống kê so với bác sĩ đa khoa (accuracy 78,5%, kappa = 0,669, $p < 0,001$, kiểm định McNemar). Trong 93 trường hợp, có 16 ca mô hình AI phân loại đúng trong khi bác sĩ đa khoa phân loại sai, và không có trường hợp nào theo chiều ngược lại ($b = 0$, $c = 16$). Kết quả này tương đồng với các nghiên cứu quốc

tế: Sato và cộng sự (2021) phát triển hệ thống CAD dựa trên EfficientNet-B4 từ hơn 10.000 ảnh đa trung tâm, đạt accuracy 96,1% và AUC 0,99, đồng thời chứng minh rằng hệ thống AI cải thiện đáng kể khả năng chẩn đoán của bác sĩ nội trú.⁶ Hsieh và cộng sự (2023) phát triển mô hình DAFDNet đạt accuracy > 94,8% trong phát hiện gãy Garden I-II và vượt trội cả bác sĩ đa khoa lẫn bác sĩ chỉnh hình trong định vị vùng gãy.¹¹ Yang và cộng sự (2025) xây dựng mô hình phân loại Garden trên X-quang sử dụng CT 3D làm tiêu chuẩn vàng, đạt accuracy 90,6% và Dice similarity coefficient 88,6%.¹²

Hệ số kappa = 0,936 của mô hình có ý nghĩa lâm sàng quan trọng. Phân loại Garden vốn có độ nhất quán liên bác sĩ rất thấp, với kappa chỉ 0,03 - 0,56 trên Xquang thường, và ngay cả khi đơn giản hóa thành hệ nhị phân (di lệch/không di lệch) cũng chỉ đạt 0,67 - 0,77.⁴ Tương tự, phân loại AO cho gãy liên mấu chuyển cũng ghi nhận mức đồng thuận trung bình đến khá giữa các bác sĩ.¹³ Trong bối cảnh đó, việc mô hình AI đạt kappa = 0,936 cho thấy tiềm năng chuẩn hóa quy trình phân loại, giảm phụ thuộc vào yếu tố chủ quan và kinh nghiệm người đọc phim.

Khoảng cách đáng kể giữa AI (95,7%) và bác sĩ đa khoa (78,5%) nhấn mạnh vai trò tiềm năng của công cụ hỗ trợ AI tại các cơ sở y tế tuyến cơ sở hoặc vùng thiếu bác sĩ chuyên khoa chỉnh hình, thực trạng phổ biến tại nhiều địa phương ở Việt Nam. Gãy Garden I-II được điều trị bằng kết hợp xương, trong khi Garden III-IV ở người cao tuổi thường cần thay khớp háng; gãy liên mấu chuyển vững và không vững cũng có các lựa chọn phương tiện kết hợp xương khác nhau.¹³ Phân loại sai dẫn đến lựa chọn điều trị không phù hợp, kéo dài thời gian phục hồi và gia tăng biến chứng. Với accuracy 95,7% và kappa 0,936, mô hình RF-DETR có tiềm năng hỗ trợ bác sĩ đa khoa trong sàng lọc ban đầu và định hướng điều trị.

Tuy nhiên, nghiên cứu có một số hạn chế cần lưu ý. Tập đánh giá ngoài chỉ gồm 93 ảnh từ một cơ sở y tế, hạn chế khả năng tổng quát hóa sang các bệnh viện khác với thiết bị và quy trình chụp phim khác nhau. Phân bố mẫu không cân bằng (nhóm 1 chỉ 6,5%) có thể ảnh hưởng đến ước tính hiệu năng trên một số phân lớp. Ngoài ra, việc dùng X-quang để phân loại garden mang ý kiến chủ quan của người đọc, tốt nhất nên có phim CT scan để kết luận chẩn đoán sau cùng, tuy nhiên ở Việt Nam không có nhiều bệnh viện làm điều này. Những hạn chế này cần được giải quyết trong các nghiên cứu tiếp theo với cỡ mẫu lớn hơn, thiết kế đa trung tâm và so sánh trực tiếp với kết quả CT scan.

V. KẾT LUẬN

Mô hình AI có khả năng phân loại gãy xương chính xác và ổn định, có tiềm năng ứng dụng như một công cụ hỗ trợ bác sĩ tại các cơ sở y tế, đặc biệt trong bối cảnh thiếu bác sĩ chuyên khoa chấn thương chỉnh hình.

TÀI LIỆU THAM KHẢO

1. Lewiecki EM, Wright NC, Curtis JR, et al. Hip fracture trends in the United States, 2002 to 2015. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*. Mar 2018; 29(3): 717-722. doi:10.1007/s00198-017-4345-0.
2. Cannon J, Silvestri S, Munro M. Imaging choices in occult hip fracture. *The Journal of emergency medicine*. Aug 2009; 37(2): 144-52. doi:10.1016/j.jemermed.2007.12.039.
3. Dominguez S, Liu P, Roberts C, Mandell M, Richman PB. Prevalence of traumatic hip and pelvic fractures in patients with suspected hip fracture and negative initial standard radiographs-a study of emergency department

- patients. *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine*. Apr 2005; 12(4): 366-9. doi:10.1197/j.aem.2004.10.024.
4. Kazley JM, Banerjee S, Abousayed MM, Rosenbaum AJ. Classifications in Brief: Garden Classification of Femoral Neck Fractures. *Clinical orthopaedics and related research*. Feb 2018; 476(2): 441-445. doi:10.1007/s11999.00000000000000066.
5. Krogue JD, Cheng KV, Hwang KM, et al. Automatic Hip Fracture Identification and Functional Subclassification with Deep Learning. *Radiology Artificial intelligence*. Mar 2020; 2(2): e190023. doi:10.1148/ryai.2020190023.
6. Sato Y, Takegami Y, Asamoto T, et al. Artificial intelligence improves the accuracy of residents in the diagnosis of hip fractures: a multicenter study. *BMC musculoskeletal disorders*. May 3 2021; 22(1): 407. doi:10.1186/s12891-021-04260-2.
7. Robinson I, Robicheaux P, Popov M, Ramanan D, Peri NJapa. RF-DETR: neural architecture search for real-time detection transformers. 2025; doi:10.48550/arXiv.2511.09554.
8. González G, Galant J, Salinas JM, et al. Classification and segmentation of hip fractures in x-rays: highlighting fracture regions for interpretable diagnosis. *Insights into imaging*. Apr 15 2025; 16(1): 86. doi:10.1186/s13244-025-01958-y.
9. Beyaz S, Açıcı K, Sümer E. Femoral neck fracture detection in X-ray images using deep learning and genetic algorithm approaches. *Joint diseases and related surgery*. 2020; 31(2): 175-183. doi:10.5606/ehc.2020.72163.
10. Sapkota R, Cheppally RH, Sharda A, Karkee MJapa. Rf-detr object detection vs yolov12: A study of transformer-based and cnn-based architectures for single-class and multi-class greenfruit detection in complex orchard environments under label ambiguity. 2025; doi:10.48550/arXiv.2504.13099.
11. Hsieh SL, Chiang JL, Chuang CH, Chen YY, Hsu CJ. A Computer-Assisted Diagnostic Method for Accurate Detection of Early Nondisplaced Fractures of the Femoral Neck. *Biomedicine*. 2023; 11(11): 3100. doi:10.3390/biomedicine11113100.
12. Yang J, Park J, Park K, et al. Garden classification of femoral neck fracture using deep-learning algorithm. *Scientific Reports*. 2025/12/16 2025; 15(1): 43926. doi:10.1038/s41598-025-27766-4.
13. Chang SM, Hou ZY, Hu SJ, Du SC. Intertrochanteric Femur Fracture Treatment in Asia: What We Know and What the World Can Learn. *The Orthopedic clinics of North America*. Apr 2020; 51(2): 189-205. doi:10.1016/j.ocl.2019.11.011.

Summary

APPLICATION OF A DEEP LEARNING MODEL FOR DETECTION AND CLASSIFICATION OF FEMORAL NECK AND INTERTROCHANTERIC FRACTURES ON RADIOGRAPHS

Femoral neck and intertrochanteric fractures are common injuries in hip trauma. Accurate detection and classification play an important role in selecting appropriate treatment methods but largely depend on physician experience. This study utilized the RF-DETR model as the foundation for developing a deep learning model to detect and classify femoral neck fractures and intertrochanteric fractures on radiographic images. The RF-DETR model was trained and internally validated on 1,483 publicly available radiographic images. External evaluation was performed on 93 hospital radiographs, with the reference standard defined by the consensus of two orthopedic specialists. On the internal validation set, the model achieved an mAP@50 of 93.9%, precision of 91.8%, and recall of 90.6%. On the hospital dataset, the model achieved an accuracy of 95.7% (95% CI: 89.5 – 98.3%), higher compared to general practitioners at 78.5% (95% CI: 69.1 – 85.6%), $p < 0.001$ (McNemar test), with a kappa coefficient of 0.936. Aforementioned results suggest the RF-DETR deep learning model demonstrated high performance and shows potential for supporting clinical diagnosis.

Keywords: Artificial intelligence, femoral neck fracture, intertrochanteric fracture, radiograph, RF-DETR.