

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
KHOA CÔNG NGHỆ THÔNG TIN

Trần Hữu Nam Thiên - Lê Ân Thạch

PHÂN LỚP ẢNH UNG THƯ XƯƠNG  
DỰA TRÊN MÔ HÌNH BỘ CHUYỂN ĐỔI

KHÓA LUẬN TỐT NGHIỆP CỦ NHÂN  
CHƯƠNG TRÌNH CHUẨN

Tp. Hồ Chí Minh, tháng 07/2024

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
KHOA CÔNG NGHỆ THÔNG TIN

Trần Hữu Nam Thiên - 20120193  
Lê Ân Thạch - 20120186

PHÂN LỐP ẢNH UNG THƯ XƯƠNG  
DỰA TRÊN MÔ HÌNH BỘ CHUYỂN ĐỔI

KHÓA LUẬN TỐT NGHIỆP CỦ NHÂN  
CHƯƠNG TRÌNH CHUẨN

GIẢNG VIÊN HƯỚNG DẪN  
PGS. TS. Lý Quốc Ngọc

Tp. Hồ Chí Minh, tháng 07/2024

# **Lời cam đoan**

Khóa luận này là công trình nghiên cứu của riêng chúng tôi, được thực hiện dưới sự hướng dẫn của PGS. TS. Lý Quốc Ngọc

Các số liệu, kết quả nghiên cứu, phân tích được trình bày trong khóa luận này là trung thực và chưa từng được công bố trong bất kỳ công trình nghiên cứu nào khác.

Nếu có điều gì không đúng trong lời cam đoan trên, chúng tôi xin hoàn toàn chịu trách nhiệm.

# Lời cảm ơn

Nhóm khóa luận xin gửi lời cảm ơn chân thành và sâu sắc nhất đến những người đã hỗ trợ và đồng hành cùng khóa luận trong suốt quá trình nghiên cứu và hoàn thành khóa luận này.

Trước hết, nhóm xin bày tỏ lòng biết ơn đến Thầy Lý Quốc Ngọc, người đã tận tình chỉ dẫn, hỗ trợ và chia sẻ những kinh nghiệm quý báu trong suốt quá trình thực hiện khóa luận. Sự chỉ dẫn và hỗ trợ của Thầy đã giúp nhóm hoàn thành công trình này một cách tốt nhất.

Nhóm khóa luận xin cảm ơn bạn bè, người thân đã luôn ở bên cạnh, động viên và khích lệ nhóm vượt qua những khó khăn trong suốt quá trình nghiên cứu và viết khóa luận.

Đồng thời, nhóm xin cảm ơn Bác sĩ Lý Thanh Trang (Phó trưởng khoa Chẩn đoán hình ảnh Bệnh viện Nguyễn Tri Phương) đã dành thời gian quý báu để hỗ trợ kiểm chứng kết quả của khóa luận.

Đặc biệt, nhóm xin bày tỏ lòng biết ơn sâu sắc đến Khoa Công nghệ Thông tin trường Đại học Khoa học Tự nhiên đã tạo điều kiện thuận lợi, cung cấp máy chủ cần thiết cho quá trình nghiên cứu của nhóm.

Cuối cùng, nhóm khóa luận xin cảm ơn tất cả những ai đã trực tiếp hoặc gián tiếp giúp đỡ, hỗ trợ nhóm trong suốt quá trình thực hiện khóa luận này. Khóa luận này là kết quả của sự nỗ lực và cố gắng không ngừng của cả nhóm, đồng thời cũng là sự giúp đỡ quý báu của rất nhiều người. Nhóm hy vọng rằng những kết quả và kiến thức từ khóa luận này sẽ đóng góp một phần nhỏ vào sự phát triển của lĩnh vực Thị giác máy tính, đặc biệt là Thị giác máy tính trong y khoa.

Xin chân thành cảm ơn!

# Đề cương chi tiết



**fit@hcmus**

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN  
KHOA CÔNG NGHỆ THÔNG TIN

ĐỀ CƯƠNG KHOÁ LUẬN TỐT NGHIỆP

# PHÂN LỚP ẢNH UNG THƯ XƯƠNG Dựa trên mô hình bộ chuyển đổi

(*Bone tumor images classification based on Transformer model*)

## 1 THÔNG TIN CHUNG

**Người hướng dẫn:**

- PGS. TS. Lý Quốc Ngọc (Khoa Công nghệ Thông tin)

**[Nhóm] Sinh viên thực hiện:**

1. Trần Hữu Nam Thiên (MSSV: 20120193 )
2. Lê Ân Thạch (MSSV: 20120186 )

**Loại đề tài:** Nghiên cứu

**Thời gian thực hiện:** Từ 12/2023 đến 07/2024

## 2 NỘI DUNG THỰC HIỆN

### 2.1 Giới thiệu về đề tài

Ung thư xương là một loại bệnh hiếm, nhưng có thể gây ra những hậu quả nghiêm trọng nếu không phát hiện và điều trị kịp thời. Việc chẩn đoán sớm và chính xác là quyết định quan trọng để cứu sống và cải thiện chất lượng cuộc sống của những người mắc bệnh. Đặc điểm khó chẩn đoán của ung thư xương là một thách thức đối với các chuyên gia y tế. Do đó, việc phát triển một hệ thống phân lớp ảnh y khoa xương có thể giúp cung cấp thông tin hữu ích và hỗ trợ quyết định chẩn đoán.

Việc lựa chọn đề tài "Phân lớp ảnh ung thư xương dựa trên mô hình Bộ chuyển đổi" là hợp lý trong điều kiện khoa học công nghệ và thị giác máy tính hiện nay. Tầm quan trọng của chẩn đoán sớm, tiềm năng của công nghệ học máy và thị giác máy tính, đa dạng và phức tạp của ảnh y khoa xương là những yếu tố chính đã thúc đẩy quyết định này. Việc này có thể mang lại nhiều lợi ích quan trọng trong việc phát hiện sớm và điều trị bệnh, từ đó cứu sống và cải thiện chất lượng cuộc sống của những người mắc bệnh ung thư xương.

### 2.2 Mục tiêu đề tài

Mục tiêu của đề tài "Phân lớp ảnh ung thư xương dựa trên mô hình Bộ chuyển đổi" là phát triển một hệ thống có khả năng phân loại ảnh ung thư xương thành các nhóm bệnh ung thư xương cụ thể. Ngoài ra, mục tiêu phụ là giải thích được lý do vì sao lại có kết quả phân loại đây. Mục tiêu này nhằm cung cấp thông tin hữu ích để hỗ trợ quyết định chẩn đoán của các chuyên gia y tế.

### 2.3 Phạm vi của đề tài

Đối tượng nghiên cứu: Đối tượng nghiên cứu chính là các ảnh X-quang ung thư xương có nguồn từ bệnh viện. Các ảnh này sẽ được sử dụng để xây dựng và đánh

giá hệ thống phân lớp ảnh ung thư xương.

Thực thể liên quan

- Các phương pháp học máy và học sâu: Bộ chuyển đổi, Mạng nơ-ron tích chập, Mô hình kết hợp.
- Các phương pháp xử lý ảnh: trích xuất đặc trưng, phân lớp ảnh, tăng cường ảnh và thay đổi kích thước ảnh.

Tập dữ liệu:

- Tập dữ liệu ảnh y khoa MedMNIST dùng để tiền huấn luyện mô hình.
- Tập dữ liệu ảnh ung thư xương có nguồn từ bệnh viện Chấn thương Chính hình đã được xử lý để huấn luyện mô hình.

Giới hạn và ràng buộc

- Hạn chế về tài nguyên: Máy không đủ điều kiện để huấn luyện, môi trường không thích hợp để huấn luyện mô hình.
- Hạn chế của dữ liệu: Lượng dữ liệu không đủ để đạt hiệu suất cao, có ảnh nhiễu làm ảnh hưởng đến quá trình huấn luyện, ảnh chất lượng thấp.
- Ràng buộc về luật pháp và đạo đức: Đảm bảo ứng dụng tuân thủ luật pháp và đạo đức, không sử dụng dữ liệu bệnh nhân ngoài mục đích nghiên cứu đề tài này.

## 2.4 Phát biểu bài toán

Bài toán phân lớp ảnh ung thư xương dựa trên mô hình Bộ chuyển đổi bao gồm:

- Đầu vào: Ảnh X-quang của bộ phận cụ thể chứa khối u.
- Đầu ra: Ảnh đã được phân lớp vào các lớp bệnh cụ thể ứng với từng vị trí xương:
  - Cánh tay: Bọc xương, U sụn, Sarcoma, U đại bào.

- Đùi: U sụn, Di căn, Sarcoma, U đại bào.
- Chày: Sarcoma, U đại bào.

## 2.5 Cách tiếp cận dự kiến

### a) Nghiên cứu liên quan

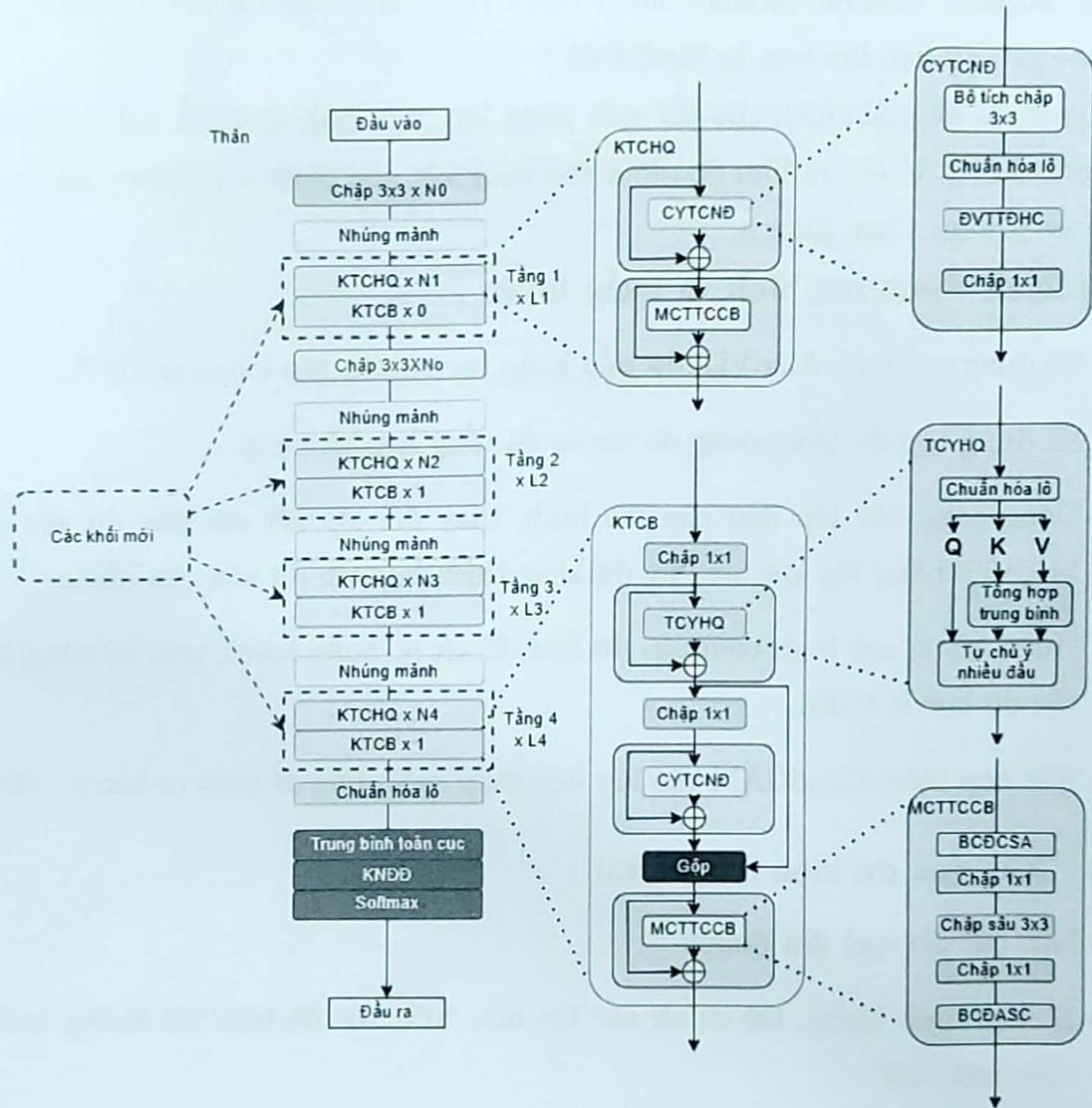
Trong lĩnh vực phân loại ảnh y khoa, nghiên cứu gần đây đã chứng minh sức mạnh của mô hình Bộ chuyển đổi. Dưới đây là một số nghiên cứu tiêu biểu dựa trên các tiêu chí: Có trí tuệ nhân tạo khả thi (TTNTKD), Phương pháp học, Hiệu suất theo độ chính xác.

STT	Tên model	TTNTKD	Phương pháp học	Tập dữ liệu	Hiệu suất (độ chính xác)
[1]	TransMIL	Có	Tự giám sát yếu	TCGA-RCC	0.946
[2]	RadioTransformer	Có	Tự giám sát	Radiography	0.9998
[3]	HatNet	Có	Tự giám sát	breast biopsy	0.71
[4]	MedViT	Không	Tự giám sát	ChestMNIST	0.956
[5]	TransMed	Không	Tự giám sát	Hospital	0.889

Bảng 1: Bảng so sánh các mô hình Bộ chuyển đổi trong phân lớp ảnh y khoa

### b) Phương pháp lựa chọn dự kiến

Dựa trên những nghiên cứu trước đó, mô hình dự kiến cho đề tài phân loại ảnh ung thư xương dựa trên mô hình Bộ chuyển đổi sẽ là: **MedViT**



Hình 1: Kiến trúc tổng quan của mô hình MedViT đã được việt hóa lại tên các thành phần.

MedViT là một mô hình kết hợp giữa Mạng nơ-ron tích chập và Bộ chuyển đổi, giới thiệu 2 khối mới là khối tích chập hiệu quả và khối chuyển đổi cục bộ. Hai khối này giúp cho mô hình có được khả năng học sâu của Mạng nơ-ron tích chập

và khả năng tự chú ý của Bộ chuyển đổi, từ đây mang lại hiệu suất cao trong việc phân loại và phát hiện các bất thường y khoa từ hình ảnh.

Phương pháp huấn luyện được sử dụng là phương pháp học chuyển tiếp, sử dụng mô hình MedViT đã được huấn luyện trước trên tập dữ liệu ChestMNIST (thuộc tập dữ liệu lớn hơn là MedMNIST [6]).

Ngoài ra, để giải thích cho kết quả phân lớp, mô hình sẽ được tích hợp thêm GradCAM [7] để tạo ra bản đồ nhiệt cho từng kết quả phân lớp. Vùng màu nóng (đỏ) sẽ là vùng chưa khối u.

### c) Tinh chỉnh mô hình và huấn luyện

- Sử dụng mô hình MedViT đã tiền huấn luyện trên tập ChestMNIST.
- Sử dụng tập dữ liệu xương do nhóm đã tổng hợp và xử lý.
- Đóng băng các lớp đầu của mô hình, thay đổi lớp kết nối đầy đủ gốc của MedViT bằng lớp kết nối đầy đủ khác thích hợp với dữ liệu của nhóm.
- Huấn luyện mô hình trên tập dữ liệu đã xử lý, huấn luyện trên 10 vòng lặp, tốc độ học là 0.005.
- Kết hợp thêm GradCAM vào lớp tích chập cuối cùng để xuất ra bản đồ nhiệt.

## 2.6 Kết quả dự kiến của đề tài

Dự kiến các kết quả đạt được:

- Số liệu định lượng: Độ chính xác lớn hơn 70%, tốc độ thực thi không quá 3s trên một ảnh.
- Sản phẩm đầu ra: Một mô hình có thể phân lớp ảnh chứa bệnh ung thư xương cũng như giải thích lý do phân lớp dựa vào bản đồ nhiệt.

## 2.7 Kế hoạch thực hiện

STT	Nội dung thực hiện	Thời gian thực hiện
1	Nghiên cứu về mục tiêu và phạm vi của đề tài	Từ 01/12/2023 đến 20/01/2024
2	Tìm hiểu các công trình liên quan	
3	Đối chiếu so sánh các công trình và lập bảng thống kê	
4	Tiến hành thu thập dữ liệu ảnh y khoa	Từ 21/01/2024 đến 28/02/2024
5	Nghiên cứu và tìm hiểu sâu mô hình dự kiến: MedViT	Từ 28/02/2024 đến 31/03/2024
6	Tiến hành tinh chỉnh và huấn luyện mô hình	Từ 31/03/2024 đến 30/04/2024
7	Kiểm thử mô hình	Từ 01/05/2024 đến 15/05/2024
8	Dánh giá và tổng kết	Từ 16/05/2024 đến 31/05/2024
9	Hoàn tất nghiên cứu và báo cáo	Từ 01/06/2024 đến 28/06/2024

Bảng 2: Bảng kế hoạch thực hiện của nhóm

## Tài liệu

- [1] Z. Shao, H. Bian, Y. Chen, Y. Wang, J. Zhang, X. Ji, and Y. Zhang†, “Transmil: Transformer based correlated multiple instance learning for whole slide image classification,” 2021.
- [2] M. Bhattacharya, S. Jain, and P. Prasanna, “Radiotransformer: A cascaded global-focal transformer for visual attention-guided disease classification,” 2022.
- [3] S. Mehta, X. Lu, D. Weaver†, J. G. Elmore, H. Hajishirzi, and L. Shapiro, “Hatnet: An end-to-end holistic attention network for diagnosis of breast biopsy images,” 2020.
- [4] O. N. Manzaria, H. Ahmadabadia, H. Kashianib, S. B. Shokouhia, and A. Ayatollahia, “Medvit: A robust vision transformer for generalized medical image classification,” 2023.
- [5] Y. Dai and Y. Gao, “Transmed: Transformers advance multi-modal medical image classification,” 2021.
- [6] J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister, and B. Ni, “Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification,” *Scientific Data*, vol. 10, no. 1, p. 41, 2023.
- [7] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

XÁC NHẬN  
CỦA NGƯỜI HƯỚNG DẪN  
(Ký và ghi rõ họ tên)

Nguyễn  
Lý Quí Ký

TP. Hồ Chí Minh, ngày 01 tháng 07 năm 2024  
NHÓM SINH VIÊN THỰC HIỆN  
(Ký và ghi rõ họ tên)

Lê Ân Thạch

Đinh

Trần Huân Nam Thiên

# Mục lục

Lời cam đoan	i
Lời cảm ơn	ii
Đề cương	iii
Mục lục	xiii
Danh sách hình	xvi
Danh sách bảng	xviii
Tóm tắt	xx
<b>1 Giới thiệu</b>	<b>1</b>
1.1 Bối cảnh chung . . . . .	1
1.2 Động lực nghiên cứu . . . . .	2
1.2.1 Ý nghĩa khoa học . . . . .	2
1.2.2 Ý nghĩa thực tiễn . . . . .	2
1.3 Phát biểu bài toán . . . . .	3
1.4 Thách thức của bài toán . . . . .	7
1.5 Đóng góp . . . . .	8
1.6 Cấu trúc của luận văn . . . . .	8
<b>2 Công trình liên quan</b>	<b>11</b>
2.1 Mạng nơ-ron tích chập . . . . .	11
2.2 Bộ chuyển đổi và các thành phần . . . . .	13

2.3	Mô hình kết hợp Bộ chuyển đổi và Mạng nơ-ron tích chập	16
2.4	Trí tuệ nhân tạo khả diễn	18
2.5	Các mô hình Bộ chuyển đổi dùng trong phân lớp ảnh y khoa	21
2.6	Phân loại vị trí ung thư xương trong y khoa	25
<b>3</b>	<b>Phương pháp</b>	<b>31</b>
3.1	Mô hình đề xuất	31
3.1.1	Kiến trúc Mô hình	31
3.1.2	Khối tích chập hiệu quả	34
3.1.3	Khối chuyển đổi chú ý hiệu quả	36
3.2	Các mô hình sử dụng trong luận văn	40
3.2.1	Mô hình phân lớp ứng với vị trí xương cánh tay	40
3.2.2	Mô hình phân lớp ứng với vị trí xương đùi	41
3.2.3	Mô hình phân lớp các ứng với vị trí xương chày	41
3.3	Mô hình trong giai đoạn ngoại tuyến	42
3.3.1	Mô-đun rút trích đặc trưng	46
3.3.2	Mô-đun phân lớp	46
3.3.2.1	Tính toán ở hàm mắt mát	46
3.3.2.2	Lan truyền ngược để tính toán độ dốc hàm mắt mát	47
3.3.2.3	Cập nhật trọng số của mô hình	48
3.4	Mô hình trong giai đoạn trực tuyến	48
3.4.1	Mô-đun rút trích đặc trưng	49
3.4.2	Mô-đun phân lớp	50
3.4.2.1	Tính toán ở lớp Kết nối đầy đủ	50
3.4.2.2	Đưa ra kết quả phân lớp cuối cùng	50
3.5	Giải thích kết quả đầu ra của mô hình	51
<b>4</b>	<b>Thực nghiệm</b>	<b>53</b>
4.1	Huấn luyện	53
4.1.1	Tập dữ liệu xương	53
4.1.2	Tiền xử lý dữ liệu	55
4.1.3	Tiền huấn luyện	56

4.1.4	Tinh chỉnh mô hình . . . . .	57
4.1.5	Cài đặt tham số huấn luyện và giải quyết quá khớp	58
4.2	Kết quả huấn luyện . . . . .	58
4.2.1	Độ đo kết quả huấn luyện . . . . .	58
4.2.1.1	Độ chính xác . . . . .	58
4.2.1.2	Diện tích dưới đường cong ROC . . . . .	59
4.2.2	Kết quả huấn luyện trên tập huấn luyện của các mô hình . . . . .	60
4.2.2.1	Mô hình xương cánh tay . . . . .	60
4.2.2.2	Mô hình xương đùi . . . . .	62
4.2.2.3	Mô hình xương chày . . . . .	64
4.2.3	Kết quả huấn luyện sau khi đã kiểm tra trên các tập dữ liệu minh chứng cho đóng góp số 1 . . . . .	65
4.2.4	Đánh giá kết quả huấn luyện . . . . .	68
4.3	Thực nghiệm phân lớp ảnh ung thư xương . . . . .	68
4.3.1	Mục đích thực nghiệm . . . . .	68
4.3.2	Kết quả thực nghiệm với các ảnh trong tập kiểm thử lần 1 . . . . .	68
4.3.3	Kết quả thực nghiệm với các ảnh trong tập kiểm thử lần 2 . . . . .	77
4.3.4	Nhận xét . . . . .	85
<b>5</b>	<b>Kết luận và hướng phát triển</b>	<b>86</b>
5.1	Kết luận . . . . .	86
5.2	Hướng phát triển . . . . .	88
<b>Tài liệu tham khảo</b>		<b>91</b>
<b>A</b>	<b>Tên thuật ngữ và viết tắt tiếng Việt, tiếng Anh</b>	<b>97</b>
<b>B</b>	<b>Kiểm chứng từ bác sĩ có chuyên môn</b>	<b>104</b>

# Danh sách hình

1.1	Sơ đồ ở giai đoạn ngoại tuyến . . . . .	5
1.2	Sơ đồ ở giai đoạn trực tuyến . . . . .	6
2.1	Bộ xương con người . . . . .	26
3.1	Kiến trúc tổng quan của BoneViT với tên các thành phần đã được việt hóa. . . . .	32
3.2	Đổi chiều BoneViT với mô hình MedViT gốc và mô hình Bộ chuyển đổi trực quan . . . . .	33
3.3	Mang mang cơ chế hiệu quả cục bộ vào Bộ chuyển đổi . .	35
3.4	Sơ đồ hoạt động của mô hình phân lớp xương cánh tay ở giai đoạn ngoại tuyến . . . . .	43
3.5	Sơ đồ hoạt động của mô hình phân lớp xương đùi ở giai đoạn ngoại tuyến . . . . .	44
3.6	Sơ đồ hoạt động của mô hình phân lớp xương chày ở giai đoạn ngoại tuyến . . . . .	45
3.7	Sơ đồ hoạt động của các mô hình trong tác vụ phân lớp ở giai đoạn trực tuyến . . . . .	49
4.1	Ví dụ một số mẫu của các tập dữ liệu . . . . .	55
4.2	Ví dụ ảnh được tăng cường dữ liệu . . . . .	56
4.3	Các tập dữ liệu nhỏ trong tập MedMNIST. Tập ChestM- NIST được lựa chọn để tiền huấn luyện mô hình . . . . .	57
4.4	Biểu đồ kết quả huấn luyện mô hình xương cánh tay . . .	61
4.5	Biểu đồ kết quả huấn luyện mô hình xương đùi . . . . .	63
4.6	Biểu đồ kết quả huấn luyện mô hình xương chày . . . . .	65

4.7	Kết quả phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương cánh tay . . . . .	71
4.8	Ví dụ một mẫu phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương cánh tay. a) ảnh gốc; b) ảnh có bản đồ nhiệt; c) ảnh có vùng chứa bệnh do bác sĩ đánh dấu . . . . .	72
4.9	Kết quả phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương đùi . . . . .	73
4.10	Ví dụ một mẫu phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương đùi. a) ảnh gốc; b) ảnh có bản đồ nhiệt; c) ảnh có vùng chứa bệnh do bác sĩ đánh dấu . . . . .	74
4.11	Kết quả phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương chày . . . . .	75
4.12	Ví dụ một mẫu phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương chày. a) ảnh gốc; b) ảnh có bản đồ nhiệt; c) ảnh có vùng chứa bệnh do bác sĩ đánh dấu . . . . .	76
4.13	Kết quả phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương cánh tay . . . . .	79
4.14	Ví dụ một mẫu phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương cánh tay . . . . .	80
4.15	Kết quả phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương đùi . . . . .	81
4.16	Ví dụ một mẫu phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương đùi . . . . .	82
4.17	Kết quả phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương chày . . . . .	83
4.18	Ví dụ một mẫu phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương chày . . . . .	84

# Danh sách bảng

2.1	Một số mô hình MNTC dùng trong bài toán phân lớp . . . . .	12
2.2	Một số mô hình MNTC dùng trong bài toán phân lớp (tt) . . . . .	13
2.3	Một số mô hình bộ chuyển đổi dùng trong bài toán phân lớp . . . . .	15
2.4	Một số mô hình bộ chuyển đổi dùng trong bài toán phân lớp (tt) . . . . .	16
2.5	Một số mô hình kết hợp giữa MNTC và BCĐ nổi bật . . . . .	17
2.6	Một số trí tuệ nhân tạo khả thi trên MNTC . . . . .	19
2.7	Một số trí tuệ nhân tạo khả thi trên BCĐ . . . . .	20
2.8	Các mô hình Bộ chuyển đổi trong phân lớp ảnh y khoa . . . . .	21
2.9	Hiệu suất của các mô hình Bộ chuyển đổi trong phân lớp ảnh y khoa . . . . .	22
2.10	Ưu điểm và nhược điểm của các mô hình Bộ chuyển đổi trong phân lớp ảnh y khoa . . . . .	22
2.11	Ưu điểm và nhược điểm của các mô hình Bộ chuyển đổi trong phân lớp ảnh y khoa (tt) . . . . .	23
2.12	Ưu điểm và nhược điểm của các mô hình Bộ chuyển đổi trong phân lớp ảnh y khoa (tt) . . . . .	24
2.13	Ung thư lành tính trên phần mềm . . . . .	27
2.14	Ung thư ác tính trên phần mềm (tt) . . . . .	28
2.15	Ung thư lành tính trên xương (tt) . . . . .	29
2.16	Ung thư ác tính trên xương (tt) . . . . .	30
3.1	Thông tin dữ liệu sử dụng trong mô hình phân lớp ở vị trí xương cánh tay . . . . .	40

3.2	Thông tin dữ liệu sử dụng trong mô hình phân lớp ở vị trí xương đùi . . . . .	41
3.3	Thông tin dữ liệu sử dụng trong mô hình phân lớp ở vị trí xương chày . . . . .	41
4.1	Thông tin về từng tập dữ liệu nhỏ trong tập Bone_Saigon-OTH_NNHLT_1.0 . . . . .	54
4.2	Thông tin số mẫu của từng tập huấn luyện, tập kiểm tra và tập kiểm thử . . . . .	55
4.3	Thông tin số mẫu của từng tập huấn luyện sau khi được tăng cường . . . . .	56
4.4	Kết quả huấn luyện của mô hình xương cánh tay . . . . .	60
4.5	Kết quả huấn luyện trên tập huấn luyện của mô hình xương đùi . . . . .	62
4.6	Kết quả huấn luyện của mô hình xương chày . . . . .	64
4.7	Kết quả huấn luyện của các mô hình trên 3 tập dữ liệu xương cánh tay, xương đùi và xương chày . . . . .	66
4.8	Kết quả huấn luyện của các mô hình trên 3 tập dữ liệu xương cánh tay, xương đùi và xương chày theo thứ hạng .	67
4.9	Kết quả kiểm thử tính theo độ chính xác (%) của các mô hình trên 3 tập dữ liệu xương cánh tay, xương đùi và xương chày . . . . .	69
4.10	Kết quả kiểm thử của các mô hình trên 3 tập dữ liệu xương cánh tay, xương đùi và xương chày theo thứ hạng . . . . .	70
4.11	Kết quả kiểm thử tính theo độ chính xác (%) của các mô hình trên 3 tập dữ liệu xương cánh tay, xương đùi và xương chày . . . . .	77
4.12	Kết quả kiểm thử của các mô hình trên 3 tập dữ liệu xương cánh tay, xương đùi và xương chày theo thứ hạng . . . . .	78
A.1	Bảng các thuật ngữ và viết tắt trong luận văn . . . . .	97
A.2	Bảng các thuật ngữ và viết tắt trong luận văn (tt) . . . . .	98
A.3	Bảng các thuật ngữ và viết tắt trong luận văn (tt) . . . . .	99
A.4	Bảng các thuật ngữ và viết tắt trong luận văn (tt) . . . . .	100

A.5	Bảng các thuật ngữ và viết tắt trong luận văn (tt) . . . . .	101
A.6	Bảng các thuật ngữ và viết tắt trong luận văn (tt) . . . . .	102
A.7	Bảng các thuật ngữ và viết tắt trong luận văn (tt) . . . . .	103

# Tóm tắt

Các Mạng nơ-ron tích chập đã có những bước tiến đáng kể trong lĩnh vực hình ảnh y tế cho việc chẩn đoán bệnh tự động, bao gồm cả việc phát hiện ung thư xương. Mặc dù có những tiến bộ này, vẫn tồn tại những lo ngại về độ tin cậy và khả năng giải thích của các mô hình học sâu trong chẩn đoán y tế. Nghiên cứu này đề xuất một mô hình kết hợp giữa mạng nơ-ron tích chập với bộ chuyển đổi tên là BoneViT đính kèm với kỹ thuật Grad-CAM để vừa cho ra mô hình có hiệu suất phân lớp ảnh ung thư xương cao vừa cung cấp khả năng giải thích kết quả phân lớp.

# Chương 1

## Giới thiệu

### 1.1 Bối cảnh chung

Ung thư xương là một loại bệnh hiếm, nhưng có thể gây ra những hậu quả nghiêm trọng nếu không phát hiện và điều trị kịp thời. Việc chẩn đoán sớm và chính xác là quyết định quan trọng để cứu sống và cải thiện chất lượng cuộc sống của những người mắc bệnh. Đặc điểm khó chẩn đoán của ung thư xương là một thách thức đối với các chuyên gia y tế. Do đó, việc phát triển một hệ thống phân lớp ảnh y khoa xương có thể giúp cung cấp thông tin hữu ích và hỗ trợ quyết định chẩn đoán.

Công nghệ học máy và thị giác máy tính đã đạt được sự tiến bộ đáng kể trong những năm gần đây, và việc áp dụng các phương pháp này trong lĩnh vực y khoa đã mang lại nhiều tiềm năng. Việc sử dụng học máy và thị giác máy tính trong phân loại ảnh y khoa đã cho thấy khả năng ẩn tượng trong việc chẩn đoán bệnh. Điều này tạo ra một cơ hội để áp dụng các thuật toán học máy và mô hình thị giác máy tính vào việc phân lớp ảnh y khoa xương, từ đó tạo ra một công cụ hỗ trợ chẩn đoán mạnh mẽ và chính xác.

Ảnh y khoa xương có tính chất phức tạp và đa dạng. Việc phân loại chính xác các hình ảnh này đòi hỏi kiến thức sâu rộng về cấu trúc xương và các biến thể bệnh lý. Một hệ thống phân lớp ảnh y khoa xương có thể giúp trích xuất các đặc trưng quan trọng từ ảnh, như hình dạng, kích thước và sự biến đổi cấu trúc xương, và từ đó phân loại chúng vào các

nhóm bệnh ung thư xương và không ung thư. Điều này sẽ giúp cung cấp thông tin hữu ích và hỗ trợ quyết định chẩn đoán của các chuyên gia y tế, đồng thời tăng khả năng phát hiện sớm và điều trị ung thư xương.

Ngoài ra, sự phát triển của công nghệ hình ảnh y khoa, bao gồm X-quang, Chụp cắt lớp vi tính và Chụp cộng hưởng từ, đã mở ra nhiều cơ hội trong việc nắm bắt thông tin từ ảnh y khoa xương. Công nghệ này cung cấp hình ảnh chất lượng cao về cấu trúc xương và khối u, đồng thời hỗ trợ trong việc xác định vị trí và phạm vi của khối u. Sự kết hợp giữa công nghệ hình ảnh y khoa và công nghệ học máy sẽ tạo ra một hệ thống phân lớp ảnh y khoa xương mạnh mẽ và hiệu quả, đồng thời đáng tin cậy trong việc hỗ trợ chẩn đoán bệnh ung thư xương.

## 1.2 Động lực nghiên cứu

### 1.2.1 Ý nghĩa khoa học

Việc lựa chọn đề tài "Phân lớp ảnh ung thư xương dựa trên mô hình Bộ chuyển đổi" là hoàn toàn hợp lý trong bối cảnh khoa học công nghệ và thị giác máy tính hiện nay. Sự phát triển nhanh chóng của các công nghệ học máy và thị giác máy tính đã mở ra nhiều tiềm năng mới cho các ứng dụng trong y học. Trong lĩnh vực chẩn đoán y khoa, đặc biệt là chẩn đoán ung thư, các mô hình học máy đã chứng minh được hiệu quả cao trong việc phân loại và phát hiện bệnh từ các hình ảnh y khoa. Điều này không chỉ giúp nâng cao độ chính xác trong chẩn đoán mà còn giảm thiểu thời gian và chi phí cho quá trình này. Sử dụng các mô hình Bộ chuyển đổi, vốn đã được chứng minh là hiệu quả trong nhiều nhiệm vụ khác nhau, là một hướng tiếp cận đầy hứa hẹn để xử lý và phân loại các hình ảnh y khoa phức tạp.

### 1.2.2 Ý nghĩa thực tiễn

Trí tuệ nhân tạo không thể trực tiếp thay thế bác sĩ, nhưng bác sĩ không sử dụng trí tuệ nhân tạo có thể sẽ bị thay thế bởi bác sĩ sử dụng

trí tuệ nhân tạo. Vì vậy, đề tài nghiên cứu này hướng đến nhiều ý nghĩa thực tiễn quan trọng.

Đầu tiên, việc áp dụng mô hình học máy tiên tiến vào chẩn đoán y khoa có thể giúp phát hiện sớm các dấu hiệu của ung thư xương, từ đó tăng cường khả năng điều trị kịp thời và hiệu quả. Phát hiện sớm không chỉ tăng cơ hội sống sót cho bệnh nhân mà còn giảm thiểu các chi phí và thời gian điều trị, góp phần giảm gánh nặng lên hệ thống y tế.

Thứ hai, mô hình Bộ chuyển đổi có khả năng xử lý và phân tích dữ liệu hình ảnh y khoa phức tạp với độ chính xác cao. Điều này giúp cải thiện chất lượng chẩn đoán, giảm thiểu sai sót do con người gây ra và hỗ trợ các bác sĩ trong việc đưa ra quyết định điều trị chính xác.

Cuối cùng, việc phát triển các công cụ chẩn đoán dựa trên công nghệ học máy và thị giác máy tính còn góp phần nâng cao năng lực và hiệu quả làm việc của các cơ sở y tế, đặc biệt là trong bối cảnh nguồn lực y tế hạn chế.

Nghiên cứu này có tiềm năng mang lại những tiến bộ quan trọng trong y học, cải thiện đáng kể chất lượng cuộc sống và cơ hội sống sót cho bệnh nhân mắc ung thư xương.

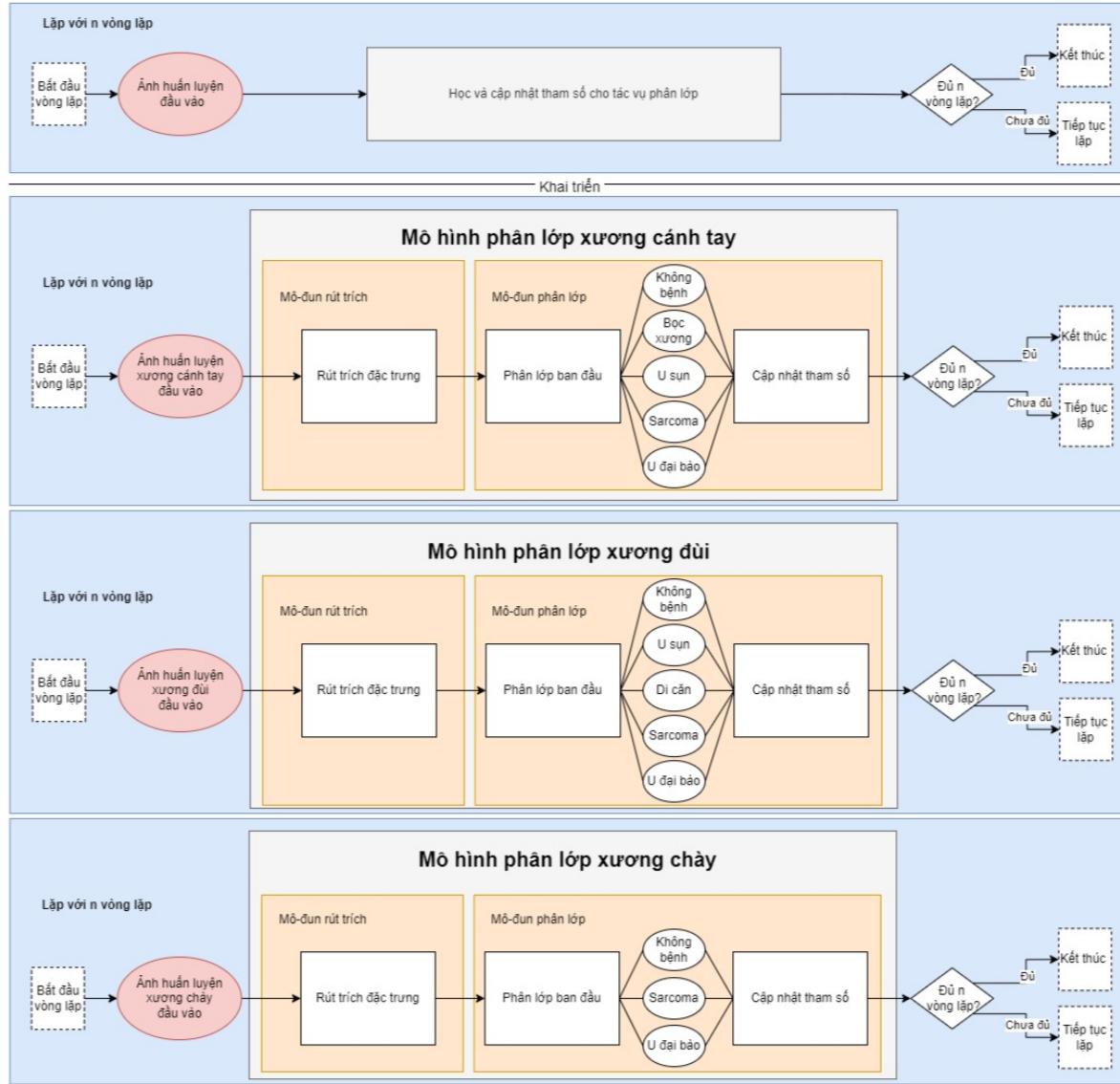
### 1.3 Phát biểu bài toán

- Đầu vào: ảnh X-quang của bộ phận cụ thể.
- Đầu ra: ảnh đã được phân lớp vào các lớp bệnh cụ thể ứng với từng vị trí xương:
  - Cánh tay: Không bệnh, Bọc xương (lành tính), U sụn (lành tính), Sarcoma (ác tính), U đai bào (ác tính).
  - Đùi: Không bệnh, U sụn (lành tính), Di căn (ác tính), Sarcoma (ác tính), U đai bào (ác tính).
  - Chày: Không bệnh, Sarcoma (ác tính), U đai bào (ác tính).
- Ân số bài toán: Tìm kiếm và xây dựng kiến trúc mạng thích hợp cho các mô-đun rút trích và phân lớp.

- Giới hạn bài toán: - Bài toán được thực hiện trên dữ liệu ảnh X-quang với số mẫu ít (dưới 1000 mẫu). - Trong 1 ảnh chỉ có duy nhất 1 bệnh, không có trường hợp nhiều bệnh.

## Khung mô hình:

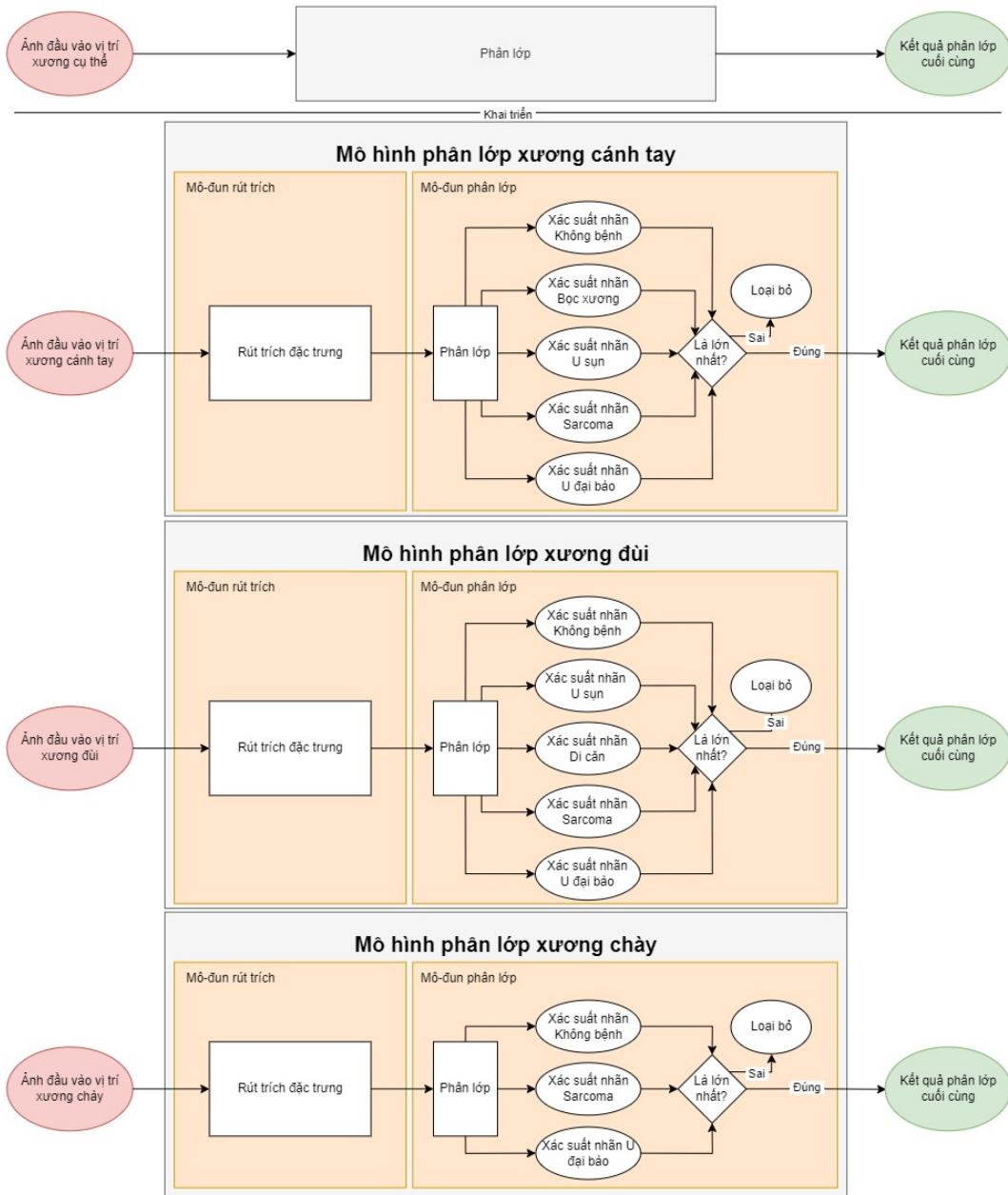
- Giai đoạn ngoại tuyến:



Hình 1.1: Sơ đồ ở giai đoạn ngoại tuyến

Ảnh đầu vào lần lượt đi qua các lớp của mô hình. Các đặc trưng của ảnh sẽ được rút trích ở mô-đun rút trích. Các đặc trưng này sẽ được chuyển tới mô-đun phân lớp để đưa dự đoán ban đầu. Sau đó, mô hình điều chỉnh các trọng số và độ lệch, từ đó tối ưu hóa mô hình. Quá trình này lặp đi lặp lại qua nhiều vòng lặp huấn luyện, giúp mô hình học cách trích xuất và sử dụng các đặc trưng một cách hiệu quả hơn.

- Giai đoạn trực tuyến:



Hình 1.2: Sơ đồ ở giai đoạn trực tuyến

Ảnh đầu vào ứng với từng bộ phận sẽ được đưa vào mô hình phân lớp bộ phận tương ứng. Các đặc trưng sẽ được rút trích ở mô-đun rút trích và chuyển tới mô-đun phân lớp. Mô-đun phân lớp sẽ cho ra xác suất của từng lớp phân loại. Xác suất cao nhất sẽ là kết quả phân lớp cuối cùng.

## 1.4 Thách thức của bài toán

- Hạn chế của dữ liệu: Lượng dữ liệu không đủ để đạt hiệu suất cao; có ảnh chứa các đối tượng làm ảnh hưởng đến quá trình huấn luyện; ảnh chất lượng thấp. Ngoài ra, tập dữ liệu cũng bị mất cân bằng số mẫu giữa các lớp. Khoảng cách dị biệt giữa từng phần tử trong một lớp là lớn và khoảng cách dị biệt giữa các phần tử trong hai lớp khác nhau là bé.
- Hạn chế về tài nguyên: Máy không đủ điều kiện để huấn luyện, môi trường không thích hợp để huấn luyện mô hình.
- Ràng buộc về luật pháp và đạo đức: Đảm bảo ứng dụng tuân thủ luật pháp và đạo đức, không sử dụng dữ liệu bệnh nhân ngoài mục đích nghiên cứu đề tài này.

Cách xử lý khó khăn:

- Hạn chế của dữ liệu: Thực hiện các phương pháp xử lý dữ liệu, loại bỏ các ảnh gây nhiễu, xóa các đối tượng cản trở huấn luyện, tăng cường ảnh để tăng số mẫu huấn luyện.
- Loại bỏ những bức ảnh có yếu tố gây nhiễu loạn thông tin cho mô hình khi huấn luyện.
- Hạn chế về tài nguyên: Xin quyền sử dụng máy chủ nhà trường để phục vụ mục đích huấn luyện. Cũng như thuê máy chủ để tiến hành thực hiện khóa luận
- Ràng buộc về luật pháp và đạo đức: Tuân thủ luật pháp, không tiết lộ thông tin bệnh nhân hoặc dùng thông tin bệnh nhân để trực lợi cho mục đích cá nhân.

Khóa luận hướng đến việc tìm ra mô hình có thể đạt hiệu suất cao với tập dữ liệu có số mẫu ít và bị mất cân bằng, có khoảng cách dị biệt giữa từng phần tử trong một lớp là lớn và khoảng cách dị biệt giữa các phần tử trong hai lớp khác nhau là bé.

## 1.5 Đóng góp

Trong đề tài này, luận văn có các đóng góp sau:

1. Đề xuất mô hình mạng kết hợp giữa Bộ chuyển đổi và Mạng nơ-ron tích chập đạt kết quả cao hơn các phương pháp tiên tiến khi huấn luyện trên tập dữ liệu ít mẫu và mất cân bằng.
2. Thành công kiểm thử đạt kết quả phân lớp ảnh ung thư xương tốt.
3. Tổng hợp tập dữ liệu ảnh X-quang Bone\_Saigon-OTH\_NNHTL\_1.0 của 3 vị trí xương gồm xương cánh tay, xương đùi và xương chày. Tập này đã được điều chỉnh và loại bỏ các tác nhân gây nhiễu loạn nhưng vẫn giữ nguyên được thông tin của ảnh. Cụ thể:
  - Chia ảnh nhận được từ bệnh viện vào các thư mục của bộ phận cụ thể.
  - Tiến hành gán nhãn cho từng ảnh.
  - Xóa bỏ các tác nhân gây nhiễu loạn thông tin như dị vật nhỏ, các dòng chữ ghi thông tin về bệnh viện và bệnh nhân.

## 1.6 Cấu trúc của luận văn

Cấu trúc luận văn được chia thành 4 chương, nội dung chính của từng chương là:

- **Chương 1: Giới thiệu**

Chương này giới thiệu về bối cảnh chung của vấn đề nghiên cứu, trình bày động lực và lý do chọn đề tài. Bối cảnh chung sẽ làm rõ tầm quan trọng của việc chẩn đoán và phân lớp ảnh ung thư xương trong y học. Tiếp theo, động lực nghiên cứu sẽ giải thích tại sao việc sử dụng các phương pháp mới như bộ chuyển đổi trong phân lớp ảnh y khoa là cần thiết. Phát biểu bài toán sẽ xác định rõ vấn đề cụ thể mà luận văn muốn giải quyết. Sau đó, các thách thức của bài toán sẽ được trình bày, bao gồm những khó khăn và trở ngại trong quá trình nghiên cứu. Phần đóng góp sẽ liệt kê những điểm mới và giá

trí mà luận văn mang lại. Cuối cùng, cấu trúc của luận văn sẽ được mô tả ngắn gọn, giúp người đọc có cái nhìn tổng quan về nội dung và thứ tự các chương.

- **Chương 2: Công trình liên quan**

Chương này tổng quan các công trình và nghiên cứu liên quan. Đầu tiên, giới thiệu về mạng nơ-ron tích chập và một số mô hình được sử dụng trong phân lớp ảnh. Tiếp theo là nói về bộ chuyển đổi và các thành phần của nó cũng như một số mô hình bộ chuyển đổi được dùng trong phân lớp. Sau đó, mô hình kết hợp giữa bộ chuyển đổi và mạng nơ-ron tích chập sẽ được mô tả và nêu ra một số mô hình nổi bật. Trí tuệ nhân tạo khả thi trong y khoa sẽ giải thích về việc sử dụng trí tuệ nhân tạo trong chẩn đoán và điều trị y khoa. Một số trí tuệ nhân tạo khả thi nổi tiếng cũng sẽ được trình bày. Tiếp đến, các mô hình bộ chuyển đổi dùng trong phân lớp ảnh y khoa sẽ được thảo luận, tập trung vào những tiến bộ và ứng dụng mới nhất. Cuối cùng là phần phân loại vị trí bệnh ung thư xương trên cơ thể con người.

- **Chương 3: Phương pháp**

Chương này trình bày chi tiết về phương pháp nghiên cứu. Đầu tiên, mô hình đề xuất được giới thiệu, bao gồm kiến trúc mô hình, khái tích chập hiệu quả và khái chuyển đổi chú ý hiệu quả. Sau đó, các mô hình sử dụng trong luận văn được liệt kê, bao gồm mô hình phân lớp ứng với các vị trí xương khác nhau như xương cánh tay, xương đùi và xương chày. Tiếp theo, các mô hình trong giai đoạn ngoại tuyến sẽ được mô tả, bao gồm các bước rút trích đặc trưng, tính toán hàm mắt mát, lan truyền ngược để tính toán độ dốc hàm mắt mát và cập nhật trọng số của mô hình. Các mô hình trong giai đoạn trực tuyến cũng sẽ được trình bày, tập trung vào các bước rút trích đặc trưng, tính toán ở lớp kết nối đầy đủ và đưa ra kết quả phân lớp, cũng như giải thích kết quả đầu ra của mô hình.

- **Chương 4: Thực nghiệm**

Chương này mô tả chi tiết quá trình thực nghiệm và kết quả đạt được nhằm minh chứng cho những đóng góp đã nêu ở trên. Đầu tiên, quá trình huấn luyện được trình bày, bao gồm tập dữ liệu xương, tiền xử lý dữ liệu, tiền huấn luyện và tinh chỉnh mô hình, cùng với các tham số huấn luyện. Kết quả huấn luyện sẽ được trình bày chi tiết, bao gồm độ đo kết quả huấn luyện, kết quả huấn luyện trên các tập dữ liệu và đánh giá kết quả huấn luyện. Phần thực nghiệm sẽ mô tả mục đích thực nghiệm và quá trình thực nghiệm phân lớp ảnh ung thư xương.

- **Chương 5: Kết luận**

Chương này tóm tắt lại những ý chính của khóa luận và minh chứng cho phần đóng góp. Đồng thời nhấn mạnh những kết quả đạt được và ý nghĩa của chúng trong thực tế. Cuối cùng, hướng phát triển sẽ đề xuất các hướng nghiên cứu và phát triển tiếp theo, giúp mở rộng và cải thiện công trình nghiên cứu hiện tại.

## Chương 2

# Công trình liên quan

### 2.1 Mạng nơ-ron tích chập

Mạng nơ-ron tích chập (MNTC) là một kiến trúc Mạng nơ-ron, chủ yếu được sử dụng trong xử lý hình ảnh. MNTC bao gồm các lớp tích chập và lớp gộp, cho phép mạng tự động học các đặc trưng không gian từ dữ liệu đầu vào mà không cần sự can thiệp của con người. Cấu trúc của MNTC cho phép nó học và trích xuất các đặc trưng phân cấp từ thấp đến cao, giúp mạng có thể nhận dạng và phân loại hình ảnh một cách hiệu quả. MNTC đã đạt được hiệu suất vượt trội trong nhiều nhiệm vụ thị giác máy tính, bao gồm nhận dạng hình ảnh, phát hiện đối tượng và phân loại hình ảnh, và trở thành lựa chọn phổ biến cho nhiều ứng dụng trong lĩnh vực này.

Bảng 2.1: Một số mô hình MNTC dùng trong bài toán phân lớp

Tên mô hình	Đối tượng phân lớp	Điểm nổi trội	Cải tiến so với MNTC trước đó
LeNET [1] (1998)	Chữ viết tay (MNIST)	Hiệu quả với các bài toán phân lớp hình ảnh đơn giản.	Một trong những MNTC đầu tiên, giúp thiết lập nền tảng cho các MNTC sau này.
AlexNET [2] (2012)	Hình ảnh tổng quát (ImageNet)	Kiến trúc sâu hơn LeNET, sử dụng DVTTĐHC và LBQNN để cải thiện hiệu suất và giảm quá khứp.	Sử dụng nhiều bộ xử lý đồ họa để tăng tốc độ huấn luyện.
VGG [3] (2014)	Hình ảnh tổng quát (ImageNet)	Sử dụng nhiều lớp tích chập nhỏ ( $3 \times 3$ ) để tăng độ sâu, giúp cải thiện độ chính xác mà không tăng quá nhiều tham số.	Cải thiện độ sâu của mạng, giúp tăng cường khả năng học đặc trưng phức tạp hơn so với AlexNET.
GoogleNET [4](2015)	Hình ảnh tổng quát (ImageNet)	Sử dụng kiến trúc Inception, kết hợp nhiều kích thước bộ lọc trong một mô-đun để thu thập thông tin ở các mức độ khác nhau.	Giảm số lượng tham số so với VGG, tăng hiệu suất và độ chính xác nhờ vào việc tận dụng tối đa các nguồn thông tin khác nhau.
ResNET [5] (2016)	Hình ảnh tổng quát (ImageNet)	Giới thiệu cơ chế học dư thừa, cho phép xây dựng các mạng rất sâu mà không gặp vấn đề về biến mất độ dốc.	Mở ra khả năng xây dựng các mạng sâu hơn nhiều (lên đến hàng trăm lớp) mà không gặp phải vấn đề biến mất độ dốc.

Bảng 2.2: Một số mô hình MNTC dùng trong bài toán phân lớp (tt)

Tên mô hình	Đối tượng phân lớp	Điểm nổi trội	Cải tiến so với MNTC trước đó
DenseNET[6] (2017)	Hình ảnh tổng quát (ImageNet)	Mỗi lớp kết nối trực tiếp với tất cả các lớp trước đó, giúp giảm số lượng tham số và cải thiện khả năng truyền thông tin qua các lớp.	Tăng hiệu quả sử dụng thông tin và độ dốc nhờ vào kết nối dày đặc giữa các lớp, giảm quá khứ và cải thiện độ tổng quát.
EfficientNET [7](2019)	Hình ảnh tổng quát (ImageNet)	Sử dụng phương pháp phóng to các chiều của mạng một cách cân bằng để đạt hiệu suất cao hơn.	Cung cấp một cách tiếp cận có hệ thống và hiệu quả trong việc mở rộng mô hình, giúp cải thiện độ chính xác với cùng một lượng tài nguyên tính toán.

## 2.2 Bộ chuyển đổi và các thành phần

Bộ chuyển đổi [8] (BCD) là một kiến trúc Mạng nơ-ron được giới thiệu bởi Vaswani và đồng nghiệp, chủ yếu được sử dụng trong xử lý ngôn ngữ tự nhiên (XLNNTN). Nó là một khối xây dựng bao gồm các lớp chú ý, cho phép mạng tổng hợp thông tin từ toàn bộ chuỗi đầu vào, không phụ thuộc vào các bước tính toán tuần tự như trong các mô hình Mạng nơ-ron hồi quy truyền thống. BCD đã đạt được hiệu suất tốt và trở thành lựa chọn phổ biến cho nhiều nhiệm vụ XLNNTN, và sau đó đã được áp dụng thành công trong lĩnh vực thị giác máy tính, được biết đến với tên gọi là Bộ chuyển đổi trực quan.

Bộ chuyển đổi trực quan [9] (BCDTQ) là một dạng của BCD được áp dụng trong lĩnh vực thị giác máy tính. Được giới thiệu bởi Dosovitskiy và đồng nghiệp, BCDTQ xử lý hình ảnh bằng cách tách nó thành các phần nhỏ gọi là các mảnh ảnh và sử dụng một mô hình BCD tiêu chuẩn để xử lý các mảnh ảnh này. Cách tiếp cận này giúp mô hình nắm bắt được ngữ cảnh toàn cục của hình ảnh và đã đạt được kết quả ấn tượng trong nhiều nhiệm vụ thị giác máy tính.

Cơ chế tự chú ý (TCY) đã được coi là một trong những yếu tố quan trọng góp phần vào thành công của các mô hình BCD, nhờ khả năng của nó trong việc mô hình hóa các phụ thuộc xa. Ý tưởng chính của cơ chế TCY là học cách tự cân bằng, tức là xác định sự quan trọng tương đối của một mã thẻ (phần nhúng mảnh ảnh) so với tất cả các mã thẻ khác trong chuỗi.

Tự chú ý nhiều đầu (TCYNĐ) là một khối gồm nhiều khối tự chú ý được nối tiếp với nhau theo chiều kinh để mô hình hóa các phụ thuộc phức tạp giữa các phần tử khác nhau trong chuỗi đầu vào.

Bảng 2.3: Một số mô hình bộ chuyển đổi dùng trong bài toán phân lớp

Tên mô hình	Đối tượng phân lớp	Điểm nổi trội	Cải tiến so với mô hình BCĐ trước đó
BCDTQ [9] (2020)	Hình ảnh	BCDTQ là mô hình đầu tiên sử dụng kiến trúc BCĐ cho phân lớp hình ảnh, thay vì MNTC truyền thống. BCDTQ chia hình ảnh thành các mảnh nhỏ và xử lý chúng như các từ trong XLNNTN.	
DeiT [10] (2021)	Hình ảnh	DeiT được thiết kế để hiệu quả hơn về mặt dữ liệu, tức là có thể huấn luyện với lượng dữ liệu ít hơn so với BCDTQ.	DeiT giới thiệu thêm các chiến lược huấn luyện mới, sử dụng một mô hình huấn luyện với một "giáo viên" MNTC để hướng dẫn mô hình DeiT.
Swin Transformer [11](2021)	Hình ảnh	Swin Transformer sử dụng các cửa sổ nhỏ dịch chuyển và không chồng chéo để tạo nên một kiến trúc có khả năng nắm bắt thông tin toàn cục và cục bộ tốt hơn.	Swin Transformer giải quyết vấn đề về độ phức tạp tính toán của BCĐ bằng cách giảm kích thước đầu vào của mỗi lớp, sử dụng TCY trong các cửa sổ nhỏ.

Bảng 2.4: Một số mô hình bộ chuyển đổi dùng trong bài toán phân lớp (tt)

Tên mô hình	Đối tượng phân lớp	Điểm nổi trội	Cải tiến so với mô hình BCĐ trước đó
CrossViT [12] (2021)	Hình ảnh	CrossViT kết hợp hai nhánh BCĐ với các kích thước mảnh ảnh khác nhau để nắm bắt được thông tin ở cả mức độ chi tiết và mức độ tổng quát.	Cung cấp một phương pháp hiệu quả để kết hợp thông tin từ các kích thước mảnh ảnh khác nhau, giúp cải thiện khả năng phân lớp hình ảnh.
EATFormer [13] (2024)	Hình ảnh	EATFormer sử dụng một cơ chế chú ý hiệu quả hơn về mặt tính toán, giúp giảm thiểu yêu cầu về tài nguyên phần cứng mà vẫn duy trì hiệu suất cao.	EATFormer tối ưu hóa các bước tính toán trong TCY để giảm độ phức tạp từ $\mathcal{O}(N^2)$ xuống gần $\mathcal{O}(N)$ , giúp mô hình dễ dàng áp dụng trên các thiết bị có tài nguyên hạn chế hơn.

## 2.3 Mô hình kết hợp Bộ chuyển đổi và Mạng nơ-ron tích chập

Mô hình kết hợp BCĐ và MNTC là một nhánh nghiên cứu đang nhận được sự quan tâm lớn trong cộng đồng học máy và thị giác máy tính. Việc kết hợp hai loại kiến trúc này mang lại những lợi ích riêng biệt từ cả hai phương diện: BCĐ với khả năng xử lý thông tin toàn cục và học biểu diễn tốt, và MNTC với khả năng trích xuất đặc trưng cục bộ và sự phổ biến trong việc xử lý hình ảnh.

Bảng 2.5: Một số mô hình kết hợp giữa MNTC và BCD nổi bật

Tên mô hình	Điểm nổi trội
MBoTNet [14] (2021)	Sử dụng một biến thể nhỏ của cơ chế TCY trong ba khối cuối cùng của ResNet, tăng cường khả năng chú ý cục bộ và toàn cục.
CvT [15] (2021)	Chèn tích chập theo điểm và theo chiều sâu trước cơ chế TCY, tăng cường khả năng trích xuất đặc trưng.
LeViT [16] (2021)	Sử dụng lớp tích chập gốc để thay thế khối nhúng mảnh ảnh và đạt được phân loại hình ảnh nhanh hơn, giảm thời gian suy luận.
MobileViT [17] (2021)	Giới thiệu một BCDTQ nhẹ bằng cách kết hợp các khối BCD với khối MobileNetV2 [18] theo chuỗi, tạo ra một mô hình nhẹ và hiệu quả.
CMT [19] (2022)	Chứa các lớp tích chập theo chiều sâu dựa trên đơn vị nhận thức cục bộ và một khối BCD nhẹ, kết hợp chú ý cục bộ với các khối tích chập nhẹ.
Mobile-Former [20](2022)	Thiết lập một cầu nối hai chiều giữa MNTC và BCD để tận dụng lợi thế của các khái niệm toàn cục và cục bộ, cải thiện độ chính xác phân loại.

## 2.4 Trí tuệ nhân tạo khả diễn

Trí tuệ nhân tạo khả diễn (TTNTKD) [21] lần đầu được giới thiệu bởi Ribeiro và đồng nghiệp, là một lĩnh vực nghiên cứu trong lĩnh vực Trí tuệ nhân tạo (TTNT) tập trung vào việc phát triển các phương pháp và công cụ để làm cho quá trình ra quyết định của các mô hình máy học và học sâu trở nên dễ hiểu và minh bạch hơn đối với con người. Trong thời đại của những mô hình TTNT ngày càng phức tạp và mạnh mẽ, việc có khả năng giải thích các quyết định của chúng trở nên ngày càng quan trọng.

TTNTKD đóng vai trò quan trọng trong phân loại hình ảnh y khoa bằng cách giúp hiểu rõ hơn về cách mà các mô hình TTNT đưa ra quyết định trong việc chẩn đoán các bệnh lý dựa trên hình ảnh y khoa. Trong lĩnh vực y học, việc đưa ra các dự đoán chính xác và minh bạch là vô cùng quan trọng để đảm bảo chất lượng chăm sóc sức khỏe cho bệnh nhân. TTNTKD cung cấp các phương tiện để giải thích quyết định của các mô hình TTNT, giúp các chuyên gia y tế và nhà nghiên cứu hiểu rõ hơn về cách mà các mô hình này phân loại và đưa ra dự đoán.

Bảng 2.6: Một số trí tuệ nhân tạo khả thi trên MNTC

Tên TTNTKD	Điểm nổi trội	Cải tiến so với công trình trước đó	Hạn chế
CAM [22] (2016)	Cho phép xác định các khu vực quan trọng trên ảnh mà mạng quan tâm.	Phát triển cơ chế chú ý dựa trên lớp cuối của mạng tích hợp.	Chỉ hoạt động với các mạng có kiến trúc đặc biệt (có lớp KNDD).
GradCAM [23] (2017)	Sử dụng độ dốc của lớp đầu ra để tạo bản đồ nhiệt, giúp dễ dàng xác định các vùng quan trọng trên ảnh.	Cải tiến từ CAM bằng cách sử dụng độ dốc của lớp đầu ra, cho phép tạo bản đồ nhiệt chính xác hơn.	Hiệu suất giảm đối với các mạng rất sâu và có nhiều lớp phức tạp.
GradCAM++ [24](2018)	Sử dụng độ dốc trọng số và kỹ thuật phân bổ tốt hơn để tạo bản đồ nhiệt chi tiết hơn, phản ánh chính xác hơn các đặc điểm quan trọng trong ảnh.	Cải thiện từ GradCAM bằng cách sử dụng độ dốc trọng số và các kỹ thuật để tạo bản đồ nhiệt chi tiết hơn, giúp phát hiện các đặc điểm nhỏ và quan trọng.	Tính toán phức tạp hơn và yêu cầu tài nguyên xử lý lớn hơn so với GradCAM gốc.

Bảng 2.7: Một số trí tuệ nhân tạo khả diễm trên BCD

Tên TTNTKD	Điểm nổi trội	Cải tiến so với công trình trước đó	Hạn chế
Attention Rollout [25] (2020)	Giúp hiểu rõ hơn về cách các mô hình BCD ra quyết định bằng cách tổng hợp trọng số chú ý qua nhiều lớp.	Phát triển phương pháp tổng hợp chú ý từ nhiều lớp để cung cấp cái nhìn sâu hơn vào quyết định của mô hình.	Có thể bị ảnh hưởng bởi sự phức tạp của mô hình, yêu cầu tính toán cao.
Transformer Interpretability Beyond Attention Visualization [26] (2021)	Mở rộng phạm vi đánh giá của các mô hình BCD bằng cách không chỉ tập trung vào biểu đồ chú ý mà còn đánh giá các yếu tố khác của các lớp.	Dưa ra phương pháp phân tích các khía cạnh khác ngoài chú ý, giúp tăng khả năng giải thích của mô hình BCD.	Cần nhiều tài nguyên tính toán hơn do phải thực hiện nhiều phương pháp phân tích khác nhau.

## 2.5 Các mô hình Bộ chuyển đổi dùng trong phân lớp ảnh y khoa

Luận văn đã tham khảo qua các cuộc khảo sát, bài tổng hợp [27] [28] và tập trung phân tích kỹ hơn về các mô hình phân lớp ảnh y khoa có sử dụng bộ chuyển đổi sau đây theo các tiêu chí: Kiến trúc mạng cơ sở, phương pháp học, có tích hợp TTNTKD không, hiệu suất tính theo độ chính xác cũng như các ưu nhược điểm và tính thích hợp với đề tài.

Bảng 2.8: Các mô hình Bộ chuyển đổi trong phân lớp ảnh y khoa

Tên mô hình	Kiến trúc mạng cơ sở	Phương pháp học	TTNTKD
TransMIL [29] (2021)	Bộ chuyển đổi	Có giám sát yếu	Có
RadioTransformer [30] (2022)	Bộ chuyển đổi	Có giám sát	Có
HATNet [31] (2020)	Kết hợp Bộ chuyển đổi và Mạng nơ-ron tích chập	Có giám sát	Có
TransMed [32] (2021)	Kết hợp Bộ chuyển đổi và Mạng nơ-ron tích chập	Có giám sát	Không
MedViT [33] (2023)	Kết hợp Bộ chuyển đổi và Mạng nơ-ron tích chập	Có giám sát	Không

Bảng 2.9: Hiệu suất của các mô hình Bộ chuyển đổi trong phân lớp ảnh y khoa

Tên mô hình	Tập dữ liệu	Kết quả đo (Độ chính xác)
TransMIL	TCGA-RCC	0.946
RadioTransformer	Radiography	0.9943
HATNet	breast biopsy	0.71
TransMed	Hospital	0.889
MedVit	ChestMNIST	0.959

Bảng 2.10: Ưu điểm và nhược điểm của các mô hình Bộ chuyển đổi trong phân lớp ảnh y khoa

Tên mô hình	Ưu điểm	Nhược điểm	Nhận xét tính ứng dụng trong đề tài
TransMIL (2021)	Giải quyết vấn đề dữ liệu không cân bằng/cân bằng trong phân loại nhị phân/nhiều lớp. Hiệu quả trong việc phân loại với diện tích dưới đường cong ROC cao. Có tính trực quan và giải thích tốt.	Tính phức tạp tính toán cao. Yêu cầu tài nguyên cao. Có thể gặp khó khăn khi đối mặt với các trường hợp hiếm gặp. Việc điều chỉnh và cải tiến mô hình dễ gặp khó khăn. Phụ thuộc vào dữ liệu chất lượng cao.	Mô hình này phù hợp với phân loại các ảnh có sự kết hợp thông tin hình thái và không gian. Tuy nhiên, nó lại phụ thuộc vào giả thuyết phân phối độc lập và đồng nhất của các thực thể.

Bảng 2.11: Ưu điểm và nhược điểm của các mô hình Bộ chuyển đổi trong phân lớp ảnh y khoa (tt)

Tên mô hình	Ưu điểm	Nhược điểm	Nhận xét tính ứng dụng trong đề tài
Radio-Transformer (2022)	Kết hợp thông tin từ quá trình tìm kiếm thị giác của các chuyên gia y tế. Sử dụng các mô-đun toàn cục và tiêu điểm, mô hình này có thể nắm bắt được cả các đặc điểm hình ảnh toàn cục và cục bộ. Điều này giúp cải thiện khả năng phân loại bằng cách cung cấp một cái nhìn toàn diện và chi tiết về hình ảnh y tế.	Phụ thuộc vào dữ liệu theo dõi ánh mắt, điều này có thể không khả thi trong nhiều trường hợp.	Khả năng học từ mẫu tìm kiếm hình ảnh của bác sĩ có thể không cần thiết cho phân lớp ảnh ung thư xương, làm cho mô hình này ít phù hợp cho nhiệm vụ phân lớp ảnh ung thư xương.
HATNet (2020)	Có độ chính xác cao hơn so với các mô hình trước đó như Y-net. Khả năng học từ các cấu trúc mô học có liên quan lâm sàng.	Khó khăn trong việc xử lý các ảnh lớn mà không chia thành các mảnh ảnh nhỏ, điều này có thể giới hạn thông tin bối cảnh.	Mô hình này có thể không phù hợp do nó tập trung vào các ảnh sinh thiết kích thước rất lớn (lên đến hàng tỷ điểm ảnh) và khả năng xử lý thông tin từ các cấu trúc mô học lớn, trong khi đó ảnh ung thư xương có thể không yêu cầu điều này.

Bảng 2.12: Ưu điểm và nhược điểm của các mô hình Bộ chuyển đổi trong phân lớp ảnh y khoa (tt)

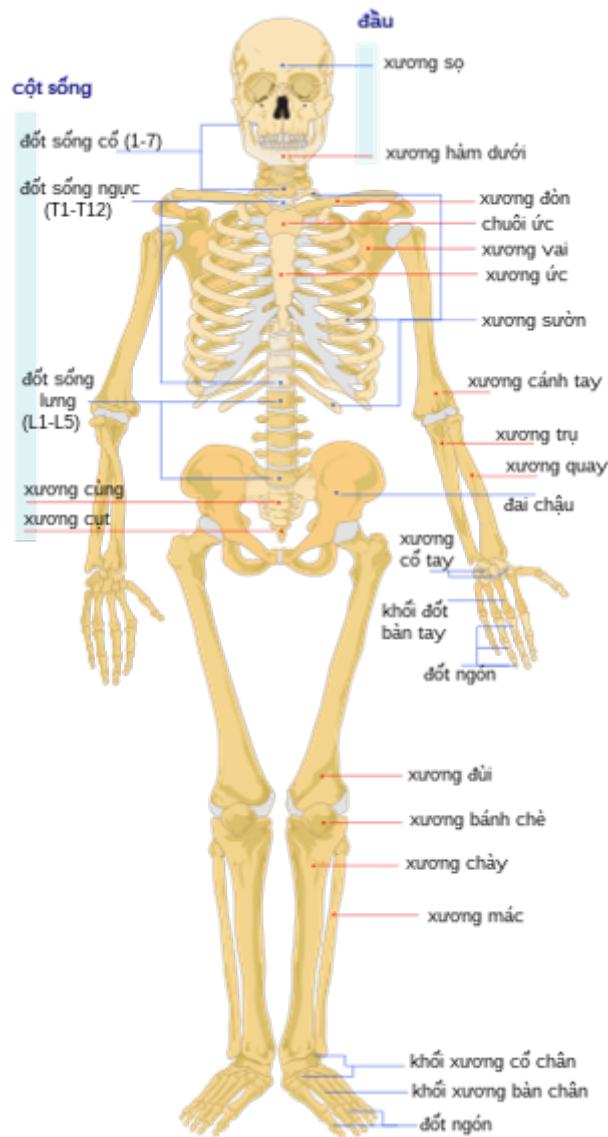
Tên mô hình	Ưu điểm	Nhược điểm	Nhận xét tính ứng dụng trong đề tài
TransMed (2021)	Kết hợp ưu điểm của Mạng nơ-ron tích chập và Bộ chuyển đổi. Khả năng trích xuất các đặc trưng cục bộ tốt. Hiệu quả cao trên các tập dữ liệu y tế nhỏ.	Yêu cầu dữ liệu đa phương thức, điều này có thể không khả thi trong nhiều trường hợp	Mô hình có thể phù hợp khi đạt hiệu quả tốt trên tập y tế nhỏ, có khả kết hợp các điểm tốt của Mạng nơ-ron tích chập và Bộ chuyển đổi. Thế nhưng việc yêu cầu dữ liệu đa phương tiện lại là một hạn chế lớn.
MedVit (2023)	Kết hợp tính cục bộ của Mạng nơ-ron tích chập và tính kết nối toàn cục của Bộ chuyển đổi. Sử dụng cơ chế chú ý với độ phức tạp tính toán thấp. Khả năng chống lại các tấn công học máy đối nghịch tốt. Không yêu cầu lượng dữ liệu quá lớn.	Phức tạp hơn các mô hình thuần Mạng nơ-ron tích chập hoặc Bộ chuyển đổi	Mô hình có thể phù hợp do có khả năng tổng quát hóa cao, cũng như độ phức tạp tính toán thấp.

## 2.6 Phân loại vị trí ung thư xương trong y khoa

Cơ thể con người có một hệ thống xương phức tạp, được chia thành nhiều nhóm và bộ phận khác nhau. Các nhóm xương chính được chia theo bộ phận cơ thể gồm có:

- Xương đầu: Sọ, Hàm dưới.
- Xương thân: Cột sống, Đòn, Vai, Ức, Sườn, Cùng, Cụt.
- Xương chi dưới: Đầu gối, Bánh chè, Chày, Mác, Cổ chân, Bàn chân, Đốt ngón chân.
- Xương chi trên: Cánh tay, Trụ, Quay, Cổ tay, Bàn tay, Đốt ngón tay.

Ung thư xương là một nhóm bệnh đặc trưng bởi sự phát triển không kiểm soát của các tế bào bất thường quanh vùng xương trong cơ thể. Những tế bào này có khả năng xâm nhập và phá hủy các mô bình thường xung quanh, đồng thời có thể lan rộng đến các phần khác của cơ thể thông qua hệ thống tuần hoàn hoặc bạch huyết. Có nhiều loại ung thư xương khác nhau, mỗi loại có các đặc điểm và cách điều trị riêng. Việc phân loại các bệnh ung thư ứng với từng vị trí xương trong cơ thể sẽ hỗ trợ rất nhiều trong việc xây dựng mô hình phân lớp. Dưới đây là các bảng phân loại những bệnh ung thư xương thường gặp theo vị trí xương được tham khảo từ sách *Bướu xương* [34].



Hình 2.1: Bộ xương con người

Bảng 2.13: Ung thư lành tính trên phần mềm

	<b>U mạch máu</b>	<b>U mỡ</b>	<b>Các u lành sợi</b>
Sọ	x	x	
Hàm dưới	x		
Cột sống	x		x
Đòn	x		
Vai		x	x
Úc			
Sườn	x	x	x
Cùng			
Cụt			
Chậu	x		
Đùi	x	x	x
Bánh chè			
Chày	x	x	x
Mác		x	
Cổ chân			
Bàn chân		x	x
Đốt ngón chân		x	
Cánh tay	x	x	x
Trụ	x	x	x
Quay	x	x	x
Cổ tay			
Bàn tay		x	
Đốt ngón tay		x	x

Bảng 2.14: Ung thư ác tính trên phần mềm (tt)

	<b>Sarcoma cơ vân</b>	<b>Các u sợi ác tính</b>
Sọ		
Hàm dưới		
Cột sống	x	x
Dòn		
Vai		x
Úc		
Sườn		x
Cùng		
Cụt		
Chậu		
Đùi	x	x
Bánh chè		
Chày	x	x
Mác		
Cổ chân		
Bàn chân		
Đốt ngón chân		
Cánh tay	x	x
Trụ	x	x
Quay	x	x
Cổ tay		
Bàn tay		
Đốt ngón tay		

Bảng 2.15: Ung thư lành tính trên xương (tt)

	<b>U sụn xương</b>	<b>U đại bào</b>	<b>U nội sụn</b>	<b>U sợi</b>
Sọ				
Hàm dưới				
Cột sống				
Dòn	x		x	
Vai	x		x	
Úc			x	
Sườn	x		x	x
Cùng		x		
Cụt				
Chậu	x	x		
Dùi	x	x	x	x
Bánh chè	x	x		
Chày	x	x	x	x
Mác	x	x	x	x
Cổ chân				
Bàn chân			x	
Đốt ngón chân	x		x	
Cánh tay	x	x	x	x
Trụ	x	x		x
Quay	x	x	x	x
Cổ tay				
Bàn tay			x	
Đốt ngón tay	x		x	

Bảng 2.16: Ung thư ác tính trên xương (tt)

	<b>Osteosarcoma tạo xương</b>	<b>Sarcoma Ewing</b>	<b>U xương di căn</b>
Sọ			x
Hàm dưới			
Cột sống			x
Dòn			
Vai			x
Úc			
Sườn		x	x
Cùng			
Cụt			
Chậu		x	x
Dùi	x	x	x
Bánh chè	x		
Chày	x	x	x
Máć	x	x	x
Cổ chân			
Bàn chân			
Dốt ngón chân			
Cánh tay		x	x
Trụ			
Quay		x	
Cổ tay			
Bàn tay			
Dốt ngón tay			

# Chương 3

## Phương pháp

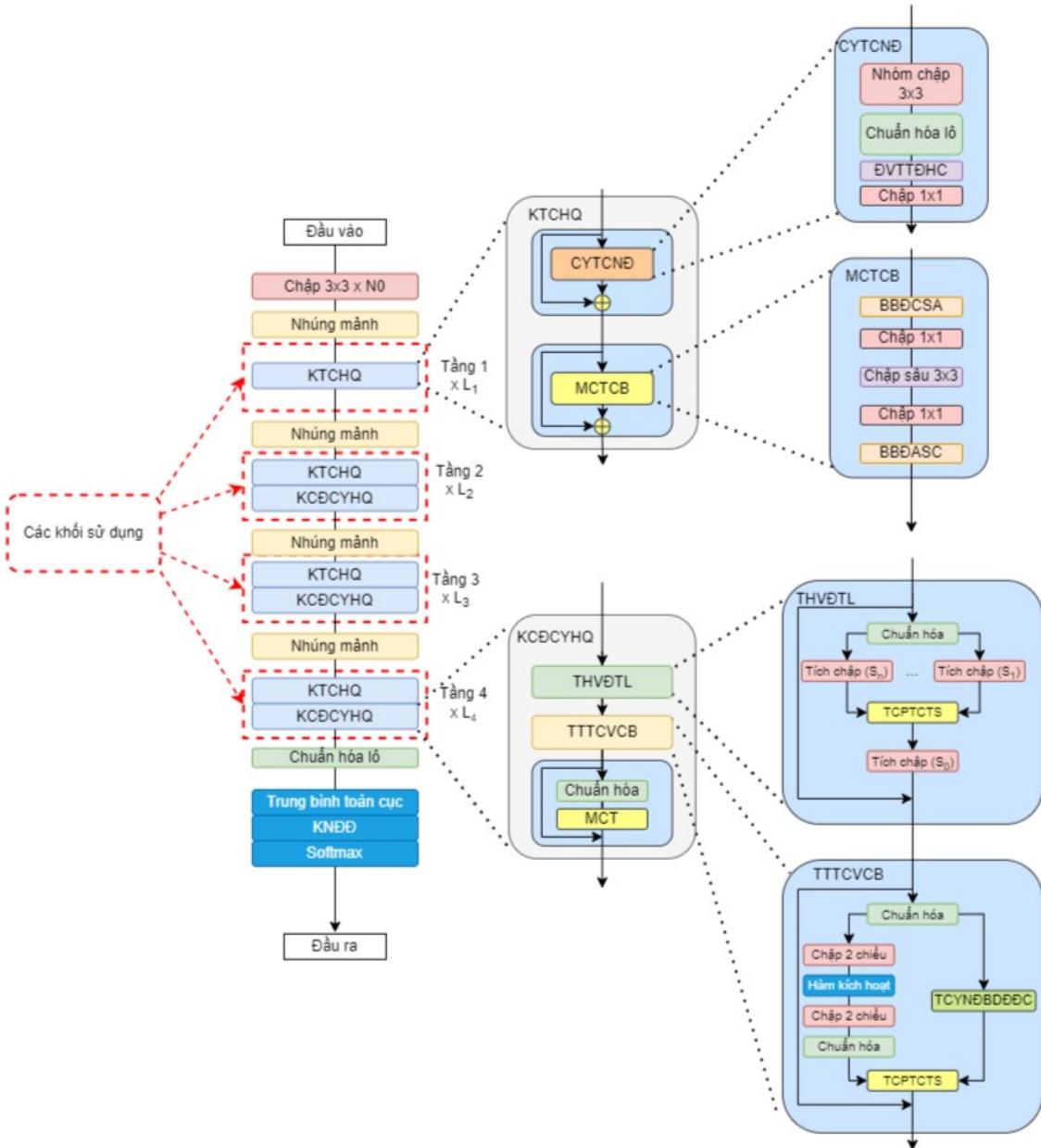
### 3.1 Mô hình đề xuất

Xét theo độ đơn giản trong việc cài đặt và vận hành, chất lượng hiện có của tập dữ liệu xương tự chuẩn bị và các tiêu chí đã nêu trên. Luận văn quyết định chọn việc kết hợp kiến trúc của MedViT và khối chuyển đổi từ EATFormer là phương pháp chính trong tác vụ phân lớp ảnh ung thư xương.

Trong luận văn này, khối tích chập của MedViT và khối chuyển đổi EATFormer đã được cải tiến cho việc sử dụng trên ảnh y khoa [35] được kết hợp để tạo thành mô hình đề xuất gọi là BoneViT. Ngoài ra, mô hình cũng tích hợp được thêm GradCAM để thực hiện giải thích cho kết quả phân lớp.

#### 3.1.1 Kiến trúc Mô hình

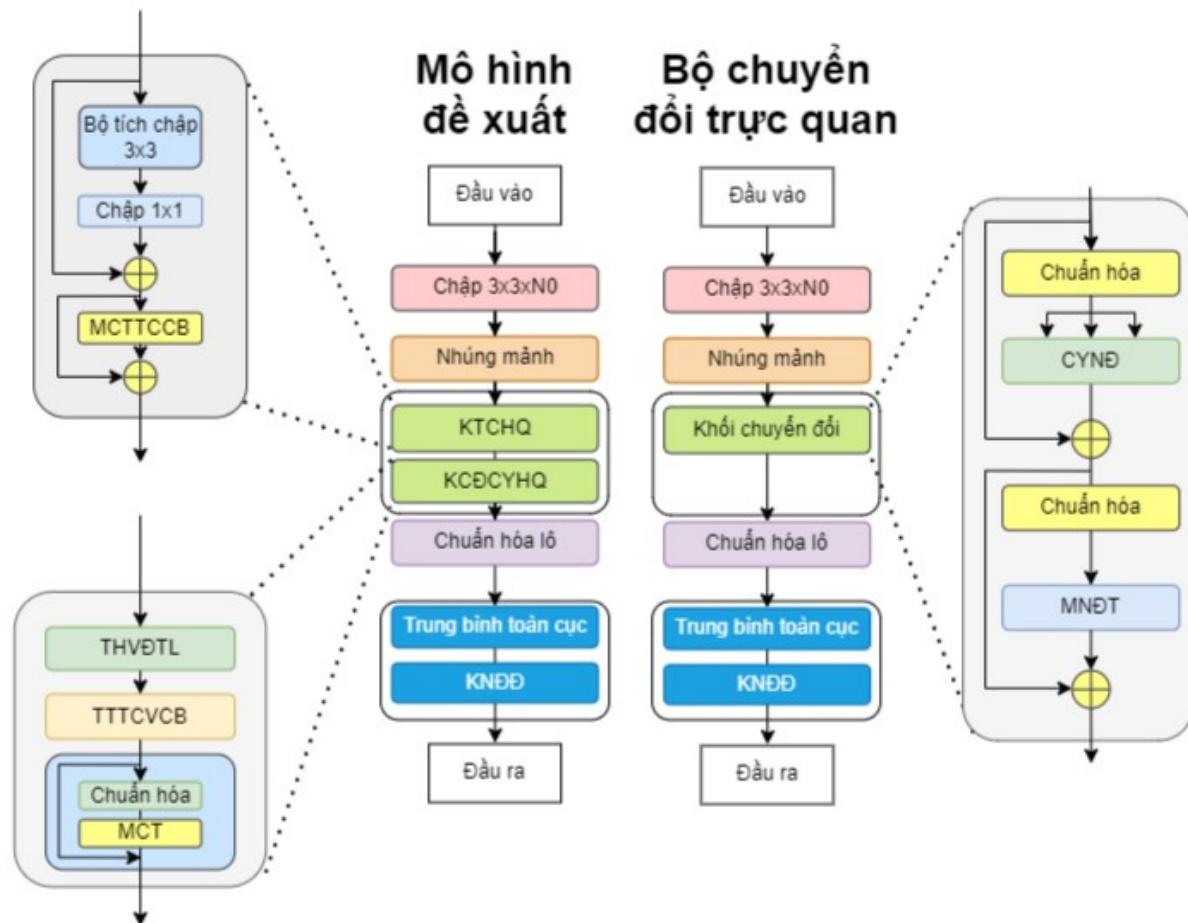
Kiến trúc mà luận văn đề xuất để xây dựng mô hình là kết hợp giữa khối tích chập hiệu quả (KTCHQ) của MedViT và khối chuyển đổi chú ý hiệu quả (KCĐCYHQ) của EATFormer đã được cải tiến. Kiến trúc này kết hợp khả năng học sâu của MNTC với khả năng tự chú ý toàn cục của BCD, mang lại hiệu suất cao trong việc phân loại và phát hiện các bất thường y khoa từ hình ảnh.



Hình 3.1: Kiến trúc tổng quan của BoneViT với tên các thành phần đã được việt hóa.

Kiến trúc của mô hình tương tự như kiến trúc của MedViT, là sự kết hợp giữa các khối tích chập và các khối chuyển đổi theo cách tiếp cận mới để đạt được kiến trúc lai mạnh mẽ cho phân loại hình ảnh y tế. Tuy nhiên, KCĐCYHQ của EATFormer được sử dụng thay thế cho khối chuyển đổi cục bộ của MedViT nhằm giảm độ phức tạp của mô hình, cũng như tăng cường khả năng nắm bắt đặc trưng toàn cục của mô hình. Như hình 3.1 đã cho thấy, BoneViT bao gồm một lớp nhúng mảnh ảnh, các khối tích chập và các khối chuyển đổi xếp chồng lên nhau (trừ tầng đầu tiên) tuân

theo kiến trúc kim tự tháp phân cấp truyền thống.



Hình 3.2: Đổi chiểu BoneViT với mô hình MedViT gốc và mô hình Bộ chuyển đổi trực quan

### 3.1.2 Khối tích chập hiệu quả

ResNET [5] đã giới thiệu các kết nối dư thừa vào MNTC và giảm thiểu vấn đề biến mất đạo hàm, điều này đảm bảo rằng mô hình có thể xây dựng sâu hơn để nắm bắt các đặc trưng cấp cao cho việc phân loại hình ảnh, và trong một thời gian dài, nó đã trở thành lựa chọn hàng đầu cho nhiều nhiệm vụ nhận diện hình ảnh, nhờ vào khả năng tương thích và ứng dụng linh hoạt trong các tình huống thực tế. Tuy nhiên, khối dư thừa không đạt hiệu suất cao như mong đợi so với khối chuyển đổi. Khối ConvNext [36] được xây dựng từ khối dư thừa bằng cách tuân theo thiết kế của khối chuyển đổi nhưng không sử dụng sự chú ý riêng. Mặc dù khối ConvNext đã cải thiện hiệu suất của mạng ở một mức độ nhất định, nhưng các thành phần không hiệu quả như tích chập sâu  $7 \times 7$ , Đơn vị tuyến tính lõi Gauss và chuẩn hóa lớp làm cho mô hình khó nắm bắt các cấu trúc cấp cao.

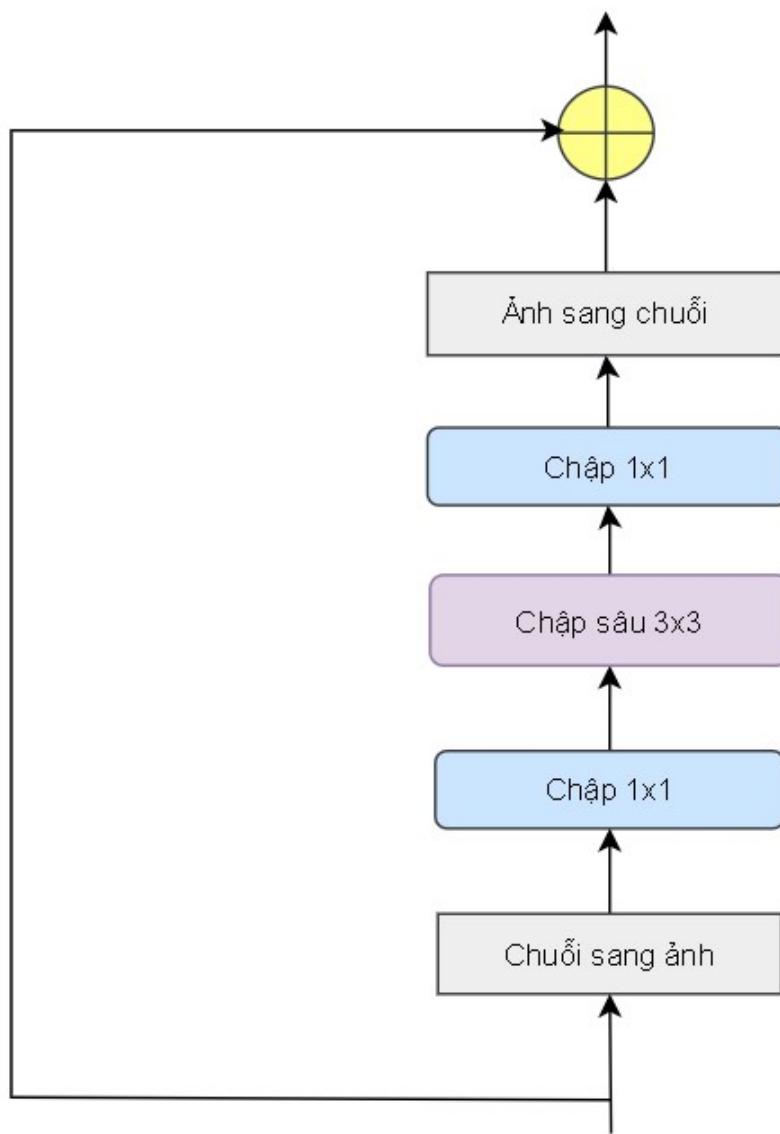
Để giải quyết vấn đề này, khối chuyển đổi đã được đề xuất để nắm bắt các cấu trúc cấp cao và đã đạt được kết quả xuất sắc trong nhiều nhiệm vụ thị giác máy tính. Ưu thế ngầm của khối chuyển đổi đến từ hoạt động trộn mã thẻ dựa trên sự chú ý [11]. Tuy nhiên, những phương pháp này chỉ tập trung vào độ phức tạp của mô hình và độ chính xác tiêu chuẩn, trong khi các kết quả cho thấy mô hình dễ bị tấn công học máy đối nghịch [37] [38], điều này không thể chấp nhận được trong các ứng dụng y tế liên quan đến lâm sàng.

Để phân loại y tế chính xác và chống lại tấn công học máy đối nghịch, Khối Tích chập Hiệu quả (KTCHQ) được tạo ra. KTCHQ đạt được hiệu suất vượt trội như một khối dựa trên BCD trong khi vẫn giữ được lợi thế triển khai của khối dư thừa. Như được minh họa trong Hình 3.1, KTCHQ theo kiến trúc lai, điều này đã được xác nhận là cần thiết để sử dụng thông tin đa tỷ lệ. Trong khi đó, một mô-đun trộn mã thẻ dựa trên cơ chế chú ý hiệu quả cũng quan trọng không kém. Mạng chuyển tiền tăng cường cục bộ (MCTCB) được thiết kế như một cách hiệu quả để giới thiệu tính cục bộ vào mạng với khối tích chập sâu và khối chú ý tích chập nhiều đầu (CYTCND) như một bộ trộn mã thẻ hiệu quả. KTCHQ được xây dựng bằng cách kết hợp khối MCTCB và CYTCND trong mô hình mạnh mẽ.

KTCHQ được đề xuất có thể được công thức hóa như sau:

$$\begin{aligned}\tilde{z}^l &= \text{CYTCND}(z^{l-1}) + z^{l-1} \\ z^l &= \text{MCTCB}(\tilde{z}^l) + \tilde{z}^l\end{aligned}\quad (3.1)$$

trong đó  $z^{l-1}$  là đầu vào từ khối  $l - 1$ ,  $\tilde{z}^l$  và  $z^l$  là đầu ra của CYTCND và KTCHQ thứ  $l$ .



Hình 3.3: Mạng mang cơ chế hiệu quả cục bộ vào Bộ chuyển đổi

Để nói sâu về MCTCB, nó được áp dụng theo vị trí cho một chuỗi các mã thẻ  $Z^r$ , có thể được biểu diễn chính xác bằng cách sắp xếp lại chuỗi

các mã thẻ thành một mạng 2 chiều được trình bày trong Hình 3.3. Kết quả là các đặc trưng được định hình lại được biểu diễn như sau:

$$Z^r = \text{BBDCSA}(Z), Z^r \in \mathbb{R}^{h \times w \times d} \quad (3.2)$$

trong đó  $h = H/p$  và  $w = W/p$ . Trong đó  $H$  và  $W$  là chiều cao và chiều rộng của ảnh gốc,  $p$  là hệ số chia tỉ lệ. Bộ biến đổi chuỗi sang ảnh (BBDCSA) lấy một chuỗi và chuyển đổi nó thành một bản đồ đặc trưng có thể được trực quan hóa. Mỗi mã thẻ tương ứng với một điểm ảnh được đặt ở các vị trí của điểm ảnh ấy trên bản đồ đặc trưng. Qua góc nhìn này, có thể giới thiệu tính cục bộ vào mạng bằng cách khôi phục sự gần kề giữa các mã thẻ. Các lớp kết nối đầy đủ có thể được thay thế bằng các lớp tích chập  $1 \times 1$ , tức là

$$\begin{aligned} Y^r &= f(Z^r \circledast W_1^r) \circledast W_2^r \\ Y &= \text{BBDASC}(Y^r) \end{aligned} \quad (3.3)$$

trong đó  $W_1^r \in \mathbb{R}^{d \times \gamma d \times 1 \times 1}$ ,  $W_2^r \in \mathbb{R}^{\gamma d \times d \times 1 \times 1}$  được định hình lại từ  $W_1$ ,  $W_2$  và là các hạt nhân của các lớp tích chập. Với Bộ biến đổi ảnh sang chuỗi (BBDASC), bản đồ đặc trưng hình ảnh được chuyển đổi lại thành một chuỗi mã thẻ, sau đó được sử dụng bởi lớp tự chú ý tiếp theo bằng cách chuyển đổi nó thành mã thẻ hợp nhất.

### 3.1.3 Khôi chuyển đổi chú ý hiệu quả

Trong khi việc biểu diễn cục bộ đã được học hiệu quả thông qua KTCHQ, việc nắm bắt thông tin toàn cục là cấp thiết và cần được giải quyết. Các khôi chuyển đổi đã được biết đến với khả năng nắm bắt các tín hiệu tần số thấp, rất hữu ích để nắm bắt thông tin toàn cục (ví dụ, các hình dạng và cấu trúc toàn cục). Tuy nhiên, đã có một số nghiên cứu liên quan [39] chỉ ra rằng các khôi chuyển đổi có xu hướng làm suy giảm thông tin tần số cao, chẳng hạn như thông tin về kết cấu cục bộ của các

đối tượng ở mức độ nào đó.

Để giải quyết vấn đề trên, khối chuyển đổi chú ý hiệu quả (KCĐCYHQ) được sử dụng, nhờ vào khối tương tác toàn cục và cục bộ, KCĐCYHQ có khả năng nắm bắt các tín hiệu tần số thấp và đồng thời giữ lại các đặc trưng cục bộ và toàn cục cần thiết. Đồng thời, nhờ vào các lớp chú ý hiệu quả, KCĐCYHQ mang lại hiệu suất tính toán tốt hơn, khả năng trích xuất đặc trưng tốt hơn và linh hoạt hơn so với các khối chuyển đổi đơn giản. Các khối chuyển đổi trong KCĐCYHQ được thiết kế để không chỉ tăng cường khả năng xử lý thông tin cục bộ và toàn cục mà còn tối ưu hóa việc sử dụng thông tin từ nhiều vùng khác nhau trong dữ liệu hình ảnh. Kiến trúc KCĐCYHQ bao gồm hai thành phần chính: Tập hợp vùng đa tỷ lệ (THVĐTL) và Tương tác toàn cục và cục bộ (TTTCVCB) cùng với một số cải tiến quan trọng để nâng cao khả năng mô hình hóa toàn cục và hiệu suất phân loại.

**Mô-đun THVĐTL** sử dụng các phép toán tích chập với các bước khác nhau để tổng hợp thông tin từ các vùng kích thước khác nhau của dữ liệu hình ảnh. Điều này cho phép khối THVĐTL thu thập thông tin từ nhiều kích thước vùng khác nhau mà không cần thêm các phép toán nhúng vị trí. Cụ thể, nếu  $x$  là bản đồ đặc trưng đầu vào, phép toán tích chập với các bước khác nhau được biểu diễn như sau:

$$o_n(x) = \text{Tích chập}_{S_n}(\text{Chuẩn hóa}(x)) \quad (3.4)$$

trong đó  $n$  là chỉ số của các phép toán tích chập và Tích chập  $S_n$  là phép toán tích chập với bước  $S_n$ . Tất cả các phép toán này được kết hợp bằng cơ chế Trộn các phép toán có trọng số (TCPCTCS):

$$x_o = \sum_{n=1}^N \frac{e^{\alpha_n}}{\sum_{n'=1}^N e^{\alpha_{n'}}} o_n(x) \quad (3.5)$$

trong đó  $\alpha_n$  là các trọng số học được. Nhìn chung, KCĐCYHQ kết hợp các thành phần tiên tiến để cải thiện khả năng xử lý và mô hình hóa của các mô hình BCD, cung cấp một giải pháp linh hoạt và mạnh mẽ cho

các tác vụ phân loại và phân tích hình ảnh.

**Mô-đun TTTCVVCB** tích hợp hai đường đi: một đường cho thông tin toàn cục và một cho thông tin cục bộ. Đường đi cục bộ tập trung vào việc khai thác thông tin chi tiết từ các vùng nhỏ của dữ liệu, trong khi đường đi toàn cục duy trì mối liên hệ toàn diện hơn. Các đặc trưng đầu vào được phân chia thành các đặc trưng toàn cục và cục bộ, sau đó được đưa qua hai đường tương ứng và được xử lý độc lập và kết hợp lại với nhau để tạo ra một biểu diễn hoàn chỉnh. Công thức cho sự kết hợp này được mô tả như sau:

$$\text{Đầu ra} = \text{Đầu ra toàn cục } \frac{e^{\alpha_g}}{e^{\alpha_g} + e^{\alpha_l}} + \text{Đầu ra cục bộ } \frac{e^{\alpha_l}}{e^{\alpha_g} + e^{\alpha_l}} \quad (3.6)$$

trong đó  $\alpha_g$  là trọng số học được cho nhánh toàn cục.  $\frac{e^{\alpha_g}}{e^{\alpha_g} + e^{\alpha_l}}$  là hệ số chuẩn hóa cho đầu ra từ nhánh toàn cục (Đầu ra toàn cục). Hệ số này đảm bảo rằng trọng số được tính dựa trên hàm SoftMax, phân bổ trọng số cho Đầu ra toàn cục theo một cách mềm mại và liên tục.  $\alpha_l$  là trọng số học được cho nhánh cục bộ.  $\frac{e^{\alpha_l}}{e^{\alpha_g} + e^{\alpha_l}}$  là hệ số chuẩn hóa cho đầu ra từ nhánh cục bộ (Đầu ra cục bộ). Hệ số này cũng được tính dựa trên hàm SoftMax, phân bổ trọng số cho Đầu ra cục bộ.

Ở đường đi toàn cục, xuất hiện khối Tự chú ý nhiều đầu biến dạng đã điều chỉnh (TCYNDBDDC). Khối này sẽ thực hiện điều chỉnh các tọa độ và trọng số của các mảng không gian để tối ưu hóa việc lấy thông tin từ các vùng biến dạng. Sự khác biệt chính giữa TCYNDBDDC và TCYND truyền thống là khả năng điều chỉnh các trọng số của các mảnh ảnh không gian thông qua các phép toán biến dạng và điều chỉnh. Trong mô hình tự chú ý truyền thống, các đặc trưng đầu vào  $X$  được biến đổi thành ba ma trận Truy vấn ( $\mathbf{Q}$ ), Khóa ( $\mathbf{K}$ ) và Giá trị ( $\mathbf{V}$ ) sử dụng hàm  $f_{qkv}$ , biểu diễn là:

$$QKV = f_{qkv}(X) = f_q(X) \oplus f_k(X) \oplus f_v(X) \quad (3.7)$$

trong đó  $Q = f_q(X)$ ,  $K = f_k(X)$ ,  $V = f_v(X)$ , phép toán  $\oplus$  là phép gộp các hàm  $f_q$ ,  $f_k$ , và  $f_v$ .

Điểm khác biệt chính giữa TCYNDBDDC và TCYNĐ truyền thống là khả năng điều chỉnh các trọng số của các mảnh ảnh không gian thông qua các phép toán biến dạng và điều chỉnh, được thể hiện trong công thức:

$$\Delta_l, \Delta_m = f_{md}(Q) \quad (3.8)$$

trong đó  $\Delta_l$  là các độ dời và  $\Delta_m$  là các hệ số điều chỉnh. Sau đó, các đặc trưng  $\mathbf{KV}$  được lấy từ bản đồ đặc trưng mới:

$$\hat{X}_l = S(X_l, \Delta_l) \cdot \Delta_m \quad (3.9)$$

với  $S$  là hàm nội suy song tuyến tính. Trong công thức này,  $\Delta_l$  đại diện cho các tọa độ tương đối với phạm vi không giới hạn cho vị trí thứ  $l$ , trong khi  $\Delta_m$  bị giới hạn trong khoảng  $(0, 1)$ .

Các đặc trưng  $\mathbf{KV}$  này giúp cơ chế chú ý chọn lọc và tổng hợp thông tin từ các phần tử đầu vào, giúp mô hình học sâu xử lý thông tin một cách hiệu quả hơn.

## 3.2 Các mô hình sử dụng trong luận văn

Luận văn đã sử dụng kiến trúc đề xuất cho 3 mô hình phân lớp ứng với từng vị trí xương cụ thể và tinh chỉnh các mô hình này để phù hợp với các tập dữ liệu đã chuẩn bị sẵn. Ngoài ra, phương thức học chuyển tiếp cũng được áp dụng cho cả 3 mô hình để giảm thời gian huấn luyện, tiết kiệm tài nguyên và tăng cường hiệu suất trên các tập dữ liệu đã chuẩn bị.

### 3.2.1 Mô hình phân lớp ứng với vị trí xương cánh tay

Mô hình này mục đích dùng để phân loại dữ liệu thành 5 lớp là "Không bệnh", "Bọc xương", "U sụn", "Sarcoma" và "U đai bào" dựa theo tập dữ liệu đã tổng hợp được từ bệnh viện. Mục tiêu của mô hình là phân lớp được các loại bệnh ung thư xảy ra trên xương cánh tay nhằm đưa ra kết quả dự đoán hỗ trợ cho y bác sĩ trong chẩn đoán bệnh ung thư xương tại vị trí xương cánh tay.

Bảng 3.1: Thông tin dữ liệu sử dụng trong mô hình phân lớp ở vị trí xương cánh tay

Nhân	Không bệnh	Bọc xương	U sụn	Sarcoma	U đai bào
<b>Nguồn gốc</b>	Ảnh được thu thập từ bệnh viện				
<b>Loại dữ liệu</b>	Ảnh X-quang				
<b>Kích thước dữ liệu (mẫu)</b>	49	44	49	20	38

### 3.2.2 Mô hình phân lớp ứng với vị trí xương đùi

Mô hình này mục đích dùng để phân loại dữ liệu thành 5 lớp là "Không bệnh", "U sụn", "Di căn", "Sarcoma" và "U đại bào" dựa theo tập dữ liệu đã tổng hợp được từ bệnh viện. Mục tiêu của mô hình là phân lớp được các loại bệnh ung thư xảy ra trên xương đùi nhằm đưa ra kết quả dự đoán hỗ trợ cho y bác sĩ trong chẩn đoán bệnh ung thư xương tại vị trí xương đùi.

Bảng 3.2: Thông tin dữ liệu sử dụng trong mô hình phân lớp ở vị trí xương đùi

Nhãn	Không bệnh	U sụn	Di căn	Sarcoma	U đại bào
Nguồn gốc	Ảnh được thu thập từ bệnh viện				
Loại dữ liệu	Ảnh X-quang				
Kích thước dữ liệu (mẫu)	113	52	30	99	115

### 3.2.3 Mô hình phân lớp các ứng với vị trí xương chày

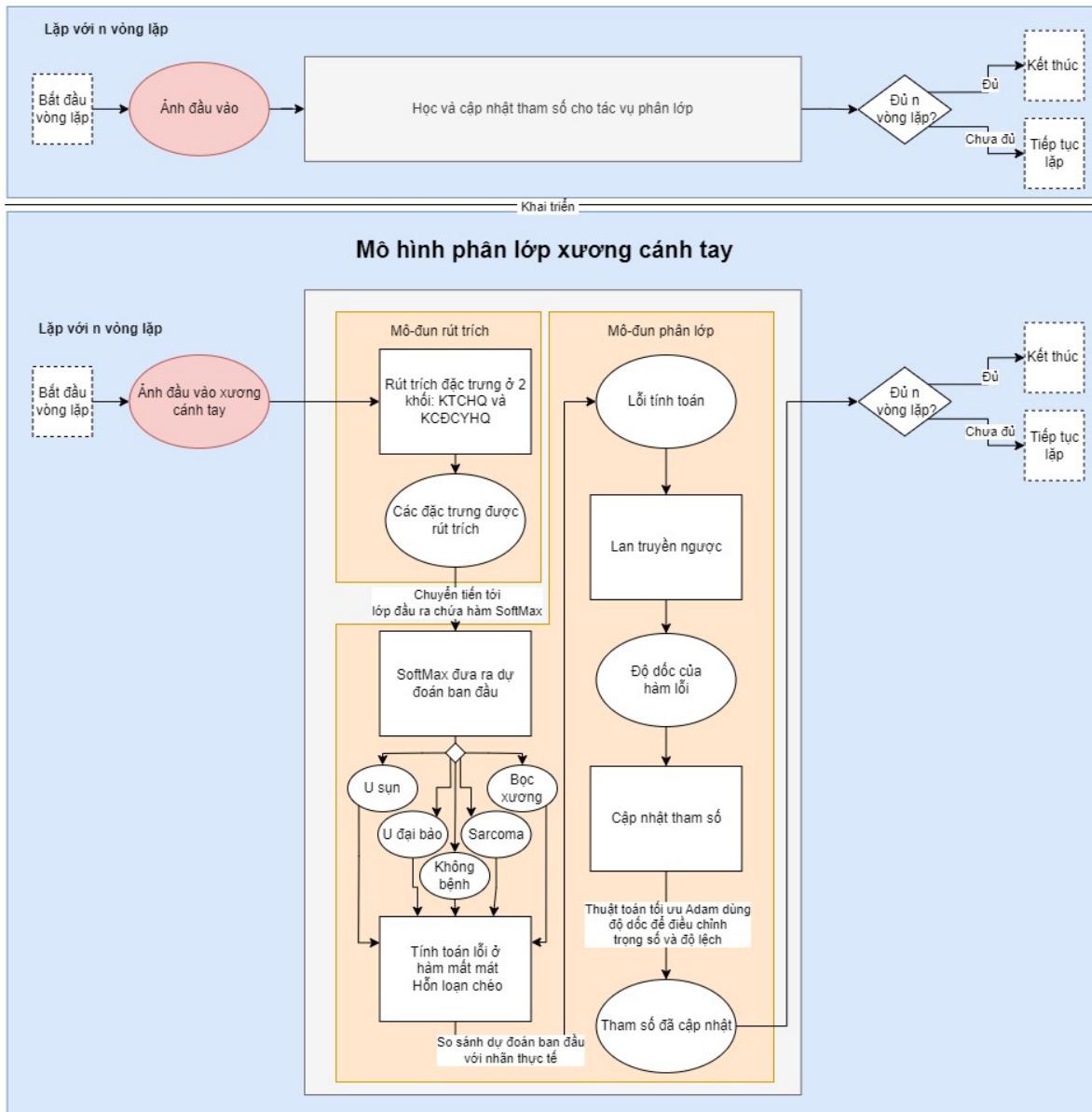
Mô hình này mục đích dùng để phân loại dữ liệu thành 3 lớp là "Không bệnh", "Sarcoma" và "U đại bào" dựa theo tập dữ liệu đã tổng hợp được từ bệnh viện. Mục tiêu của mô hình là phân lớp được các loại bệnh ung thư xảy ra trên xương chày nhằm đưa ra kết quả dự đoán hỗ trợ cho y bác sĩ trong chẩn đoán bệnh ung thư xương tại vị trí xương chày.

Bảng 3.3: Thông tin dữ liệu sử dụng trong mô hình phân lớp ở vị trí xương chày

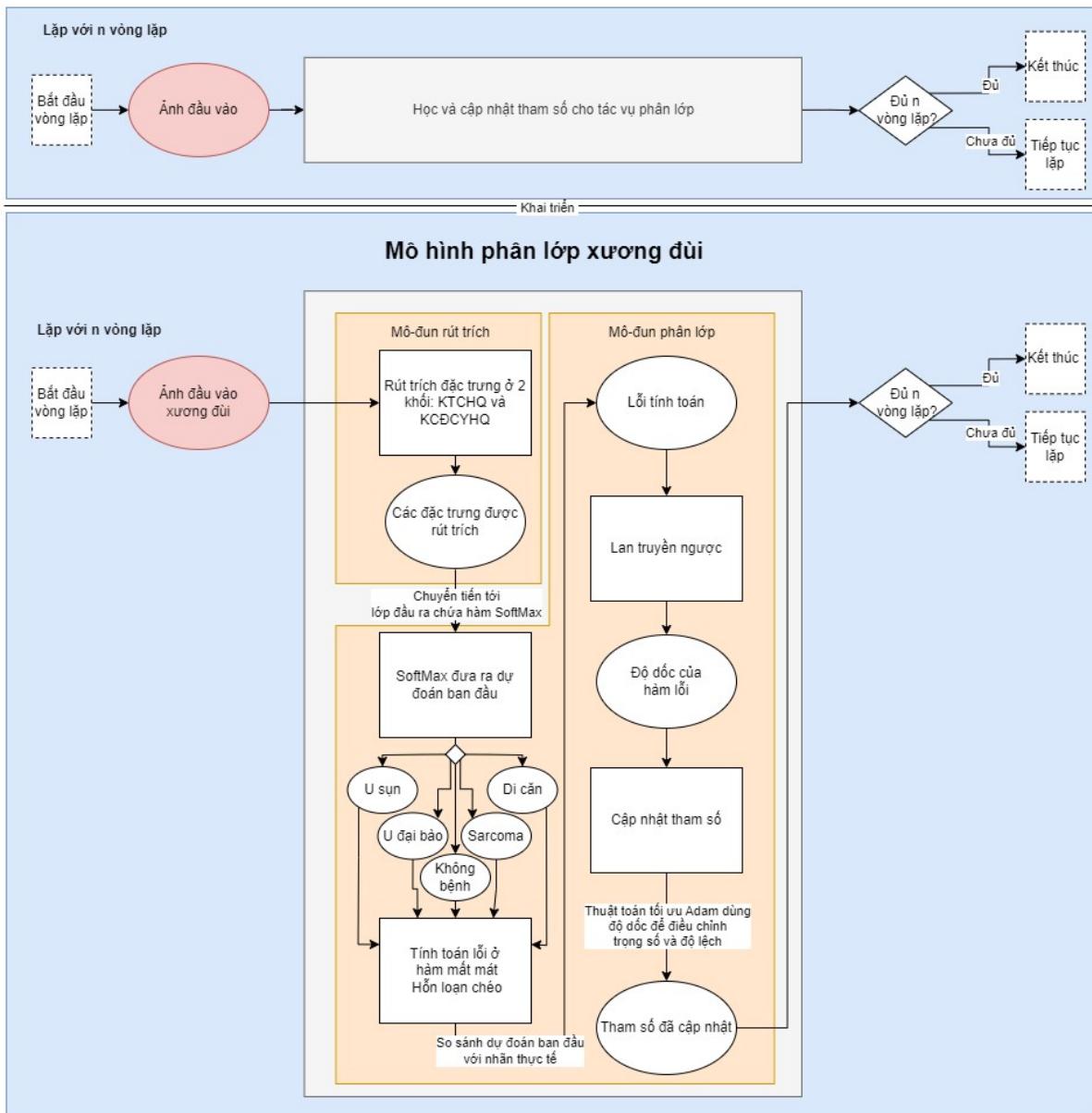
Nhãn	Không bệnh	Sarcoma	U đại bào
Nguồn gốc	Ảnh được thu thập từ bệnh viện		
Loại dữ liệu	Ảnh X-quang		
Kích thước dữ liệu (mẫu)	91	56	92

### **3.3 Mô hình trong giai đoạn ngoại tuyến**

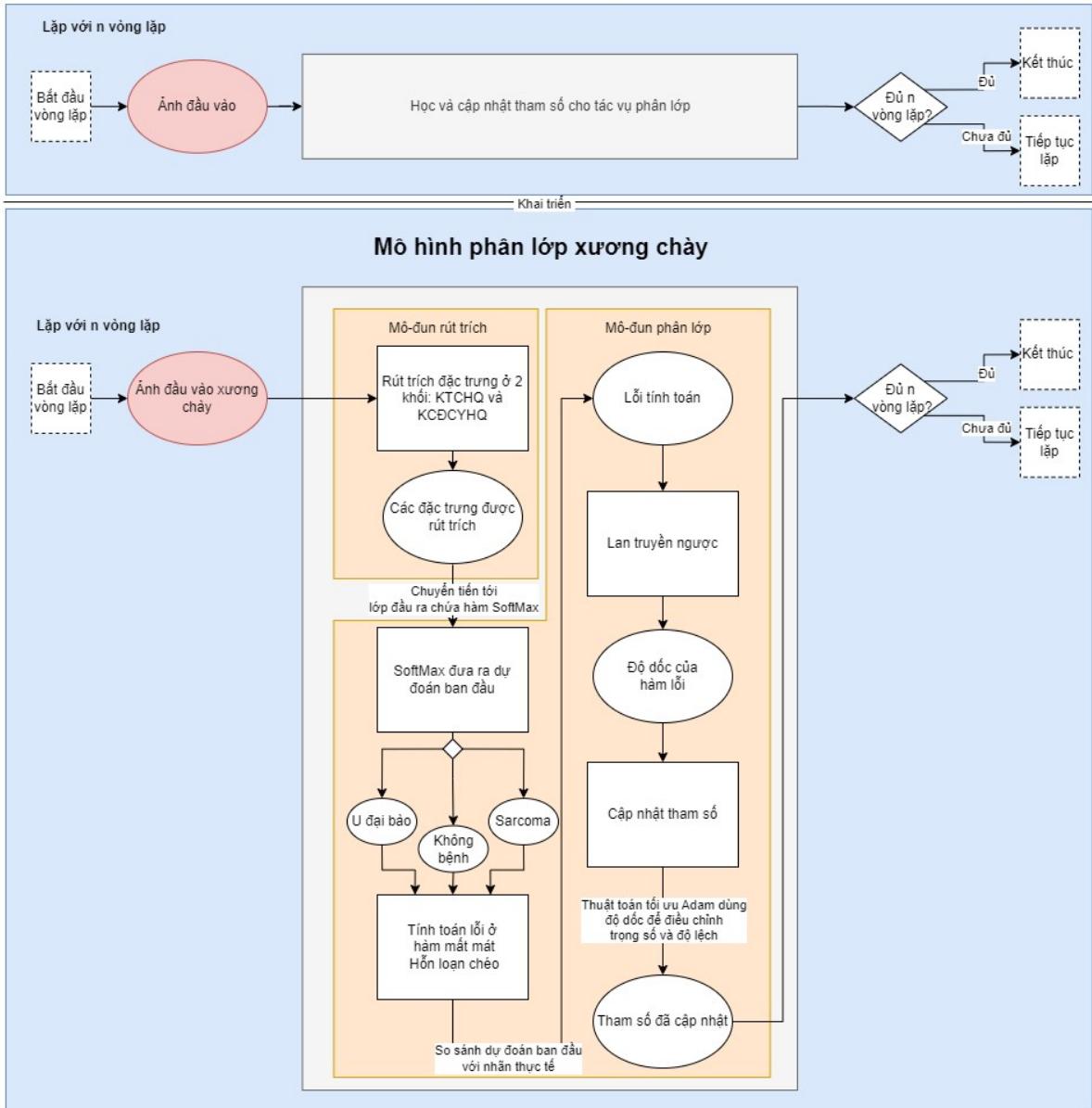
Trong giai đoạn ngoại tuyến, các mô hình của từng vị trí xương được huấn luyện với ảnh đầu vào của vị trí xương tương ứng. Đối với phương thức huấn luyện thông thường, quá trình huấn luyện bao gồm các giai đoạn: rút trích đặc trưng ở mô-đun rút trích và tính toán ở hàm măt măt, lan truyền ngược, cập nhật tham số ở mô-đun phân lớp. Các giai đoạn này sẽ được lặp lại nhiều lần để mô hình học cách trích xuất và sử dụng các đặc trưng một cách hiệu quả, nhằm đạt được hiệu suất tốt nhất.



Hình 3.4: Sơ đồ hoạt động của mô hình phân lớp xương cánh tay  
ở giai đoạn ngoại tuyến



Hình 3.5: Sơ đồ hoạt động của mô hình phân lớp xương đùi  
ở giai đoạn ngoại tuyến



Hình 3.6: Sơ đồ hoạt động của mô hình phân lớp xương chày  
ở giai đoạn ngoại tuyến

### 3.3.1 Mô-đun rút trích đặc trưng

Ảnh đầu vào sau khi chuyển thành các mảnh ảnh sẽ đi qua mô hình và được rút trích đặc trưng tại 2 khối của mô hình.

Tại Khối tích chập hiệu quả: Mô-đun CYTCND tập trung vào nhận diện các cấu trúc quan trọng của khối  $u$ , mô-đun này giúp kết hợp thông tin từ nhiều vùng khác nhau của hình ảnh nhằm tăng cường khả năng phân tích và biểu diễn của mô hình. Sau đó, KTCHQ sử dụng MCTCB để xử lý các đặc trưng đã được tập trung, sử dụng các phép tính 3.2 3.3 để cải thiện độ chính xác và hiệu suất của mô hình. Kết quả đầu ra của KTCHQ là các đặc trưng cục bộ và đặc điểm chi tiết của khối  $u$ .

Tại Khối chuyển đổi chú ý hiệu quả: Sau khi nhận được các đặc trưng từ KTCHQ, hình ảnh được truyền qua lớp THVDTL, nơi mà nó trải qua một chuỗi các phép toán tích chập với các kích thước khác nhau (phép toán 3.4) để trích xuất các đặc trưng đa tỷ lệ. Lớp này cũng áp dụng chuẩn hóa và kết hợp các đầu ra từ các phép toán khác nhau bằng cách sử dụng cơ chế TCPTCTS (phép toán 3.5). Sau đó, đặc trưng thu được từ lớp THVDTL được đưa vào lớp TTTCVCB, nơi các phép toán như 3.6 và TCYNDBDDDC (phép toán 3.8 3.9) được áp dụng để xử lý và tinh chỉnh các đặc trưng. Lớp TTTCVCB cũng có khả năng kết hợp các đầu ra từ các phép toán này dựa trên trọng số học được, nhờ vào cơ chế chú ý và các kỹ thuật tối ưu hóa khác. Sau khi đi qua lớp TTTCVCB, các đặc trưng sẽ được tổng hợp và chuẩn bị đi đến bước tiếp theo.

### 3.3.2 Mô-đun phân lớp

#### 3.3.2.1 Tính toán ở hàm măt mát

Sau khi đặc trưng được rút trích, chúng được chuyển tiến tới lớp đầu ra chứa hàm SoftMax để đưa ra dự đoán ban đầu là véc-tơ dự đoán  $\mathbf{y} = (y_1, y_2, \dots, y_K)$  với  $K$  là số nhãn của từng mô hình.

Tiếp theo, các mô hình tính toán lõi bằng cách so sánh dự đoán ban đầu với nhãn thực thế thông qua hàm măt mát Hỗn loạn chéo bằng công thức:

$$\mathcal{L}_{CE} = - \sum_{i=1}^K t_i \log(y_i) \quad (3.10)$$

Trong đó:

- $\mathcal{L}_{CE}$  là hàm mất mát Hỗn loạn chéo,
- $t_i$  biểu thị nhãn mục tiêu cho lớp  $i$ ,
- $y_i$  đại diện cho xác suất dự đoán cho lớp  $i$ .

### 3.3.2.2 Lan truyền ngược để tính toán độ dốc hàm mất mát

Sau khi tính toán hàm mất mát, quá trình lan truyền ngược được áp dụng để tính toán độ dốc của hàm mất mát đối với từng tham số của mô hình. Độ dốc này cho biết hướng và độ lớn của sự thay đổi cần thiết để giảm giá trị của hàm mất mát. Độ dốc được lan truyền ngược qua các lớp của mô hình, bắt đầu từ lớp cuối cùng và đi ngược lại đến lớp đầu tiên theo công thức:

$$\frac{\partial y}{\partial x} = \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial x} \quad (3.11)$$

Trong đó:

- $\frac{\partial y}{\partial x}$  là đạo hàm của  $y$  theo  $x$ , tức là độ dốc của hàm mất mát theo các tham số của mô hình,
- $\frac{\partial f}{\partial g}$  là đạo hàm của hàm mất mát  $f$  theo đầu ra của mô hình  $g$ ,
- $\frac{\partial g}{\partial x}$  là đạo hàm của đầu ra  $g$  theo các tham số  $x$  của mô hình.

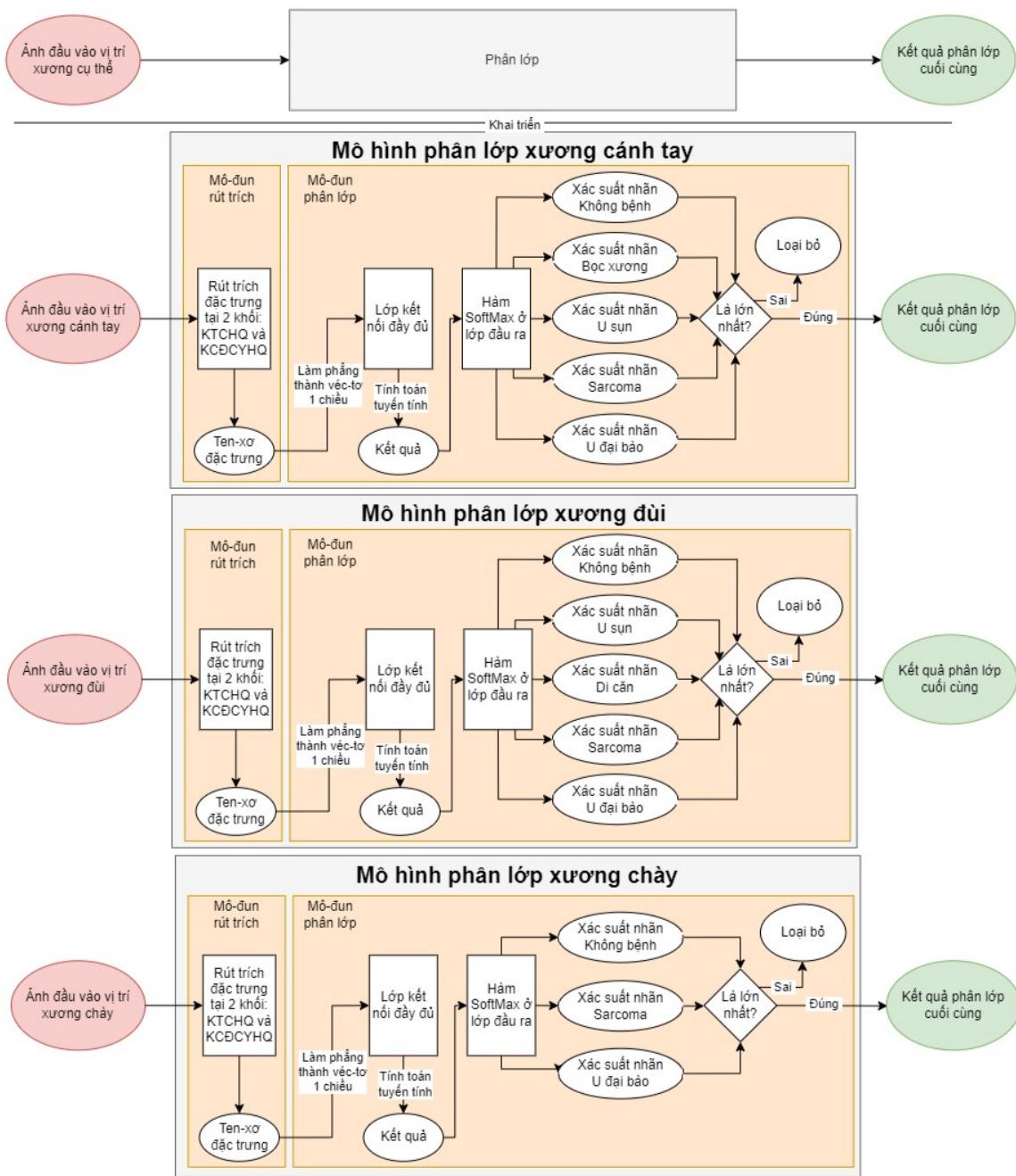
Do các mô hình trong khóa luận sử dụng phương thức học chuyển tiếp, nên các lớp đầu tiên của mô hình sẽ được đóng băng và chỉ có các lớp KNĐĐ cuối mới được điều chỉnh. Quá trình lan truyền ngược vẫn được thực hiện nhưng chỉ áp dụng cho các lớp không bị đóng băng để cập nhật trọng số của các lớp này.

### **3.3.2.3 Cập nhật trọng số của mô hình**

Sau khi tính toán độ dốc của mô hình, thuật toán tối ưu hóa Adam được áp dụng để điều chỉnh các trọng số và độ lệch, cụ thể là chỉ điều chỉnh lớp KNĐĐ trong mạng. Adam sử dụng hai trung bình cộng của độ dốc và bình phương độ dốc để tự động điều chỉnh tốc độ học cho từng tham số của mô hình. Quá trình này giúp tăng tốc độ hội tụ và cải thiện độ chính xác của mô hình khi nó được huấn luyện qua nhiều lần lặp. Mục tiêu là tối ưu hóa liên tục mô hình, nâng cao khả năng phân loại hoặc dự đoán chính xác trên tập dữ liệu, bao gồm cả việc điều chỉnh các tham số trong lớp KNĐĐ để phù hợp với nhiệm vụ phân lớp của mô hình.

## **3.4 Mô hình trong giai đoạn trực tuyến**

Trong giai đoạn trực tuyến, các mô hình của từng vị trí xương thực hiện tác vụ phân lớp ảnh đầu vào của vị trí xương tương ứng. Quá trình phân lớp diễn ra qua các giai đoạn là: rút trích đặc trưng tại mô-đun rút trích, tính toán tuyến tính ở lớp KNĐĐ và đưa ra kết quả phân lớp cuối cùng tại mô-đun phân lớp.



Hình 3.7: Sơ đồ hoạt động của các mô hình trong tác vụ phân lớp ở giai đoạn trực tuyến

### 3.4.1 Mô-đun rút trích đặc trưng

Quá trình rút trích đặc trưng ở đây tương tự như giai đoạn ngoại tuyến. Sau khi đi qua 2 khối, các đặc trưng hình ảnh được lưu lại dưới dạng ten-xo 3 chiều và di chuyển tới lớp KNDD.

## 3.4.2 Mô-đun phân lớp

### 3.4.2.1 Tính toán ở lớp Kết nối đầy đủ

Trước khi đi vào lớp KNĐĐ, các ten-xơ 3 chiều sẽ được làm phẳng thành véc-tơ đặc trưng 1 chiều có kích thước phù hợp với đầu vào của lớp KNĐĐ.

Lớp KNĐĐ gồm các đơn vị chứa bộ trọng số và độ lệch có được từ quá trình huấn luyện, các bộ trọng số và độ lệch này dùng để tính toán tuyến tính đầu ra của lớp KNĐĐ.

$$z = Wx + b \quad (3.12)$$

Trong đó:

- $z$  là véc-tơ đầu ra của lớp KNĐĐ
- $x$  là véc-tơ đặc trưng đã làm phẳng,
- $W$  là ma trận trọng số của lớp KNĐĐ,
- $b$  là véc-tơ điều chỉnh (độ lệch).

### 3.4.2.2 Đưa ra kết quả phân lớp cuối cùng

Sau khi kết quả được tính toán tuyến tính từ lớp KNĐĐ, nó được đưa vào hàm kích hoạt SoftMax để tiến hành phân lớp theo công thức:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \quad \text{với } i = 1, 2, \dots, n \quad (3.13)$$

Trong đó:

- $\sigma(z)_i$  là xác suất của nhãn  $i$  sau khi áp dụng SoftMax,
- $z_i$  là giá trị đầu ra từ lớp KNĐĐ cho nhãn  $i$ ,

- $n$  là tổng số lớp.

Giá trị  $\sigma(\mathbf{z})_i$  nào lớn nhất thì đó là kết quả phân lớp cuối cùng của mô hình.

### 3.5 Giải thích kết quả đầu ra của mô hình

Cuối cùng, để có thể hỗ trợ các bác sĩ, Grad-CAM được sử dụng nhằm giải thích kết quả phân lớp của các mô hình. Đây là một kỹ thuật hữu ích để hiểu rõ hơn về cách mà mô hình dự đoán bằng cách tạo ra các bản đồ nhiệt của các vùng quan trọng trong ảnh đầu vào. Không có Grad-CAM thì tác vụ phân lớp của các mô hình vẫn hoạt động bình thường và không bị ảnh hưởng gì.

Grad-CAM được kết nối đến lớp tích chập gần cuối của các mô hình, có kích thước đầu ra vừa phải. Đầu vào của Grad-CAM là các bản đồ đặc trưng từ lớp tích chập ấy, nơi mà thông tin không gian của hình ảnh vẫn được bảo tồn. Trong các mô hình phân lớp ung thư xương, bản đồ đặc trưng là một ma trận ba chiều (gồm kích thước lô, chiều cao và chiều rộng ảnh đầu vào).

Đầu tiên, tính độ dốc của điểm số cho lớp cụ thể (trước khi áp dụng SoftMax) đối với các kích hoạt của bản đồ đặc trưng từ lớp tích chập cuối cùng. Điều này cho biết mức độ thay đổi của điểm số lớp khi các kích hoạt trong bản đồ đặc trưng thay đổi:

$$\nabla_{A_{ij}^k} y_c = \frac{\partial y_c}{\partial A_{ij}^k} \quad (3.14)$$

Trong đó:

- $y_c$  là điểm số của lớp  $c$ ,
- $A_{ij}^k$  là kích hoạt tại vị trí hàng  $i$  cột  $j$  của bản đồ đặc trưng thứ  $k$ ,
- $\nabla_{A_{ij}^k} y_c$  là độ dốc của điểm số  $y_c$  đối với kích hoạt  $A_{ij}^k$ .

Tiếp theo, sử dụng phép tính trung bình toàn cục trên độ dốc đã tính được để thu được trọng số  $\alpha_{c,k}$ . Trọng số này biểu thị mức độ quan trọng của từng nơ-ron trong bản đồ đặc trưng đối với lớp cụ thể:

$$\alpha_{c,k} = \frac{1}{Z} \sum_i \sum_j \frac{\partial y_c}{\partial A_k} \quad (3.15)$$

Trong đó,  $Z$  là hằng số chuẩn hóa để tổng các trọng số bằng 1.

Kết hợp có trọng số của các bản đồ đặc trưng với trọng số  $\alpha_{c,k}$  để tạo ra một bản đồ nhiệt. Áp dụng hàm DVTTĐHC để chỉ giữ lại các giá trị dương:

$$L_c^{Grad-CAM} = DVTTDH \left( \sum_k \alpha_{c,k} A_k \right) \quad (3.16)$$

Bản đồ nhiệt  $L_c^{Grad-CAM}$  được tạo ra có kích thước tương tự như các bản đồ đặc trưng cuối cùng trong mô hình. Mỗi điểm ảnh trên Bản đồ nhiệt được tô màu dựa trên mức độ quan trọng đã được xác định bởi  $\alpha_{c,k}$ . Các điểm ảnh có giá trị cao hơn sẽ được tô màu sáng hơn, thể hiện vị trí và mức độ ảnh hưởng của chúng đối với quyết định của mô hình.

# Chương 4

## Thực nghiệm

### 4.1 Huấn luyện

#### 4.1.1 Tập dữ liệu xương

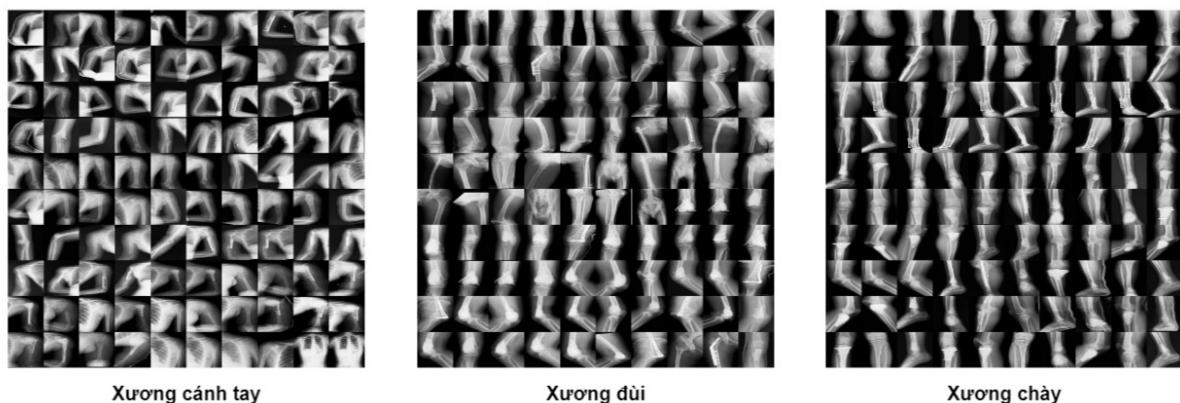
Tập dữ liệu Bone\_Saigon-OTH\_NNHLT\_1.0 gồm các ảnh X-quang xương được thu thập từ bệnh viện chấn thương chỉnh hình do nhóm nghiên cứu của thầy Lý Quốc Ngọc và Th.S Võ Thế Hào thực hiện. Trong đó, mỗi ảnh tùy thuộc vào vị trí sẽ được gắn nhãn ứng với từng bệnh có thể xảy ra ở vị trí đấy. Các ảnh này được chia thành 3 tập nhỏ (tỉ lệ mỗi tập là 70% để huấn luyện, 20% để kiểm tra và 10% để kiểm thử) ứng với 3 vị trí xương cụ thể là: xương cánh tay, xương đùi và xương chày. Cụ thể:

Bảng 4.1: Thông tin về từng tập dữ liệu nhỏ trong tập  
Bone\_Saigon-OTH\_NNHLT\_1.0

Tập dữ liệu	Nhãn	Nguồn gốc	Loại dữ liệu	Kích thước (mẫu)
Xương cánh tay	Không bệnh	Ảnh được thu thập từ bệnh viện	Ảnh X-quang	49
	Bọc xương (lành)			44
	U sụn (lành)			49
	Sarcoma (ác)			20
	U đai bào (ác)			38
<hr/>				
Xương đùi	Không bệnh	Ảnh được thu thập từ bệnh viện	Ảnh X-quang	113
	U sụn (lành)			52
	Di căn (ác)			30
	Sarcoma (ác)			99
	U đai bào (ác)			115
<hr/>				
Xương chày	Không bệnh	Ảnh được thu thập từ bệnh viện	Ảnh X-quang	91
	Sarcoma (ác)			56
	U đai bào (ác)			92

Bảng 4.2: Thông tin số mẫu của từng tập huấn luyện, tập kiểm tra và tập kiểm thử

Tập dữ liệu	Tổng số mẫu gốc	Số mẫu tập huấn luyện	Số mẫu tập kiểm tra	Số mẫu tập kiểm thử
Xương cánh tay	200	140	40	20
Xương đùi	409	286	82	41
Xương chày	239	167	48	24



Hình 4.1: Ví dụ một số mẫu của các tập dữ liệu

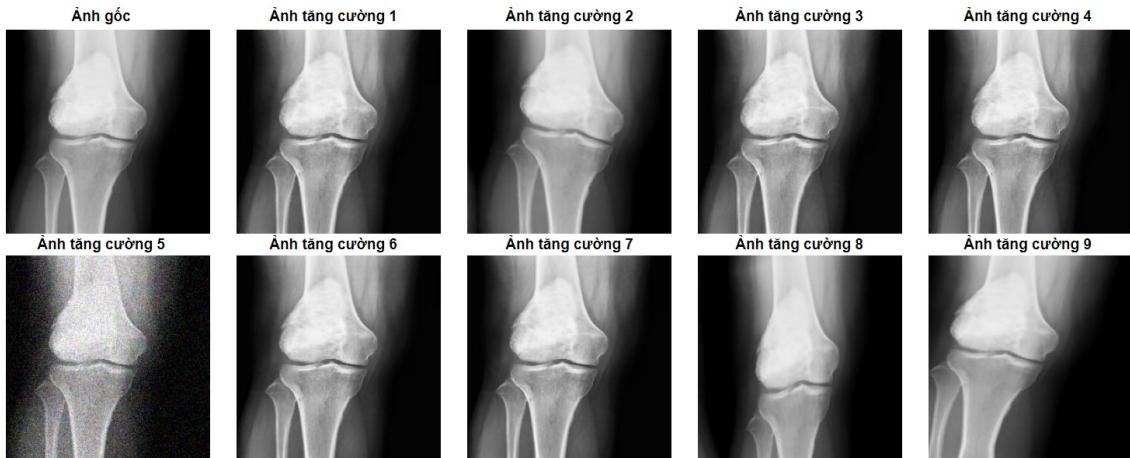
### 4.1.2 Tiề̂n xử lý dữ liệu

Trước khi đưa vào mô hình, các ảnh được tiền xử lý bao gồm thay đổi định dạng ảnh và tăng cường dữ liệu.

- Thay đổi định dạng ảnh bao gồm thay đổi kích thước các ảnh về kích thước chuẩn  $224 \times 224$  để phù hợp với mô hình. Chuyển đổi màu sắc từ Xanh dương - Xanh lá - Đỏ sang Đỏ - Xanh lá - Xanh dương và chuẩn hóa dữ liệu thành ma trận đa chiều.
- Các kỹ thuật tăng cường ảnh bao gồm: Thay đổi độ tương phản, độ sáng, xoay ảnh, lật ảnh ngang dọc, xóa một phần thông tin ảnh ngẫu nhiên, biến đổi đòn hồi, thêm nhiễu Gauss sao cho 1 ảnh gốc tạo thành 10 ảnh nhầm tăng độ đa dạng của dữ liệu.

Bảng 4.3: Thông tin số mẫu của từng tập huấn luyện sau khi được tăng cường

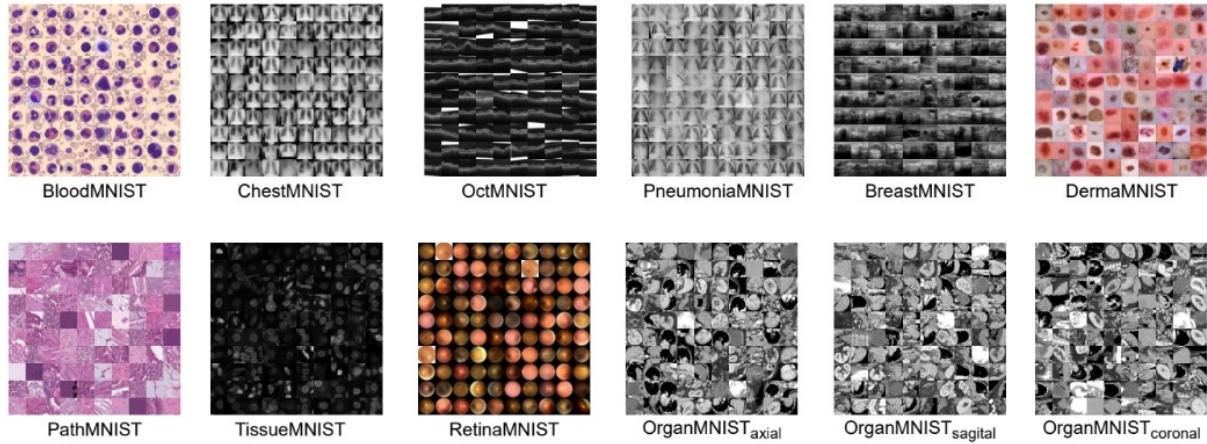
Tập dữ liệu	Tổng số mẫu gốc tập huấn luyện	Tổng số mẫu sau tăng cường
Xương cánh tay	140	1400
Xương đùi	286	2860
Xương chày	167	1670



Hình 4.2: Ví dụ ảnh được tăng cường dữ liệu

### 4.1.3 Tiền huấn luyện

Mô hình được tiến hành tiền huấn luyện trên tập dữ liệu ChestMNIST (thuộc tập dữ liệu lớn hơn là MedMNIST [40]) do đây là tập dữ liệu ảnh X-quang xương để giúp tăng độ chính xác khi huấn luyện trên tập dữ liệu ảnh ung thư xương đã được chuẩn bị.



Hình 4.3: Các tập dữ liệu nhỏ trong tập MedMNIST. Tập ChestMNIST được lựa chọn để tiền huấn luyện mô hình

#### 4.1.4 Tinh chỉnh mô hình

Sau khi tiền huấn luyện trên tập dữ liệu ChestMNIST. Tiến hành học chuyển tiếp mô hình cho tác vụ phân lớp ảnh ung thư xương của tập dữ liệu đã chuẩn bị, các trọng số của các lớp đầu được giữ nguyên và chỉ huấn luyện lại các lớp cuối cùng. Cụ thể, thay thế lớp đầu ra ban đầu bằng một lớp đầu ra mới phù hợp với số lượng lớp phân loại trong bộ dữ liệu ung thư xương. Các lớp KNĐĐ cuối được thay thế bằng một lớp KNĐĐ khác với các đơn vị đầu ra ứng với số lớp, sử dụng hàm SoftMax để tính xác suất cho mỗi lớp. Để tận dụng các trọng số đã được học từ quá trình tiền huấn luyện trước đó, các trọng số của các lớp đầu tiên trong mô hình sẽ bị đóng băng. Các trọng số này sẽ không được cập nhật trong quá trình huấn luyện lại, giúp tiết kiệm thời gian và tài nguyên.

#### 4.1.5 Cài đặt tham số huấn luyện và giải quyết quá khớp

Quá trình huấn luyện mô hình được thực hiện với các tham số sau:

- **Số lần lặp:** 100
- **Kích thước lô:** 10
- **Tốc độ học:** 0.005
- **Hàm mất mát:** Hàm Hỗn loạn chéo để phân loại nhiều lớp.
- **Bộ tối ưu hóa:** Thuật toán tối ưu hóa Adam.

Về vấn đề giải quyết quá khớp, do mô hình học theo phương thức học chuyển tiếp nên các lớp phía trước đã được đóng băng. Điều này giúp giảm số lượng tham số cần tối ưu hóa và hạn chế khả năng mô hình học quá mức từ dữ liệu huấn luyện. Cuối cùng là thêm lớp bỏ qua ngẫu nhiên (LBQNN) sau 2 khối nhằm giảm khả năng quá khớp.

## 4.2 Kết quả huấn luyện

### 4.2.1 Độ đo kết quả huấn luyện

#### 4.2.1.1 Độ chính xác

Đối với bài toán phân lớp, Độ chính xác (DCX) là độ đo được sử dụng phổ biến nhất. Công thức tính Độ chính xác như sau:

$$\text{Độ chính xác} = \frac{\text{Số mẫu phân lớp đúng}}{\text{Tổng số mẫu}} \quad (4.1)$$

Độ chính xác cho biết tỷ lệ phần trăm các mẫu được phân loại đúng so với tổng số mẫu. Nó cung cấp một cái nhìn tổng quan về khả năng dự đoán của mô hình.

Tuy nhiên, độ chính xác có thể không phản ánh chính xác hiệu suất của mô hình trong những trường hợp mất cân bằng dữ liệu , khi mà số lượng mẫu các nhãn không cân đối nhau. Trong các trường hợp này, độ đo diện tích dưới đường cong được sử dụng bổ sung để đánh giá chi tiết hơn về hiệu suất của mô hình.

#### 4.2.1.2 Diện tích dưới đường cong ROC

Diện tích dưới đường cong ROC (DTDDCR) là một độ đo quan trọng để đánh giá hiệu suất của mô hình phân loại, đặc biệt trong các tình huống mất cân bằng dữ liệu. Công thức tính Diện tích dưới đường cong ROC:

$$\text{DTDDCR} = \frac{1}{N(N - 1)} \sum_{i \neq j} \text{DTDDCR}_{ij} \quad (4.2)$$

Trong đó:

- $N$  là số lớp,
- $\text{DTDDCR}_{ij}$  là diện tích dưới đường cong ROC giữa các lớp  $i$  và  $j$ .

DTDDCR cho biết khả năng phân biệt của mô hình giữa các lớp khác nhau. Giá trị DTDDCR càng cao, mô hình càng tốt trong việc phân loại.

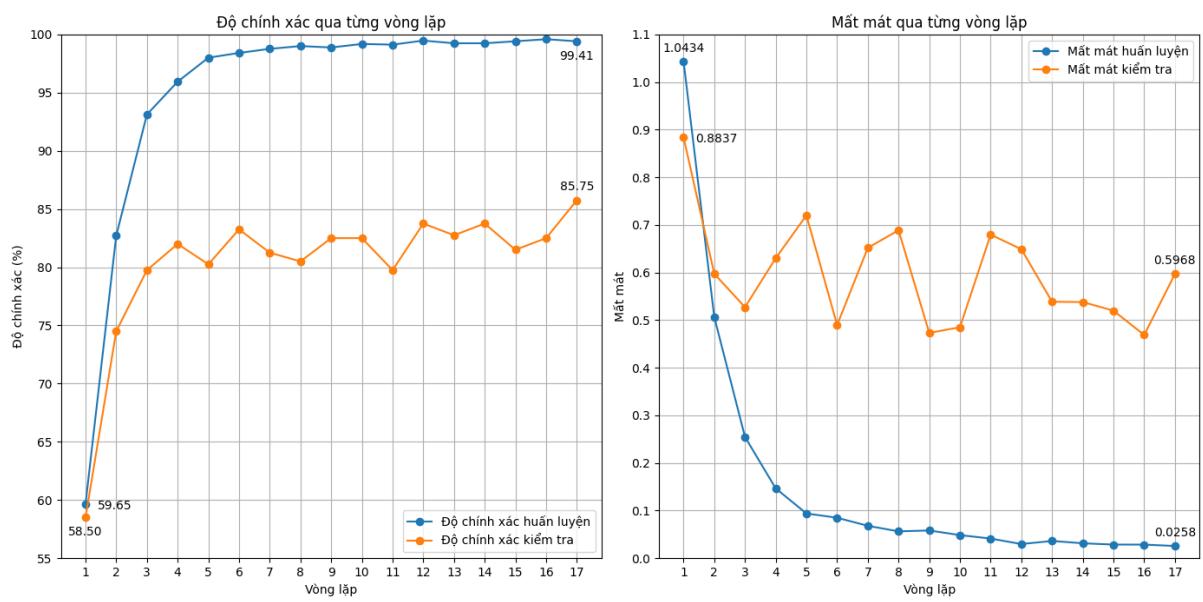
DTDDCR đặc biệt hữu ích khi dữ liệu mất cân bằng, vì nó không bị ảnh hưởng bởi tỷ lệ các lớp như độ chính xác. Trong các tình huống mất cân bằng dữ liệu, một mô hình có thể đạt độ chính xác cao bằng cách chỉ dự đoán lớp chiếm đa số, nhưng DTDDCR sẽ phản ánh rõ hơn khả năng phân biệt giữa các lớp của mô hình. Điều này giúp đánh giá hiệu suất của mô hình một cách toàn diện hơn.

## 4.2.2 Kết quả huấn luyện trên tập huấn luyện của các mô hình

### 4.2.2.1 Mô hình xương cánh tay

Bảng 4.4: Kết quả huấn luyện của mô hình xương cánh tay

Vòng lặp	Độ chính xác huấn luyện	Mất mát huấn luyện	Độ chính xác kiểm tra	Mất mát kiểm tra
1	59.60%	1.0434	58.50%	0.8837
2	82.71%	0.5063	74.50%	0.5973
3	93.12%	0.2547	79.75%	0.5272
4	95.94%	0.1467	82.00%	0.6301
5	98.00%	0.0940	80.25%	0.7200
6	98.41%	0.0850	83.25%	0.4897
7	98.76%	0.0681	81.25%	0.6518
8	99.00%	0.0564	80.50%	0.6888
9	98.88%	0.0583	82.50%	0.4732
10	99.18%	0.0486	82.50%	0.4851
11	99.12%	0.0413	79.75%	0.6796
12	99.47%	0.0298	83.75%	0.6489
13	99.24%	0.0363	82.75%	0.5386
14	99.24%	0.0314	83.75%	0.5380
15	99.41%	0.0287	81.50%	0.5199
16	99.59%	0.0286	82.50%	0.4688
17	99.41%	0.0258	85.75%	0.5968

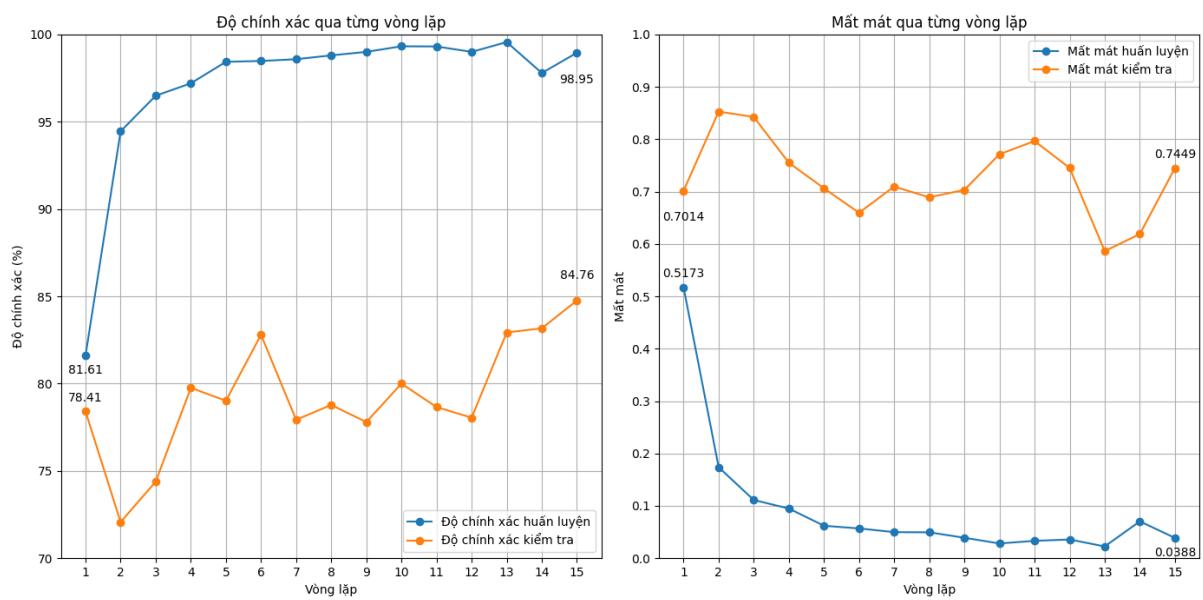


Hình 4.4: Biểu đồ kết quả huấn luyện mô hình xương cánh tay

#### 4.2.2.2 Mô hình xương đùi

Bảng 4.5: Kết quả huấn luyện trên tập huấn luyện của mô hình xương đùi

Vòng lặp	Độ chính xác huấn luyện	Mất mát huấn luyện	Độ chính xác kiểm tra	Mất mát kiểm tra
1	81.61%	0.5173	78.41%	0.7014
2	94.45%	0.1737	72.07%	0.8527
3	96.49%	0.1112	74.39%	0.8424
4	97.20%	0.0949	79.76%	0.7550
5	98.43%	0.0620	79.02%	0.7063
6	98.48%	0.0569	82.80%	0.6595
7	98.58%	0.0498	77.93%	0.7096
8	98.80%	0.0495	78.78%	0.6890
9	99.00%	0.0391	77.80%	0.7032
10	99.32%	0.0282	80.00%	0.7712
11	99.31%	0.0332	78.66%	0.7964
12	99.00%	0.0358	78.05%	0.7451
13	99.56%	0.0225	82.93%	0.5863
14	97.79%	0.0704	83.17%	0.6189
15	98.95%	0.0388	84.76%	0.7449

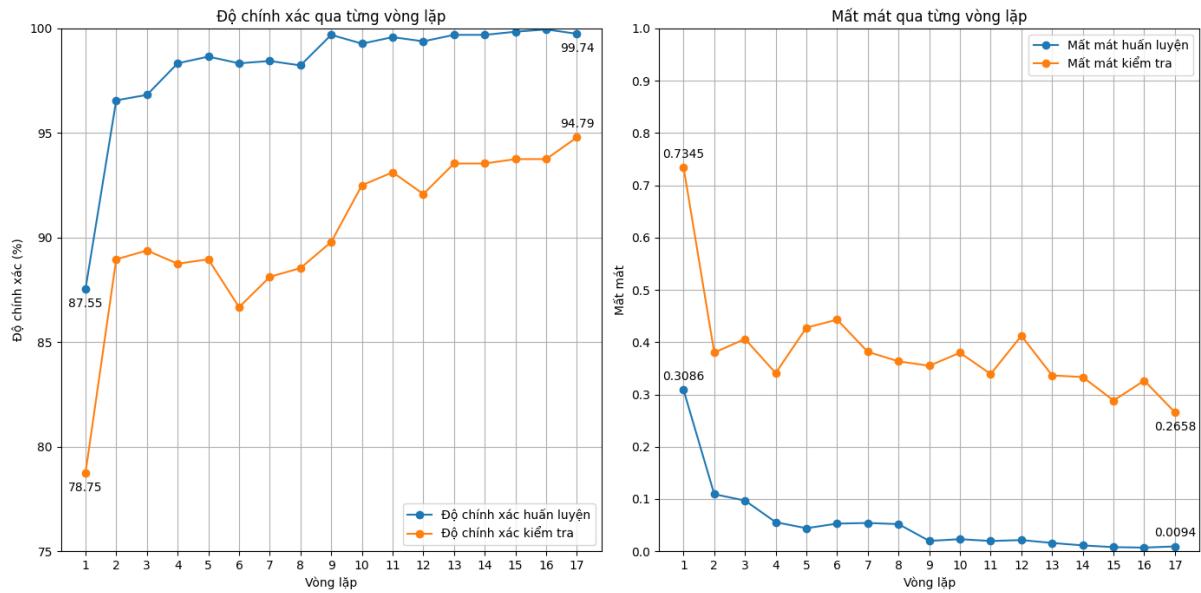


Hình 4.5: Biểu đồ kết quả huấn luyện mô hình xương đùi

#### 4.2.2.3 Mô hình xương chày

Bảng 4.6: Kết quả huấn luyện của mô hình xương chày

Vòng lặp	Độ chính xác huấn luyện	Mất mát huấn luyện	Độ chính xác kiểm tra	Mất mát kiểm tra
1	87.55%	0.3086	78.75%	0.7345
2	96.56%	0.1092	88.96%	0.3800
3	96.82%	0.0971	89.38%	0.4061
4	98.33%	0.0557	88.75%	0.3408
5	98.65%	0.0442	88.96%	0.4277
6	98.33%	0.0530	86.67%	0.4430
7	98.44%	0.0542	88.12%	0.3814
8	98.23%	0.0520	88.54%	0.3633
9	99.69%	0.0199	89.79%	0.3548
10	99.27%	0.0232	92.50%	0.3798
11	99.58%	0.0197	93.12%	0.3390
12	99.38%	0.0215	92.08%	0.4122
13	99.69%	0.0160	93.54%	0.3364
14	99.69%	0.0113	93.54%	0.3332
15	99.84%	0.0080	93.75%	0.2883
16	99.95%	0.0071	93.75%	0.3264
17	99.74%	0.0094	94.79%	0.2658



Hình 4.6: Biểu đồ kết quả huấn luyện mô hình xương chày

#### 4.2.3 Kết quả huấn luyện sau khi đã kiểm tra trên các tập dữ liệu minh chứng cho đóng góp số 1

Bảng 4.7 thể hiện so sánh về Độ chính xác (%) và Diện tích dưới đường cong ROC (%) trên tập dữ liệu Bone\_Saigon-OTH\_NNHLT\_1.0 giữa các phương pháp tiên tiến trước đây là ResNet, DenseNET, Bộ chuyển đổi trực quan, DeiT, EATFormer MedViT và mô hình đề xuất BoneViT theo vị trí xương tương ứng với từng tập dữ liệu (Ví dụ mô hình xương cánh tay sẽ tương ứng với tập dữ liệu xương cánh tay).

Bảng 4.7: Kết quả huấn luyện của các mô hình trên 3 tập dữ liệu xương cánh tay, xương đùi và xương chày

Mô hình	Tập xương cánh tay		Tập xương đùi		Tập xương chày	
	ĐCX	DTDDCR	ĐCX	DTDDCR	ĐCX	DTDDCR
BoneViT (MNTC và BCD)	85.75	96.38	84.26	93.97	94.79	99.02
MedViT (MNTC và BCD)	70.32	86.14	72.42	80.23	88.11	94.63
ResNet18 (MNTC)	83.62	95.19	83.75	92.22	94.83	99.14
DenseNet (MNTC)	83.88	96.53	84.14	93.52	96.66	99.75
BCDTQ (BCD)	53.62	76.82	47.00	77.92	76.50	90.70
Deit (BCD)	56.24	75.51	48.6	74.66	71.75	87.60
EATFormer (BCD)	69.75	83.41	70.60	89.96	83.95	95.74

Bảng 4.8: Kết quả huấn luyện của các mô hình trên 3 tập dữ liệu xương cánh tay, xương đùi và xương chày theo thứ hạng

Mô hình	Tập xương cánh tay		Tập xương đùi		Tập xương chày	
	ĐCX	DTDDCR	ĐCX	DTDDCR	ĐCX	DTDDCR
BoneViT (MNTC và BCĐ)	(1)	(2)	(1)	(1)	(3)	(3)
MedViT (MNTC và BCĐ)	(4)	(4)	(4)	(5)	(4)	(5)
ResNet18 (MNTC)	(3)	(3)	(3)	(3)	(2)	(2)
DenseNet (MNTC)	(2)	(1)	(2)	(2)	(1)	(1)
BCDTQ (BCĐ)	(7)	(6)	(7)	(6)	(6)	(6)
Deit (BCĐ)	(6)	(7)	(6)	(7)	(7)	(7)
EATFormer (BCĐ)	(5)	(5)	(5)	(4)	(5)	(4)

#### **4.2.4 Đánh giá kết quả huấn luyện**

Có thể thấy rõ độ chính xác và diện tích dưới đường cong ROC của mô hình xương cánh tay và mô hình xương đùi thấp hơn so với mô hình xương chày. Điều này có thể được dự đoán trước do vấn đề về dữ liệu. Trong khi xương cánh tay, xương đùi có tận 5 lớp thì xương chày chỉ có 3 lớp. Tổng số mẫu cho mỗi tập dữ liệu cũng không quá nhiều để các mô hình có thể học đủ sâu đến khi có thể nắm bắt được hết tất cả đặc trưng của từng vị trí xương.

Nhưng khi so sánh với các phương pháp tiên tiến, BoneViT vẫn có số liệu cao. Nghĩa là khả năng học trên tập dữ liệu ít và mất cân bằng của BoneViT là tốt. Điều này cũng đồng thời thành công minh chứng cho đóng góp số 1.

### **4.3 Thực nghiệm phân lớp ảnh ung thư xương**

#### **4.3.1 Mục đích thực nghiệm**

Mục đích thực nghiệm để kiểm tra khả năng phân lớp cũng như khả năng diễn giải kết quả phân lớp bằng bản đồ nhiệt của từng mô hình ở các vị trí xương trên các tập dữ liệu ngoài tập huấn luyện nhằm minh chứng cho đóng góp 2.

#### **4.3.2 Kết quả thực nghiệm với các ảnh trong tập kiểm thử lần 1**

Chọn các ảnh ung thư xương của 3 vị trí xương cánh tay, xương đùi, xương chày không nằm trong tập ảnh huấn luyện, tiến hành thực hiện biến đổi hình học (vẫn giữ nguyên các thông tin chính của ảnh). Tiếp đó là đưa vào mô hình tương ứng với vị trí xương để tiến hành phân lớp.

Dưới đây là kết quả phân lớp đúng (đã được bác sĩ chuyên môn kiểm chứng tại phần phụ lục B) khi thực nghiệm kèm theo bản đồ nhiệt giải thích cho kết quả phân lớp:

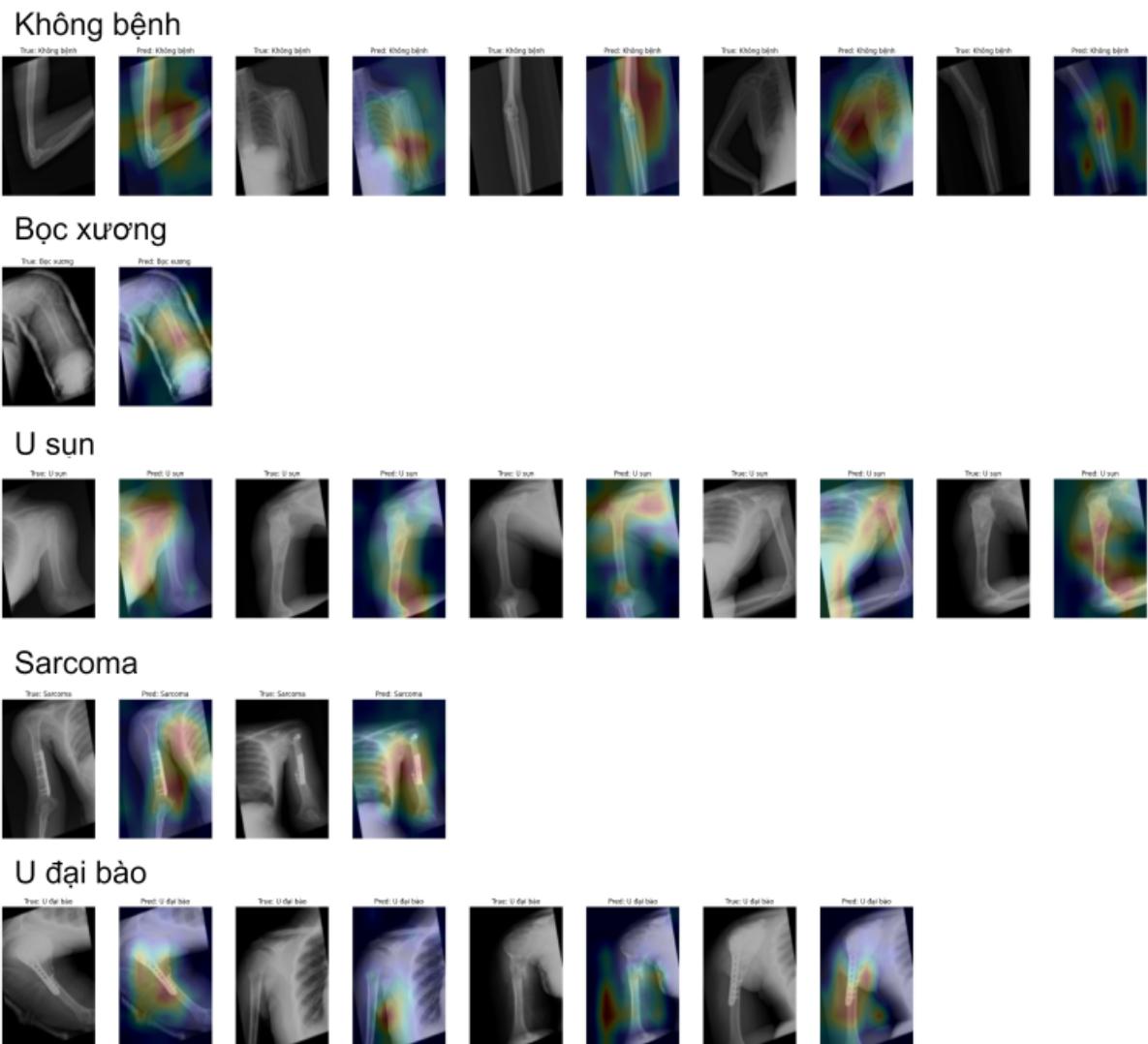
Bảng 4.9: Kết quả kiểm thử tính theo độ chính xác (%) của các mô hình trên 3 tập dữ liệu xương cánh tay, xương đùi và xương chày

Mô hình	Tập xương cánh tay	Tập xương đùi	Tập xương chày
BoneViT (MNTC và BCD)	85.00	48.78	83.33
MedViT (MNTC và BCD)	80.00	46.34	62.50
ResNet18 (MNTC)	85.00	60.98	79.17
DenseNet (MNTC)	85.00	60.98	87.50
BCDTQ (BCD)	65.00	36.59	58.33
DeiT (BCD)	75.00	43.90	54.17
EATFormer (BCD)	80.00	48.78	75.00

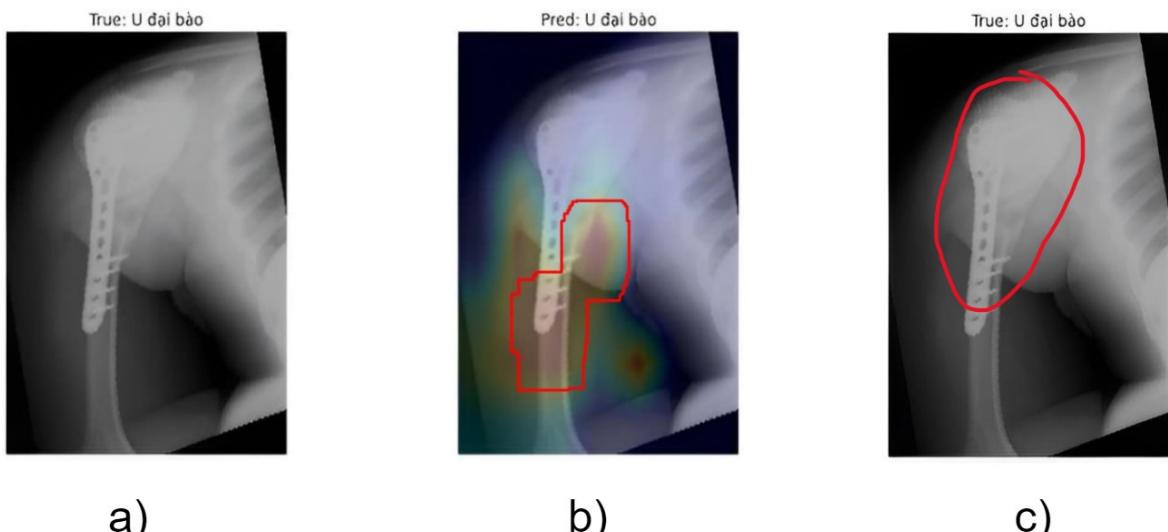
Bảng 4.10: Kết quả kiểm thử của các mô hình trên 3 tập dữ liệu xương cánh tay, xương đùi và xương chày theo thứ hạng

Mô hình	Tập xương cánh tay	Tập xương đùi	Tập xương chày
BoneViT (MNTC và BCD)	(1)	(3)	(3)
MedViT (MNTC và BCD)	(3)	(3)	(5)
ResNet18 (MNTC)	(3)	(1)	(2)
DenseNet (MNTC)	(1)	(1)	(1)
BCDTQ (BCD)	(7)	(7)	(7)
DeiT (BCD)	(6)	(6)	(6)
EATFormer (BCD)	(5)	(5)	(4)

Phân lớp đúng ở xương cánh tay: 17/20 ảnh



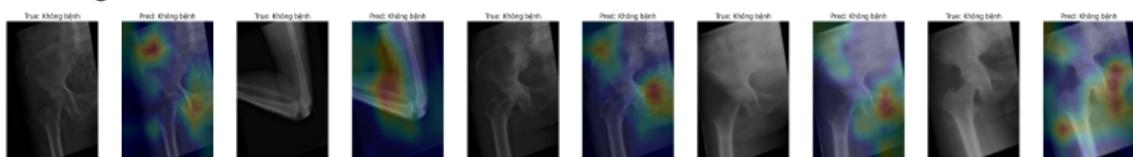
Hình 4.7: Kết quả phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương cánh tay



Hình 4.8: Ví dụ một mẫu phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương cánh tay. a) ảnh gốc; b) ảnh có bản đồ nhiệt; c) ảnh có vùng chứa bệnh do bác sĩ đánh dấu

Phân lớp đúng ở xương đùi: 20/41

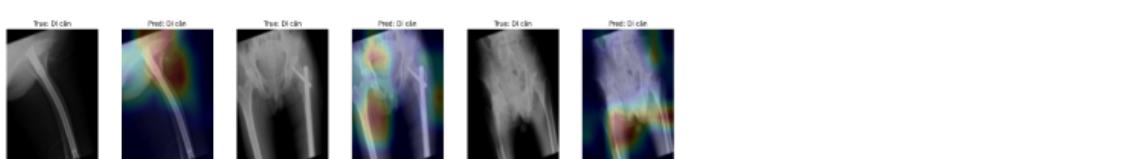
### Không bệnh



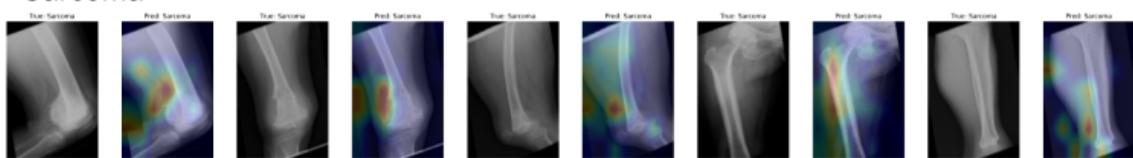
### U sụn



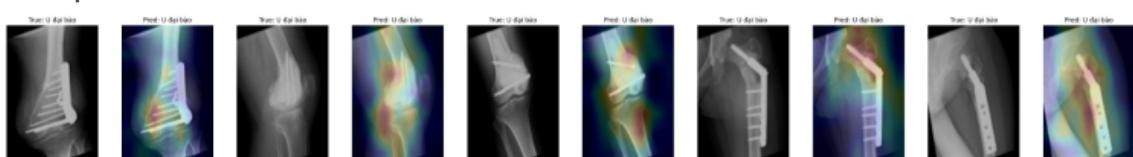
### Di căn



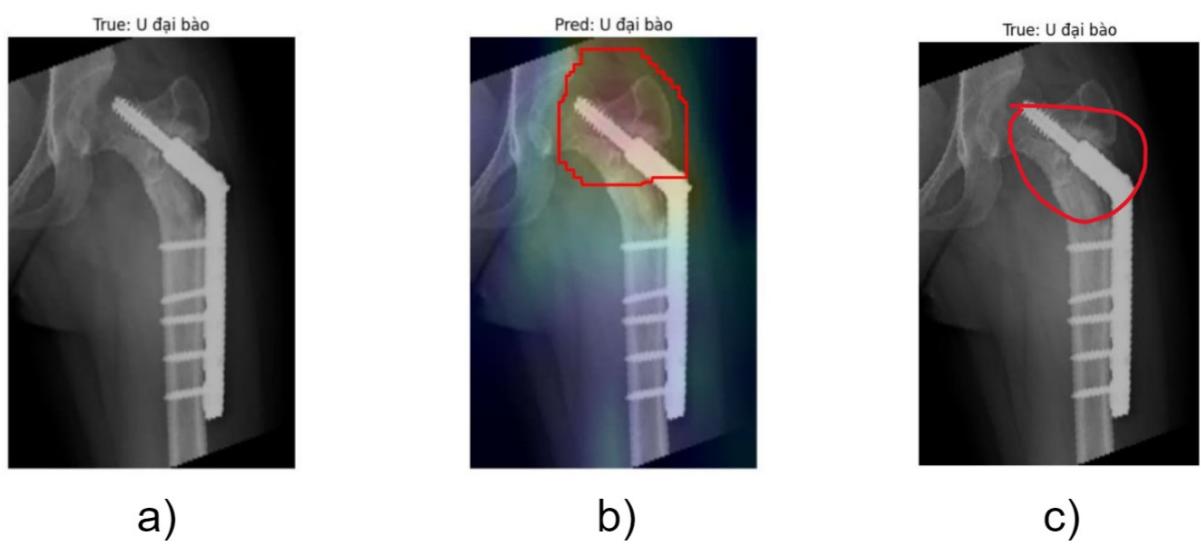
### Sarcoma



### U đại bào



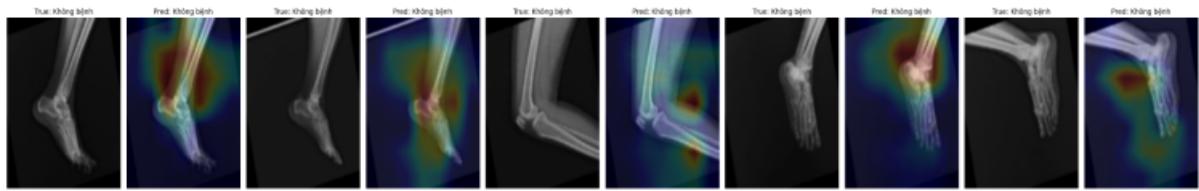
Hình 4.9: Kết quả phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương đùi



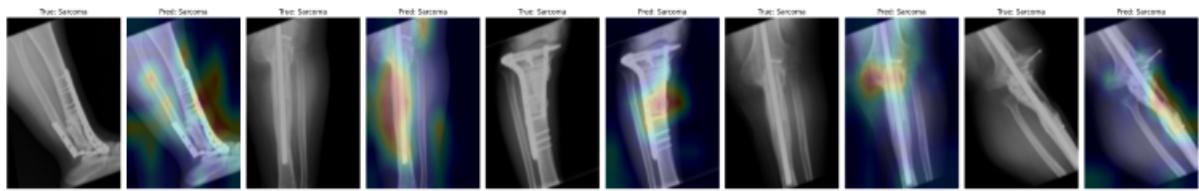
Hình 4.10: Ví dụ một mẫu phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương đùi. a) ảnh gốc; b) ảnh có bản đồ nhiệt; c) ảnh có vùng chứa bệnh do bác sĩ đánh dấu

Phân lớp đúng ở xương chày: 15/24

### Không bệnh



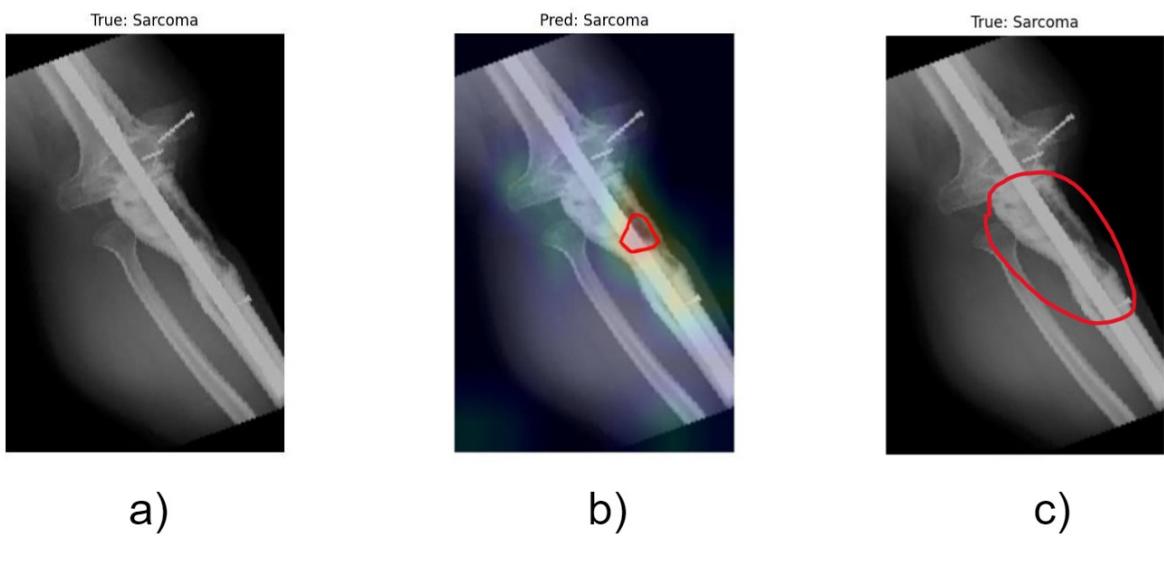
### Sarcoma



### U đại bào



Hình 4.11: Kết quả phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương chày



Hình 4.12: Ví dụ một mẫu phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương chày. a) ảnh gốc; b) ảnh có bản đồ nhiệt; c) ảnh có vùng chứa bệnh do bác sĩ đánh dấu

### 4.3.3 Kết quả thực nghiệm với các ảnh trong tập kiểm thử lần 2

Chọn các ảnh ung thư xương của 3 vị trí xương cánh tay, xương đùi, xương chày không nằm trong tập ảnh huấn luyện và đưa vào mô hình tương ứng với vị trí xương để tiến hành phân lớp.

Dưới đây là kết quả phân lớp đúng khi thực nghiệm kèm theo bản đồ nhiệt giải thích cho kết quả phân lớp:

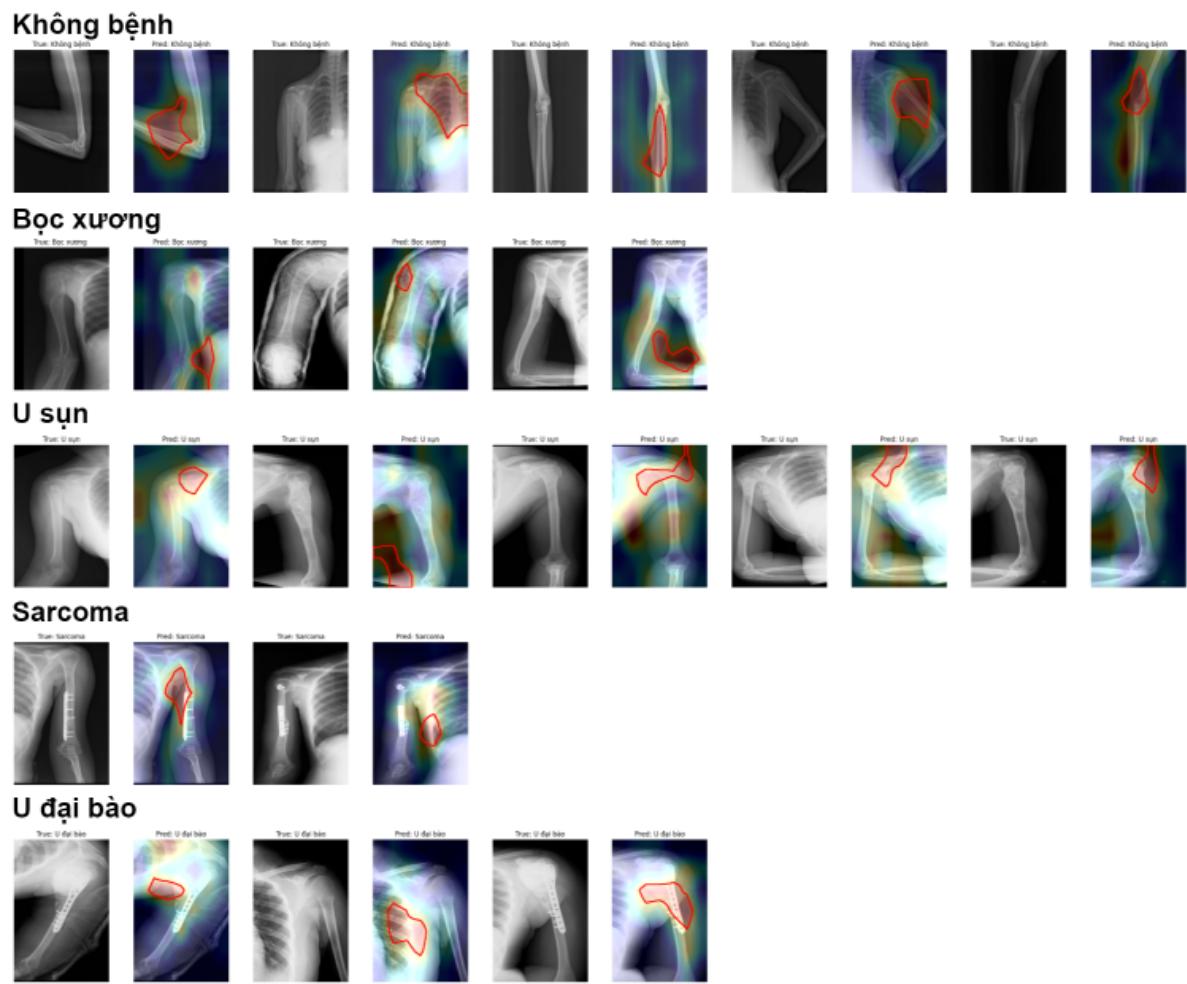
Bảng 4.11: Kết quả kiểm thử tính theo độ chính xác (%) của các mô hình trên 3 tập dữ liệu xương cánh tay, xương đùi và xương chày

Mô hình	Tập xương cánh tay	Tập xương đùi	Tập xương chày
BoneViT (MNTC và BCD)	90.00	85.37	83.33
MedViT (MNTC và BCD)	80.00	75.61	62.50
ResNet18 (MNTC)	85.00	80.49	79.17
DenseNet (MNTC)	85.00	80.49	87.50
BCDTQ (BCD)	70.00	63.41	58.33
Deit (BCD)	75.00	60.98	54.17
EATFormer (BCD)	80.00	73.17	75.00

Bảng 4.12: Kết quả kiểm thử của các mô hình trên 3 tập dữ liệu xương cánh tay, xương đùi và xương chày theo thứ hạng

Mô hình	Tập xương cánh tay	Tập xương đùi	Tập xương chày
BoneViT (MNTC và BCD)	(1)	(1)	(2)
MedViT (MNTC và BCD)	(4)	(4)	(5)
ResNet18 (MNTC)	(2)	(2)	(3)
DenseNet (MNTC)	(2)	(2)	(1)
BCDTQ (BCD)	(7)	(6)	(6)
Deit (BCD)	(6)	(7)	(7)
EATFormer (BCD)	(4)	(5)	(4)

Phân lớp đúng ở xương cánh tay: 18/20 ảnh



Hình 4.13: Kết quả phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương cánh tay

True: U đại bào



True: U đại bào

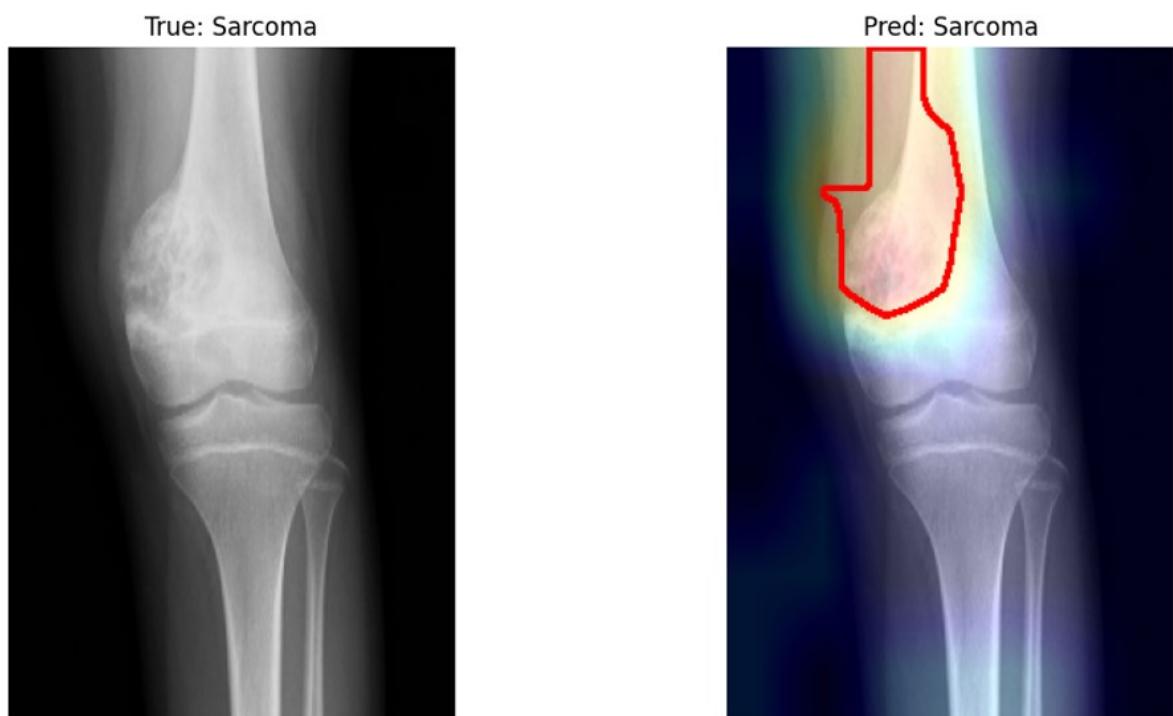


Hình 4.14: Ví dụ một mẫu phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương cánh tay

## Phân lớp đúng ở xương đùi: 35/41

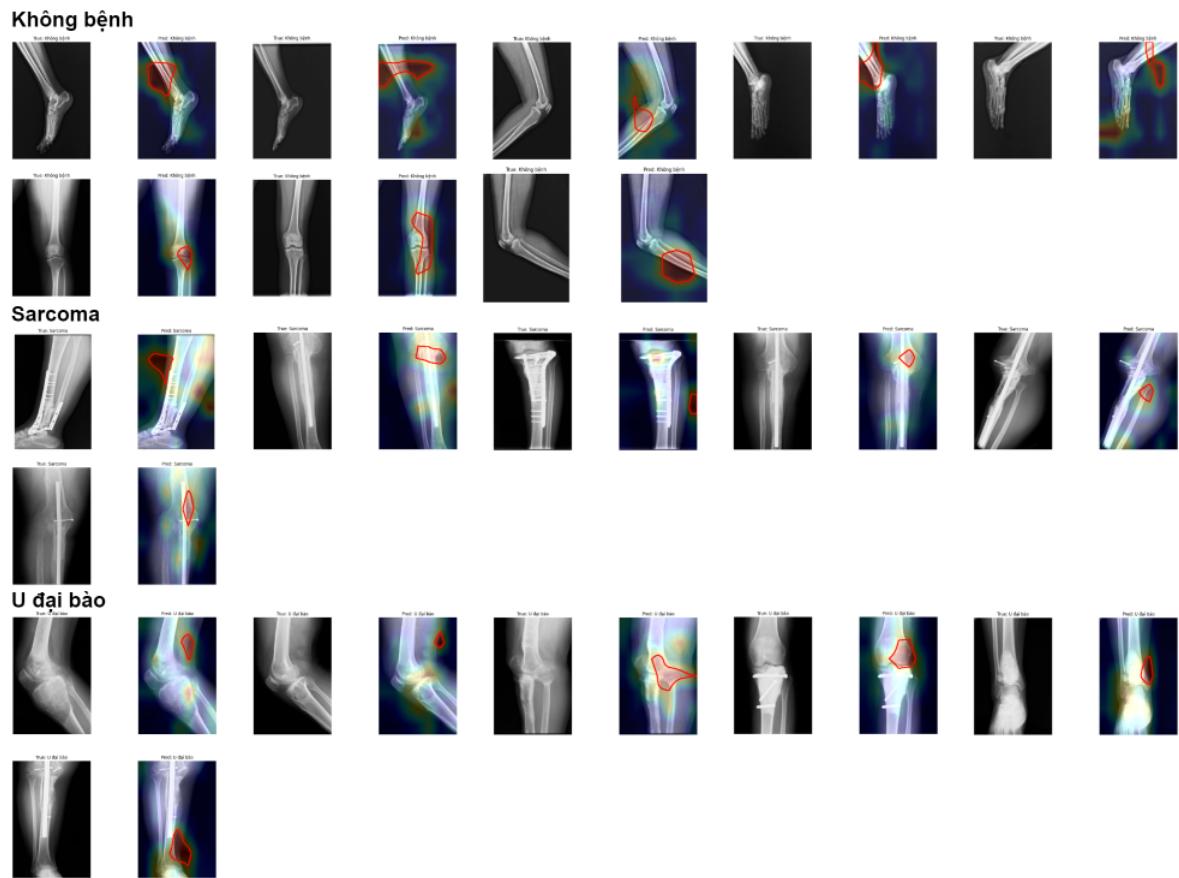


Hình 4.15: Kết quả phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương đùi



Hình 4.16: Ví dụ một mẫu phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương đùi

Phân lớp đúng ở xương chày: 20/24



Hình 4.17: Kết quả phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương chày



Hình 4.18: Ví dụ một mẫu phân lớp đúng khi kiểm thử mô hình phân lớp ung thư xương chày

#### **4.3.4 Nhận xét**

Kết quả thực nghiệm cho thấy mô hình phân lớp ung thư xương trên cả 3 vị trí xương hoạt động khá hiệu quả trên các ảnh không có trong tập huấn luyện. Mô hình đã phân lớp chính xác đa phần các ảnh, nhưng vẫn tồn tại trường hợp phân lớp sai. Đối với các nhãn bệnh có các ảnh thể hiện dấu hiệu bệnh không rõ ràng, dấu hiệu bệnh gần giống với dấu hiệu bệnh của nhãn khác, lượng ảnh huấn luyện chưa nhiều thì kết quả thực nghiệm của mô hình không tốt lắm (ví dụ ảnh thuộc nhãn nhán Sarcoma ở tập xương cánh tay). Các bản đồ nhiệt đi kèm có trường hợp vẫn cho thấy các vùng quan trọng mà mô hình tập trung vào khi đưa ra quyết định phân lớp (thể hiện rõ nhất là trong mô hình phân lớp của xương đùi). Tuy nhiên, bản đồ nhiệt vẫn còn bị nhiễu và chưa tập trung hoàn toàn và vùng quan trọng. Lý do đơn giản nhất để giải thích cho việc này là vì tập dữ liệu quá ít khiến cho mô hình chưa nắm bắt được thật tốt các đặc trưng. Nhưng khi được huấn luyện trên tập dữ liệu lớn hơn, đầy đủ hơn, mô hình chắc chắn sẽ thực hiện tác vụ phân lớp tốt hơn, cho ra các bản đồ nhiệt trực quan hơn.

# Chương 5

## Kết luận và hướng phát triển

Trong chương này, khóa luận tổng kết lại những thành tựu đạt được từ nghiên cứu, đánh giá những đóng góp quan trọng của mô hình BoneViT trong việc phân loại ảnh y khoa ung thư xương, đồng thời đề xuất những hướng phát triển tiềm năng nhằm nâng cao hiệu suất và tính ứng dụng của mô hình trong tương lai.

### 5.1 Kết luận

Ung thư xương đang trở thành một trong những vấn đề sức khỏe nghiêm trọng trong xã hội hiện đại. Việc nhận diện và phát hiện các loại bệnh ung thư xương từ sớm không chỉ giúp nâng cao tỷ lệ sống sót mà còn giảm thiểu những ảnh hưởng xấu của căn bệnh đối với chất lượng cuộc sống của bệnh nhân. Đồng thời, hiện nay trong lĩnh vực học máy, mô hình bộ chuyển đổi nổi lên như một phương pháp tiềm năng trong việc phân lớp ảnh y khoa. Thứ trước nay bị các mô hình tích chập thống trị trong một thời gian dài. Việc phân lớp ảnh y khoa dựa trên mô hình bộ chuyển đổi đang là một lĩnh vực nghiên cứu tiềm năng, .... Trở thành một công cụ để hỗ trợ các bác sĩ trong các công đoạn y khoa có tính lặp lại liên tục. Đề tài “phân lớp ảnh x-quang ung thư xương bằng mô hình bộ chuyển đổi” được lựa chọn như một biện pháp để hỗ trợ các bác sĩ trong việc nhận diện các loại bệnh ung thư trên từng ảnh x-quang ung thư xương. Mục đích chính của đề tài là tìm ra một mô hình bộ chuyển đổi thích hợp để phân

lớp đúng các loại bệnh, đồng thời trực quan hóa các vùng có dấu hiệu bị bệnh trên ảnh bằng Grad-CAM. Qua đó có thể tăng tốc quá trình phát hiện bệnh ung thư xương trên một lượng dữ liệu lớn và giảm thiểu công việc của bác sĩ. Từ đó cải thiện tốc độ làm việc của bệnh viện và đóng góp vào sự phát triển của mô hình bộ chuyển đổi trong lĩnh vực y khoa nói riêng và trên lĩnh vực học máy nói chung.

Khóa luận đã đề xuất mô hình BoneViT, là mô hình kết hợp giữa khối tích chập hiệu quả từ mô hình MedViT và khối EATBlock từ EATFormer. Khối tích chập hiệu quả từ MedViT mang đến khả năng trích xuất đặc trưng hiệu quả nhờ vào thiết kế tối ưu và khả năng giảm thiểu tính toán, trong khi EATBlock từ EATFormer cung cấp các khả năng chú ý nâng cao giúp cải thiện khả năng phân loại. Sự kết hợp này tạo ra một mô hình mạnh mẽ hơn, có khả năng tận dụng cả các kỹ thuật học sâu dựa trên tích chập và các phương pháp chú ý tiên tiến để cải thiện kết quả phân loại và trực quan hóa trong các ứng dụng y khoa. Khóa luận đã áp dụng kiến trúc BoneViT để xây dựng ba mô hình phân lớp ảnh y khoa ung thư xương tại ba vị trí khác nhau: xương cánh tay, xương chày và xương đùi. Mỗi loại có từ 3 đến 5 nhãn bệnh, từ đó có thể so sánh và chỉ ra ưu điểm của mô hình qua từng tập dữ liệu. Phương pháp học của đề tài được chọn là phương pháp học chuyển giao. Mô hình đã được 4 lần huấn luyện trên tập ChestMNIST với lượng dữ liệu 78000 mẫu và 14 lớp. Sau đó được sử dụng để huấn luyện lại lớp cuối cùng trên bộ dữ liệu được cung cấp, từ đó cải thiện hiệu quả kết quả phân lớp của mô hình. Kết quả huấn luyện dựa trên độ chính xác và diện tích dưới đường cong ROC (AUC) cho thấy mô hình không chỉ đạt hiệu suất cao mà còn vượt trội so với một số mô hình tiên tiến hiện có, ví dụ như mô hình Vit. Đối với dữ liệu có nhiều nhãn, lượng dữ liệu đủ lớn, mô hình có ưu thế hơn các mô hình khác. Điều này minh chứng cho khả năng của mô hình trong việc học và phát hiện các đặc trưng quan trọng từ dữ liệu. Nhất là trong bối cảnh tập dữ liệu sử dụng trong nghiên cứu rất ít mẫu và mất cân bằng nghiêm trọng giữa các lớp, từ đó thể hiện được khả năng tổng quát hóa tốt, đặc biệt ở một số nhãn cụ thể, số lượng dự đoán chính xác cao hơn kỳ vọng. Qua đây, khóa luận không chỉ cho thấy tính khả thi của mô hình mà còn mở ra cơ hội

áp dụng vào thực tiễn lâm sàng, nơi mà việc phân lớp chính xác và nhanh chóng có thể giúp hỗ trợ các bác sĩ trong quá trình chẩn đoán và điều trị, từ đây minh chứng cho đóng góp số 1 của nghiên cứu. Mô hình cũng đã phân lớp được các nhãn ứng với 3 vị trí xương gồm xương cánh tay, xương chày, xương đùi. Mặc dù được huấn luyện trên tập dữ liệu rất ít và bị mất cân bằng nặng nhưng mô hình vẫn cho ra kết quả tốt, số lượng phân lớp đúng vẫn cao ở một số nhãn cụ thể, ngoài ra bản đồ nhiệt đi kèm cũng thể hiện được một phần vùng quan trọng khi mô hình đưa ra kết quả phân lớp nhưng vẫn cần phải cải thiện thêm. Điều này đã minh chứng cho đóng góp số 2. Một điểm đáng lưu ý là khả năng trực quan hóa của mô hình thông qua phương pháp GradCAM. Bản đồ nhiệt GradCAM giúp làm nổi bật các vùng mà mô hình cho là quan trọng trong việc đưa ra dự đoán, từ đó cung cấp cái nhìn rõ ràng hơn về vị trí khối u tiềm năng. Tuy nhiên, vẫn còn một số hạn chế trong việc làm nổi bật các vùng chứa khối u. Mặc dù bản đồ nhiệt đã thể hiện được một phần vùng quan trọng, nhưng để đáp ứng yêu cầu cao hơn trong chẩn đoán y khoa, cần phải cải thiện chất lượng của các bản đồ này. Kết quả nghiên cứu này không chỉ góp phần vào việc phát triển các phương pháp phân lớp ảnh y khoa, mà còn có ý nghĩa lớn trong việc ứng dụng thực tiễn. Mô hình BoneViT có tiềm năng trở thành công cụ hỗ trợ đắc lực trong quá trình chẩn đoán ung thư xương, giúp nâng cao độ chính xác và hiệu quả trong chẩn đoán bệnh. Tuy nhiên, vẫn còn nhiều khuyết điểm là chưa phân lớp tốt ở những nhãn có số mẫu huấn luyện ít, trực quan hóa kết quả phân lớp dùng heatmap chưa được tốt. Tuy nhiên, nhiệm vụ chính của mô hình vẫn là hỗ trợ bác sĩ chẩn đoán bệnh, không thể thay thế bác sĩ

## 5.2 Hướng phát triển

Những kết quả đạt được từ nghiên cứu này đã đặt nền móng vững chắc cho các nghiên cứu trong tương lai, với mục tiêu khắc phục những hạn chế hiện tại và mở rộng phạm vi ứng dụng của mô hình BoneViT. Dưới đây là những hướng phát triển tiềm năng:

- Nâng cao khả năng rút trích đặc trưng: Để cải thiện hiệu suất của mô hình, cần xem xét việc nâng cấp các kỹ thuật rút trích đặc trưng. Đồng thời, việc giảm độ phức tạp của mô hình cũng là một hướng đi quan trọng, nhằm tăng cường khả năng tính toán và giảm thời gian huấn luyện mà không làm giảm chất lượng dự đoán.

- Cải thiện chất lượng bản đồ nhiệt GradCAM: Hiện tại, các bản đồ nhiệt của mô hình vẫn chưa thể hiện được đầy đủ các vùng quan trọng cần chú ý trong chẩn đoán. Việc cải thiện chất lượng của các bản đồ nhiệt này có thể thực hiện thông qua việc tinh chỉnh các tham số của mô hình hoặc kết hợp với các phương pháp trực quan hóa khác. Mục tiêu là tạo ra các bản đồ nhiệt trực quan hơn, chính xác hơn, và tập trung vào các vùng chứa khối u, từ đó hỗ trợ tốt hơn cho quá trình chẩn đoán.

- Khắc phục vấn đề mất cân bằng dữ liệu: Mất cân bằng dữ liệu là một thách thức lớn trong lĩnh vực học sâu, đặc biệt là trong lĩnh vực y khoa. Để giải quyết vấn đề này, có thể áp dụng các phương pháp như cân bằng lại dữ liệu bằng cách tăng cường mẫu ở các lớp thiểu số, hoặc sử dụng các trọng số lớp để điều chỉnh quá trình huấn luyện. Những giải pháp này sẽ giúp mô hình chú ý nhiều hơn đến các lớp có ít mẫu, từ đó cải thiện hiệu suất tổng thể.

- Nghiên cứu các phương pháp phân lớp tiên tiến Ngành y học liên tục tiến bộ, và với sự phát triển của các mô hình học sâu, có rất nhiều phương pháp phân lớp mới đang được nghiên cứu. Để đảm bảo mô hình BoneViT luôn ở mức tiên tiến, cần liên tục cập nhật và thử nghiệm các phương pháp mới, như các mô hình học sâu dựa trên bộ chuyển đổi, hoặc các phương pháp kết hợp giữa các mô hình khác nhau. Những nghiên cứu này không chỉ giúp cải thiện hiệu suất mô hình mà còn mở rộng phạm vi ứng dụng trong các bài toán khác nhau.

- Phát triển ứng dụng thực tiễn: Một hướng phát triển quan trọng là chuyển kết quả nghiên cứu thành ứng dụng thực tế. Điều này đòi hỏi sự

hợp tác với các bệnh viện, trung tâm y tế để triển khai mô hình BoneViT trong môi trường lâm sàng. Việc thử nghiệm và điều chỉnh mô hình trong thực tế sẽ giúp phát hiện và khắc phục những hạn chế chưa được lường trước trong quá trình nghiên cứu, đồng thời đảm bảo rằng mô hình có thể đáp ứng được các yêu cầu khắt khe của ngành y tế.

- Tích hợp trí tuệ nhân tạo vào hệ thống hỗ trợ quyết định lâm sàng: Một trong những ứng dụng tiềm năng của mô hình BoneViT là tích hợp vào các hệ thống hỗ trợ quyết định lâm sàng. Các hệ thống này giúp cung cấp thông tin và gợi ý chẩn đoán cho các bác sĩ dựa trên kết quả phân tích từ mô hình, giúp nâng cao chất lượng chẩn đoán và tiết kiệm thời gian cho bác sĩ. Việc tích hợp mô hình vào hệ thống sẽ giúp phát huy tối đa tiềm năng của trí tuệ nhân tạo trong y tế, từ đó cải thiện hiệu quả chăm sóc bệnh nhân.

Những hướng phát triển này không chỉ giúp nâng cao chất lượng của nghiên cứu hiện tại mà còn mở ra nhiều cơ hội mới cho các ứng dụng trong tương lai. Các kết quả nghiên cứu và ứng dụng thành công của mô hình BoneViT sẽ đóng góp tích cực vào sự phát triển của ngành y tế, đặc biệt trong việc chẩn đoán và điều trị bệnh ung thư xương.

# Tài liệu tham khảo

## Tiếng Anh

- [1] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [2] Krizhevsky, A., Sutskever, I., and Hinton, G. E., “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, 2012.
- [3] Simonyan, K. and Zisserman, A., “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [4] Szegedy, C., Liu, W., Jia, Y., et al., “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [5] He, K., Zhang, X., Ren, S., and Sun, J., “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [6] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q., “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708. [Online]. Available: [https://openaccess.thecvf.com/content\\_cvpr\\_2017/papers/Huang\\_Densely\\_Connected\\_Convolutional\\_CVPR\\_2017\\_paper.pdf](https://openaccess.thecvf.com/content_cvpr_2017/papers/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.pdf).

- [7] Tan, M. and Le, Q., “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International conference on machine learning*, PMLR, 2019, pp. 6105–6114.
- [8] Vaswani, A., Shazeer, N., Parmar, N., *et al.*, “Attention is all you need,” *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, pp. 5998–6008, 2017. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fb053c1c4a845aa-Paper.pdf>.
- [9] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.
- [10] Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., and Jégou, H., “Training data-efficient image transformers & distillation through attention,” in *International conference on machine learning*, PMLR, 2021, pp. 10347–10357.
- [11] Liu, Z., Lin, Y., Cao, Y., *et al.*, “Swin transformer: Hierarchical vision transformer using shifted windows,” *Arxiv*, 2021. [Online]. Available: <https://arxiv.org/pdf/2103.14030>.
- [12] Chen, C.-F. R., Fan, Q., and Panda, R., “Crossvit: Cross-attention multi-scale vision transformer for image classification,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 357–366.
- [13] Zhang, J., Li, X., Wang, Y., *et al.*, “Eatformer: Improving vision transformer inspired by evolutionary algorithm,” *International Journal of Computer Vision*, pp. 1–28, 2024.
- [14] Srinivas, A., Lin, T.-Y., Parmar, N., Shlens, J., Abbeel, P., and Vaswani, A., “Bottleneck transformers for visual recognition,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 16519–16529. [Online]. Available: [https://openaccess.thecvf.com/content/CVPR2021/papers/Srinivas\\_](https://openaccess.thecvf.com/content/CVPR2021/papers/Srinivas_)

Bottleneck\_Transformers\_for\_Visual\_Recognition\_CVPR\_2021\_paper.pdf.

- [15] Haiping, W., Xiao, B., Codella, N., *et al.*, “Cvt: Introducing convolutions to vision transformers,” 2021. [Online]. Available: [https://openaccess.thecvf.com/content/ICCV2021/papers/Wu\\_CvT\\_Introducing\\_Convolutions\\_to\\_Vision\\_Transformers\\_ICCV\\_2021\\_paper.pdf](https://openaccess.thecvf.com/content/ICCV2021/papers/Wu_CvT_Introducing_Convolutions_to_Vision_Transformers_ICCV_2021_paper.pdf).
- [16] Graham, B., El-Nouby, A., Touvron, H., *et al.*, “Levit: A vision transformer in convnet’s clothing for faster inference,” in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 12 259–12 269. [Online]. Available: [https://openaccess.thecvf.com/content/ICCV2021/papers/Graham\\_LeViT\\_A\\_Vision\\_Transformer\\_in\\_ConvNets\\_Clothing\\_for\\_Faster\\_Inference\\_ICCV\\_2021\\_paper.pdf](https://openaccess.thecvf.com/content/ICCV2021/papers/Graham_LeViT_A_Vision_Transformer_in_ConvNets_Clothing_for_Faster_Inference_ICCV_2021_paper.pdf).
- [17] Mehta, S. and Rastegari, M., “Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer,” *arXiv preprint arXiv:2110.02178*, 2021. [Online]. Available: <https://arxiv.org/pdf/2110.02178.pdf>.
- [18] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C., “Mobilenetv2: Inverted residuals and linear bottlenecks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520. [Online]. Available: [https://openaccess.thecvf.com/content\\_cvpr\\_2018/papers/Sandler\\_MobileNetV2\\_Inverted\\_Residuals\\_CVPR\\_2018\\_paper.pdf](https://openaccess.thecvf.com/content_cvpr_2018/papers/Sandler_MobileNetV2_Inverted_Residuals_CVPR_2018_paper.pdf).
- [19] Guo, J., Han, K., Wu, H., *et al.*, “Cmt: Convolutional neural networks meet vision transformers,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 12 175–12 185. [Online]. Available: [https://openaccess.thecvf.com/content/CVPR2022/papers/Guo\\_CMT\\_Convolutional\\_Neural\\_Networks\\_Meet\\_Vision\\_Transformers\\_CVPR\\_2022\\_paper.pdf](https://openaccess.thecvf.com/content/CVPR2022/papers/Guo_CMT_Convolutional_Neural_Networks_Meet_Vision_Transformers_CVPR_2022_paper.pdf).
- [20] Chen, Y., Dai, X., Chen, D., *et al.*, “Mobile-former: Bridging mobilenet and transformer,” in *Proceedings of the IEEE/CVF confer-*

- ence on computer vision and pattern recognition, 2022, pp. 5270–5279. [Online]. Available: [https://openaccess.thecvf.com/content/CVPR2022/papers/Chen\\_Mobile-Former\\_Bridging\\_MobileNet\\_and\\_Transformer\\_CVPR\\_2022\\_paper.pdf](https://openaccess.thecvf.com/content/CVPR2022/papers/Chen_Mobile-Former_Bridging_MobileNet_and_Transformer_CVPR_2022_paper.pdf).
- [21] Ribeiro, M. T., Singh, S., and Guestrin, C., “"why should i trust you?": Explaining the predictions of any classifier,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2016, pp. 1135–1144. [Online]. Available: <https://arxiv.org/abs/1602.04938>.
- [22] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A., “Learning deep features for discriminative localization,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921–2929.
- [23] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D., “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626. [Online]. Available: [https://openaccess.thecvf.com/content\\_ICCV\\_2017/papers/Selvaraju\\_Grad-CAM\\_Visual\\_Explanations\\_ICCV\\_2017\\_paper.pdf](https://openaccess.thecvf.com/content_ICCV_2017/papers/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.pdf).
- [24] Chattopadhyay, A., Sarkar, A., Howlader, P., and Balasubramanian, V. N., “Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks,” in *2018 IEEE winter conference on applications of computer vision (WACV)*, IEEE, 2018, pp. 839–847.
- [25] Abnar, S. and Zuidema, W., “Quantifying attention flow in transformers,” *arXiv preprint arXiv:2005.00928*, 2020.
- [26] Chefer, H., Gur, S., and Wolf, L., “Transformer interpretability beyond attention visualization,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 782–791.

- [27] Shamshad, F., Khan, S., Zamir, S. W., *et al.*, “Transformers in medical imaging: A survey,” *arXiv preprint arXiv:2201.09873*, 2022. [Online]. Available: <https://arxiv.org/pdf/2201.09873>.
- [28] Azad, R., Kazerouni, A., Heidari, M., *et al.*, “Advances in medical image analysis with vision transformers: A comprehensive review,” 2023. [Online]. Available: <https://arxiv.org/pdf/2301.03505>.
- [29] Shao, Z., Bian, H., Chen, Y., *et al.*, “Transmil: Transformer based correlated multiple instance learning for whole slide image classification,” 2021. [Online]. Available: [https://proceedings.neurips.cc/paper\\_files/paper/2021/file/10c272d06794d3e5785d5e7c5356e9ff-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2021/file/10c272d06794d3e5785d5e7c5356e9ff-Paper.pdf).
- [30] Bhattacharya, M., Jain, S., and Prassanna, P., “Radiotransformer: A cascaded global-focal transformer for visual attention-guided disease classification,” 2022. [Online]. Available: <https://arxiv.org/pdf/2202.11781>.
- [31] Mehta, S., Lu, X., Weaver†, D., Elmore, J. G., Hajishirzi, H., and Shapiro, L., “Hatnet: An end-to-end holistic attention network for diagnosis of breast biopsy images,” 2020. [Online]. Available: <https://arxiv.org/pdf/2007.13007>.
- [32] Dai, Y. and Gao, Y., “Transmed: Transformers advance multi-modal medical image classification,” 2021. [Online]. Available: <https://arxiv.org/pdf/2103.05940>.
- [33] Manzari, O. N., Ahmadabadi, H., Kashiani, H., Shokouhi, S. B., and Ayatollahi, A., “Medvit: A robust vision transformer for generalized medical image classification,” *Computers in Biology and Medicine*, vol. 157, p. 106 791, 2023.
- [34] Dũng, L. C., *Bướu xương*. Nhà xuất bản Y Học, 2003.
- [35] Shisu, Y., Mingwin, S., Wanwag, Y., Chenso, Z., and Huing, S., “Improved eatformer: A vision transformer for medical image classification,” *arXiv preprint arXiv:2403.13167*, 2024. [Online]. Available: <https://arxiv.org/abs/2403.13167>.

- [36] Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S., “A convnet for the 2020s,” *Arxiv*, 2022. [Online]. Available: <https://arxiv.org/pdf/2201.03545>.
- [37] Bortsova, G., González-Gonzalo, C., Wetstein, S. C., *et al.*, “Adversarial attack vulnerability of medical image analysis systems: Unexplored factors,” *National Library of Medicine*, 2021. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/34246850/>.
- [38] Bortsova, G., González-Gonzalo, C., Wetstein, S. C., *et al.*, “Adversarial attack vulnerability of medical image analysis systems: Unexplored factors,” *Medical Image Analysis*, vol. 73, p. 102141, 2021.
- [39] Park, N. and Kim, S., “How do vision transformers work?” *Arxiv*, 2022. [Online]. Available: <https://arxiv.org/pdf/2202.06709>.
- [40] Yang, J., Shi, R., Wei, D., *et al.*, “Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification,” *Scientific Data*, vol. 10, no. 1, p. 41, 2023. [Online]. Available: <https://medmnist.com/>.

## Phụ lục A

# Tên thuật ngữ và viết tắt tiếng Việt, tiếng Anh

Bảng A.1: Bảng các thuật ngữ và viết tắt trong luận văn

STT	Tên đầy đủ		Viết tắt	
	Tiếng Việt	Tiếng Anh	Tiếng Việt	Tiếng Anh
01	Bản đồ đặc trưng	Feature Map		
02	Bản đồ nhiệt	Heat Map		
03	Bộ biến đổi ảnh sang chuỗi	Image to Sequence	BBDASC	Img2Seq
04	Bộ biến đổi chuỗi sang ảnh	Sequence to Image	BBDCSA	Seq2Img
05	Bộ chuyển đổi	Transformer	BCĐ	
06	Bộ chuyển đổi trực quan	Vision Transformer	BCĐTQ	ViT
07	Bộ mã hóa chuyển đổi	Transformer Encoder		

Bảng A.2: Bảng các thuật ngữ và viết tắt trong luận văn (tt)

	Tên đầy đủ		Viết tắt	
STT	Tiếng Việt	Tiếng Anh	Tiếng Việt	Tiếng Anh
08	Bộ xử lý đồ họa	Graphics Processing Unit		GPU
09	Chiếu	Projection		
10	Chiếu tuyến tính của các mảnh phẳng	Linear Projection of Flattened Patches		
11	Chú ý theo tích vô hướng có tỉ lệ	Scaled Dot Production Attention		
12	Chú ý tích chập nhiều đầu	Multi-head convolutional attention	CYTCND	MHCA
13	Chuẩn hóa lô	Batch Normalization		
14	Chuẩn hóa lớp	Layer Normalization		
15	Chụp cắt lớp vi tính	Computed Tomography Scan		CT Scan
16	Chụp cộng hưởng từ	Magnetic Resonance Imaging		MRI
17	Chuyển tiến	Feedforward		
18	Có giám sát	Supervised		
19	Có giám sát yếu	Weakly Supervised		

Bảng A.3: Bảng các thuật ngữ và viết tắt trong luận văn (tt)

	Tên đầy đủ		Viết tắt	
STT	Tiếng Việt	Tiếng Anh	Tiếng Việt	Tiếng Anh
20	Dịch chuyển	Shifting		
21	Diện tích dưới đường cong ROC	Area under the ROC curve	DTDĐCR	AUC
22	Đỏ - Xanh lá - Xanh dương	Red - Green - Blue		RGB
23	Độ chính xác	Accuracy	ĐCX	ACC
24	Độ dốc	Gradient		
25	Độ dời	off-sets		
26	Độ lệch	Bias		
27	Đơn vị tuyến tính được hiệu chỉnh	Rectified Linear Unit	ĐVTTĐHC	ReLU
28	Đơn vị tuyến tính lỗi Gauss	Gaussian Error Linear Unit		GELU
29	Dư thừa	Residual		
30	Giá trị	Value		
31	Gộp	Concat		
32	Hàm mất mát	Loss Function		
33	Hàm mất mát Hỗn loạn chéo	Cross-Entropy Loss Function		
34	Hệ số điều chỉnh	Modulation scalars		

Bảng A.4: Bảng các thuật ngữ và viết tắt trong luận văn (tt)

	Tên đầy đủ		Viết tắt	
STT	Tiếng Việt	Tiếng Anh	Tiếng Việt	Tiếng Anh
35	Học chuyển tiếp	Transfer Learning		
36	Kết nối đầy đủ	Fully connected	KNDD	FC
37	Khóa	Key		
38	Khối chuyển đổi	Transformer block		
39	Khối chuyển đổi chú ý hiệu quả Efficient Attention	Transformer Block	KCDCYHQ	EATBlock
40	Khối chuyển đổi cục bộ	Local Transformer Block	KCDCB	LTB
41	Khối nhúng mảnh ảnh	Patch Embedding Block		
42	Khối tích chập	Convolutional Block		
43	Khối tích chập hiệu quả	Efficient Convolutional Block	KTCHQ	ECB
44	Không chồng chéo	Non-overlapping		
45	Kích thước lô	Batchsize		

Bảng A.5: Bảng các thuật ngữ và viết tắt trong luận văn (tt)

	Tên đầy đủ		Viết tắt	
STT	Tiếng Việt	Tiếng Anh	Tiếng Việt	Tiếng Anh
46	Lan truyền ngược	Backpropagation		
47	Lớp bỏ qua ngẫu nhiên	Dropout	LBQNN	
48	Lớp nhúng mảnh ảnh	Patch Embedding Layer		
49	Mã thẻ	Token		
50	Mạng chuyển tiến cục bộ	Locally Feed-Forward Network	MCTCB	LFFN
51	Mạng nơ-ron đa tầng	Multi-layer Perceptron	MNDT	MLP
52	Mạng nơ-ron hồi quy	Recurrent Neural Network		RNN
53	Mạng nơ-ron tích chập	Convolutional Neural Network	MNTC	CNN
54	Mảnh ảnh	Patch		
55	Mô hình kết hợp	Hybrid model	MHKH	
56	Nhân ma trận	Matrix Multiplication		MatMul
57	Số lần lặp	Epoch		

Bảng A.6: Bảng các thuật ngữ và viết tắt trong luận văn (tt)

	Tên đầy đủ		Viết tắt	
STT	Tiếng Việt	Tiếng Anh	Tiếng Việt	Tiếng Anh
58	Song tuyến tính	Bilinear		
59	Tấn công học máy đối nghịch	Adversarial Attack		
60	Tập hợp vùng đa tỷ lệ	Multi-Scale Region Aggregation	THVĐTL	MSRA
61	Tích chập gốc	Stem convolutional		
62	Tốc độ học	Learning rate		
63	Tổng hợp trung bình	Average Pooling		
64	Trí tuệ nhân tạo	Artificial Intelligence	TTNT	AI
65	Trí tuệ nhân tạo khả diễnn	eXplainable Artificial Intelligence	TNTTKD	XAI
66	Trộn các phép toán có trọng số	Weighted Operation Mixing	TCPTCTS	WOM
67	Trọng số	Weight		
68	Trung bình toàn cục	Global Average Pooling	TBTC	GAP
69	Truy vấn	Querry		

Bảng A.7: Bảng các thuật ngữ và viết tắt trong luận văn (tt)

	Tên đầy đủ		Viết tắt	
STT	Tiếng Việt	Tiếng Anh	Tiếng Việt	Tiếng Anh
70	Tự chú ý	Self Attention	TCY	
71	Tự chú ý nhiều đầu	Multi-head Self Attention	TCYNĐ	MHSA
72	Tự chú ý nhiều đầu biến dạng đã điều chỉnh	Modulated Deformable MSA	TCYNĐBD-ĐDC	MD-MSA
73	Tương tác toàn cục và cục bộ	Global and Local Interaction	TTTCVCB	GLI
74	Tuyến tính	Linear		
75	Xanh dương - Xanh lá - Đỏ	Blue - Green - Red		BGR
76	X-quang	X ray		
77	Xử lý ngôn ngữ tự nhiên	Natural Language Processing	XLNNTN	NLP

## **Phụ lục B**

# **Kiểm chứng từ bác sĩ có chuyên môn**

## KIỂM CHỨNG KẾT QUẢ PHÂN LỚP ẢNH UNG THƯ XƯƠNG

Nhóm khóa luận kính nhờ Bác sĩ kiểm tra kết quả phân lớp. Nội dung của các bức ảnh như sau: Ảnh đầu tiên là **anh gốc**, ảnh thứ hai là **anh có chúa heatmap** với vùng màu nóng là vùng mà AI dựa vào để đưa ra kết quả phân lớp, ảnh thứ ba là **anh gốc** để Bác sĩ đánh dấu (khoanh tròn) vào những nơi nghi ngờ có bệnh lý để đưa ra chẩn đoán.

Nếu sử dụng Word để xem file .docx, Bác sĩ bấm vào phần Enable editing trên thanh thông báo màu vàng như hình dưới, sau đó dùng chức năng Draw trên thanh công cụ (sử dụng màu đỏ) của Word để đánh dấu vào ảnh. Ở phần đúng/sai chỉ cần nhấp chuột vào ô vuông.

 PROTECTED VIEW Be careful—files from the Internet can contain viruses. Unless you need to edit, it's safer to stay in Protected View. [Enable Editing](#)

Nếu sử dụng Foxit Reader để xem file .pdf, Bác sĩ vào phần comment trên thanh công cụ, chọn pencil để đánh dấu vào ảnh cũng như đánh dấu vào ô đúng sai.

Nếu kết quả phân lớp đúng Bác sĩ tích vào ô đúng, nếu sai Bác sĩ tích vào ô sai và nếu có ghi chú gì thì Bác sĩ vui lòng ghi thêm để nhóm cải thiện tính chính xác của mô hình.

Sau khi đã hoàn thành kiểm chứng, Bác sĩ vui lòng lưu file này lại và tải lên thư mục kết quả trên link google drive:

<https://drive.google.com/drive/folders/1U12zOoA1gzF5tAcdUd3v-eRgVov7cVon?usp=sharing>

Cuối cùng kính nhờ Bác sĩ viết giấy xác nhận đã kiểm chứng kết quả phân lớp của mô hình.

Nhóm khóa luận xin chân thành cảm ơn Bác sĩ đã dành thời gian quý báu hỗ trợ nhóm kiểm chứng kết quả mà mô hình của nhóm đã phân lớp.

## 1. Xương cánh tay:

### a. Không bệnh:



Đúng

Sai

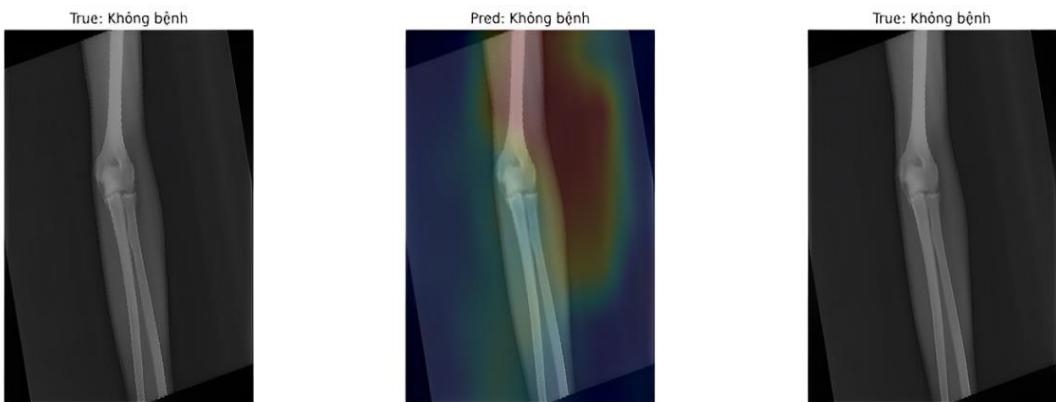
Ghi chú (nếu có):



Đúng

Sai

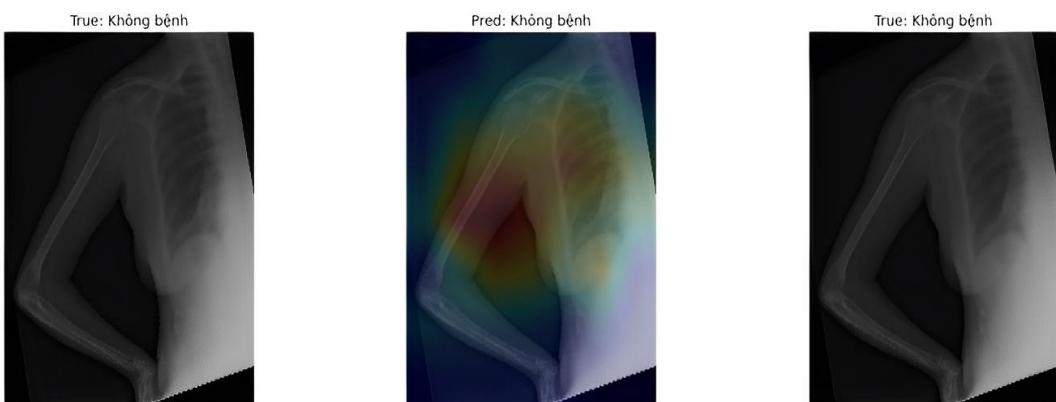
Ghi chú (nếu có):



Đúng

Sai

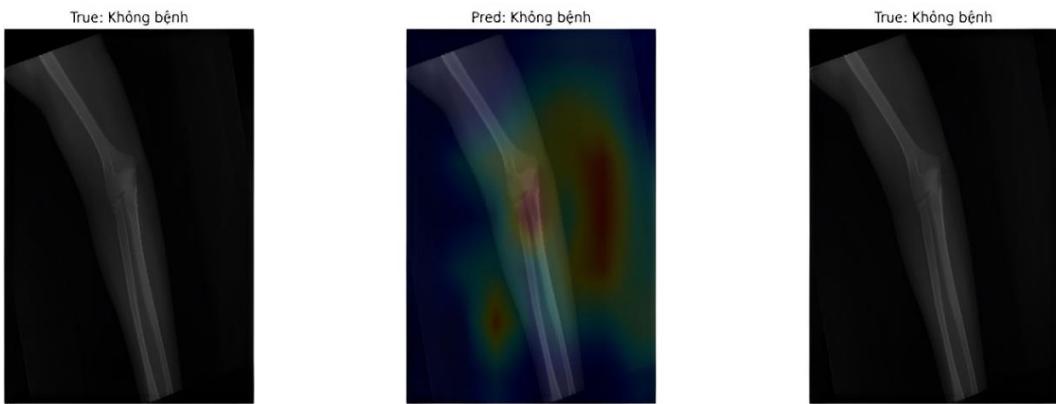
Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):

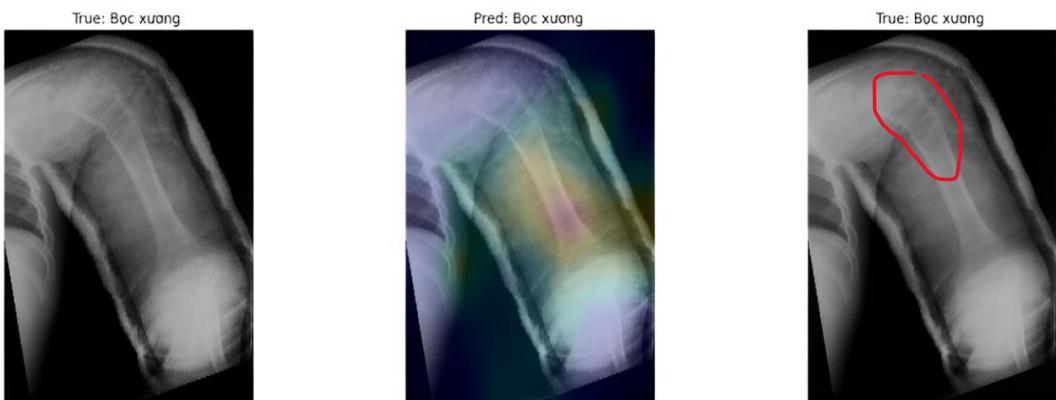


Đúng

Sai

Ghi chú (nếu có):

**b. Bọc xương (lành tính):**



Đúng

Sai

Ghi chú (nếu có):

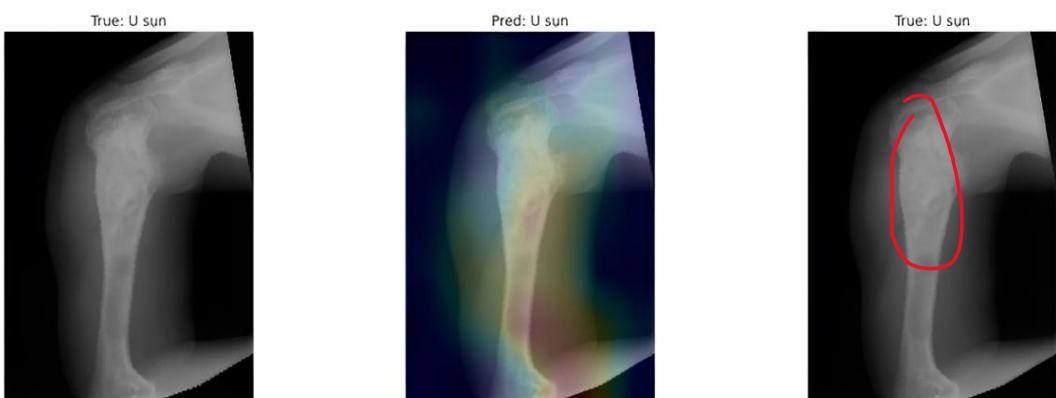
**c. U sụn (lành tính):**



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

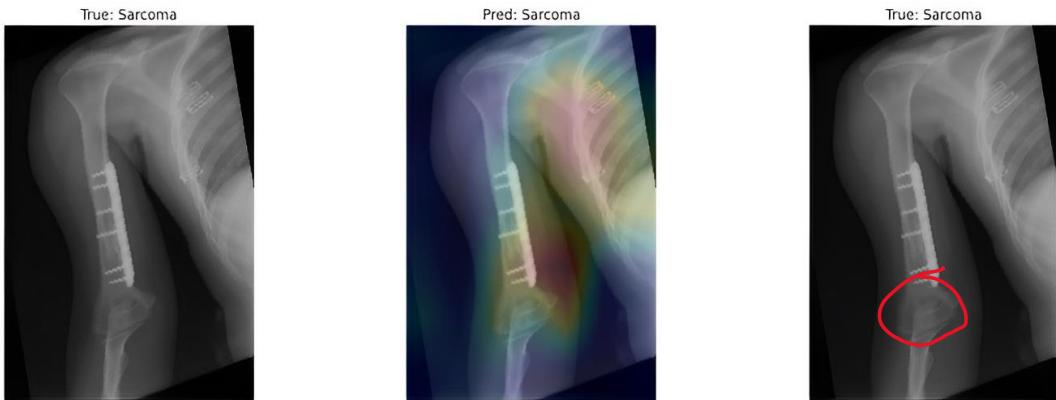
Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):

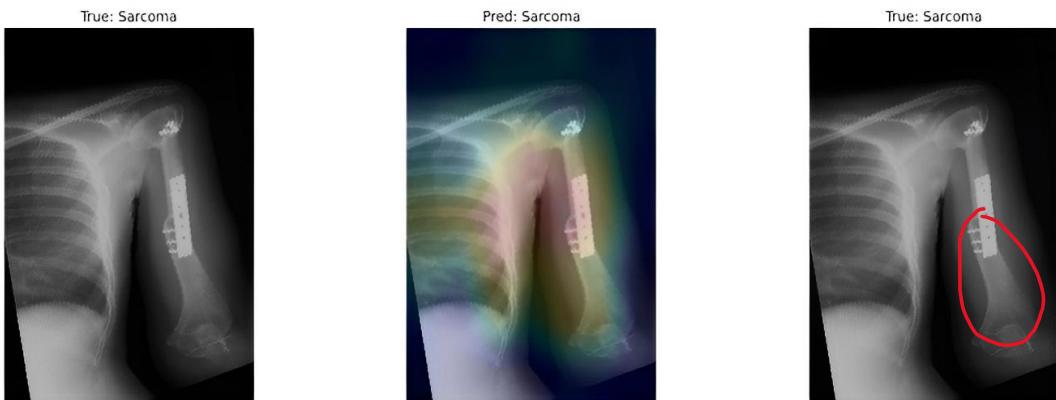
**d. Sarcoma (ác tính):**



Đúng

Sai

Ghi chú (nếu có):

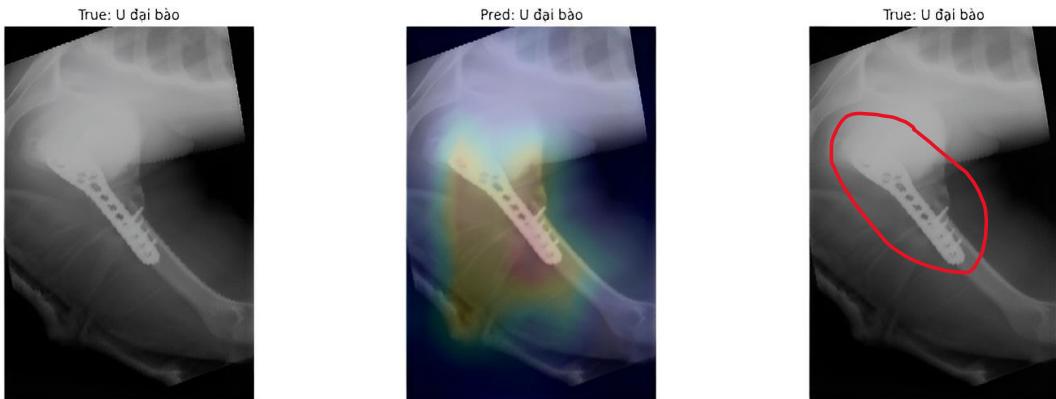


Đúng

Sai

Ghi chú (nếu có):

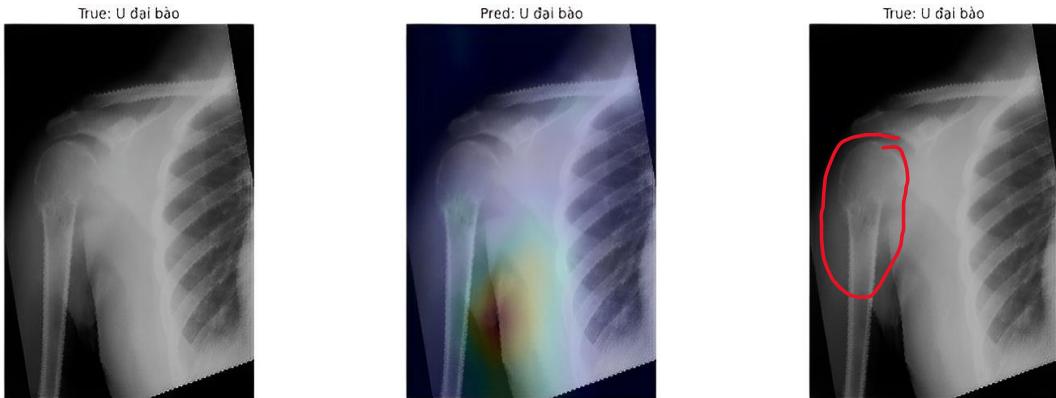
#### e. U đai bào (ác tính):



Đúng

Sai

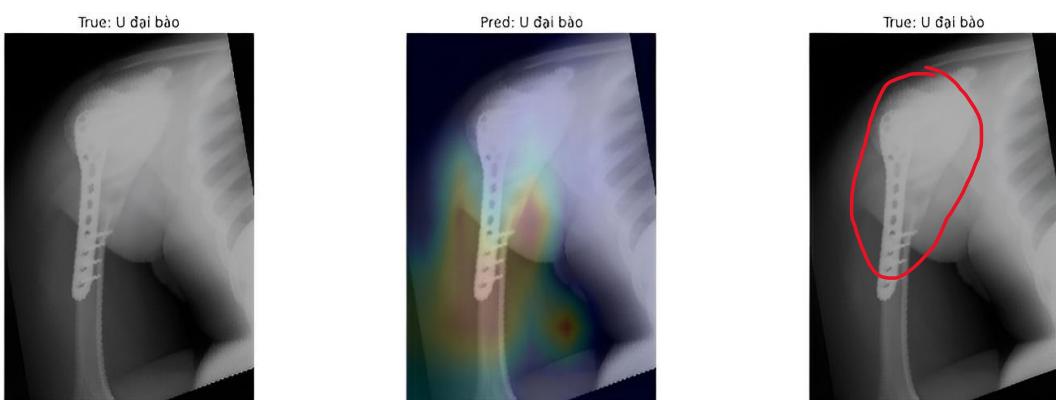
Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):

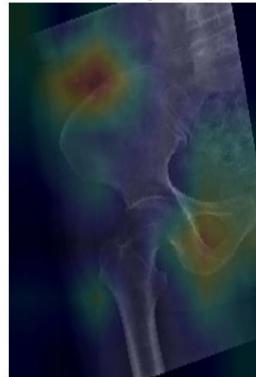
## 2. Xương đùi:

### a. Không bệnh:

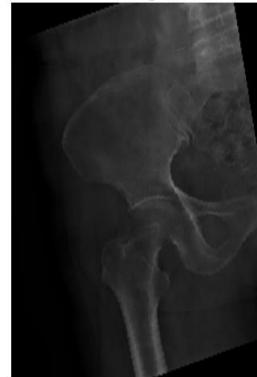
True: Không bệnh



Pred: Không bệnh



True: Không bệnh



Đúng

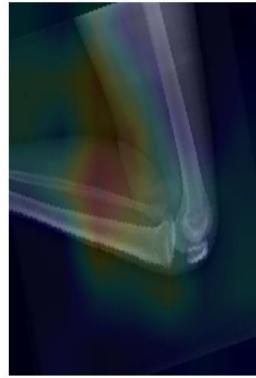
Sai

Ghi chú (nếu có):

True: Không bệnh



Pred: Không bệnh



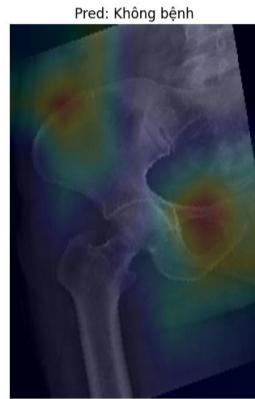
True: Không bệnh



Đúng

Sai

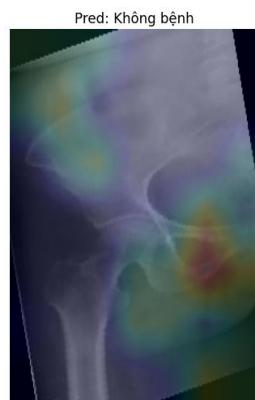
Ghi chú (nếu có):



Đúng

Sai

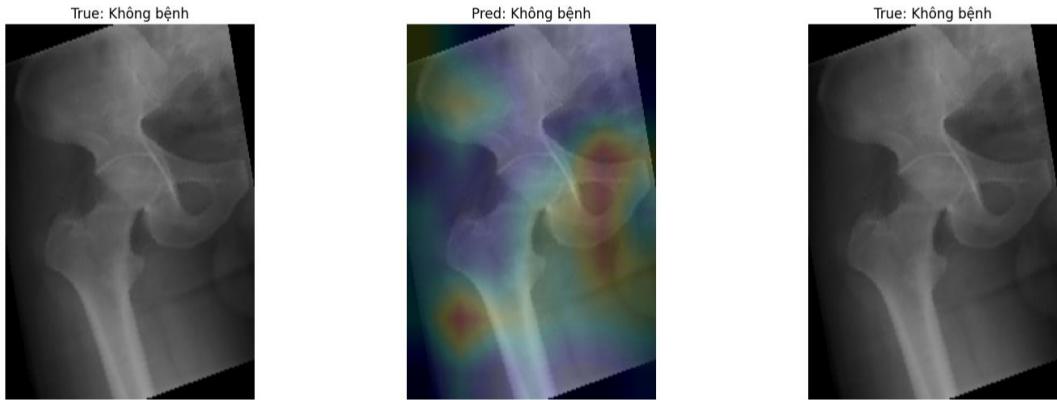
Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):

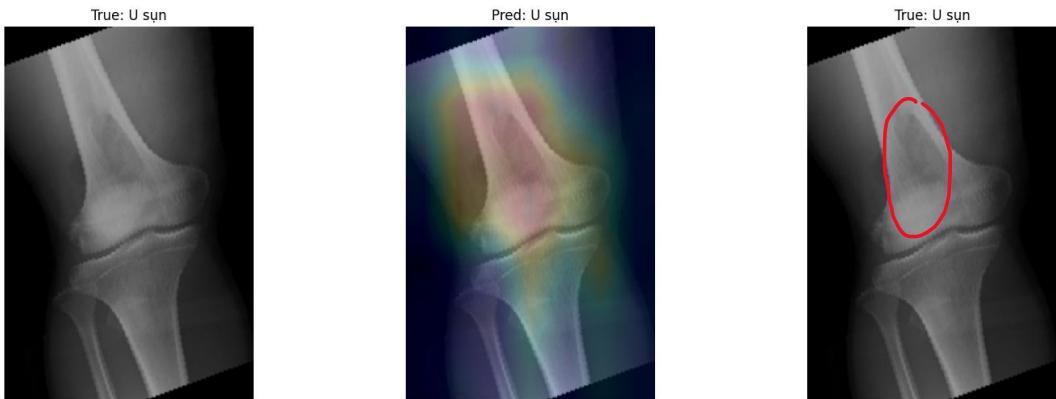
**b. U sụn (lành tính):**



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):

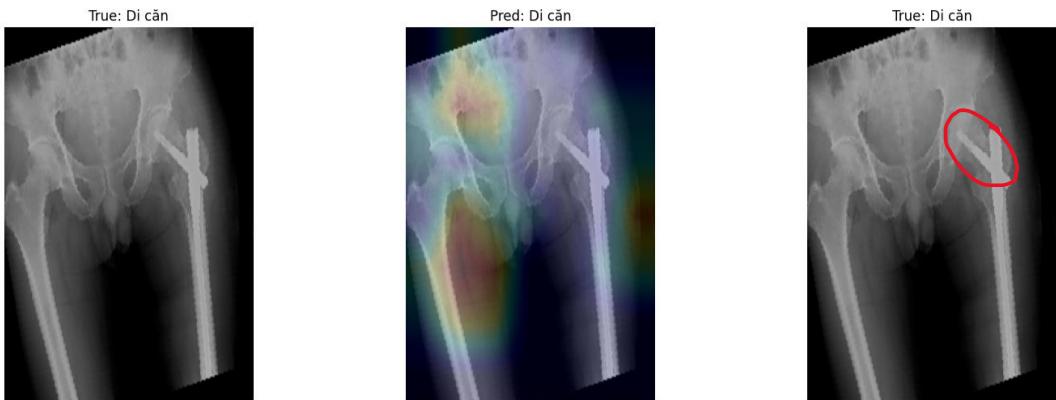
**c. Di căn (ác tính):**



Đúng

Sai

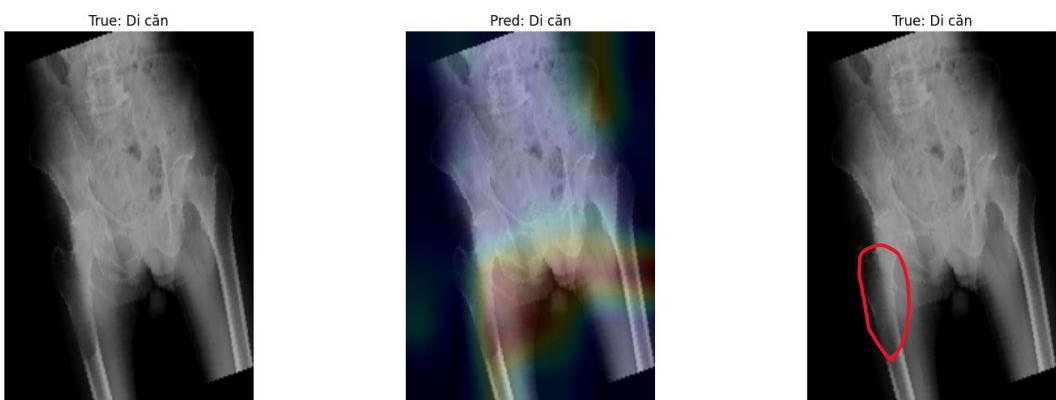
Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):

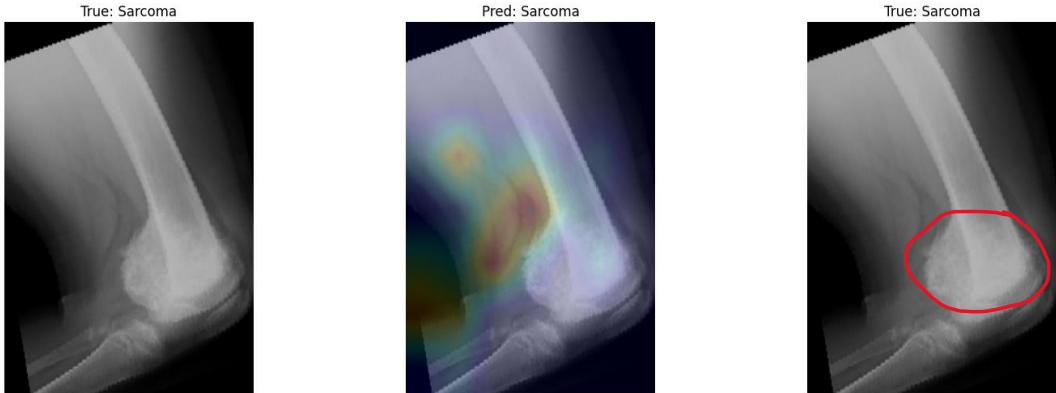


Đúng

Sai

Ghi chú (nếu có):

**d. Sarcoma (ác tính):**



Đúng

Sai

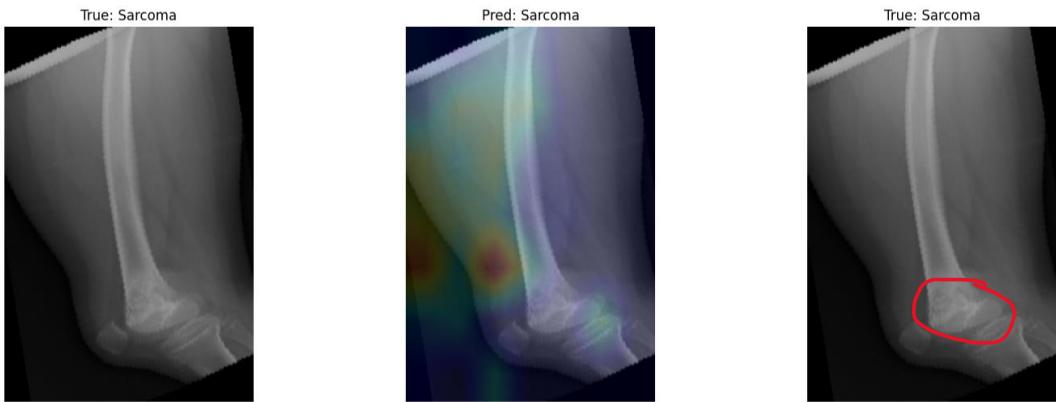
Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

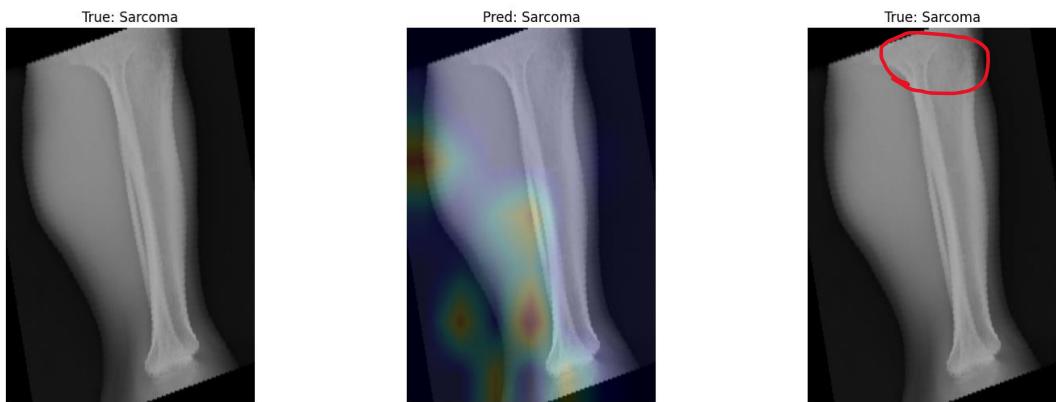
Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):

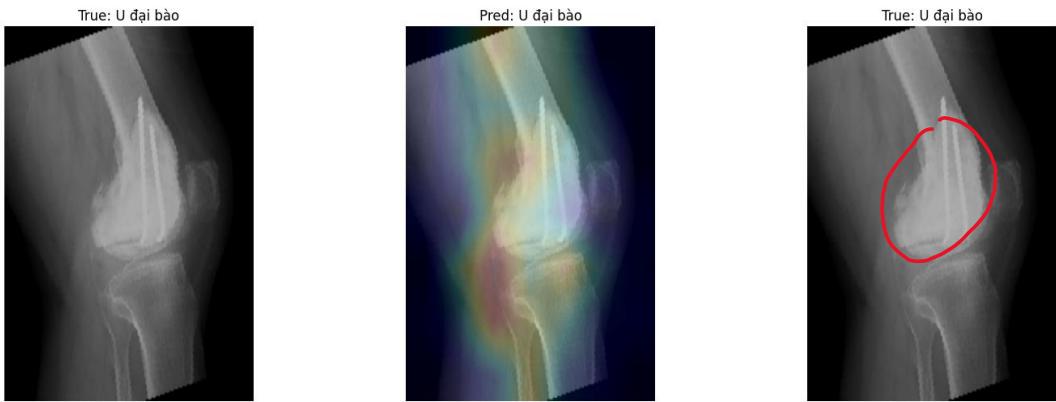
#### e. U đại bào (ác tính):



Đúng

Sai

Ghi chú (nếu có):



Đúng ✕

Sai □

Ghi chú (nếu có):



Đúng ✕

Sai □

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):

**3. Xương chày:**  
**a. Không bệnh:**



Đúng

Sai

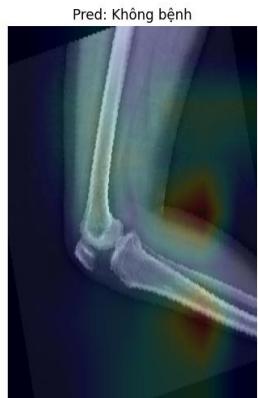
Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

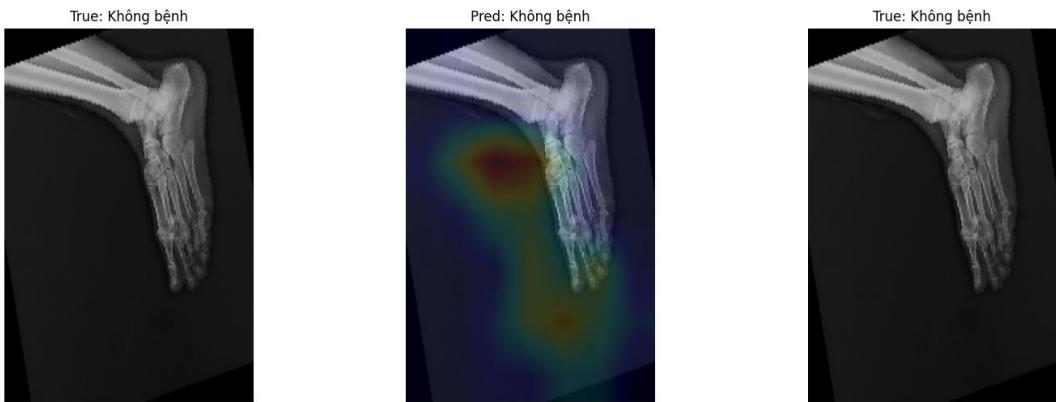
Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):

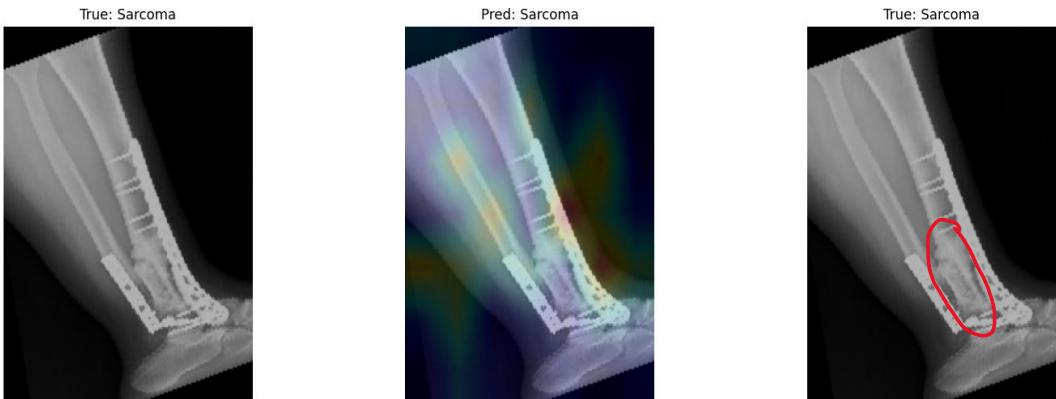


Đúng

Sai

Ghi chú (nếu có):

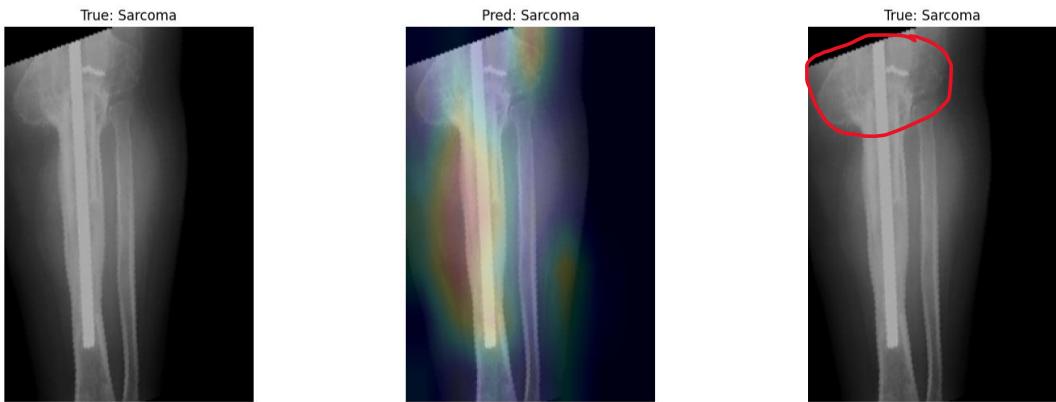
### b. Sarcoma (ác tính)



Đúng

Sai

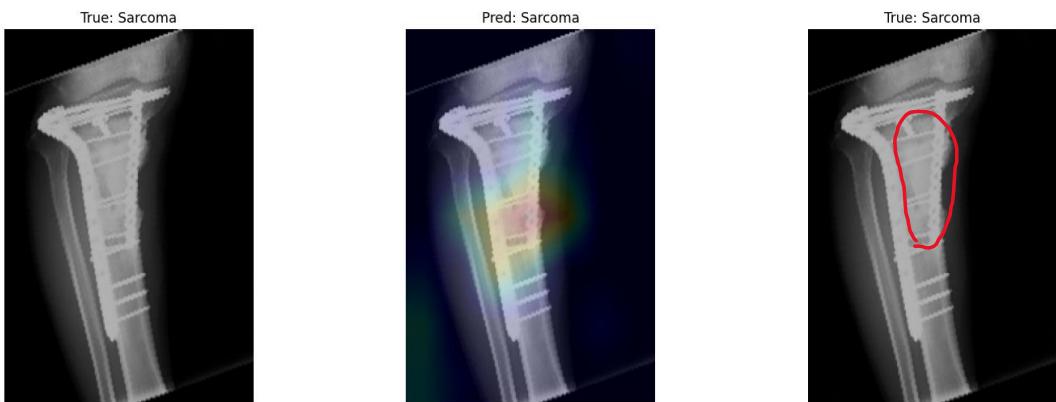
Ghi chú (nếu có):



Đúng

Sai

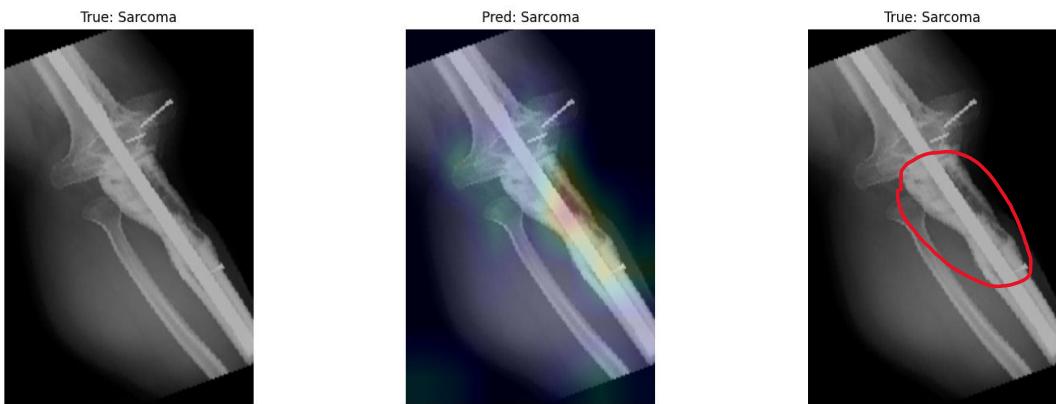
Ghi chú (nếu có):



Đúng

Sai

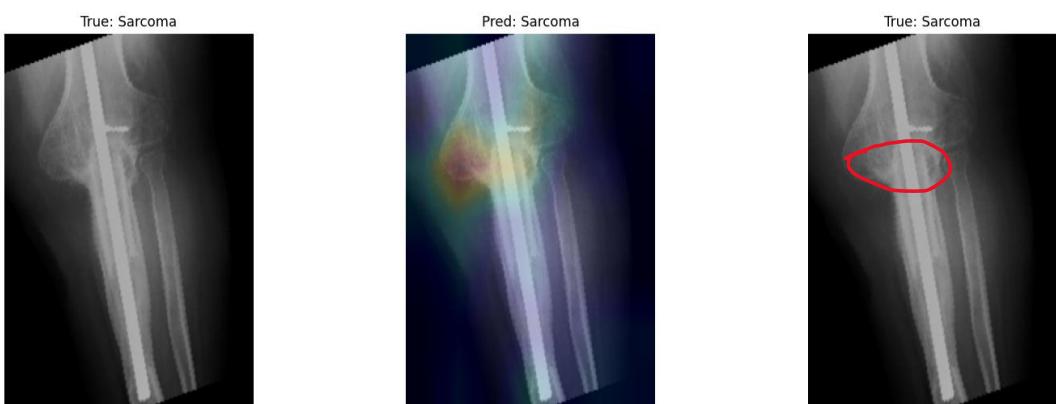
Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):

**c. U đai bào (ác tính):**



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Đúng

Sai

Ghi chú (nếu có):



Thứ ..... ngày ..... tháng ..... năm .....

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM  
TỘC LẬP TỰ DO HÀNH PHÚC.

ĐƠN KIỂM CHỨNG KẾT QUA PHÂN LỚP  
ẢNH X QUANG UNG THỦ XƯỞNG.

- Tổng số mẫu cần kiểm chứng: 52
- Tổng số mẫu phân lớp đúng: 52
- Tổng số mẫu phân lớp sai: 00.

tpHCM ngày 24 - tháng 2 - năm 2024

Minh

Bs Lý Thành Trang (Khoa CHTA)  
BV. Nguyễn Tri Phương

Yêu thương - Tôn trọng - Trung thực - Trách nhiệm

SỞ Y TẾ THÀNH PHỐ HỒ CHÍ MINH  
BỆNH VIỆN NGUYỄN TRÌ PHƯƠNG



LÝ THANH TRANG  
BÁC SĨ CHUYÊN KHOA I  
PHÓ TRƯỞNG KHOA

11100056857