

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG



NGUYỄN THỊ TÂN TIẾN

**NGHIÊN CỨU PHÁT TRIỂN MỘT SỐ MÔ HÌNH HỌC SÂU
TRONG HỖ TRỢ CHẨN ĐOÁN BỆNH BỤI PHỔI SILIC NGHỀ
NGHIỆP DỰA TRÊN ẢNH X-QUANG NGỰC**

LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

NGÀNH: KHOA HỌC MÁY TÍNH
MÃ SỐ: 9 48 01 01

THÁI NGUYÊN, NĂM 2025

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG



NGUYỄN THỊ TÂN TIẾN

**NGHIÊN CỨU PHÁT TRIỂN MỘT SỐ MÔ HÌNH HỌC SÂU
TRONG HỖ TRỢ CHẨN ĐOÁN BỆNH BỤI PHỔI SILIC NGHỀ
NGHIỆP DỰA TRÊN ẢNH X-QUANG NGỰC**

LUẬN ÁN TIẾN SĨ KHOA HỌC MÁY TÍNH

NGÀNH: KHOA HỌC MÁY TÍNH
MÃ SỐ: 9 48 01 01

TẬP THỂ HƯỚNG DẪN KHOA HỌC:

- 1. PGS.TS. PHẠM VĂN CƯỜNG**
- 2. TS. NGUYỄN VĂN TẢO**

THÁI NGUYÊN, NĂM 2025

LỜI CAM ĐOAN

Tôi xin cam đoan các kết quả trình bày trong luận án là công trình nghiên cứu của tôi dưới sự định hướng, hướng dẫn, giám sát của tập thể giáo viên hướng dẫn. Các kết quả được viết chung với các tác giả khác đều được sự đồng ý của đồng tác giả trước khi đưa vào luận án. Các kết quả trong luận án là trung thực và chưa từng được công bố trong các công trình khác.

Nghiên cứu sinh

Nguyễn Thị Tân Tiến

LỜI CẢM ƠN

Trong suốt thời gian học tập và nghiên cứu, tôi đã nhận được sự định hướng, hỗ trợ và đồng hành quý báu từ nhiều tập thể và cá nhân. Với tất cả lòng biết ơn và trân trọng, tôi xin gửi lời tri ân sâu sắc tới những người thầy, người cô, các đồng nghiệp, bạn bè và người thân đã luôn sát cánh, sẻ chia và tiếp thêm nghị lực cho tôi trên suốt hành trình này.

Trước hết, tôi xin được bày tỏ lòng biết ơn trân trọng nhất tới hai thầy PGS.TS Phạm Văn Cường và TS. Nguyễn Văn Tảo. Những chỉ dẫn tận tình, sự định hướng sáng suốt cùng những nhận xét quý báu của các thầy không chỉ giúp tôi hoàn thiện luận án mà còn mở ra cho tôi những chân trời tri thức mới, tạo nền tảng vững chắc cho con đường nghiên cứu khoa học sau này.

Tôi xin trân trọng cảm ơn Ban Giám hiệu, Khoa Công nghệ thông tin, Phòng Đào tạo, Trường Đại học Công nghệ thông tin và Truyền thông, Đại học Thái Nguyên đã tạo dựng một môi trường học tập và nghiên cứu thuận lợi để tôi có thể hoàn thành tốt luận án.

Tôi xin gửi lời cảm ơn chân thành đến các bác sĩ tại Hội Y học Lao động Thái Nguyên cùng tập thể các nhà khoa học thực hiện đề tài “Nghiên cứu đặc điểm dịch tễ học phân tử, yếu tố nguy cơ và ứng dụng kỹ thuật tiên tiến trong chẩn đoán sớm bệnh bụi phổi Silic tại Việt Nam” do GS.TS. Lê Thị Hương làm Chủ nhiệm. Sự tận tâm, nhiệt huyết của các bác sĩ và những kết quả nghiên cứu giá trị từ đề tài đã trở thành nguồn hỗ trợ thiết thực và động lực quan trọng cho tôi trong suốt quá trình thực hiện luận án.

Một lời cảm ơn chân thành và sâu sắc tôi xin gửi đến thầy cô khoa Trí tuệ nhân tạo, Học viện Công nghệ Bưu chính Viễn thông, vì những chia sẻ chuyên môn và tinh thần học thuật quý báu, đã góp phần định hình phương hướng nghiên cứu và nuôi dưỡng đam mê khoa học trong tôi.

Tôi xin chân thành cảm ơn Trường Đại học Y-Dược, Đại học Thái Nguyên nơi tôi công tác, đã tạo điều kiện thuận lợi về thời gian và cơ sở vật chất, giúp tôi chuyên tâm học tập, nghiên cứu và hoàn thành luận án.

Cuối cùng, tôi xin gửi lời cảm ơn sâu sắc tới gia đình, bạn bè và đồng

nghiệp, những người đã luôn bên cạnh, sẻ chia, động viên và tiếp thêm nghị lực. Chính sự thấu hiểu và ủng hộ bền bỉ ấy là điểm tựa tinh thần quý giá, giúp tôi kiên định, vững bước trên con đường nghiên cứu khoa học.

Xin trân trọng cảm ơn.

Mục lục

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
MỤC LỤC	iv
DANH MỤC CÁC CHỮ VIẾT TẮT	viii
DANH MỤC CÁC BẢNG	xii
DANH MỤC HÌNH VẼ	xiv
PHẦN MỞ ĐẦU	1
1. Tính cấp thiết của nội dung nghiên cứu	1
2. Mục tiêu nghiên cứu	3
3. Đối tượng và phạm vi nghiên cứu	4
4. Phương pháp nghiên cứu	5
5. Bố cục của Luận án	6
6. Đóng góp khoa học của Luận án	6
7. Ý nghĩa khoa học và ý nghĩa thực tiễn	7
7.1. Ý nghĩa khoa học	7
7.2. Ý nghĩa thực tiễn	7
1 CHƯƠNG 1: TỔNG QUAN	9
1.1 Tổng quan về bệnh bụi phổi silic nghề nghiệp	9
1.1.1 Sơ lược về bệnh nghề nghiệp	9
1.1.2 Sơ lược về bệnh bụi phổi silic nghề nghiệp	9
1.1.3 Triệu chứng, chẩn đoán bệnh bụi phổi silic nghề nghiệp	10
1.1.4 Phân loại tiêu chuẩn bệnh bụi phổi silic nghề nghiệp . .	12
1.2 Khảo sát nghiên cứu liên quan	15

1.2.1	Các phương pháp chẩn đoán lâm sàng bệnh bụi phổi silic nghề nghiệp	15
1.2.2	Các phương pháp học máy truyền thống	16
1.2.3	Các phương pháp học sâu trong chẩn đoán ảnh X-quang y tế	18
1.3	Cơ sở lý thuyết	26
1.3.1	Một số mô hình học sâu tiêu biểu	26
1.3.2	Học sâu trên đồ thị	31
1.4	Các chỉ số đánh giá hiệu năng	36
1.5	Những hạn chế của các nghiên cứu đã có và mục tiêu nghiên cứu	39
1.6	Những hướng nghiên cứu chính của luận án	40
1.7	Kết luận chương 1	41
2	CHƯƠNG 2: XÂY DỰNG TẬP DỮ LIỆU	42
2.1	Khảo sát về các tập dữ liệu	42
2.1.1	Tập dữ liệu quốc tế	42
2.1.2	Tập dữ liệu trong nước	43
2.2	Xây dựng tập dữ liệu	44
2.2.1	Nguồn gốc dữ liệu	44
2.2.2	Thách thức trong việc xây dựng và xử lý dữ liệu	45
2.2.3	Tiêu chuẩn lựa chọn và tiêu chuẩn loại trừ dữ liệu	46
2.2.4	Tiền xử lý dữ liệu	50
2.2.5	Cấu trúc và đặc điểm tập dữ liệu	53
2.3	Kết luận chương 2	57
3	CHƯƠNG 3: ĐỀ XUẤT MÔ HÌNH HỌC SÂU ÍT MẪU HỖ TRỢ PHÁT HIỆN VÀ KHOANH VÙNG BỆNH BỤI PHỔI SILIC NGHỀ NGHIỆP	58

3.1	Đặt vấn đề và động lực nghiên cứu	58
3.2	Phát biểu bài toán trong ngữ cảnh dữ liệu y tế hạn chế	59
3.3	Giải quyết bài toán với Few-Shot Learning	62
3.3.1	Mô hình hóa bài toán	62
3.3.2	Kiến trúc mô hình	64
3.4	Thiết kế thực nghiệm và cấu hình huấn luyện	68
3.4.1	Cấu hình thực nghiệm	68
3.4.2	Cấu hình huấn luyện	69
3.5	Kết quả thực nghiệm	69
3.5.1	So sánh kết quả với các phương pháp nền tảng	69
3.5.2	So sánh kết quả dựa trên số lượng lớp	71
3.5.3	So sánh hiệu suất giữa các mô hình theo số lượng mẫu	73
3.5.4	Trực quan hóa kết quả	74
3.6	Kết luận chương 3	75

4 CHƯƠNG 4: ĐỀ XUẤT MÔ HÌNH HỌC SÂU HỖ TRỢ PHÂN LOẠI BỆNH BỤI PHỔI SILIC NGHỀ NGHIỆP **77**

4.1	Ứng dụng các mô hình học sâu trong hỗ trợ phân loại bệnh bụi phổi silic nghề nghiệp	77
4.1.1	Phát biểu bài toán	77
4.1.2	Kiến trúc chung của mô hình học sâu áp dụng trong chẩn đoán ảnh X-quang	78
4.1.3	Kết quả thực nghiệm	79
4.2	Đề xuất mô hình học sâu trên đồ thị	82
4.2.1	Phát biểu bài toán	82
4.2.2	Đề xuất kiến trúc Graph Transformer Post-hoc	82
4.2.3	Tổng quan quy trình xử lý và khả năng tích hợp mô hình Graph Transformer hậu xử lý	87

4.3	Lựa chọn hàm mất mát	90
4.3.1	Hàm mất mát tiêu chuẩn	91
4.3.2	Hàm mất mát cân bằng	91
4.4	Kỹ thuật kết hợp mô hình	92
4.5	Thiết lập thực nghiệm và đánh giá	94
4.5.1	Thiết lập thực nghiệm	94
4.5.2	Kết quả thực nghiệm	94
4.6	Kết luận chương 4	106
KẾT LUẬN		108
1.	Kết quả nghiên cứu và đóng góp chính của Luận án	108
2.	Đề xuất các vấn đề cần tiếp tục nghiên cứu	108
3.	Hạn chế và hướng khắc phục	109
DANH MỤC CÔNG TRÌNH KHOA HỌC		110
TÀI LIỆU THAM KHẢO		112

DANH MỤC CÁC CHỮ VIẾT TẮT

STT	Tên viết tắt	Tên tiếng Anh	Tên tiếng Việt
1	AI	Artificial Intelligence	Trí tuệ nhân tạo
2	ANN	Artificial Neural Network	Mạng nơ-ron nhân tạo
3	AUC	Area Under the Curve	Diện tích dưới đường cong
4	AUC-ROC	Area Under the Curve Receiver Operating Characteristic	Diện tích dưới đường cong ROC
5	BalCE	Balanced Cross-Entropy	Hàm mất mát entropy chéo có cân bằng
6	BPSi	Occupational Silicosis	Bụi phổi silic nghề nghiệp
7	CAD	Computer-Aided Diagnosis	Hệ thống chẩn đoán hỗ trợ bằng máy tính
8	CART	Classification and Regression Tree	Cây phân loại và hồi quy
9	CE	Cross-Entropy	Hàm mất mát entropy chéo
10	CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
11	CXR	Chest X-ray	X-quang ngực
12	CDSS	Clinical decision support systems	Hệ thống hỗ trợ quyết định lâm sàng
13	DICOM	Digital Imaging and Communications in Medicine	Tiêu chuẩn quốc tế để truyền tải và lưu trữ ảnh y tế
14	DenseNet	Densely Connected Convolutional Network	Mạng tích chập kết nối dày đặc
15	DL	Deep Learning	Học sâu
16	DNN	Deep Neural Network	Mạng nơ-ron sâu

STT	Tên viết tắt	Tên tiếng Anh	Tên tiếng Việt
17	DT	Decision Tree	Cây quyết định
18	ER	Exact Ratio	Tỷ lệ chính xác tuyệt đối
19	F1-score	F1-score	Điểm trung bình điều hòa của <i>precision</i> và <i>recall</i>
20	FC	Fully Connected	Lớp kết nối đầy đủ
21	FN	False Negative	Âm tính giả
22	FP	False Positive	Dương tính giả
23	GAT	Graph Attention Network	Mạng nơ-ron đồ thị cơ chế chú ý
24	GCN	Graph Convolutional Network	Mạng tích chập đồ thị
25	GNN	Graph Neural Network	Mạng nơ-ron đồ thị
26	Grad-CAM	Gradient-weighted Class Activation Mapping	Trực quan hóa vùng nhận diện của mô hình
27	GTN	Graph Transformer Network	Mạng biến đổi đồ thị
28	GTP	Graph Transformer Post-hoc	
29	IG-FSL	Integrative Few-Shot Learning for Classification and Segmentation	Học ít mẫu tích hợp cho phân loại và phân đoạn
30	LR	Logistic Regression	Hồi quy logistic
31	ML	Machine Learning	Học máy
32	MLP	Multilayer Perceptron	Phương pháp học máy có giám sát nhiều lớp thuộc mạng nơ-ron nhân tạo
33	MobileNetV3-Small	Mobile Neural Network Version 3 (Small variant)	Mạng nơ-ron di động phiên bản 3 (biến thể nhỏ)

STT	Tên viết tắt	Tên tiếng Anh	Tên tiếng Việt
34	mIoU	Mean Intersection over Union	Trung bình giao cắt trên hợp (chỉ số đánh giá phân đoạn)
35	NN	Neural Network	Mạng nơ ron nhân tạo
36	KNN	K-Nearest Neighbors	K-hàng xóm gần nhất
37	PACS	Picture Archiving and Communication System	Hệ thống lưu trữ và truyền tải ảnh
38	RNN	Recurrent Neural Networks	Mạng nơ-ron hồi quy
39	SVM	Support Vector Machine	Máy vector hỗ trợ
40	SVBCX	Silicosis Viral Bacteria Chest X-ray	Bộ dữ liệu X-quang ngực gồm 4 lớp bệnh
41	ROC	Receiver Operating Characteristics	Đường cong ROC
42	RF	Random Forest	Rừng ngẫu nhiên
43	ReLU	Rectified Linear Unit	Hàm kích hoạt đơn vị tuyến tính có chỉnh sửa
44	RNN	Recurrent Neural Network	Mạng nơ-ron hồi quy
45	Swin Transformer	Shifted Windows Transformer	Mô hình Transformer với cửa sổ trượt
46	TN	True Negative	Âm tính thật
47	TP	True Positive	Dương tính thật

DANH MỤC CÁC BẢNG

Bảng 1.1	Thang phân loại mức độ tập trung của các nốt mờ nhỏ theo tiêu chuẩn ILO 2011	13
Bảng 1.2	Tổng hợp các nghiên cứu tiêu biểu về bài toán phát hiện bệnh phổi sử dụng Mạng nơ-ron tích chập (CNN) trên ảnh X-quang ngược.	19
Bảng 1.3	Tổng hợp các nghiên cứu tiêu biểu sử dụng Học sâu trên đồ thị trên ảnh X-quang.	22
Bảng 1.4	Tổng hợp các nghiên cứu tiêu biểu sử dụng Học sâu ít mẫu (FSL) trên ảnh X-quang phổi.	25
Bảng 1.5	Vai trò và kích thước đầu ra của các thành phần trong mô hình Swin Transformer	30
Bảng 2.1	Tiêu chí lựa chọn và vai trò các nhóm bệnh trong phân lớp dữ liệu	48
Bảng 2.2	Số lượng ảnh X-quang giữa các lớp bệnh trong ba tập dữ liệu nghiên cứu	53
Bảng 3.1	Phân loại lớp kỹ thuật và lớp bệnh lý trong tập dữ liệu	60
Bảng 3.2	Cấu hình huấn luyện và siêu tham số	70
Bảng 3.3	Kết quả hiệu suất của các phương pháp khác nhau trong hai thiết lập thử nghiệm	70
Bảng 3.4	So sánh các phương pháp khác nhau dựa trên số liệu phân loại (0/1 ER) và phân đoạn (mIoU)	73
Bảng 3.5	Ma trận nhầm lẫn kết quả phân loại	75
Bảng 4.1	Cấu trúc phân bố dữ liệu DNNBPSi	79
Bảng 4.2	Cấu hình huấn luyện của các mô hình học sâu	80
Bảng 4.3	Độ chính xác và hiệu suất giữa các mô hình	81
Bảng 4.4	Kết quả ma trận nhầm lẫn	81

Bảng 4.5	Cấu hình chi tiết của kiến trúc GTP trong thực nghiệm . .	89
Bảng 4.6	Kết quả tác động của hai hàm mất mát lên hiệu suất mô hình	95
Bảng 4.7	Bảng tóm tắt kết quả của các mô hình với hai chỉ số đánh giá Macro-F1 và AUC-ROC	97
Bảng 4.8	Kết quả độ chính xác phân loại (%) của các mô hình qua 5 lần kiểm chứng chéo (5-fold cross-validation)	98
Bảng 4.9	Độ nhạy và độ đặc hiệu theo từng lớp bệnh đối với các mô hình	102
Bảng 4.10	Ma trận nhầm lẫn từ kết quả chẩn đoán của chuyên gia . .	105

DANH MỤC HÌNH VẼ

Hình 1.1	Hình ảnh X-quang bệnh nhân bệnh bụi phổi silic	10
Hình 1.2	Minh họa hệ thống phân loại tổn thương phổi theo tiêu chuẩn ILO [23]	14
Hình 1.3	Kiến trúc DenseNet201 [110]	28
Hình 1.4	Kiến trúc Attention: Scaled Dot-Product và Multi-Head Attention [118]	34
Hình 1.5	Minh họa cơ chế của một lớp Graph Transformer, trong đó Attention được điều chỉnh theo cấu trúc đồ thị. [118]	35
Hình 2.1	Hình minh họa ảnh X-quang ngực được lưu ở định dạng DICOM	50
Hình 2.2	Hình minh họa tiền xử lý dữ liệu	52
Hình 2.3	Hình ảnh X-quang ngực của người không mắc bệnh bụi phổi silic nghề nghiệp	55
Hình 2.4	Hình ảnh X-quang ngực của người mắc bệnh bụi phổi silic nghề nghiệp	55
Hình 2.5	Thống kê ca bệnh bụi phổi silic nghề nghiệp theo giới tính	56
Hình 2.6	Thống kê ca bệnh bụi phổi silic nghề nghiệp theo số năm kinh nghiệm	57
Hình 3.1	Luồng tổng quan bài toán	63
Hình 3.2	Kiến trúc mạng và luồng xử lý dữ liệu của mô hình	65
Hình 3.3	Kết quả phân đoạn dựa trên số lượng lớp của FSL	71
Hình 3.4	Kết quả phân loại dựa trên số lượng lớp của FSL	72
Hình 3.5	Bản đồ Grad-CAM minh họa vùng chú ý của IG-FSL	74
Hình 3.6	Kết quả phân đoạn trong trường hợp xét hai nhãn, trong đó vùng xanh lá cây thể hiện phần xương, vùng đỏ thể hiện vùng phổi.	74

Hình 4.1	Kiến trúc khung mô hình học sâu áp dụng trong phân loại bệnh bụi phổi silic nghề nghiệp	78
Hình 4.2	Kết quả Grad-CAM của các mô hình học sâu	81
Hình 4.3	Kiến trúc tổng thể của mô hình đề xuất:	90
Hình 4.4	Tổng quan kỹ thuật kết hợp các mô hình	93
Hình 4.5	Accuracy theo từng Epoch của 3 mô hình	98
Hình 4.6	Loss theo từng Epoch của 3 mô hình	99
Hình 4.7	Ma trận nhầm lẫn	101
Hình 4.8	Độ nhạy và độ đặc hiệu theo mô hình	103
Hình 4.9	Bản đồ nhiệt Grad-CAM của mô hình DenseNet201-GTP cho 4 lớp phân loại	104
Hình 4.10	Trực quan hóa t-SNE của các đặc trưng được học bởi mô hình: DenseNet201 (Hình bên trái) và DenseNet201-GTP (Hình bên phải).	105

PHẦN MỞ ĐẦU

1. Tính cấp thiết của nội dung nghiên cứu

Bệnh nghề nghiệp là hệ quả trực tiếp của quá trình lao động trong môi trường có yếu tố nguy cơ, gây ảnh hưởng lâu dài đến sức khỏe và chất lượng sống của người lao động. Theo báo cáo của Tổng Liên đoàn Lao động Việt Nam, đến cuối năm 2023, cả nước có khoảng 33.000 lao động mắc bệnh nghề nghiệp, với trung bình mỗi năm có khoảng 7.000 trường hợp được chẩn đoán mới và 500 trường hợp được giám định để hưởng bảo hiểm xã hội [1]. Trong danh mục 35 bệnh nghề nghiệp được bảo hiểm y tế chi trả, bệnh bụi phổi silic nghề nghiệp là bệnh phổ biến, chiếm tỷ lệ cao và gây ảnh hưởng nặng nề đến hệ hô hấp [2]. Bệnh tiến triển âm thầm, không có triệu chứng điển hình ở giai đoạn đầu, nhưng theo thời gian sẽ dẫn đến xơ hóa phổi không hồi phục, suy hô hấp mạn tính, suy giảm chức năng lao động, giảm tuổi thọ, và thậm chí tử vong nếu không được phát hiện và can thiệp kịp thời [3]. Theo báo cáo của Cục Quản lý môi trường Y tế (Bộ Y tế), bệnh bụi phổi silic vẫn đang là một gánh nặng bệnh tật của Việt Nam, số người mắc và tử vong thực tế cao hơn nhiều so với số liệu báo cáo và được giám định hàng năm. Mặc dù hiện nay chưa có phương pháp điều trị đặc hiệu, nhưng bệnh hoàn toàn có thể được dự phòng nếu được phát hiện sớm và kiểm soát tốt các yếu tố nguy cơ [4,5]. Vì vậy việc dự phòng và chẩn đoán sớm bệnh bụi phổi silic nghề nghiệp rất quan trọng trong công tác chăm sóc sức khỏe người lao động.

Hiện nay, chụp X-quang ngực vẫn là phương pháp phổ biến trong chẩn đoán và sàng lọc bệnh bụi phổi silic [6]. Tuy nhiên, kỹ thuật này còn tồn tại một số hạn chế. Độ nhạy chẩn đoán ở giai đoạn sớm còn thấp, dễ bỏ sót các tổn thương nhỏ, không điển hình [7,8]. Ngoài ra, kết quả chẩn đoán phụ thuộc nhiều vào trình độ và kinh nghiệm của bác sĩ, trong khi các dấu hiệu của bệnh bụi phổi silic trên X-quang có thể bị nhầm lẫn với các bệnh lý hô hấp khác như lao phổi, viêm phổi, hoặc bệnh phổi tắc nghẽn mạn tính [9]. Sự không nhất quán trong chẩn đoán không chỉ ảnh hưởng đến hiệu quả điều trị mà còn làm chậm trễ trong công tác quản lý, giám sát và phòng ngừa bệnh tiến triển.

Trong bối cảnh chẩn đoán hình ảnh y tế đòi hỏi độ chính xác cao và tự động

hóa, trí tuệ nhân tạo (AI), đặc biệt là học sâu (Deep Learning), được ghi nhận là giải pháp để nâng cao hiệu quả chẩn đoán. Tại Việt Nam, một số nghiên cứu đã ứng dụng AI trong chẩn đoán các bệnh lý lồng ngực. Tiêu biểu như hệ thống VinDr được phát triển bởi VinBigdata từ năm 2018, hỗ trợ bác sĩ phân tích ảnh X-quang ngực để chẩn đoán các bệnh lý như viêm phổi và tổn thương liên quan đến COVID-19, với mục tiêu tăng tốc độ và độ chính xác chẩn đoán [19]. Đề tài “Nghiên cứu ứng dụng trí tuệ nhân tạo (AI) hỗ trợ chẩn đoán hình ảnh X-quang phổi tại Bệnh viện Đa khoa thành phố Vinh” do Sở Khoa học và Công nghệ Nghệ An nghiệm thu (23/10/2024) đã chứng minh khả năng cải thiện độ chính xác, tốc độ chẩn đoán các bệnh lý phổi phổ biến. Đề tài cấp Nhà nước KC4.0 “Nghiên cứu xây dựng hệ thống hỗ trợ chẩn đoán và dự báo dịch tễ địa không gian bệnh lao phổi bằng ảnh X-quang ngực ở Việt Nam”, do PGS.TS Nguyễn Viết Nhung chủ nhiệm, Bệnh viện Phổi Trung ương chủ trì, hướng tới phát triển hệ thống hỗ trợ chẩn đoán lao phổi và dự báo nguy cơ dịch tễ trên quy mô quốc gia. Những nỗ lực này minh chứng cho xu hướng ứng dụng ngày càng rộng rãi của trí tuệ nhân tạo trong lĩnh vực y tế tại Việt Nam. Tuy nhiên, chưa có nghiên cứu chuyên biệt nào tại Việt Nam tập trung vào bệnh bụi phổi silic nghề nghiệp – một bệnh lý đặc thù với tổn thương không điển hình, dữ liệu khan hiếm.

Các công trình quốc tế đã chỉ ra các mô hình học sâu có thể hỗ trợ phát hiện các tổn thương mờ hoặc khó nhận diện trên ảnh X-quang ngực, những dấu hiệu mà phương pháp chẩn đoán truyền thống đôi khi bỏ sót [11,45]. Các hướng nghiên cứu mới như học sâu trên đồ thị (Graph Neural Network) và học sâu ít mẫu (Few-Shot Learning FSL) mở ra triển vọng trong việc nâng cao hiệu quả chẩn đoán trong điều kiện dữ liệu khan hiếm và mất cân bằng. Việc ứng dụng các mô hình này không chỉ giúp cải thiện khả năng phát hiện sớm bệnh lý mà còn góp phần giảm phụ thuộc vào đánh giá chủ quan, tăng tính tin cậy và khả năng triển khai tại các cơ sở y tế có nguồn lực hạn chế. Do đó, việc nghiên cứu và ứng dụng các mô hình học sâu hiện đại nhằm hỗ trợ chẩn đoán bệnh bụi phổi silic không chỉ mang tính cấp thiết về mặt y tế, mà còn phù hợp với định hướng Nghị quyết số 57-NQ/TW ngày 22/12/2024 của Bộ Chính trị về đột phá trong phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia, trong đó đặc biệt nhấn mạnh đến việc thúc đẩy ứng dụng trí tuệ nhân tạo trong lĩnh vực y tế và chăm sóc sức khỏe cộng đồng [12].

Xuất phát từ những cơ sở thực tiễn và định hướng chiến lược nêu trên, đề

tài luận án tiến sĩ: “*Nghiên cứu phát triển một số mô hình học sâu trong hỗ trợ chẩn đoán bệnh bụi phổi silic nghề nghiệp dựa trên ảnh X-quang ngực*” là hoàn toàn phù hợp với yêu cầu thực tiễn, mang ý nghĩa khoa học, tính ứng dụng cao và góp phần giải quyết một vấn đề cấp bách trong chăm sóc sức khỏe nghề nghiệp hiện nay.

2. Mục tiêu nghiên cứu

2.1. Mục tiêu tổng quát

Nghiên cứu phát triển các mô hình học sâu nhằm hỗ trợ chẩn đoán bệnh bụi phổi silic nghề nghiệp từ ảnh X-quang ngực, dựa trên bộ dữ liệu đa lớp được xây dựng và chuẩn hóa, kết hợp hai hướng tiếp cận học ít mẫu và học sâu trên đồ thị nhằm nâng cao độ chính xác và khả năng tổng quát của mô hình.

2.2. Mục tiêu cụ thể

Để đạt được mục tiêu tổng quát nêu trên, luận án đặt ra các mục tiêu cụ thể như sau:

Nghiên cứu và đánh giá tổng quan: phân tích, hệ thống hóa các phương pháp chẩn đoán bệnh bụi phổi silic từ ảnh X-quang ngực, đặc biệt là các kỹ thuật ứng dụng trí tuệ nhân tạo như học máy, học sâu tiên tiến; xác định hướng nghiên cứu để làm cơ sở đề xuất mô hình mới.

Xây dựng tập dữ liệu: đề xuất, tăng cường và hệ thống hóa bộ dữ liệu ảnh X-quang ngực có gán nhãn gồm bốn nhóm: bụi phổi silic nghề nghiệp, viêm phổi do virus, viêm phổi do vi khuẩn và đối tượng bình thường; đảm bảo dữ liệu được chuẩn hóa, cân bằng và có giá trị phục vụ huấn luyện mô hình học sâu.

Phát triển mô hình: nghiên cứu và đề xuất mô hình học sâu cải tiến cho bài toán phân vùng và phân loại tổn thương phổi, dựa trên hai hướng tiếp cận:

(i) Đề xuất ứng dụng phương pháp học ít mẫu (Few-Shot Learning) cho phép đồng thời thực hiện phân đoạn và phân loại trong điều kiện dữ liệu hạn chế.

(ii) Phát triển kiến trúc Graph Transformer Post-hoc (GTP) kết hợp hàm mất mát cân bằng (Balanced Loss) và kỹ thuật học tổng hợp (Ensemble Learn-

ing) để khai thác mối quan hệ không gian giữa các vùng đặc trưng, nâng cao độ ổn định và hiệu năng phân loại.

Đánh giá và ứng dụng: Đánh giá mô hình theo các chỉ số y sinh (độ nhạy, độ đặc hiệu, AUC, F1-score...), trực quan hóa kết quả và so sánh với chẩn đoán của chuyên gia nhằm xác minh khả năng ứng dụng lâm sàng.

Những mục tiêu cụ thể này sẽ được triển khai tương ứng theo bố cục nội dung các chương trong luận án, từ đó làm rõ giá trị khoa học và tính ứng dụng thực tiễn của luận án.

3. Đối tượng và phạm vi nghiên cứu

3.1. Đối tượng nghiên cứu

Luận án tập trung vào hai nhóm đối tượng nghiên cứu chính:

Tập dữ liệu ảnh X-quang ngực có gán nhãn, được xây dựng phục vụ bài toán chẩn đoán bệnh bụi phổi silic nghề nghiệp. Tập dữ liệu bao gồm bốn nhóm: bệnh bụi phổi silic nghề nghiệp, bệnh viêm phổi do vi khuẩn, bệnh viêm phổi do virus và nhóm bình thường.

Các mô hình học sâu ứng dụng trong chẩn đoán hình ảnh y tế, bao gồm mạng nơ-ron tích chập (CNN), mạng học sâu trên đồ thị (Graph Neural Networks – GNN) và mạng biến đổi (Transformers). Luận án tập trung triển khai các hướng tiếp cận chính: mô hình học sâu ít mẫu (Few-Shot Learning); mô hình Graph Transformer Post-hoc (GTP); các kỹ thuật học tổng hợp (Ensemble Learning). Đặc biệt, mô hình được tối ưu hóa với hàm mất mát cân bằng BalCE nhằm cải thiện hiệu quả phân loại trong điều kiện dữ liệu mất cân bằng.

3.2. Phạm vi nghiên cứu

Phạm vi nghiên cứu của luận án được xác định trong giới hạn cụ thể nhằm tập trung giải quyết bài toán chẩn đoán bệnh bụi phổi silic từ ảnh X-quang ngực dưới góc nhìn của khoa học dữ liệu và học sâu.

Về dữ liệu: luận án sử dụng bộ dữ liệu ảnh X-quang ngực được thu thập từ nguồn trong nước và quốc tế, bao gồm bốn nhóm chính: bụi phổi silic nghề nghiệp, viêm phổi do vi khuẩn, viêm phổi do virus và nhóm bình thường. Dữ liệu được xử lý và chuẩn hóa để đảm bảo đầu vào thống nhất cho các mô hình

học sâu. Luận án không mở rộng sang các loại dữ liệu y học khác như CT.

Về bài toán nghiên cứu: luận án tập trung vào hai hướng chính: (i) Phân loại ảnh X-quang ngực đa lớp trong điều kiện dữ liệu mất cân bằng; và (ii) Hỗ trợ chẩn đoán với số lượng mẫu hạn chế, tiếp cận theo hướng học sâu ít mẫu. Các bài toán được xây dựng trên cơ sở phù hợp với đặc điểm của dữ liệu thực tế và năng lực xử lý của mô hình đề xuất.

Về mô hình triển khai: luận án tập trung nghiên cứu các mô hình học sâu có giám sát, bao gồm các kiến trúc mạng nơ-ron tích chập như DenseNet và MobileNet, cùng với mô hình Transformer hiện đại là Swin Transformer, phù hợp với dữ liệu ảnh y tế.

Ngoài ra, luận án triển khai kiến trúc mô hình Few-Shot Learning (FS-CS) để giải quyết bài toán học từ số lượng mẫu hạn chế, và kiến trúc mô hình Graph Transformer Post-hoc thuộc nhóm mạng học sâu trên đồ thị nhằm nâng cao hiệu quả phân loại.

Luận án không mở rộng sang các phương pháp học không giám sát, học tăng cường, hoặc các hệ thống đa phương thức, nhằm tập trung vào khai thác hiệu quả các mô hình có giám sát trong điều kiện dữ liệu y tế thực tế.

Về đánh giá: Hiệu quả của mô hình được đánh giá chủ yếu thông qua các chỉ số định lượng như độ chính xác, độ nhạy, độ đặc hiệu, F1-score và AUC. Ngoài ra, luận án cũng thực hiện trực quan hóa đầu ra mô hình và phân tích ma trận nhầm lẫn nhằm hỗ trợ việc giải thích kết quả. Việc triển khai lâm sàng hoặc thử nghiệm trên hệ thống thực tế không nằm trong phạm vi nghiên cứu của đề tài.

4. Phương pháp nghiên cứu

Luận án sử dụng phương pháp nghiên cứu ứng dụng, kết hợp giữa phân tích lý thuyết, thực nghiệm kỹ thuật và tham vấn chuyên gia:

Phương pháp nghiên cứu lý thuyết: tổng quan và hệ thống hóa các công trình liên quan đến bệnh bụi phổi silic và các mô hình học sâu nhằm xác định hướng nghiên cứu, định hướng đề xuất mô hình phù hợp với bài toán.

Phương pháp thực nghiệm kỹ thuật: thu thập, xử lý và gán nhãn dữ liệu ảnh X-quang ngực; xây dựng tập dữ liệu; triển khai và huấn luyện các mô hình học sâu với nhiều kiến trúc khác nhau, kết hợp thử nghiệm hàm mất mát cân

bằng và kỹ thuật học tổng hợp.

Phương pháp đánh giá và kiểm chứng: sử dụng các chỉ số chuẩn như độ chính xác, độ nhạy, độ đặc hiệu, Macro-F1, AUC-ROC; trực quan hóa với Grad-CAM và kiểm định chéo để đảm bảo độ tin cậy.

Phương pháp tham vấn chuyên gia: Tiếp nhận ý kiến của bác sĩ và chuyên gia trong suốt quá trình gán nhãn, xác thực mô hình và đánh giá khả năng ứng dụng thực tiễn trong chẩn đoán lâm sàng.

5. Bố cục của Luận án

Luận án với tên đề tài “*Nghiên cứu phát triển một số mô hình học sâu trong hỗ trợ phát hiện bệnh bụi phổi silic nghề nghiệp dựa trên ảnh X-quang ngực*” được cấu trúc gồm bốn chương, cụ thể như sau:

Chương 1 – Tổng quan: Trình bày tổng quan về bệnh bụi phổi silic nghề nghiệp, các phương pháp chẩn đoán hiện nay, và khảo sát hệ thống các nghiên cứu liên quan trong lĩnh vực học máy và học sâu. Trên cơ sở đó, chương xác định hướng nghiên cứu và định hướng tiếp cận phù hợp cho luận án.

Chương 2 – Xây dựng tập cơ sở dữ liệu: Mô tả quy trình thu thập, xử lý và chuẩn hóa tập dữ liệu ảnh X-quang ngực, đồng thời phân tích đặc điểm dữ liệu và định dạng đầu vào cho các mô hình học sâu.

Chương 3 – Đề xuất mô hình học sâu ít mẫu hỗ trợ chẩn đoán bệnh bụi phổi silic nghề nghiệp: Trình bày mô hình Few-Shot Learning với cấu trúc phân đoạn – phân loại kết hợp (FS-CS), được thiết kế để xử lý bài toán thiếu mẫu. Chương cũng trình bày thiết lập thí nghiệm và kết quả đánh giá hiệu năng mô hình.

Chương 4 – Đề xuất mô hình học sâu hỗ trợ phân loại bệnh bụi phổi silic nghề nghiệp: Giới thiệu mô hình Graph Transformer Post-hoc (GTP), kết hợp hàm mất mát cân bằng và kỹ thuật học tổng hợp. Chương này trình bày chi tiết quy trình xây dựng mô hình, thực nghiệm so sánh với các phương pháp nền, đánh giá về hiệu suất và khả năng ứng dụng mô hình.

6. Đóng góp khoa học của Luận án

Luận án có một số đóng góp khoa học chính như sau:

1. Đề xuất, tăng cường và hệ thống hóa được bộ dữ liệu.

Đề xuất, tăng cường và hệ thống hóa được bộ dữ liệu ảnh X-quang ngực có gán nhãn bao gồm 4 nhóm: bệnh bụi phổi silic nghề nghiệp, viêm phổi do virus, viêm phổi do vi khuẩn và đối tượng bình thường. Việc sử dụng bộ dữ liệu này để huấn luyện các mô hình học sâu cho mục đích hỗ trợ đoán bệnh bụi phổi silic nghề nghiệp. Nội dung được trình bày tại Chương 2.

2. Đề xuất cải tiến mô hình phân vùng và phân loại tổn thương bệnh bụi phổi silic nghề nghiệp từ ảnh X-quang ngực dựa trên hai hướng tiếp cận:

(1) Điều chỉnh và ứng dụng phương pháp học ít mẫu cho phép đồng thời phân đoạn và phân loại. Nội dung được trình bày tại Chương 3.

(2) Phát triển kiến trúc Graph Transformer Post-hoc (GTP) kết hợp hàm mất mát cân bằng (Balanced Loss), kết hợp kỹ thuật học tổng hợp (Ensemble Learning) nhằm khai thác mối quan hệ không gian giữa các vùng đặc trưng, cải thiện độ ổn định, nâng cao hiệu quả mô hình phân loại. Nội dung được trình bày tại Chương 4 .

7. Ý nghĩa khoa học và ý nghĩa thực tiễn

7.1. Ý nghĩa khoa học

Luận án góp phần mở rộng và điều chỉnh một số phương pháp học sâu hiện có, phù hợp với đặc thù của dữ liệu y tế, vốn thường khan hiếm và mất cân bằng. Cụ thể, việc đề xuất mô hình Graph Transformer Post-hoc cho phép khai thác quan hệ liên ảnh dưới dạng đồ thị, mở rộng khả năng ứng dụng của học sâu trên đồ thị trong chẩn đoán hình ảnh. Bên cạnh đó, mô hình học sâu ít mẫu (Few-Shot Learning) tích hợp phân loại và phân đoạn cũng giúp làm rõ phạm vi ứng dụng của các phương pháp meta-learning trong bài toán y học thực tế.

7.2. Ý nghĩa thực tiễn

Các kết quả nghiên cứu của luận án có tiềm năng ứng dụng trong lĩnh vực y học nghề nghiệp, với các giá trị ứng dụng cụ thể như sau:

Kết quả nghiên cứu có thể hỗ trợ bước đầu trong việc phát triển hệ thống hỗ trợ chẩn đoán bệnh bụi phổi silic, đặc biệt phù hợp khám sàng lọc đối với cơ sở y tế thiếu nhân lực.

Các mô hình đề xuất có tính khả chuyển, có thể được mở rộng cho các bệnh

lý phổi khác có biểu hiện ảnh học tương đồng, góp phần thúc đẩy xu hướng ứng dụng AI trong hỗ trợ chẩn đoán hình ảnh y tế tại Việt Nam.

CHƯƠNG 1: TỔNG QUAN

1.1 Tổng quan về bệnh bụi phổi silic nghề nghiệp

1.1.1 Sơ lược về bệnh nghề nghiệp

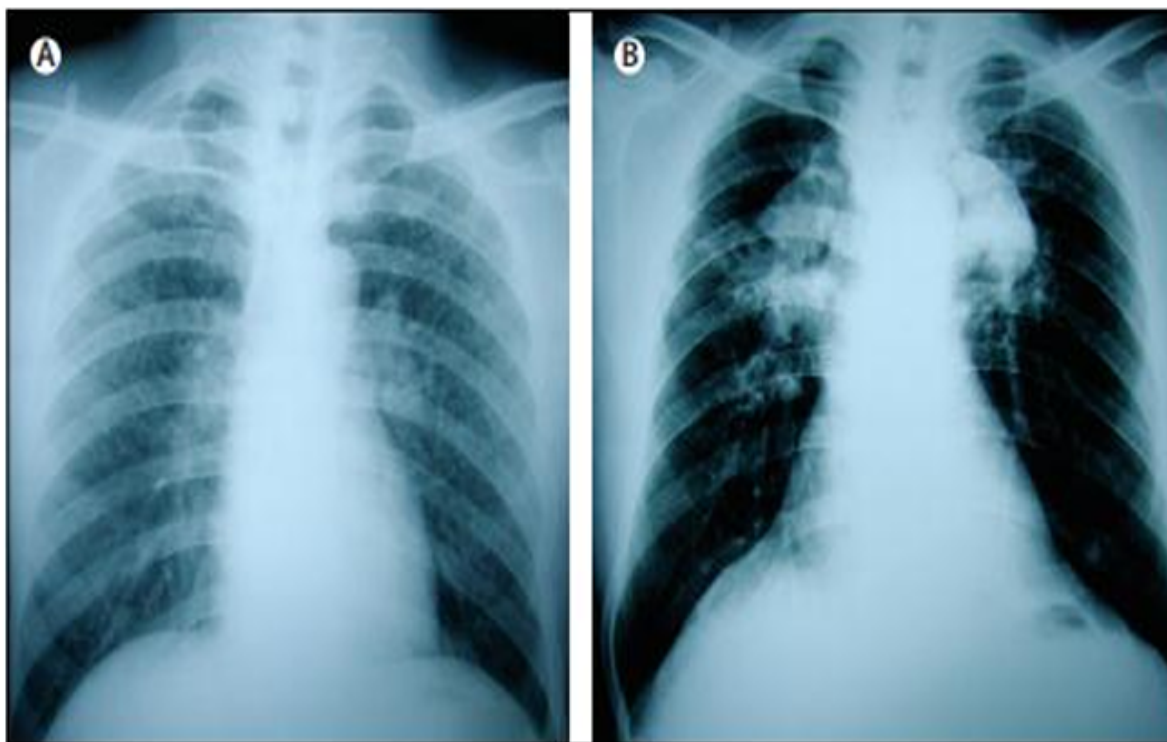
Theo Luật An toàn, vệ sinh lao động số 84/2015/QH13 của Quốc hội (2015), “bệnh nghề nghiệp là bệnh phát sinh do điều kiện lao động có hại của nghề nghiệp tác động đối với người lao động”. Những căn bệnh này không chỉ tiềm ẩn nhiều rủi ro, làm suy giảm thể chất và tinh thần, mà còn ảnh hưởng trực tiếp đến thu nhập của người lao động. Đồng thời, doanh nghiệp cũng phải đối mặt với các chi phí liên quan đến bồi thường và chăm sóc y tế cho người mắc bệnh. Do đó, việc chăm sóc y tế kỹ lưỡng, thực hiện khám sức khỏe định kỳ và duy trì hồ sơ theo dõi riêng biệt là rất cần thiết để đảm bảo an toàn cho người lao động.

Hiện nay, có 35 loại bệnh nghề nghiệp được bảo hiểm xã hội (BHXH) hỗ trợ khi người lao động đáp ứng đủ điều kiện [13]. Trong số đó, bệnh bụi phổi silic là một trong những căn bệnh nghiêm trọng nhất, với tỷ lệ mắc cao và gây ra tổn thương nghiêm trọng.

1.1.2 Sơ lược về bệnh bụi phổi silic nghề nghiệp

Bệnh bụi phổi là một bệnh phổi xơ hóa không hồi phục, hình thành do người lao động hít phải bụi chứa tinh thể silic dioxit trong quá trình lao động nghề nghiệp [14]. Khi đi vào hệ hô hấp, các hạt bụi kích thước nhỏ này gây tổn thương nhu mô phổi, dẫn đến phản ứng viêm kéo dài và xơ hóa mô phổi – một quá trình tiến triển mạn tính, không thể đảo ngược. Bệnh bụi phổi silic đã tạo ra gánh nặng tài chính rất lớn cho người lao động, gia đình của họ, cũng như tác động tiêu cực đến phát triển kinh tế xã hội [3, 15–17]. Theo thống kê của Tổ chức lao động Quốc tế (ILO), ước tính mỗi năm có khoảng gần 3 triệu người tử vong có nguyên nhân từ bệnh nghề nghiệp, con số này tương đương với khoảng 5.500 người chết mỗi ngày [18].

Tại Việt Nam, bệnh bụi phổi silic đã được nhận diện từ nhiều thập kỷ và luôn nằm trong nhóm bệnh nghề nghiệp trọng điểm cần giám sát. Một nghiên cứu hồi cứu dựa trên bộ số liệu của 63 tỉnh/thành phố giai đoạn 2016–2020 cho thấy bệnh bụi phổi silic chiếm khoảng 11,9% tổng số ca bệnh nghề nghiệp



Hình 1.1 Hình ảnh X-quang bệnh nhân bệnh bụi phổi silic
Hình 1.1 A là X-quang nốt xơ bệnh bụi phổi silic đơn thuần,
Hình 1.1 B là X-quang bệnh xơ hóa tiến triển nhanh [19]

được ghi nhận, tuy nhiên đây mới chỉ là số liệu từ một cơ sở dữ liệu y tế nhất định [20]. Nghiên cứu dịch tễ học cắt ngang tại một số nơi làm việc có nguy cơ đã xác định tỷ lệ mắc bệnh bụi phổi silic hiện mắc đáng kể [21, 22]. Bệnh bụi phổi silic chưa có thuốc điều trị đặc hiệu (không chữa được) nhưng có thể dự phòng được. Do vậy, việc chẩn đoán phát hiện bệnh sớm khi chưa có biểu hiện lâm sàng rõ ràng, xác định được tỷ lệ mắc ở mỗi thời điểm, tại từng vị trí, từ đó đề xuất được các biện pháp quản lý, theo dõi và điều trị là yêu cầu đặc biệt cấp thiết trong chăm sóc sức khỏe nghề nghiệp.

1.1.3 Triệu chứng, chẩn đoán bệnh bụi phổi silic nghề nghiệp

Theo Thông tư 15/2016/TT-BYT về bệnh nghề nghiệp được hưởng bảo hiểm xã hội, hướng dẫn chẩn đoán bệnh bụi phổi silic được thực hiện dựa trên các yếu tố lâm sàng, cận lâm sàng, tiền sử tiếp xúc nghề nghiệp và loại trừ bệnh khác:

- Lâm sàng: Bệnh nhân có thể gặp các triệu chứng: Khó thở khi gắng sức,

dần trở thành khó thở thường xuyên. Đau tức ngực, ho, khạc đờm. Nghe phổi có thể phát hiện ran nổ, ran ẩm (thể cấp).

- Cận lâm sàng:

+ Với khả năng ứng dụng rộng rãi, X-quang ngực được coi là phương pháp chẩn đoán chính của bệnh bụi phổi silic, giúp nhận diện các đặc điểm tổn thương đặc trưng. Theo hướng dẫn của Tổ chức Lao động Quốc tế (ILO) năm 2011 [24], hình ảnh X-quang có thể biểu hiện:

Các nốt mờ nhỏ, tròn, đều (ký hiệu p, q, r) hoặc các đám mờ lớn hơn (ký hiệu A, B, C), được so sánh và đánh giá theo bộ phim mẫu kỹ thuật số ILO 2011.

Hình ảnh có thể kèm theo các dấu hiệu như khí phế thũng, hoại tử khoang, hoặc các vôi hóa dạng vỏ trứng.

+ Chụp cắt lớp vi tính (CT): Được chỉ định khi cần thiết để cung cấp hình ảnh chi tiết hơn về tổn thương phổi, hỗ trợ trong các trường hợp X-quang chưa đủ rõ ràng hoặc cần xác định thêm mức độ tổn thương.

+ Đánh giá chức năng hô hấp: Thường được thực hiện để xác định mức độ rối loạn chức năng phổi, bao gồm rối loạn thông khí hạn chế, tắc nghẽn hoặc hỗn hợp.

- Tiền sử tiếp xúc nghề nghiệp: Người lao động làm việc trong môi trường có nồng độ, số lượng, kích thước hạt bụi và hàm lượng silic tự do vượt quá giới hạn cho phép.

Mức độ tiếp xúc: Tiếp xúc với bụi chứa silic tự do ở nồng độ $> 0,1 \text{ mg/m}^3$ trong thời gian làm việc 8 giờ/ngày.

Thời gian tiếp xúc: Ít nhất 3 tháng cho các trường hợp cấp tính và 5 năm đối với các trường hợp mãn tính.

Thời gian phát bệnh: Đối với bệnh bụi phổi silic cấp tính, bệnh thường phát triển trong vòng 1 năm sau khi ngừng tiếp xúc. Trong trường hợp mãn tính, thời gian phát bệnh có thể kéo dài đến 35 năm sau khi ngừng tiếp xúc.

Trong thực tế lâm sàng, ảnh X-quang ngực vẫn là công cụ chủ đạo trong quy trình chẩn đoán nhờ tính phổ biến, chi phí thấp và khả năng nhận diện đặc trưng hình thái bệnh. Tuy nhiên, việc đọc và diễn giải hình ảnh X-quang phụ thuộc nhiều vào kinh nghiệm của bác sĩ, dễ xảy ra sai sót trong các trường hợp

biểu hiện không điển hình hoặc bệnh lý chồng lấp. Do đó, việc kết hợp X-quang với các kỹ thuật hiện đại như chụp CT scanner và đánh giá chức năng hô hấp giúp tăng độ chính xác trong chẩn đoán, hỗ trợ phân loại mức độ tổn thương. Đây cũng là cơ sở quan trọng để xây dựng và phát triển các hệ thống chẩn đoán hỗ trợ bằng trí tuệ nhân tạo, góp phần tăng tính nhất quán, giảm sai lệch chủ quan và nâng cao độ tin cậy trong thực hành y khoa hiện nay

1.1.4 Phân loại tiêu chuẩn bệnh bụi phổi silic nghề nghiệp

Trên ảnh X-quang ngực, các bất thường do bụi phổi silic thường biểu hiện qua sự thay đổi mật độ phổi, phản ánh mức độ cản tia X. Những vùng tổn thương này xuất hiện dưới dạng các vùng mờ phổi, bao gồm: đông đặc, xẹp phổi, và đặc biệt là tổn thương kẽ – dạng điển hình trong bệnh bụi phổi silic.

Năm 2011, Tổ chức Lao động Quốc tế (ILO) đã đưa ra hướng dẫn phân loại trên phim X-quang ngực cho bệnh bụi phổi silic nghề nghiệp [25]. Các bất thường trên X-quang ngực được phân loại thành hai nhóm chính: Tổn thương nhu mô phổi (nốt mờ nhỏ, nốt mờ lớn) và tổn thương màng phổi.

- Đối với nốt mờ nhỏ ($< 1\text{ cm}$):

+ Phân loại theo mức độ tập trung: ILO đã đề xuất hệ thống phân loại bệnh bụi phổi theo mức độ bốn cấp độ: từ 0 đến 3, với cấp 0 là bình thường và cấp 3 là nghiêm trọng nhất.

Thang điểm phân loại cấp độ nốt mờ nhỏ gồm 12 phân cấp từ nhẹ đến nặng: 0/-, 0/0, 0/1, 1/0, 1/1, 1/2, 2/1, 2/2, 2/3, 3/2, 3/3, 3/+.

Trong hệ thống phân loại ILO, mỗi cặp ký hiệu X/Y có ý nghĩa như sau: X: cấp độ chính (primary category), phản ánh mức độ tập trung nốt mờ chủ yếu. Y: cấp độ phụ (secondary category), phản ánh mức độ kế cận nếu hình ảnh có đặc điểm dao động giữa hai mức.

+ Phân loại theo hình dạng và kích thước, nốt mờ nhỏ được chia thành hai nhóm: Hình dạng nốt tròn và không tròn.

Hình dạng nốt tròn:

p: các nốt mờ nhỏ có đường kính $\leq 1,5\text{ mm}$ (nhóm nhỏ nhất).

q: các nốt mờ nhỏ có đường kính trong khoảng từ 1,5 mm đến 3mm (nhóm

Bảng 1.1 Thang phân loại mức độ tập trung của các nốt mờ nhỏ theo tiêu chuẩn ILO 2011

Cấp độ	Phân cấp	Ý nghĩa
0	0/-, 0/0, 0/1	Không có hoặc rất ít nốt
1	1/0, 1/1, 1/2	Mật độ nhẹ
2	2/1, 2/2, 2/3	Mật độ trung bình
3	3/2, 3/3, 3/+	Mật độ cao (tổn thương nặng, lan tỏa)

trung bình).

r: các nốt mờ nhỏ có đường kính trong khoảng từ 3 mm đến 10mm (nhóm lớn nhất trong các nốt mờ nhỏ).

Hình dạng nốt không tròn: Những nốt mờ này được quan sát và đánh giá tại sáu vùng chính của phổi gồm ba vùng (trên, giữa, dưới) của mỗi bên phổi trái và phải như minh họa trong Hình 1.2. Ngoài cách phân loại theo ký hiệu (p, q, r), kích thước các nốt mờ nhỏ còn được phân loại cụ thể dựa trên tiêu chuẩn đo đạc trực tiếp từ hình ảnh X-quang như sau:

Nhóm nhỏ nhất s: đường kính $\leq 1,5$ mm.

Nhóm trung bình t: đường kính từ trên 1,5 mm đến 3 mm.

Nhóm lớn nhất u (trong các nốt nhỏ): đường kính từ trên 3 mm đến 10 mm.

- Đối với các nốt mờ lớn: Các nốt mờ có đường kính lớn hơn 10 mm được xếp vào nhóm nốt mờ lớn và tiếp tục được phân loại chi tiết thành ba nhóm dựa trên diện tích bề mặt:

Nhóm A: diện tích từ 0 đến 50 mm².

Nhóm B: diện tích từ 50 đến 125 mm².

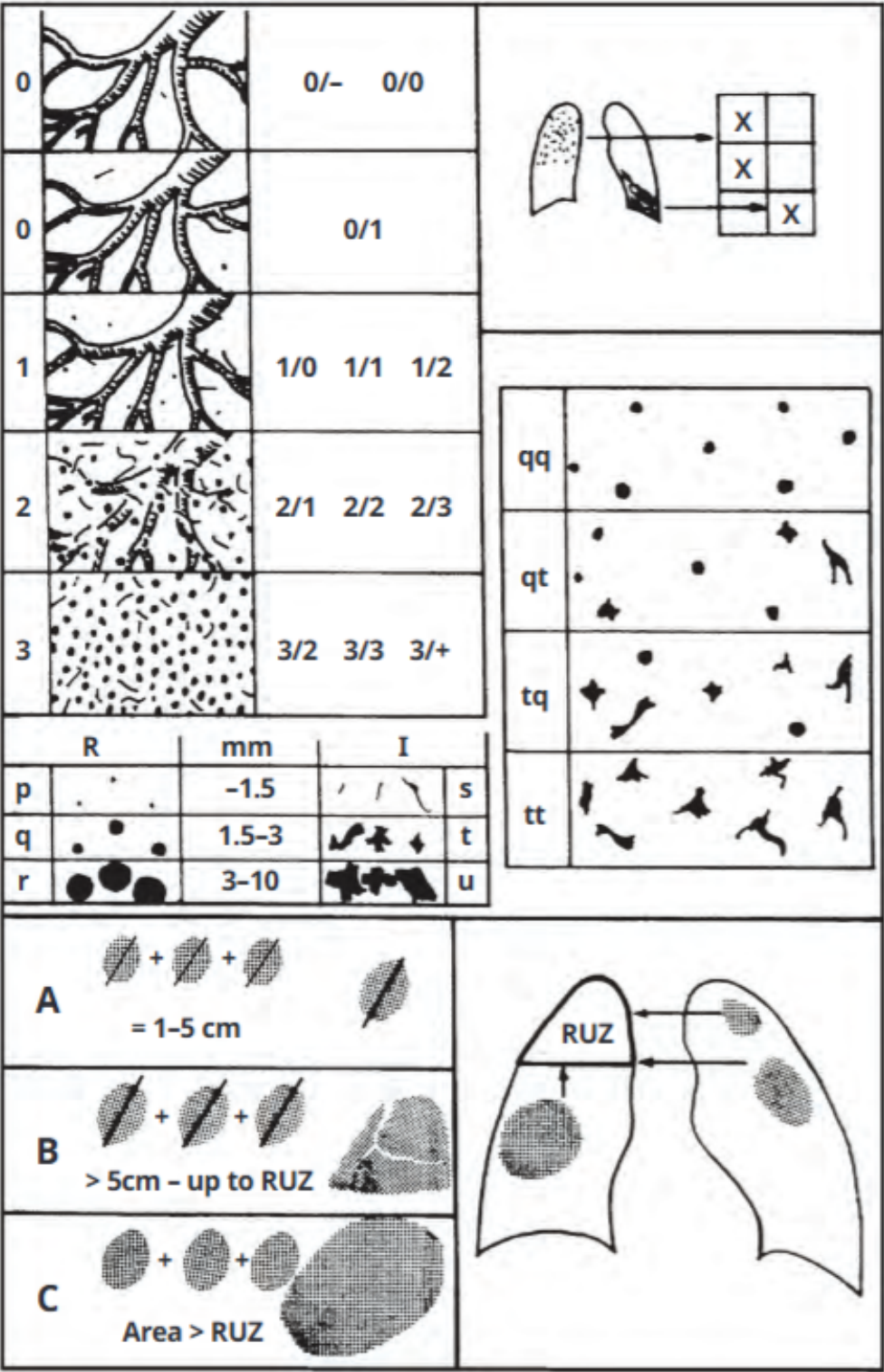
Nhóm C: diện tích lớn hơn 125 mm².

Dựa trên thang điểm trên, mức độ tổn thương có thể được phân loại như sau:

Nhẹ: các nốt mờ nhỏ ở mức 0/- đến 1/2.

Trung bình: nốt mờ nhỏ mức 2/1 đến 2/3.

Nặng: nốt mờ nhỏ từ 3/2 đến 3/+ hoặc có xuất hiện nốt mờ lớn (A, B, C).



Hình 1.2 Minh họa hệ thống phân loại tổn thương phổi theo tiêu chuẩn ILO [23]

Hình 1.2 mô tả mức độ tập trung nốt mờ nhỏ (profusion), hình dạng – kích

thước nốt tròn và không tròn (p, q, r, s, t, u), và phân loại nốt lớn (A, B, C) theo vùng tổn thương trên ảnh X-quang ngực [25].

Mặc dù hệ thống ILO mang tính tiêu chuẩn, nhưng trong thực tế, việc xác định chính xác các nốt tổn thương nhỏ vẫn gặp nhiều khó khăn: Nốt có kích thước nhỏ, mờ, phân bố thưa thớt. Có thể bị che lấp bởi xương sườn, mạch máu, tim. Tổn thương màng phổi dễ nhầm lẫn với các bệnh lý khác như viêm phổi, lao phổi, bệnh bụi amiăng hoặc các bệnh phổi nghề nghiệp khác.

Do đó, việc nghiên cứu và phát triển các mô hình học máy để tích hợp vào các hệ thống chẩn đoán hỗ trợ bằng máy tính (CAD) là rất cần thiết nhằm nâng cao độ chính xác trong việc phát hiện bệnh bụi phổi silic (BPSi) nghề nghiệp. Các mô hình học sâu như CNN, Transformer có khả năng xử lý và phân tích các đặc trưng phức tạp từ hình ảnh X-quang ngực [26–29], giúp phát hiện sớm và chính xác hơn các dấu hiệu của BPSi. Việc tích hợp các mô hình học máy vào hệ thống CAD không chỉ giúp giảm thiểu sai sót trong quá trình chẩn đoán mà còn hỗ trợ các bác sĩ đưa ra quyết định điều trị hiệu quả và kịp thời, đặc biệt trong các trường hợp bệnh BPSi khó nhận biết ở giai đoạn đầu.

1.2 Khảo sát nghiên cứu liên quan

1.2.1 Các phương pháp chẩn đoán lâm sàng bệnh bụi phổi silic nghề nghiệp

Hiện nay, chẩn đoán bệnh bụi phổi silic nghề nghiệp vẫn đặt trọng tâm vào khai thác tiền sử phơi nhiễm nghề nghiệp kết hợp thăm khám lâm sàng. Theo hướng dẫn của Bộ Y tế, người lao động cần có thời gian tiếp xúc đủ dài với bụi chứa silic tự do, các triệu chứng thường gặp bao gồm khó thở khi gắng sức, ho khan, đau ngực hoặc biểu hiện hội chứng hạn chế/tắc nghẽn hô hấp. Tuy nhiên, các triệu chứng này thường xuất hiện muộn; giai đoạn sớm của bệnh thường nghèo nàn biểu hiện và nhiều trường hợp được phát hiện tình cờ qua khám sức khỏe định kỳ.

Các phương pháp cận lâm sàng giữ vai trò quan trọng trong việc xác định chẩn đoán, phân biệt với bệnh lý khác và đánh giá biến chứng. Bao gồm:

Chẩn đoán hình ảnh: X-quang phổi tiêu chuẩn (phim thẳng) là công cụ sàng lọc chủ đạo trong khám bệnh nghề nghiệp [3]. Trên phim X-quang ngực, thể mạn tính thường biểu hiện bằng các nốt mờ nhỏ kích thước 1–3 mm, phân bố chủ yếu ở các thùy trên của hai phổi; kèm theo đó, vôi hóa hạch rốn phổi và

trung thất là những dấu hiệu gợi ý quan trọng. Khi bệnh tiến triển, các nốt này có thể hợp nhất tạo thành khối xơ hóa lớn. Tuy nhiên, hình ảnh tương tự cũng có thể gặp trong các bệnh lý khác như bụi phổi than, lao kê, sarcoidosis, viêm phổi kẽ, nhiễm nấm hoặc ung thư di căn phổi.

Chụp cắt lớp vi tính (CT), đặc biệt là CT độ phân giải cao: Có độ nhạy cao hơn trong phát hiện các tổn thương nhỏ và theo dõi tiến triển bệnh. CT hỗ trợ phân biệt bụi phổi silic với các bệnh phổi khác. Hạn chế chính là chi phí cao, liều phơi nhiễm tia X lớn nên không phù hợp sàng lọc diện rộng.

Thăm dò chức năng hô hấp: Được sử dụng để đánh giá rối loạn thông khí (hạn chế hoặc tắc nghẽn). Tuy không đặc hiệu cho bụi phổi silic, nhưng có giá trị trong theo dõi tiến triển bệnh và đánh giá mức độ suy giảm chức năng phổi.

Sinh thiết phổi: Hiếm khi được chỉ định, chỉ thực hiện khi hình ảnh học và tiền sử phơi nhiễm không điển hình, nhằm loại trừ các bệnh lý khác.

Xét nghiệm sinh học phân tử và dấu ấn sinh học: Hiện vẫn chưa có xét nghiệm sinh học đặc hiệu; một số nghiên cứu đang khảo sát các dấu ấn sinh học, chẳng hạn như biến đổi gen $TNF-\alpha$, với kỳ vọng hỗ trợ chẩn đoán sớm và dự báo nguy cơ [30].

Tóm lại, việc chẩn đoán bệnh bụi phổi silic nghề nghiệp hiện vẫn chủ yếu dựa vào phim X-quang ngực tiêu chuẩn kết hợp khai thác tiền sử phơi nhiễm, đây là phương pháp đơn giản, chi phí thấp và phù hợp cho công tác tầm soát trên quy mô lớn ở nhóm lao động nguy cơ cao. Tuy nhiên, hạn chế lớn là tính chủ quan trong chẩn đoán, kết quả phụ thuộc nhiều vào trình độ, kỹ năng, kinh nghiệm của bác sĩ. Chính vì vậy, nhu cầu ứng dụng các công nghệ hỗ trợ chẩn đoán dựa trên trí tuệ nhân tạo, đặc biệt là học sâu được ngày càng trở nên cấp thiết, nhằm nâng cao độ chính xác, tăng tính nhất quán giữa các bác sĩ và cải thiện khả năng phát hiện sớm được các tổn thương trên ảnh X-quang bụi phổi silic nghề nghiệp.

1.2.2 Các phương pháp học máy truyền thống

Trước khi các mô hình học sâu được phát triển rộng rãi, các phương pháp học máy truyền thống thường được triển khai trong phân tích ảnh y tế, bao gồm ảnh X-quang ngực, nhằm hỗ trợ phát hiện các bệnh phổi nghề nghiệp. Là một phân nhánh của trí tuệ nhân tạo, học máy được ứng dụng trong nhiều bài toán

lâm sàng trên ảnh y tế như dự đoán mức độ xâm lấn mô học và khả năng sống sót ở ung thư phổi [31], phân loại nốt phổi [32], cũng như chẩn đoán và phân loại COVID-19 [33]. Tuy vậy, so với nhiều bệnh phổi khác, nghiên cứu về bệnh bụi phổi silic nghề nghiệp còn hạn chế, dẫn đến thiếu hụt các công cụ AI chuyên biệt; thực trạng này bắt nguồn từ đặc điểm bệnh học phức tạp và sự khan hiếm bộ dữ liệu X-quang ngực được gán nhãn phục vụ huấn luyện.

Việc phát hiện bệnh phổi từ ảnh X-quang ngực nói chung, đặc biệt là viêm phổi, vẫn là một thách thức lớn do hạn chế về số lượng mẫu dữ liệu có nhãn [34, 35]. Các kỹ thuật trích xuất đặc trưng có kết cấu (texture features) như biểu đồ cường độ và ma trận đồng xuất hiện mức xám (GLCM), kết hợp bộ phân loại dựa trên mạng nơ-ron, đã được sử dụng để phân biệt bệnh bụi phổi (pneumoconiosis) với hình ảnh phổi bình thường [36]. Bên cạnh đó, mô hình cây quyết định (Decision Tree) và rừng ngẫu nhiên (Random Forest) cũng được áp dụng nhằm nhận diện các dấu hiệu tổn thương đặc trưng, hỗ trợ bác sĩ phân biệt nhanh và giảm sai lệch chủ quan [37].

Ở mức độ nghiên cứu cụ thể, Chandra và cộng sự [38] trích xuất các đặc trưng thống kê (cường độ trung bình, độ lệch chuẩn và đặc trưng kết cấu) từ ảnh X-quang để phân loại viêm phổi bằng mô hình MLP, đạt độ chính xác 95,39% trên 412 ảnh. Năm 2019, Kuo và cộng sự [39] đã kết hợp 11 đặc trưng lâm sàng với ảnh X-quang và mô hình cây quyết định để chẩn đoán viêm phổi ở bệnh nhân tâm thần phân liệt, đạt độ chính xác 94,5%. Ngoài ra, Yue và đồng sự [40] sử dụng sáu đặc trưng radiomics bậc hai trích xuất từ ảnh CT để dự đoán thời gian nằm viện của bệnh nhân COVID-19, cho AUC lên tới 0,97.

Mặc dù các phương pháp học máy truyền thống có ưu điểm về sự đơn giản và khả năng tính toán hiệu quả trong bối cảnh dữ liệu huấn luyện hạn chế, chúng vẫn bộc lộ nhiều giới hạn đáng kể. Trước hết, hiệu năng của các mô hình này phụ thuộc lớn vào việc trích xuất đặc trưng thủ công — một quá trình đòi hỏi nhiều thời gian và kinh nghiệm chuyên môn — đồng thời dễ bỏ sót các tín hiệu tinh vi, phức tạp trong dữ liệu ảnh [41]. Bên cạnh đó, hiệu suất phân loại của chúng thường suy giảm khi phải xử lý các bộ dữ liệu có quy mô lớn, chứa nhiều nhiễu hoặc mất cân bằng mẫu [42, 43]. Chính vì vậy, các nghiên cứu gần đây có xu hướng chuyển dịch sang học sâu nhằm tận dụng khả năng tự động học đặc trưng từ dữ liệu phức tạp.

Dẫu vậy, học máy truyền thống vẫn giữ một vai trò nhất định trong nghiên cứu y học, đặc biệt khi dữ liệu bị giới hạn hoặc khi cần các mô hình có khả năng diễn giải và kiểm chứng. Ngoài ra, các phương pháp này còn đóng vai trò là nền tảng để đối chiếu, so sánh và đánh giá hiệu quả của những mô hình học sâu tiên tiến hơn [42, 44].

1.2.3 Các phương pháp học sâu trong chẩn đoán ảnh X-quang y tế

1.2.3.1 Mạng nơ-ron tích chập (Convolutional Neural Networks - CNNs)

Sự xuất hiện của Mạng nơ-ron tích chập (CNN) đã đánh dấu một bước tiến trong lĩnh vực thị giác máy tính và phân tích hình ảnh y tế. Khác với các phương pháp học máy truyền thống, CNN có khả năng tự động trích xuất các đặc trưng phân cấp trực tiếp từ dữ liệu ảnh thô, nhờ đó loại bỏ hoàn toàn giai đoạn trích xuất đặc trưng thủ công vốn tốn nhiều công sức và phụ thuộc vào chuyên gia. Với quy trình học “từ đầu đến cuối”, mô hình có thể nhận diện những mẫu hình ảnh phức tạp và tinh vi mà con người dễ bỏ qua, mang lại những cải thiện về hiệu quả chẩn đoán.

Thành công của mạng nơ-ron tích chập CNN trong chẩn đoán ảnh X-quang phổi được minh chứng rõ nét qua nghiên cứu kinh điển về CheXNet của Rajpurkar và các cộng sự vào năm 2017 [45]. Bằng cách sử dụng kiến trúc DenseNet-121 và kỹ thuật học chuyển giao (Transfer Learning), mô hình này đã đạt được hiệu suất phát hiện bệnh viêm phổi có thể sánh ngang với các bác sĩ chẩn đoán hình ảnh X-quang, khởi đầu cho một làn sóng ứng dụng CNN vào chẩn đoán y tế. Từ đó đến nay, học chuyển giao trở thành một chiến lược phổ biến, giúp các mô hình CNN đạt hiệu quả cao ngay cả với các bộ dữ liệu y tế có số lượng mẫu hạn chế. Các công trình nghiên cứu của Rahman và cộng sự (2020) [49] cùng với Liang & Zheng (2020) [50] đã khẳng định sức mạnh cũng như độ hiệu quả của phương pháp này trong việc chẩn đoán viêm phổi nói chung và viêm phổi cho bệnh nhi nói riêng.

Tính linh hoạt giúp mạng CNN có thể được áp dụng để giải quyết nhiều bệnh lý về phổi khác nhau, từ viêm phổi [46], COVID-19 [51] cho đến các bệnh lý ít phổ biến hơn như bệnh bụi phổi silic [54]. Các phương pháp tiếp cận đa dạng, từ việc so sánh các bộ phân loại khác nhau [47] đến việc phân tích các yếu tố ảnh hưởng đến độ tin cậy của mô hình [52], đều cho thấy tiềm năng to

lớn của CNN trong xử lý ảnh y tế, đặc biệt là trong lĩnh vực phát hiện bệnh phổi dựa trên ảnh X quang.

Tuy nhiên, các mô hình mạng CNN cũng tồn tại những thách thức đáng kể. Thứ nhất, chúng thường được coi là các "hộp đen" (black box), khiến việc diễn giải, giải thích quyết định của mô hình trở nên khó khăn, đây là một trong những rào cản lớn trong lĩnh vực y khoa, nơi tính minh bạch là yếu tố tối quan trọng. Thứ hai, các mô hình này đòi hỏi một lượng lớn dữ liệu được gán nhãn chất lượng cao cho quá trình huấn luyện. Thứ ba, hiệu suất của chúng có thể bị ảnh hưởng bởi các yếu tố không liên quan đến bệnh lý, hay còn gọi là các "đường tắt" (shortcuts), thay vì học các dấu hiệu lâm sàng thực sự [52]. Những hạn chế này thúc đẩy sự phát triển của các phương pháp học sâu tiên tiến hơn như học sâu trên đồ thị và học sâu ít mẫu, nhằm giải quyết các bài toán về tính diễn giải và sự khan hiếm dữ liệu.

Bảng 1.2 Tổng hợp các nghiên cứu tiêu biểu về bài toán phát hiện bệnh phổi sử dụng Mạng nơ-ron tích chập (CNN) trên ảnh X-quang ngược.

Nghiên cứu	Phương pháp	Bộ dữ liệu	Kết quả	Ưu điểm	Hạn chế
Rajpurkar et al., 2017 [45]	DenseNet-121 với Học chuyển giao.	ChestX-ray14 (NIH).	AUC: 0.768 (viêm phổi).	Chứng minh CNN có thể đạt hiệu suất ngang tầm chuyên gia.	Tính diễn giải của mô hình còn hạn chế.
Stephen et al., 2019 [46]	Mô hình CNN tùy chỉnh.	Bộ dữ liệu X-quang ngực nhi khoa (Kaggle).	Độ chính xác: 93.73%	Xây dựng mô hình CNN hiệu quả cho phân loại viêm phổi.	Mô hình có thể không tổng quát tốt trên các bộ dữ liệu người lớn.
Al Mamlook et al., 2020 [47]	So sánh nhiều mô hình (bao gồm CNN).	Bộ dữ liệu X-quang ngực (Kaggle).	CNN đạt hiệu suất cao nhất.	Cung cấp một so sánh toàn diện giữa các bộ phân loại khác nhau.	Không tập trung sâu vào việc tối ưu hóa một kiến trúc CNN cụ thể.

Nghiên cứu	Phương pháp	Bộ dữ liệu	Kết quả	Ưu điểm	Hạn chế
Zhang et al., 2020 [48]	Phát hiện bất thường dựa trên CNN.	Bộ dữ liệu riêng và công khai (COVID-19, viêm phổi).	Sàng lọc hiệu quả viêm phổi virus.	Tiếp cận theo hướng phát hiện bất thường, hữu ích khi dữ liệu bệnh hiếm.	Hiệu quả phụ thuộc vào việc định nghĩa "bình thường".
Rahman et al., 2020 [49]	Học chuyển giao với các CNN phổ biến.	Bộ dữ liệu X-quang ngực (Kaggle).	Độ chính xác: 98.0%	Chứng minh hiệu quả vượt trội của học chuyển giao.	Hiệu suất có thể bị giảm khi áp dụng cho dữ liệu lâm sàng thực tế.
Liang & Zheng, 2020 [50]	Học chuyển giao với ResNet.	Bộ dữ liệu X-quang ngực nhi khoa.	Độ chính xác: 96.8%	Áp dụng thành công ResNet cho dữ liệu nhi khoa.	Tập trung vào một nhóm đối tượng cụ thể (nhi khoa).
Ozturk et al., 2020 [51]	Mạng nơ-ron sâu (DarkCovidNet).	Bộ dữ liệu X-quang ngực (COVID-19 và bình thường).	Độ chính xác: 98.08% (phân loại 2 lớp).	Phản ứng nhanh với đại dịch, đề xuất một kiến trúc chuyên biệt.	Bộ dữ liệu ban đầu còn nhỏ, cần kiểm chứng trên tập lớn hơn.
DeGrave et al., 2021 [52]	Phân tích các mô hình AI (CNN).	Ảnh X-quang COVID-19.	AI học các "đường tắt" thay vì dấu hiệu bệnh.	Nghiên cứu nền tảng, vạch trần vấn đề học 'đường tắt' (shortcuts) của AI.	Cho thấy sự nguy hiểm của việc triển khai AI 'hộp đen' mà không kiểm tra kỹ lưỡng.

Nghiên cứu	Phương pháp	Bộ dữ liệu	Kết quả	Ưu điểm	Hạn chế
Tuncer et al., 2021 [53]	CNN + F-transform để trích xuất đặc trưng.	Bộ dữ liệu X-quang ngực (COVID-19, viêm phổi, bình thường).	Độ chính xác: 98.14%	Kết hợp CNN với một kỹ thuật xử lý tín hiệu (F-transform) để cải thiện hiệu suất.	Cấu trúc mô hình phức tạp, khó triển khai thực tế trên quy mô lớn.
Sharma et al., 2024 [54]	Mô hình Học chuyển giao mở rộng.	Ảnh X-quang bệnh bụi phổi silic.	Phát hiện bệnh bụi phổi silic hiệu quả.	Mở rộng ứng dụng CNN sang một bệnh phổi nghề nghiệp cụ thể.	Cần xem xét chi tiết về "Extended" model. Có thể dữ liệu không công khai.

1.2.3.2 Học sâu trên đồ thị

Để khắc phục những hạn chế về khả năng diễn giải, mô hình hóa các mối quan hệ phức tạp của CNN, một hướng tiếp cận mới là Học sâu trên đồ thị, đặc biệt là sử dụng Mạng nơ-ron đồ thị (Graph Neural Networks - GNNs). Thay vì xử lý từng ảnh X-quang đơn thuần và riêng lẻ, phương pháp này biểu diễn dữ liệu đầu vào dưới dạng một đồ thị, trong đó các nút (nodes) đại diện cho các vùng giải phẫu hoặc các đặc trưng cục bộ, và các cạnh (edges) mã hóa mối quan hệ không gian hoặc ngữ cảnh giữa chúng. Cách tiếp cận này cho phép mô hình học được các cấu trúc quan hệ tinh vi, vốn là yếu tố quan trọng trong chẩn đoán y tế [55, 56].

Các ứng dụng cốt lõi của GNN trong chẩn đoán X-quang ngực thường tập trung vào việc xây dựng đồ thị từ các vùng ảnh và học các mối tương quan giữa chúng. Chẳng hạn, một số nghiên cứu xây dựng đồ thị trong đó mỗi nút là một vùng đặc trưng được trích xuất bởi một CNN, sau đó GNN được sử dụng để tổng hợp thông tin và đưa ra chẩn đoán cuối cùng. Hướng đi này đã được triển khai hiệu quả trong việc phân loại đa bệnh lý lồng ngực, do có thể mô hình hóa mối quan hệ phức tạp giữa các bệnh lý khác nhau [57, 58].

Tuy nhiên, một hướng nghiên cứu tiềm năng khác của GNN nằm ở việc ứng dụng

vào các bài toán tăng cường khả năng diễn giải và độ tin cậy của mô hình. Một trong những cách tiếp cận mới là xây dựng đồ thị dựa trên các dữ liệu bổ trợ từ chuyên gia. Vào năm 2024, công trình nghiên cứu GazeGNN của Wang và cộng sự [59] đã tích hợp dữ liệu ánh nhìn (eye gaze) của các bác sĩ X-quang vào quá trình xây dựng đồ thị, giúp mô hình tập trung vào các vùng ảnh quan trọng mà chuyên gia quan tâm. Cũng trong năm 2024, Sultana và cộng sự [60] đã kết hợp thêm báo cáo giọng nói (speech report) cùng với dữ liệu ánh nhìn của bác sĩ để tạo ra một đồ thị đa phương thức, cho phép mô hình "nhìn qua đôi mắt của chuyên gia".

Tóm lại, ưu điểm chính của học sâu trên đồ thị là khả năng mô hình hóa tường minh các mối quan hệ cấu trúc và ngữ cảnh, điều mà CNN chưa thực sự giải thích được một rõ ràng. Điều này không chỉ có khả năng cải thiện độ chính xác mà còn mở ra cơ hội lớn cho việc diễn giải mô hình, giúp bác sĩ hiểu được "lý do" đằng sau một chẩn đoán. Mặc dù vậy, phương pháp này cũng có những hạn chế. Việc xây dựng một đồ thị có ý nghĩa từ ảnh X-quang là một thách thức không nhỏ và thường đòi hỏi tri thức chuyên môn hoặc các kỹ thuật tiền xử lý phức tạp. Hơn nữa, GNN vẫn là một lĩnh vực tương đối mới và cần thêm nhiều nghiên cứu để kiểm chứng hiệu quả trên quy mô lớn.

Bảng 1.3 Tổng hợp các nghiên cứu tiêu biểu sử dụng Học sâu trên đồ thị trên ảnh X-quang.

Nghiên cứu	Phương pháp	Bộ dữ liệu	Kết quả	Ưu điểm	Hạn chế
Ahmedt-Aristizabal et al., 2021b [55]	Tổng quan/Khảo sát các phương pháp GNN.	Tổng quan nhiều bộ dữ liệu y tế.	Tổng hợp các xu hướng và thách thức.	Cung cấp cái nhìn tổng quan, hệ thống hóa kiến thức về GNN trong y tế.	Không tập trung vào một ứng dụng cụ thể.
Sekuboyina et al., 2021 [56]	Mạng đồ thị quan hệ (Relational Graph Network).	ChestX-ray14 (NIH).	Hiệu suất vượt trội hơn các phương pháp CNN.	Mô hình hóa rõ ràng mối quan hệ giữa các vùng giải phẫu và bệnh lý.	Việc xác định các vùng giải phẫu (nút) có thể cần tiền xử lý phức tạp.

Nghiên cứu	Phương pháp	Bộ dữ liệu	Kết quả	Ưu điểm	Hạn chế
Zhang et al., 2023 [57]	Tổng quan/Khảo sát các phương pháp GNN.	Tổng quan các ứng dụng hình ảnh y tế.	Hệ thống hóa các phương pháp GNN.	Củng cố nền tảng lý thuyết và chỉ ra các hướng nghiên cứu tiềm năng.	Không đi sâu vào chi tiết kỹ thuật của một mô hình.
Raghu et al., 2024 [58]	Framework lai CNN-GCN.	Kaggle X-ray và CT scan.	Độ chính xác: 98.5%	Kết hợp sức mạnh trích xuất đặc trưng của CNN và khả năng học quan hệ của GCN.	Tính phức tạp của một framework hai giai đoạn.
Wang et al., 2024 [59]	GNN được dẫn dắt bởi ánh nhìn (Gaze-Guided GNN).	Bộ dữ liệu X-quang ngực có kèm dữ liệu ánh nhìn của bác sĩ.	Cải thiện độ chính xác và tính diễn giải.	Đột phá trong việc tích hợp tri thức chuyên gia (qua ánh nhìn) vào mô hình.	Yêu cầu bộ dữ liệu đặc thù (có kèm dữ liệu ánh nhìn), khó thu thập.
Sultana et al., 2024 [60]	Học sâu đồ thị đa phương thức (GNN + Gaze + Speech).	Bộ dữ liệu X-quang ngực đa phương thức.	Nâng cao hiệu suất và khả năng diễn giải.	Hướng đi tiên phong, kết hợp nhiều nguồn thông tin từ chuyên gia.	Độ phức tạp cao, yêu cầu bộ dữ liệu rất đặc thù và khó thu thập.

1.2.3.3 Học sâu ít mẫu (Few-Shot Learning)

Một trong những thách thức lớn nhất của các mô hình học sâu như CNN là yêu cầu cần có tập dữ liệu gán nhãn đủ lớn để bảo đảm quá trình huấn luyện hiệu quả. Trong thực tế, điều này khó khả thi đối với các bệnh lý hiếm gặp hoặc trong bối cảnh dữ liệu

y tế còn hạn chế. Để khắc phục vấn đề này, học sâu ít mẫu (Few-Shot Learning – FSL) được xem là một hướng tiếp cận tiềm năng, giúp giảm sự phụ thuộc vào dữ liệu quy mô lớn. Trong số các phương pháp gần đây. Mục tiêu của FSL là xây dựng các mô hình có khả năng học và nhận dạng các khái niệm mới chỉ từ một vài mẫu ví dụ, tương tự như khả năng học của con người. Hướng tiếp cận này đặc biệt có giá trị trong y tế, nhất là với các bài toán như chẩn đoán bệnh bụi phổi silic nghề nghiệp, nơi dữ liệu bệnh nhân cực kỳ khan hiếm [65].

Trong những năm trở lại đây, các kỹ thuật FSL đã được áp dụng phổ biến trong bài toán chẩn đoán X-quang phổi. Một trong những nhánh được ứng dụng mạnh mẽ nhất là Siêu học tập (Meta-Learning), hay còn gọi là "học cách học". Thay vì học để phân loại một tập hợp bệnh cố định, mô hình được huấn luyện qua một loạt các tác vụ phân loại nhỏ, giúp nó hình thành khả năng tổng quát hóa nhanh chóng đến các loại bệnh mới chỉ với vài ảnh X-quang [98]. Một nhánh phổ biến khác là Học theo cặp (Metric Learning), điển hình là sử dụng Mạng Siamese (Siamese Networks). Phương pháp này không học cách phân loại trực tiếp mà học một hàm đo lường sự tương đồng, cho phép so sánh một ảnh X-quang chưa biết với một vài ảnh mẫu đã biết để đưa ra chẩn đoán [63].

Bên cạnh đó, một số các phương pháp khác tiếp cận bằng cách kết hợp nhiều kỹ thuật/nhiều mô hình để tối ưu hóa hiệu suất. Ví dụ, một số công trình nghiên cứu đã sử dụng học tập hợp (Ensemble Learning) để tăng cường độ ổn định cũng như độ chính xác cho các mô hình FSL [61]. Một số phương pháp phức tạp hơn còn tích hợp thêm học tự giám sát (Self-Supervised Learning) và Siêu học tập để tận dụng tối đa lượng dữ liệu không nhãn có sẵn, trước khi tinh chỉnh trên một vài mẫu có nhãn [66]. Một trong những nghiên cứu đột phá hơn cả là các kiến trúc FSL không giám sát (Unsupervised FSL), vốn có tiềm năng chẩn đoán bệnh từ một vài mẫu mà không cần đến bất kỳ nhãn nào, giải quyết đồng thời cả hai bài toán lớn nhất là thiếu dữ liệu và thiếu nhãn [64].

Mặc dù FSL mở ra một hướng tiếp cận trong việc chẩn đoán các bệnh hiếm gặp, nó cũng có những hạn chế nhất định. Hiệu suất của các mô hình FSL thường nhạy cảm với chất lượng của một vài mẫu hỗ trợ và có thể không ổn định bằng các mô hình được huấn luyện trên dữ liệu lớn. Hơn nữa, các thuật toán FSL, đặc biệt là Siêu học tập, thường rất phức tạp để thiết kế và tối ưu hóa.

Bảng 1.4 Tổng hợp các nghiên cứu tiêu biểu sử dụng Học sâu ít mẫu (FSL) trên ảnh X-quang phổi.

Nghiên cứu	Phương pháp	Bộ dữ liệu	Kết quả	Ưu điểm	Hạn chế
Paul et al., 2021 [61]	Học tập hợp (Ensemble Learning) cho FSL.	ChestX-ray14 (NIH).	Cải thiện độ chính xác và ổn định.	Tăng cường độ ổn định và chính xác cho các mô hình FSL.	Tăng độ phức tạp tính toán do phải huấn luyện và kết hợp nhiều mô hình.
Chen et al., 2024 [62]	Siêu học tập (Meta-Learning).	Bộ dữ liệu X-quang ngực.	Xác định viêm phổi hiệu quả từ ít mẫu.	Mô hình "học cách học", có khả năng tổng quát hóa nhanh với các lớp bệnh mới.	Các thuật toán Meta-Learning thường phức tạp và khó tối ưu hóa.
Atukunda, 2024 [63]	Mạng Siamese + Học chuyển giao.	Bộ dữ liệu X-quang ngực.	Cải thiện phân loại viêm phổi từ ít mẫu.	Học một hàm đo lường sự tương đồng, mạnh mẽ hơn khi đối mặt với các lớp mới.	Hiệu suất phụ thuộc vào việc lựa chọn các cặp ảnh trong quá trình huấn luyện.
Kim et al., 2024 [64]	Học ít mẫu không giám sát (Unsupervised FSL).	Bộ dữ liệu X-quang ngực.	Chẩn đoán bệnh phổi mà không cần nhãn.	Giải quyết đồng thời vấn đề thiếu dữ liệu và thiếu nhãn gán.	Là hướng đi rất mới, hiệu suất có thể chưa bằng các phương pháp có giám sát.

Nghiên cứu	Phương pháp	Bộ dữ liệu	Kết quả	Ưu điểm	Hạn chế
Atukunda & Mwangi, 2025 [66]	Học tập hợp + Tự giám sát + Siêu học tập.	Bộ dữ liệu X-quang ngực.	Nâng cao hiệu suất phát hiện viêm phổi.	Kết hợp nhiều kỹ thuật tiên tiến để tận dụng tối đa dữ liệu có và không có nhãn.	Mô hình rất phức tạp, đòi hỏi tài nguyên tính toán lớn.

Các nghiên cứu về Few-Shot Learning (FSL) trong bảng chủ yếu tập trung vào bài toán phân loại và phát hiện viêm phổi trên ảnh X-quang ngực. Kết quả cho thấy khi số lượng mẫu huấn luyện bị giới hạn, độ chính xác của mô hình suy giảm đáng kể so với kịch bản học trên tập dữ liệu đầy đủ. Tuy vậy, với sự hỗ trợ của các kỹ thuật như Meta-Learning, Siamese Network hoặc học tự giám sát, mức suy giảm này được kiểm soát, giúp mô hình duy trì hiệu quả chẩn đoán ở mức chấp nhận được ngay cả trong điều kiện dữ liệu hạn chế.

1.3 Cơ sở lý thuyết

1.3.1 Một số mô hình học sâu tiêu biểu

1.3.1.1 Mô hình MobileNet

MobileNetV3 là một kiến trúc mạng nơ-ron tích chập (CNN) tiên tiến, được Google phát triển nhằm tối ưu hóa hiệu suất tính toán trên các thiết bị có tài nguyên hạn chế như thiết bị di động và hệ thống nhúng [107]. Phiên bản MobileNetV3 Small là biến thể nhẹ hơn so với MobileNetV3 Large, được thiết kế để cân bằng giữa độ chính xác và hiệu quả, đặc biệt phù hợp cho các bài toán phân loại ảnh y tế với dữ liệu có kích thước nhỏ và nhiều cao [108].

MobileNetV3 kế thừa các cải tiến từ các phiên bản trước (MobileNetV1 và V2), đồng thời tích hợp các kỹ thuật học sâu hiện đại, bao gồm: Depthwise Separable Convolution: giảm đáng kể số lượng tham số và chi phí tính toán. Hard-Swish Activation: hàm kích hoạt phi tuyến được tối ưu cho suy luận nhanh. Squeeze-and-Excitation (SE) Block: tăng cường khả năng học đặc trưng theo kênh. Neural Architecture Search (NAS): hỗ trợ tìm kiếm kiến trúc mạng tối ưu tự động.

MobileNetV3-Small có kiến trúc phân cấp, gồm các thành phần chính: Lớp đầu vào: ảnh chuẩn hóa kích thước $224 \times 224 \times 3$. Tầng tích chập ban đầu: kết hợp BatchNorm và Hard-Swish. Chuỗi khối Inverted Residual: sử dụng Depthwise Convolution, SE Block và Linear Bottleneck. Global Average Pooling: rút gọn đặc trưng không gian. Tầng fully connected: điều chỉnh để phù hợp với phân loại nhị phân. Tầng đầu ra: sử dụng sigmoid để dự đoán xác suất.

MobileNetV3-Small nổi bật với thiết kế gọn nhẹ, tốc độ suy luận nhanh và khả năng triển khai hiệu quả trên các thiết bị hạn chế về tài nguyên. Trong nghiên cứu này, mô hình được chọn làm kiến trúc học sâu tiêu chuẩn nhẹ để so sánh hiệu quả phân loại bệnh bụi phổi silic nghề nghiệp với các mô hình phức tạp hơn như DenseNet201 và hiện đại hơn như Swin Transformer.

1.3.1.2 Mô hình DenseNet201

DenseNet (Densely Connected Convolutional Network) là một kiến trúc mạng nơ-ron tích chập sâu, nổi bật với thiết kế kết nối dày đặc (dense connectivity), trong đó mỗi lớp đầu ra được truyền đến tất cả các lớp kế tiếp trong cùng một khối. Cơ chế này giúp tăng cường khả năng truyền thông tin, tái sử dụng đặc trưng hiệu quả, giảm thiểu số lượng tham số và khắc phục hiện tượng mất mát gradient trong các mạng sâu [106–109, 111]. DenseNet đã được phát triển thành nhiều biến thể như DenseNet-121, DenseNet-169, DenseNet-201 và DenseNet-264, với số lớp khác nhau để phù hợp với quy mô bài toán và tài nguyên tính toán [112–114].

Trong luận án, kiến trúc DenseNet-201 được chọn để phân loại ảnh X-quang ngực nhằm chẩn đoán bệnh bụi phổi silic nghề nghiệp, nhờ khả năng học các đặc trưng tinh tế từ các tổn thương mờ nhạt và không đồng đều. DenseNet-201 có khoảng 20 triệu tham số, với chi phí tính toán cao hơn MobileNetV3 Small, nhưng thấp hơn Swin Transformer, phù hợp để cân bằng hiệu suất và tài nguyên.

Kiến trúc DenseNet-201 bao gồm bốn Dense Block, với số lớp lần lượt là 6, 12, 48, và 32. Mỗi Dense Block gồm các bước: chuẩn hóa hàng loạt (Batch Normalization), hàm kích hoạt ReLU, và các lớp tích chập (1×1 và 3×3 xen kẽ). Giữa các Dense Block, một Transition Layer sử dụng tích chập 1×1 và gộp trung bình (Average Pooling, kích thước 2×2 , stride 2) để giảm kích thước đặc trưng và kiểm soát quá khớp. Sau Dense Block cuối, một lớp Global Average Pooling được áp dụng, kết nối với lớp phân loại sử dụng hàm Softmax để xuất kết quả phân loại (0: bình thường, 1: mắc bệnh).

Layer Label	Input Layer Dimension	Output Dimension
Input Layer		$244 \times 244 \times 3$
Convolution	112×112	Filter size = 7×7
		Strike = 2
		Padding = 3
pooling	56×56	Maximum Pooling = 3×3
		Strike = 2
Dense Block 1	56×56	$\begin{bmatrix} 1 & \times & 1 \\ 3 & \times & 3 \end{bmatrix} \times 6$
Transition Layer 1	56×56	1×1 convolution
	28×28	2×2 average pooling, stride = 2
Dense Block 2	28×28	$\begin{bmatrix} 1 & \times & 1 \\ 3 & \times & 3 \end{bmatrix} \times 12$
Transition Layer 2	28×28	1×1 convolution
	14×14	2×2 average pooling, stride = 2
Dense Block 3	14×14	$\begin{bmatrix} 1 & \times & 1 \\ 3 & \times & 3 \end{bmatrix} \times 48$
Transition Layer 3	14×14	1×1 convolution
	7×7	2×2 average pooling, stride = 2
Dense Block 4	7×7	$\begin{bmatrix} 1 & \times & 1 \\ 3 & \times & 3 \end{bmatrix} \times 32$
Pooling		Average Pooling = 7×7
		Strike = 7
		1×1
FC Layer		1000

Hình 1.3 Kiến trúc DenseNet201 [110]

1.3.1.3 Mô hình Swin Transformer

Swin Transformer [109] là một kiến trúc học sâu tiên tiến dựa trên Transformer, được tối ưu cho các bài toán thị giác máy tính, đặc biệt với dữ liệu ảnh độ phân giải cao như ảnh X-quang ngược. Khác với Vision Transformer (ViT) sử dụng self-attention toàn cục (độ phức tạp $O(n^2)$), Swin Transformer áp dụng Window-based Multi-head Self-Attention (W-MSA) và Shifted Window-based Multi-head Self-Attention (SW-MSA) trong các cửa sổ trượt, giảm độ phức tạp xuống $O(n)$ mà vẫn duy trì khả năng học ngữ cảnh hiệu quả.

Cơ chế chính:

W-MSA: Tính self-attention trong các cửa sổ cố định (kích thước 7×7), tập trung vào đặc trưng cục bộ và tiết kiệm tài nguyên tính toán.

SW-MSA: Dịch chuyển cửa sổ giữa các tầng để kết nối các vùng ảnh, mở rộng vùng tiếp nhận mà không cần attention toàn cục.

Cấu trúc phân cấp: Sử dụng Patch Merging để giảm kích thước không gian và tăng chiều biểu diễn, học đặc trưng từ chi tiết cục bộ đến tổng quát.

Chi tiết về vai trò và kích thước đầu ra của các thành phần trong mô hình Swin Transformer được trình bày cụ thể trong bảng 1.5. Sự kết hợp của các cơ chế này giúp mô hình Swin Transformer đạt được hiệu suất cao trong việc phân tích các đặc trưng tinh tế và phân bố không đồng nhất trên ảnh X-quang BPSi, đồng thời tối ưu hóa hiệu quả tính toán so với các kiến trúc Transformer toàn cục.

Ưu điểm chính của Swin Transformer so với ViT và CNN bao gồm:

(1) Cấu trúc phân cấp (Hierarchical Representation) Tương tự CNN, Swin Transformer sử dụng các tầng xử lý với kích thước không gian giảm dần (qua Patch Merging), trích xuất đặc trưng từ chi tiết cục bộ đến biểu diễn tổng quát toàn ảnh. Cấu trúc này rất phù hợp với ảnh X-quang, nơi tổn thương bụi phổi silic có thể khu trú (nốt nhỏ) hoặc lan tỏa (rải rác nhiều vùng phổi).

(2) Cơ chế self-attention cục bộ hiệu quả (Local Window-based Attention): Khác với ViT (sử dụng attention toàn cục, độ phức tạp $O(n^2)$), Swin Transformer tính attention trong các cửa sổ nhỏ (7×7), giảm độ phức tạp xuống $O(n)$. Cơ chế này giúp mô hình tập trung vào đặc trưng cục bộ, rất cần thiết để phát hiện các tổn thương mờ nhạt trên X-quang ngực của bệnh nhân mắc bệnh bụi phổi silic nghề nghiệp.

(3) Cơ chế dịch chuyển cửa sổ thông minh (Shifted Window Mechanism): Cửa sổ được dịch chuyển giữa các tầng để kết nối các vùng ảnh liên kề, mở rộng trường tiếp nhận (receptive field) mà không tăng chi phí tính toán. Điều này đặc biệt quan trọng trong các tổn thương lan tỏa, không tập trung của bệnh bụi phổi silic, đảm bảo học được ngữ cảnh liên vùng.

(4) Khả năng tổng quát hóa tốt: Nhờ kết hợp attention cục bộ và cấu trúc phân cấp, Swin Transformer đạt hiệu quả cao trên tập dữ liệu y tế cỡ vừa, phù hợp với thực tế y học nơi dữ liệu gán nhãn thường hạn chế và chi phí thu thập cao.

Kiến trúc Swin Transformer được trình bày chi tiết trên đây không chỉ tối ưu hóa khả năng trích xuất đặc trưng hiệu quả từ dữ liệu hình ảnh X-quang, mà còn cho thấy khả năng cân bằng tốt giữa độ phức tạp tính toán và hiệu suất dự đoán. Chính cơ chế

Bảng 1.5 Vai trò và kích thước đầu ra của các thành phần trong mô hình Swin Transformer

Bước	Thành phần	Vai trò chính	Kích thước đầu ra
1.	Input Tensor	Ảnh X-quang đầu vào	(32, 3, 224, 224)
2.	Conv2D + Permute	Trích xuất đặc trưng ban đầu, chia patch	(32, 128, 56, 56)
3.	LayerNorm	Chuẩn hóa đầu vào trước attention	(32, 128, 56, 56)
4.	Swin Block $\times 2$ (Stage 1)	Học đặc trưng cục bộ qua attention cửa sổ	(32, 56, 56, 128)
5.	Patch Merging 1	Giảm kích thước không gian, tăng chiều đặc trưng	(32, 28, 28, 256)
6.	Swin Block $\times 2$ (Stage 2)	Học đặc trưng sâu hơn trong không gian nhỏ hơn	(32, 28, 28, 256)
7.	Patch Merging 2	Tiếp tục giảm kích thước, tăng chiều đặc trưng	(32, 14, 14, 512)
8.	Swin Block $\times 18$ (Stage 3)	Học biểu diễn phức tạp, trừu tượng	(32, 14, 14, 512)
9.	Patch Merging 3	Giảm về kích thước nhỏ nhất, tăng chiều sâu	(32, 7, 7, 1024)
10.	Swin Block $\times 2$ (Stage 4)	Biểu diễn sâu nhất trước phân loại	(32, 7, 7, 1024)
11.	LayerNorm + Permute	Chuẩn hóa và sắp xếp lại thứ tự chiều tensor	(32, 1024, 7, 7)
12.	AdaptiveAvgPool2d	Tổng hợp đặc trưng toàn ảnh	(32, 1024, 1, 1)
13.	Flatten + Dense	Làm phẳng và phân loại nhị phân	(32, 2)
14.	Output	Dự đoán xác suất hai lớp	(32, 2)

Self-Attention cục bộ kết hợp với các kỹ thuật thu gọn kích thước không gian và tăng chiều đặc trưng đã giúp mô hình đạt hiệu suất tốt hơn, phù hợp để ứng dụng vào bài toán chẩn đoán bệnh bụi phổi silic nghề nghiệp.

1.3.2 Học sâu trên đồ thị

1.3.2.1 Tổng quan về dữ liệu dạng đồ thị

Các mô hình học sâu truyền thống như mạng neuron tích chập (CNN) thường dựa trên dữ liệu đầu vào có cấu trúc lưới đều đặn như ảnh (2D) hoặc chuỗi (1D) [111]. Tuy nhiên, những cấu trúc này lại không phù hợp để biểu diễn nhiều loại dữ liệu quan trọng trong thực tế, đặc biệt là dữ liệu trong y học, sinh học và mạng xã hội vốn mang đặc tính phi tuyến, phức tạp, và không thể dễ dàng mô tả bằng cấu trúc lưới thông thường [112]. Sự hạn chế này đặt ra yêu cầu cấp thiết phải phát triển các kỹ thuật biểu diễn dữ liệu hiệu quả hơn, trong đó dữ liệu dạng đồ thị nổi bật như một phương pháp mạnh mẽ, có khả năng phản ánh đầy đủ các mối quan hệ phức tạp giữa các thực thể [113].

Một đồ thị G được định nghĩa gồm tập các nút V (vertices hoặc nodes), tập cạnh E (edges) giữa các nút đó. Một đồ thị $G = (V, E)$ được định nghĩa bởi tập các nút V và một tập các cạnh E giữa các nút đó. Một cạnh đi từ nút $u \in V$ đến nút $v \in V$ được ký hiệu là $(u, v) \in E$.

Trong nhiều trường hợp, đối tượng nghiên cứu thường là đồ thị đơn vô hướng, tức là cạnh giữa hai nút không có hướng. Khi đó, $(u, v) \in E \Leftrightarrow (v, u) \in E$.

Cách thức biểu diễn đồ thị là thông qua một ma trận kề (adjacency matrix) $A \in \mathbb{R}^{|V| \times |V|}$. Để biểu diễn ma trận kề, người ta sắp xếp các nút trong đồ thị theo thứ tự hàng và cột, các cạnh được biểu diễn như các thực thể trong ma trận đó: $A[u, v] = 1$ nếu có cạnh giữa u, v và ngược lại thì $A[u, v] = 0$.

1.3.2.2 Mạng nơ-ron đồ thị Graph Neural Networks

Mạng nơ-ron đồ thị GNN là lớp mô hình học sâu có khả năng xử lý thông tin trên các đồ thị phi cấu trúc. Điểm nổi bật của GNN là quá trình truyền thông tin giữa các nút thông qua các kết nối trên đồ thị [114]. Phương pháp truyền tin này thường gọi là quá trình Message Passing và thường bao gồm hai bước chính:

Aggregation: Tổng hợp thông tin từ các nút lân cận.

Update: Cập nhật vector đặc trưng của nút dựa trên thông tin tổng hợp từ bước trước.

Trong mỗi lần lặp truyền thông điệp trong GNN, một đặc trưng ẩn $h_u^{(k)}$ tương ứng của mỗi nút $u \in V$ được cập nhật theo thông tin được kết tập từ các nút hàng xóm

$v \in N(u)$. Các thông điệp được cập nhật theo công thức (1.1):

$$\begin{aligned} h_u^{(l+1)} &= \text{Update}^l(h_u^{(l)}, \text{Aggregate}^{(l)}(\{h_v^{(l)}, \forall v \in N(u)\})) \\ &= \text{Update}^l(h_u^{(l)}, m_{N(u)}^{(l)}). \end{aligned} \quad (1.1)$$

Tại lớp thứ k của GNN, hàm Aggregate tổng hợp các đầu vào của nút u và sinh ra thông điệp $m_{N(u)}^{(l)}$ dựa trên các thông tin hàng xóm được kết tập của nó.

Hàm Update kết hợp thông điệp $m_{N(u)}^{(l)}$ với đặc trưng biểu diễn trước đó của nút u $h_u^{(l-1)}$ để sinh ra vec-tơ đặc trưng $h_u^{(l)}$.

Khởi tạo ban đầu $l = 0$, để thiết lập các đặc trưng đầu vào cho tất cả các nút (ví dụ: $h_u^{(0)} = x_u, \forall u \in V$).

Sau khi l lớp GNN truyền thông điệp, ta có đầu ra của lớp cuối cùng như là các embedding cho mỗi nút (ví dụ: $z_u = h_u^{(l)}, \forall u \in V$).

1.3.2.3 Một số biến thể của GNN

Kể từ khi được giới thiệu, mạng nơ-ron đồ thị đã không ngừng được cải tiến và mở rộng với nhiều kiến trúc khác nhau nhằm tăng cường khả năng biểu diễn, hiệu quả tính toán và khả năng khái quát hóa trên các dạng đồ thị đa dạng. Mỗi biến thể được phát triển đều thể hiện một cách tiếp cận riêng biệt trong việc thiết kế hàm lan truyền thông tin, tổng hợp đặc trưng từ lân cận, hoặc gán trọng số cho các mối liên kết, từ đó mở ra nhiều hướng ứng dụng thực tiễn. Dưới đây là một số kiến GNN tiêu biểu.

Graph Convolutional Networks (GCN): Được đề xuất bởi Kipf và Welling [111], GCN mở rộng khái niệm tích chập từ dữ liệu lưới sang dữ liệu đồ thị, bằng cách sử dụng phép chuẩn hóa ma trận kề để lan truyền thông tin qua các đỉnh lân cận.

Trong mô hình GCN, mỗi lớp thực hiện phép tích chập trên đồ thị bằng cách tổng hợp đặc trưng từ các nút lân cận, đồng thời chuẩn hóa theo bậc của các nút để đảm bảo sự ổn định trong huấn luyện. Công thức cập nhật đặc trưng cho một nút u tại lớp $l + 1$ như sau:

$$\mathbf{h}_u^{(l+1)} = \sigma \left(\sum_{v \in N(u) \cup \{u\}} \frac{\tilde{\mathbf{A}}_{uv}}{\sqrt{\tilde{\mathbf{D}}_{uu} \tilde{\mathbf{D}}_{vv}}} \mathbf{W}^{(l)} \mathbf{h}_v^{(l)} \right), \quad (1.2)$$

trong đó:

- $\mathbf{h}_u^{(l)}$ là vector đặc trưng của nút u tại lớp l ;
- $\mathbf{W}^{(l)}$ là ma trận trọng số cần học tại lớp l ;
- $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ là ma trận kề đã được bổ sung self-loop;

- $\tilde{\mathbf{D}}$ là ma trận bậc tương ứng với $\tilde{\mathbf{A}}$;
- $\sigma(\cdot)$ là hàm kích hoạt phi tuyến, thường là ReLU;
- $N(u)$ là tập các nút lân cận của u .

E là tập các cạnh được biểu diễn bởi hai ma trận, ma trận thuộc tính các đỉnh (X), ma trận kề biểu diễn kết nối các đỉnh (A). $X \in \mathbb{R}^{N \times F}$ thể hiện cho N nút, mỗi nút được biểu diễn bởi vector F chiều, $A \in \mathbb{R}^{N \times N}$ là ma trận vuông biểu diễn kết nối có hay không giữa hai nút.

Graph Attention Networks (GAT): Velickovic và cộng sự [115] đưa ra cơ chế attention trong GNN, cho phép mô hình học trọng số khác nhau cho từng đỉnh lân cận khi tổng hợp thông tin.

GraphSAGE: Hamilton và cộng sự [116] đề xuất GraphSAGE – một kỹ thuật học biểu diễn theo hướng suy diễn (inductive), có khả năng tạo đặc trưng cho các đỉnh chưa từng xuất hiện trong quá trình huấn luyện.

Graph Isomorphism Network (GIN): Xuất phát từ nhu cầu tăng khả năng phân biệt cấu trúc đồ thị, GIN [117] được chứng minh là có năng lực biểu diễn mạnh tương đương với kiểm định đẳng cấu đồ thị Weisfeiler-Lehman.

1.3.2.4 Graph Transformer và cơ chế Attention

Cơ chế chú ý (Attention) là một cơ chế ánh xạ một truy vấn (Query) và một tập các cặp khóa-giá trị (Key-Value) tới một đầu ra. Đầu ra được tính toán như một tổng có trọng số của các giá trị, trong đó trọng số của mỗi giá trị được xác định bởi độ tương thích giữa truy vấn và khóa tương ứng.

Kiến trúc Transformer sử dụng một cơ chế attention đặc biệt và hiệu quả, được xây dựng từ hai thành phần chính: Scaled Dot-Product Attention và Multi-Head Attention.

Scaled Dot-Product Attention

Cơ chế này tính toán attention dựa trên các ma trận đầu vào: truy vấn $\mathbf{Q}$, khóa $\mathbf{K}$ (có số chiều cột là d_k), và giá trị $\mathbf{V}$. Công thức tính toán được biểu diễn như sau:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}} \right) \mathbf{V} \quad (1.3)$$

Trong công thức này, ma trận $\mathbf{Q}\mathbf{K}^T$ chứa các điểm số tương thích giữa các truy vấn và các khóa. Việc chia cho $\sqrt{d_k}$ là một bước chia tỷ lệ (scaling) quan trọng giúp ổn định gradient trong quá trình huấn luyện.

Multi-Head Attention

Thay vì thực hiện một phép chú ý duy nhất, cơ chế chú ý đa đầu (Multi-Head Attention) cho phép mô hình cùng lúc chú ý đến thông tin từ các không gian con biểu diễn khác nhau. Cơ chế này thực hiện h phép chú ý khác nhau một cách song song. Đầu tiên, các ma trận Q, K, V được biến đổi tuyến tính h lần bằng các ma trận chiếu (projection matrices) riêng biệt:

$$\text{head}_i = \text{Attention}(\mathbf{QW}_i^Q, \mathbf{KW}_i^K, \mathbf{VW}_i^V) \quad (1.4)$$

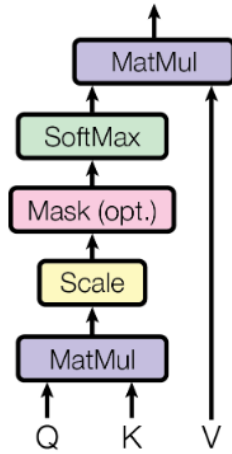
với các ma trận chiếu là các tham số có thể học được: $\mathbf{W}_i^Q, \mathbf{W}_i^K \in \mathbb{R}^{d_{\text{model}} \times d_k}$ và $\mathbf{W}_i^V \in \mathbb{R}^{d_{\text{model}} \times d_v}$.

Kết quả đầu ra từ h đầu chú ý sau đó được ghép nối (concatenated) và một lần nữa được biến đổi tuyến tính bằng một ma trận chiếu đầu ra $\mathbf{W}^O \in \mathbb{R}^{hd_v \times d_{\text{model}}}$:

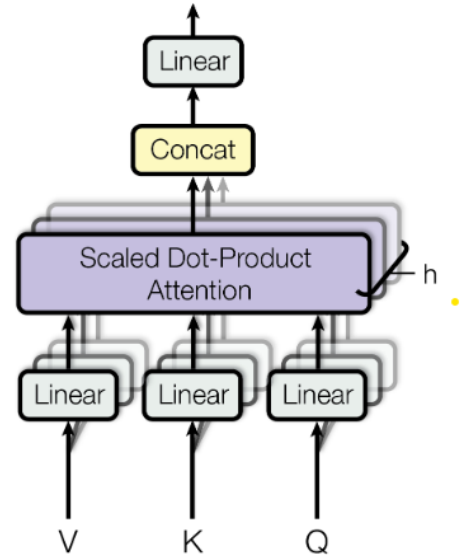
$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \mathbf{W}^O \quad (1.5)$$

Trong kiến trúc Transformer, cơ chế này được áp dụng theo dạng tự chú ý (self-attention), trong đó Q, K, V đều được tính toán từ cùng một nguồn là đầu ra của lớp trước đó.

Scaled Dot-Product Attention



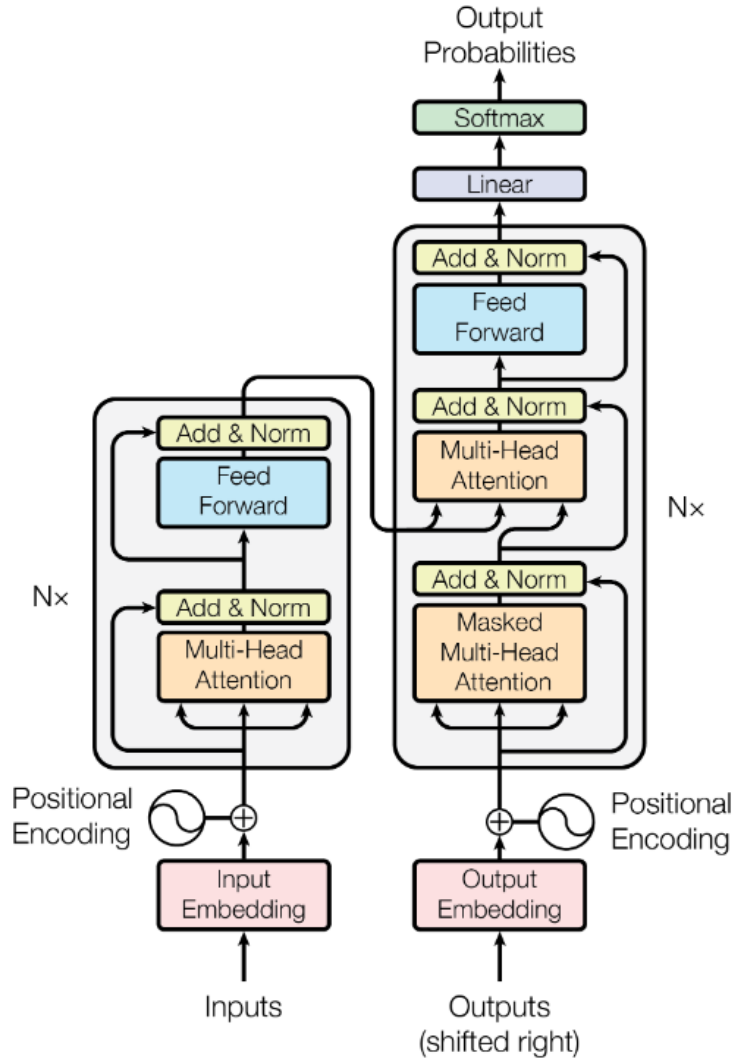
Multi-Head Attention



Hình 1.4 Kiến trúc Attention: Scaled Dot-Product và Multi-Head Attention [118]

Tích hợp cấu trúc đồ thị trong Graph Transformer

Một thách thức lớn khi áp dụng Transformer lên đồ thị là kiến trúc này vốn "mù" về cấu trúc. Để giải quyết vấn đề này, các mô hình như Graphormer đã đề xuất các kỹ



Hình 1.5 Minh họa cơ chế của một lớp Graph Transformer, trong đó Attention được điều chỉnh theo cấu trúc đồ thị. [118]

thuật hiệu quả để "tiêm" thông tin cấu trúc vào cơ chế tự chú ý. Các kỹ thuật này đóng vai trò như những thiên kiến quy nạp (inductive biases), giúp mô hình học được các biểu diễn đồ thị tốt hơn.

Mã hóa độ trung tâm (Centrality Encoding)

Để giúp mô hình nhận biết tầm quan trọng của các đỉnh, Graphormer mã hóa độ bậc (degree) của mỗi đỉnh. Các embedding có thể học được, tương ứng với bậc vào và bậc ra của đỉnh, được cộng trực tiếp vào đặc trưng đầu vào. Công thức khởi tạo biểu diễn của đỉnh v_i được sửa đổi như sau:

$$\mathbf{h}_i^{(0)} = \mathbf{x}_i + \mathbf{z}_{\deg^-(v_i)}^- + \mathbf{z}_{\deg^+(v_i)}^+ \quad (1.6)$$

trong đó $\mathbf{x}_i$ là đặc trưng ban đầu, $\mathbf{z}^-$ và $\mathbf{z}^+$ là các vector embedding có thể học được, được xác định bởi bậc vào ($\deg^-(v_i)$) và bậc ra ($\deg^+(v_i)$).

Mã hóa cấu trúc trong Attention

Graphormer đề xuất sửa đổi trực tiếp quá trình tính toán điểm số chú ý để tích hợp thông tin về quan hệ không gian và đặc trưng cạnh giữa các đỉnh. Công thức tính toán phần tử (i, j) trong ma trận chú ý được mở rộng như sau:

$$\mathbf{A}_{ij} = \frac{(\mathbf{h}_i \mathbf{W}_Q)(\mathbf{h}_j \mathbf{W}_K)^T}{\sqrt{d}} + \mathbf{b}_{\phi(\mathbf{v}_i, \mathbf{v}_j)} + c_{ij} \quad (1.7)$$

trong đó $\mathbf{A}_{ij}$ là điểm số chú ý trước khi qua hàm softmax. Hai thành phần mới được thêm vào bao gồm:

- **Mã hóa không gian (Spatial Encoding):** $\mathbf{b}_{\phi(\mathbf{v}_i, \mathbf{v}_j)}$ là một hệ số điều chỉnh (bias) vô hướng có thể học được, phụ thuộc vào quan hệ không gian $\phi(\mathbf{v}_i, \mathbf{v}_j)$ giữa hai đỉnh (ví dụ, khoảng cách đường đi ngắn nhất - SPD). Thành phần này giúp mô hình nhận biết được các đỉnh "gần" hay "xa" nhau về mặt cấu trúc.
- **Mã hóa cạnh (Edge Encoding):** c_{ij} là một hệ số điều chỉnh khác, được tính bằng cách lấy trung bình các embedding có thể học được của các đặc trưng cạnh trên đường đi ngắn nhất giữa $\mathbf{v}_i$ và $\mathbf{v}_j$. Thành phần này cho phép mô hình tận dụng thông tin từ các loại liên kết khác nhau.

Bằng cách này, ma trận chú ý không chỉ phụ thuộc vào sự tương đồng về đặc trưng mà còn được điều chỉnh bởi cấu trúc liên kết và thuộc tính của đồ thị.

1.4 Các chỉ số đánh giá hiệu năng

Ma trận nhầm lẫn là một công cụ giá trị trong việc đánh giá hiệu quả của mô hình ở các tác vụ phân loại, mang lại góc nhìn toàn diện về hiệu suất trên từng lớp cụ thể. Trong y học, việc đánh giá chính xác độ chính xác chẩn đoán là rất quan trọng; các chỉ số được rút ra từ ma trận nhầm lẫn cho phép đo lường chi tiết và có ý nghĩa về hiệu năng của mô hình.

True Positive (TP): trường hợp mắc bệnh được dự đoán đúng là bệnh.

True Negative (TN): trường hợp không mắc bệnh được dự đoán đúng là không bệnh.

False Positive (FP): trường hợp không mắc bệnh nhưng bị dự đoán nhầm thành bệnh (báo động giả).

False Negative (FN): trường hợp mắc bệnh bị phổi silic nhưng bị dự đoán nhầm là không bệnh (bỏ sót bệnh).

Từ các thông số này, một số chỉ số quan trọng được sử dụng trong luận án gồm:

(a) Độ chính xác (Accuracy)

Độ chính xác cho biết tỷ lệ các dự đoán đúng trên tổng số dự đoán

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1.8)$$

Độ chính xác cung cấp cái nhìn tổng thể về hiệu suất của mô hình nhưng có thể bị ảnh hưởng bởi sự mất cân đối giữa các lớp (khi số lượng mẫu của các lớp không tương đương).

(b) Độ chuẩn (Precision - PRE)

Đo lường độ tin cậy của các dự đoán dương tính, tức là trong số tất cả các ca được dự đoán là bệnh, có bao nhiêu ca thực sự mắc bệnh.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1.9)$$

Precision cao có nghĩa là hầu hết bệnh nhân được mô hình dự đoán là mắc bệnh thực sự bị bệnh, giúp giảm tình trạng báo động giả.

(c) Độ nhạy (Sensitivity)

Đo lường khả năng phát hiện đúng các ca mắc bệnh, tức là trong số tất cả ca bệnh thực sự, có bao nhiêu ca được dự đoán đúng.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (1.10)$$

Đây là chỉ số quan trọng trong y tế. Độ nhạy cao cho thấy mô hình có khả năng phát hiện tốt các ca bệnh. Độ nhạy thấp nghĩa là mô hình bỏ sót nhiều bệnh nhân (FN lớn).

(d) F1-score

Trung bình điều hòa giữa Precision và Recall.

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1.11)$$

Trong bài toán hỗ trợ chẩn đoán bụi phổi silic, cần đồng thời giảm thiểu báo động giả (FP) và bỏ sót bệnh (FN). Do vậy, F1 -score là chỉ số cân bằng và đáng tin cậy hơn Accuracy khi dữ liệu không cân bằng.

(e) Macro-F1

$$\text{Macro} - F1 = \frac{1}{4} \sum_{i=1}^4 \frac{2x R_i x P_i}{R_i + P_i} \quad (1.12)$$

trong đó P_i và R_i lần lượt là độ chính xác và độ hồi phục của lớp i trong bài toán phân loại đa lớp.

(f) Độ đặc hiệu (Specificity)

Đo lường tỷ lệ các trường hợp không mắc bệnh được mô hình phân loại đúng (true negative rate).

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (1.13)$$

Chỉ số này có ý nghĩa trong việc hạn chế tình trạng chẩn đoán nhầm người khỏe mạnh thành mắc bệnh, giúp giảm chi phí và lo lắng không cần thiết cho bệnh nhân.

(g) AUC (Area Under the ROC Curve)

AUC-ROC (Area Under the Receiver Operating Characteristic curve) là diện tích dưới đường cong ROC, được sử dụng để đánh giá khả năng phân biệt giữa các lớp của mô hình. Đường cong ROC được vẽ bằng cách biểu diễn tỷ lệ dương tính thật (True Positive Rate - TPR) theo tỷ lệ dương tính giả (False Positive Rate - FPR) ở các ngưỡng phân loại khác nhau.

$$AUC = \int_0^1 TPR(FPR^{-1}(x)) dx. \quad (1.14)$$

Trong đó,

$$TPR = \frac{TP}{TP + FN} \text{ và } FPR = \frac{FP}{FP + TN}.$$

Trong nghiên cứu này, AUC được sử dụng để so sánh trực tiếp hiệu năng giữa các mô hình học máy khác nhau trong chẩn đoán bệnh bụi phổi silic nghề nghiệp.

(h) Grad-CAM (Gradient-weighted Class Activation Mapping)

Grad-CAM không chỉ là chỉ số định lượng, mà là công cụ trực quan hóa được sử dụng rộng rãi để phân tích và diễn giải mô hình học sâu. Phương pháp này dựa trên việc tính toán gradient của lớp cuối cùng theo từng đặc trưng trong bản đồ kích hoạt (feature map), từ đó làm nổi bật các vùng hình ảnh mà mô hình tập trung khi đưa ra dự đoán [120]. Trong bối cảnh chẩn đoán y khoa, Grad-CAM đặc biệt hữu ích vì cho phép

xác định trực quan vùng phổi hoặc khu vực tổn thương mà mô hình xem là quan trọng, qua đó giúp bác sĩ kiểm chứng xem mô hình có thực sự dựa vào những đặc điểm lâm sàng hợp lý hay không. Đây chính là yếu tố quan trọng trong việc nâng cao tính giải thích (interpretability) và khả năng tin cậy (trustworthiness) của mô hình.

Trong luận án, Grad-CAM được sử dụng để trực quan hóa các vùng phổi mà mô hình quan tâm khi dự đoán. Nhờ đó, có thể kiểm chứng mức độ hợp lý của kết quả và so sánh trực quan giữa các mô hình khác nhau.

1.5 Những hạn chế của các nghiên cứu đã có và mục tiêu nghiên cứu

Trên cơ sở tổng hợp các công trình liên quan, NCS nhận thấy rằng nhiều nghiên cứu hiện nay đã đạt được những tiến bộ nhất định trong việc áp dụng học máy và học sâu vào các bệnh lý phổ biến như viêm phổi, ung thư phổi, lao phổi hay bệnh phổi tắc nghẽn mãn tính (COPD). Tuy nhiên, bệnh bụi phổi silic nghề nghiệp dù là một trong những bệnh nghề nghiệp nguy hiểm, tiến triển âm thầm và có tỷ lệ mắc đáng kể trong một số nhóm ngành lại chưa được nghiên cứu sâu rộng, đặc biệt là dưới góc nhìn của các mô hình học sâu hiện đại. Sự thiếu vắng này không chỉ tạo ra một khoảng trống trong lĩnh vực trí tuệ nhân tạo ứng dụng trong y học nghề nghiệp, mà còn cho thấy nhu cầu cấp thiết phải phát triển các mô hình chẩn đoán chuyên biệt, tin cậy và phù hợp với thực tiễn lâm sàng.

Từ thực tiễn đó, luận án nhận định rằng vẫn còn tồn tại một số hạn chế trong lĩnh vực nghiên cứu này, cụ thể như sau:

Thứ nhất, thiếu dữ liệu chuyên biệt và gán nhãn đầy đủ: Hầu hết các nghiên cứu sử dụng ảnh X-quang tổng quát với các bệnh phổ biến như viêm phổi, COVID-19, lao phổi... , trong khi dữ liệu về bụi phổi silic nghề nghiệp, đặc biệt theo giai đoạn bệnh còn hạn chế và chưa được chuẩn hóa.

Thứ hai, hạn chế về phương pháp: Nhiều phương pháp hiện tại vẫn dựa trên kỹ thuật trích xuất đặc trưng thủ công, vốn khó phát hiện được các dấu hiệu tổn thương mờ nhạt và lan tỏa của bệnh bụi phổi silic nghề nghiệp. Trong khi đó, các mô hình học sâu có ưu thế trong việc học đặc trưng tự động từ dữ liệu hình ảnh. Tuy nhiên, phần lớn các nghiên cứu hiện tại vẫn chưa giải quyết tốt bài toán giải thích kết quả, khiến việc kiểm chứng và thuyết phục bác sĩ trong ứng dụng lâm sàng còn nhiều thách thức.

Thứ ba, hạn chế về khả năng tổng quát hóa: Sự không đồng nhất giữa các nguồn dữ liệu, kết hợp với hiện tượng mất cân bằng và số lượng mẫu bệnh hạn chế, khiến mô hình dễ bị quá khớp với tập huấn luyện và giảm hiệu suất khi áp dụng trên các tập dữ liệu mới. Ngoài ra, do đặc điểm hình ảnh của bệnh bụi phổi silic nghề nghiệp dễ bị

nhầm lẫn với các bệnh phổi khác, mô hình càng cần có khả năng thích nghi cao để đảm bảo độ chính xác trong các tình huống thực tế đa dạng.

Từ những phân tích trên, luận án xác định mục tiêu nghiên cứu tổng quát như sau:

Nghiên cứu, phát triển và đánh giá các mô hình học máy, học sâu tiên tiến nhằm nâng cao hiệu quả phát hiện và phân loại bệnh bụi phổi silic nghề nghiệp từ hình ảnh X-quang ngực, đặc biệt trong điều kiện dữ liệu hạn chế và mất cân bằng nghiêm trọng.

1.6 Những hướng nghiên cứu chính của luận án

Xuất phát từ những bất cập đã phân tích ở Mục 1.5, luận án xác định mục tiêu tổng quát là nghiên cứu, phát triển và đánh giá các mô hình học máy, học sâu tiên tiến nhằm nâng cao hiệu quả phát hiện và phân loại bệnh bụi phổi silic nghề nghiệp từ hình ảnh X-quang ngực, đặc biệt trong điều kiện dữ liệu khan hiếm và mất cân bằng nghiêm trọng. Để hiện thực hóa mục tiêu này, luận án tập trung triển khai các hướng nghiên cứu chính sau:

(1) Xây dựng một tập dữ liệu ảnh X-quang ngực phục vụ nghiên cứu hỗ trợ chẩn đoán bệnh bụi phổi silic nghề nghiệp, có gán nhãn đáng tin cậy và tích hợp thông tin lâm sàng liên quan.

(2) Phát triển kiến trúc mô hình học sâu tích hợp với mạng nơ-ron đồ thị (Graph Neural Networks): Đề xuất và hiện thực hóa mô hình Graph Transformer Post-hoc, trong đó tích hợp Graph Transformer Plugin (GTP) để khai thác mối quan hệ ngữ nghĩa giữa các đặc trưng hình ảnh trong batch dữ liệu. Hướng tiếp cận này nhằm vượt qua giới hạn của các mô hình học sâu truyền thống vốn xử lý từng ảnh một cách độc lập và tăng cường khả năng học biểu diễn theo ngữ cảnh, giúp mô hình nhận diện chính xác hơn các biểu hiện mờ nhạt của bệnh bụi phổi silic.

(3) Ứng dụng các chiến lược học thích nghi và linh hoạt: Khai thác các kỹ thuật như Few-Shot Learning để học hiệu quả trong bối cảnh số lượng mẫu bệnh hạn chế, cũng như áp dụng kỹ thuật học tổng hợp nhằm tăng cường độ ổn định và khả năng tổng quát hóa của mô hình trên các tập dữ liệu khác nhau. Các chiến lược này đóng vai trò then chốt trong việc đối phó với mất cân bằng dữ liệu và nâng cao hiệu năng mô hình trong môi trường triển khai thực tế.

(4) Đánh giá toàn diện hiệu quả mô hình đề xuất: Thực hiện đánh giá hiệu năng mô hình theo nhiều góc độ: (i) định lượng qua các chỉ số như accuracy, recall, precision, F1-score; (ii) trực quan hóa sự chú ý của mô hình bằng kỹ thuật Grad-CAM để xác định vùng tổn thương được mô hình tập trung; và (iii) so sánh kết quả chẩn đoán của mô hình với đánh giá của bác sĩ chuyên khoa nhằm kiểm chứng tính khả thi, độ tin cậy và giá trị

ứng dụng lâm sàng của mô hình.

Thông qua các hướng nghiên cứu trên, luận án không chỉ hướng đến việc giải quyết các thách thức kỹ thuật cụ thể trong bài toán chẩn đoán bụi phổi silic từ ảnh X-quang, mà còn góp phần thúc đẩy ứng dụng thực tiễn của trí tuệ nhân tạo trong lĩnh vực y học nghề nghiệp.

1.7 Kết luận chương 1

Chương 1 đã trình bày tổng quan về bối cảnh nghiên cứu, tính cấp thiết, mục tiêu, phạm vi và phương pháp tiếp cận của luận án trong việc ứng dụng trí tuệ nhân tạo vào chẩn đoán bệnh bụi phổi silic nghề nghiệp. Thông qua phân tích có hệ thống các nghiên cứu trong và ngoài nước, luận án đã chỉ ra rằng dù các mô hình học máy và học sâu đã đạt được những kết quả đáng ghi nhận trong chẩn đoán các bệnh phổi phổ biến, nhưng việc áp dụng vào bệnh bụi phổi silic vẫn còn rất hạn chế, đặc biệt là từ góc nhìn của học sâu hiện đại.

Từ đó, luận án xác định rõ các hướng nghiên cứu cần tập trung giải quyết, bao gồm: hạn chế về dữ liệu (không đồng đều, mất cân bằng, thiếu chuẩn hóa), hạn chế về phương pháp (thiếu cơ chế học ngữ cảnh, khó diễn giải), và hạn chế về khả năng tổng quát hóa mô hình. Trên cơ sở đó, luận án đã đề xuất mục tiêu nghiên cứu tổng quát và bốn hướng nghiên cứu chính, bao gồm: (i) xây dựng và chuẩn hóa bộ dữ liệu chuyên biệt, (ii) phát triển mô hình học sâu tích hợp mạng nơ-ron đồ thị, (iii) ứng dụng các chiến lược học thích nghi và tổng hợp, và (iv) đánh giá toàn diện hiệu quả mô hình cả về định lượng và tính khả thi lâm sàng.

Những nội dung nêu trên là cơ sở định hướng cho các chương tiếp theo của luận án và được NCS công bố trong các công trình CT1, CT4, CT7. Trong chương 2, luận án sẽ trình bày cơ sở lý thuyết liên quan đến bệnh bụi phổi silic, phương pháp chẩn đoán trên ảnh X-quang ngực, cũng như các mô hình học sâu hiện đại có liên quan, làm nền tảng cho các giải pháp đề xuất trong các chương sau.

CHƯƠNG 2: XÂY DỰNG TẬP DỮ LIỆU

Chẩn đoán bệnh bụi phổi silic dựa trên ảnh X-quang ngực là một bài toán nhiều thách thức, không chỉ do các dấu hiệu tổn thương thường mờ nhạt, bệnh tiến triển âm thầm, mà còn do biểu hiện lâm sàng dễ nhầm lẫn với các bệnh phổi khác như lao, viêm phổi, COPD hay ung thư phổi.

Trong phạm vi nghiên cứu, ảnh X-quang ngực được lựa chọn làm nguồn dữ liệu chính để xây dựng và huấn luyện mô hình. Việc lựa chọn này xuất phát từ cả cơ sở y học lâm sàng và cơ sở khoa học kỹ thuật. Ảnh X-quang ngực là phương tiện chẩn đoán tiêu chuẩn được Tổ chức Lao động Quốc tế (ILO, 2011) khuyến nghị trong phân loại và đánh giá bệnh bụi phổi silic nghề nghiệp. Điều này khẳng định vai trò của ảnh X-quang ngực trong phát hiện, phân loại và theo dõi tiến triển bệnh bụi phổi silic nghề nghiệp. So với các phương pháp chẩn đoán hiện đại hơn như CT, ảnh X-quang phổ biến cao, chi phí thu nhận thấp và dễ triển khai tại các cơ sở y tế cơ sở, giúp mô hình đề xuất có khả năng ứng dụng thực tế và nhân rộng trong công tác sàng lọc bệnh nghề nghiệp.

Để hỗ trợ chẩn đoán bệnh bụi phổi silic nghề nghiệp đòi hỏi hai yếu tố: (1) bộ dữ liệu huấn luyện phải đầy đủ, chính xác và phản ánh được tính đa dạng trong thực tế lâm sàng; (2) mô hình học sâu cần được thiết kế phù hợp với đặc điểm cấu trúc của ảnh X-quang và sự phân bố không đồng đều giữa các lớp bệnh.

Trong chương này, luận án trình bày quá trình khảo sát, thu thập, xử lý, và kiểm định dữ liệu ảnh X-quang ngực phục vụ nghiên cứu. Đây là tập dữ liệu chuyên biệt cho bài toán chẩn đoán bệnh bụi phổi silic nghề nghiệp, được xây dựng thông qua sự phối hợp chặt chẽ giữa nghiên cứu sinh và các chuyên gia y tế trong nước.

2.1 Khảo sát về các tập dữ liệu

2.1.1 Tập dữ liệu quốc tế

ChestX-ray14: Bộ dữ liệu công khai có quy mô lớn, gồm 112.120 ảnh X-quang ngực của 30.805 bệnh nhân, được phát triển bởi Viện Y tế Quốc gia Hoa Kỳ (NIH). . Dữ liệu được gán nhãn bán tự động từ báo cáo chẩn đoán bằng kỹ thuật xử lý ngôn ngữ tự nhiên (NLP) với 14 loại bệnh lý phổi (như viêm phổi, tràn dịch màng phổi, xẹp phổi, giãn phế nang, nốt phổi, u phổi, xơ hóa...) [45, 83]. Tuy nhiên, không có nhãn cụ thể cho bệnh bụi phổi silic nghề nghiệp. Nhãn “xơ hóa” có thể liên quan gián tiếp nhưng không được phân loại theo tiêu chuẩn quốc tế ILO. Ngoài ra, phương pháp gán nhãn bằng NLP có thể thiếu chính xác so với chú thích trực tiếp từ bác sĩ chuyên khoa.

MIMIC-CXR: Bộ dữ liệu quy mô lớn với 377.110 ảnh X-quang ngực kèm báo cáo chẩn đoán, thu thập từ Trung tâm Y tế Beth Israel Deaconess và công khai bởi Viện Công nghệ Massachusetts. Dữ liệu được ẩn danh hóa nhằm bảo vệ quyền riêng tư của người bệnh và công khai. Tuy nhiên, không chứa nhãn cụ thể cho bệnh bụi phổi silic nghề nghiệp. [84].

LIDC-IDRI: Bộ dữ liệu gồm 1.018 ảnh CT ngực được chú thích tổn thương phổi từ bốn bác sĩ chuyên khoa [85]. Tuy nhiên, dữ liệu là ảnh CT thay vì X-quang ngực và không có nhãn liên quan đến bệnh bụi phổi silic nghề nghiệp

Silicodata: Bộ dữ liệu công khai đầu tiên tập trung vào bệnh bụi phổi silic, gồm các ảnh X-quang ngực thu thập trong vòng 3 năm từ công nhân khai thác đá tại các trạm y tế tuyến cơ sở ở Rajasthan, Ấn Độ. Dữ liệu được chú thích bởi bác sĩ chuyên khoa và bao gồm các lớp: bình thường, bụi phổi silic, bụi phổi lao và lao phổi. [86]. Bộ dữ liệu được công bố cho mục đích nghiên cứu và giáo dục, không sử dụng thương mại. Để truy cập, các tổ chức nghiên cứu cần gửi thỏa thuận cấp phép (license agreement) do người có thẩm quyền pháp lý của đơn vị ký.

Hạn chế chung của các tập dữ liệu quốc tế:

- (i) Không có nhãn cụ thể cho bệnh bụi phổi silic nghề nghiệp (ngoại trừ tập dữ liệu silicodata).
- (ii) Chưa phản ánh đặc điểm dịch tễ học và môi trường lao động tại Việt Nam.

Do đó, việc xây dựng một bộ dữ liệu nội sinh phù hợp với đặc thù bệnh tật và môi trường lao động tại Việt Nam là cần thiết.

2.1.2 Tập dữ liệu trong nước

Trong bối cảnh nghiên cứu về chẩn đoán bệnh bụi phổi silic tại Việt Nam, việc xây dựng và sử dụng các tập dữ liệu phù hợp với đặc điểm dịch tễ học và môi trường lao động trong nước là rất quan trọng. Hiện nay, có hai tập dữ liệu trong nước đáng chú ý là VinDr-CXR, tập dữ liệu từ đề tài KC.10.33/16-20 .

Về tập dữ liệu VinDr-CXR [87]: VinDr-CXR là một tập dữ liệu ảnh X-quang ngực được xây dựng bởi Viện Dữ liệu lớn VinBigdata, với mục tiêu hỗ trợ nghiên cứu và phát triển các mô hình học máy trong lĩnh vực chẩn đoán hình ảnh y khoa. Tập dữ liệu bao gồm 18.000 ảnh X-quang ngực thu thập từ Bệnh viện Trung ương Quân đội 108 và Bệnh viện Đại học Y Hà Nội trong giai đoạn từ năm 2018 đến năm 2020. Mỗi ảnh được chú thích bởi nhóm gồm 17 bác sĩ chẩn đoán hình ảnh có kinh nghiệm, với 22 nhãn cục bộ (local labels) và 6 nhãn toàn cục (global labels). Dữ liệu được chia thành hai phần: tập huấn luyện gồm 15.000 ảnh và tập kiểm tra gồm 3.000 ảnh. Quá trình

gán nhãn được thực hiện thông qua nền tảng VinDr Lab, một hệ thống dựa trên PACS (Picture Archiving and Communication System). Mỗi ảnh trong tập huấn luyện được gán nhãn độc lập bởi 3 bác sĩ, trong khi các ảnh trong tập kiểm tra được gán nhãn dựa trên sự đồng thuận của 5 bác sĩ. Mặc dù VinDr-CXR là một tập dữ liệu có chất lượng cao và được chú thích bởi các chuyên gia, nhưng hiện tại chưa có nhãn cụ thể cho bệnh bụi phổi silic.

Về tập dữ liệu KC.10.33/16-20: Tại Việt Nam, số liệu về bệnh bụi phổi silic nghề nghiệp chủ yếu được ghi nhận trong các đề tài nghiên cứu cấp bộ hoặc cấp nhà nước, với phạm vi triển khai tại các cơ sở y tế chuyên ngành lao động – bệnh nghề nghiệp. Trong đó, đề tài KC.10.33/16-20 là một trong những công trình có quy mô lớn, được Bộ Khoa học và Công nghệ phê duyệt và nghiệm thu vào năm 2021. Đề tài này có ý nghĩa đặc biệt quan trọng trong việc xây dựng nguồn dữ liệu chuẩn hóa về bệnh bụi phổi silic nghề nghiệp ở Việt Nam. Tập dữ liệu được xây dựng trong khuôn khổ đề tài này bao gồm ảnh X-quang ngực, thông tin dịch tễ học và chẩn đoán lâm sàng của 8030 công nhân. Toàn bộ dữ liệu đã được kiểm duyệt và gán nhãn bởi các bác sĩ chuyên khoa nghề nghiệp.

Nhờ vào quy trình xây dựng nghiêm ngặt và sự tham gia của các chuyên gia, tập dữ liệu này được coi là một nguồn tài nguyên đáng tin cậy. Nó đóng vai trò là cơ sở dữ liệu đầu vào cho nhiều nghiên cứu ứng dụng trí tuệ nhân tạo trong việc chẩn đoán bệnh bụi phổi silic nghề nghiệp tại Việt Nam [88].

2.2 Xây dựng tập dữ liệu

Trong bối cảnh nghiên cứu bệnh bụi phổi silic nghề nghiệp, một trong những thách thức lớn nhất là sự thiếu hụt các tập dữ liệu ảnh X-quang ngực đã được gán nhãn chính xác và tổ chức theo dạng phù hợp để huấn luyện mô hình AI. Trước tình trạng này, luận án đã kế thừa và phát triển quy trình chuẩn hóa, tổ chức và tái cấu trúc bộ dữ liệu ảnh X-quang ngực nhằm đảm bảo tính thống nhất và chất lượng phục vụ cho quá trình huấn luyện mô hình.

2.2.1 Nguồn gốc dữ liệu

Trong Luận án này, NCS đã sử dụng bộ dữ liệu, được tổng hợp từ các nguồn trong nước và quốc tế. Các nguồn dữ liệu này đều đã được gán nhãn bởi chuyên gia y tế, hoặc công bố công khai trong các nghiên cứu trước đó. Nhờ vậy, bộ dữ liệu đảm bảo tính đại diện, độ tin cậy lâm sàng, đồng thời cho phép so sánh kết quả mô hình.

(i) Nguồn dữ liệu trong nước: Dữ liệu từ đề tài khoa học cấp Nhà nước KC.10.33/16-20: Bộ dữ liệu được thu thập từ đề tài mang tên “Nghiên cứu đặc điểm dịch tễ học phân

tử, yếu tố nguy cơ và ứng dụng kỹ thuật tiên tiến trong chẩn đoán sớm bệnh bụi phổi silic tại Việt Nam” (Mã số: KC.10.33/16-20). Dữ liệu được thu thập từ các bệnh viện chuyên khoa tại các tỉnh Thái Nguyên, Hải Dương, Bình Định, Phú Yên và Đồng Nai, nơi có tỷ lệ mắc bệnh bụi phổi silic cao do hoạt động khai thác khoáng sản, xây dựng và sản xuất vật liệu xây dựng. Mỗi người lao động tham gia nghiên cứu đều được sàng lọc và chẩn đoán bởi đội ngũ bác sĩ chuyên khoa, đảm bảo rằng bộ dữ liệu không chỉ đầy đủ về số lượng mà còn đạt độ tin cậy cao về mặt lâm sàng.

(ii) Nguồn dữ liệu quốc tế: Dữ liệu này bao gồm các ảnh X-quang ngực được chú thích và xác thực, được mô tả và phân tích trong nghiên cứu "Deep learning-based classification and referral of treatable human diseases". Dữ liệu được chia thành các tập huấn luyện và kiểm tra độc lập, với các nhãn bao gồm bình thường, viêm phổi do vi khuẩn, viêm phổi do virus. Bộ dữ liệu này được công bố công khai và có thể truy cập tại: Mendeley Data [89].

Việc sử dụng dữ liệu trong nghiên cứu được thực hiện hoàn toàn theo các quy định về bảo mật, đạo đức nghiên cứu và quyền riêng tư của người bệnh. Toàn bộ dữ liệu đã được ẩn danh và chỉ phục vụ cho mục đích khoa học, tuân thủ các hướng dẫn của Bộ Y tế về đạo đức trong nghiên cứu y sinh học.

2.2.2 Thách thức trong việc xây dựng và xử lý dữ liệu

Việc chuẩn bị dữ liệu y khoa để huấn luyện mô hình học sâu đặt ra nhiều thách thức, không chỉ ở khía cạnh kỹ thuật mà còn liên quan đến tính phức tạp và không đồng nhất của dữ liệu lâm sàng. Trong phạm vi luận án, việc xây dựng và xử lý dữ liệu gặp phải một số thách thức sau:

(1) Khó khăn trong thu thập dữ liệu thực tế

Việc thu thập ảnh X-quang ngực phục vụ cho nghiên cứu bệnh bụi phổi silic gặp nhiều thách thức do tính chất đặc thù của bệnh lý này. Bệnh phân bố không đồng đều theo vùng địa lý và ngành nghề. Quá trình tiếp cận và số hóa dữ liệu từ các cơ sở y tế yêu cầu: Sự hợp tác chính thức với các bệnh viện chuyên khoa và tổ chức y tế địa phương; Sự chấp thuận từ Hội đồng đạo đức nghiên cứu y sinh học; Quy trình ẩn danh hóa thông tin cá nhân nghiêm ngặt. Ngoài ra, việc đồng bộ dữ liệu từ nhiều cơ sở (các tỉnh Thái Nguyên, Đồng Nai, Bình Định, Hải Dương...) đòi hỏi công cụ tổ chức lưu trữ, kiểm định chất lượng và thống nhất định dạng kỹ thuật số, vốn là một bước không thể thiếu trước khi đưa vào mô hình học sâu.

(2) Đa dạng trong biểu hiện bệnh

Tổn thương bụi phổi silic biểu hiện rất khác nhau giữa các bệnh nhân. Một số

trường hợp chỉ có nốt mờ nhỏ, khu trú và khó phát hiện, trong khi các ca nặng có thể xuất hiện xơ hóa lan tỏa làm biến dạng toàn bộ mô phổi. Điều này khiến cho việc huấn luyện mô hình trở nên khó khăn do nguy cơ thiếu đại diện cho các mức độ bệnh lý trong tập huấn luyện, làm suy giảm khả năng tổng quát hóa và độ ổn định của mô hình.

(3) Tương đồng hình ảnh giữa các bệnh lý

Một thách thức quan trọng khác là khả năng nhầm lẫn giữa tổn thương do bụi phổi silic và các bệnh lý phổi khác có biểu hiện tương tự trên ảnh X-quang ngực.

Do đó, mô hình học sâu cần không chỉ nhận diện đặc trưng của bụi phổi silic, mà còn phân biệt được các dạng tổn thương có hình thái tương đồng, điều này đòi hỏi tập dữ liệu phải được gán nhãn chính xác và có cấu trúc phân lớp rõ ràng.

(4) Mất cân bằng dữ liệu

Số lượng ca mắc bụi phổi silic trong tập dữ liệu thực tế luôn nhỏ hơn đáng kể so với nhóm không mắc bệnh hoặc các bệnh phổ biến hơn. Nếu không xử lý, điều này sẽ dẫn đến mô hình bị thiên lệch về lớp chiếm ưu thế, làm giảm độ nhạy trong phát hiện ca bệnh.

Những thách thức nêu trên không chỉ là rào cản trong khâu chuẩn bị dữ liệu, mà còn đặt ra yêu cầu về thiết kế kiến trúc mô hình, chiến lược huấn luyện và đánh giá kết quả. Việc nhận diện và xử lý đúng các thách thức này là điều kiện cần thiết để đảm bảo các mô hình học sâu phát triển từ dữ liệu y khoa có thể đạt được độ chính xác cao, khả năng tổng quát hóa tốt và tiềm năng ứng dụng trong môi trường lâm sàng thực tế.

2.2.3 Tiêu chuẩn lựa chọn và tiêu chuẩn loại trừ dữ liệu

Để đảm bảo chất lượng và tính đại diện của tập dữ liệu đầu vào phục vụ huấn luyện mô hình học sâu, quá trình lựa chọn ảnh X-quang và thông tin bệnh nhân được thực hiện theo một bộ tiêu chí rõ ràng, được thống nhất giữa nhóm nghiên cứu và các chuyên gia y tế tham gia thẩm định. Các tiêu chuẩn này nhằm tối ưu hóa độ tin cậy lâm sàng, đồng thời giảm thiểu và biến thiên không mong muốn trong dữ liệu đầu vào.

(1) Tiêu chí lâm sàng

Chẩn đoán xác định: Chỉ lựa chọn các trường hợp đã được bác sĩ chuyên khoa bệnh nghề nghiệp hoặc chẩn đoán hình ảnh xác định mắc hoặc không mắc bệnh bụi phổi silic, theo tiêu chuẩn phân loại của Tổ chức Lao động Quốc tế (ILO 2011).

Đối chứng rõ ràng: Các trường hợp được xếp vào nhóm bình thường (không mắc bệnh) phải có kết luận âm tính từ bác sĩ chuyên khoa.

(2) Tiêu chí kỹ thuật ảnh

Định dạng chuẩn: Chỉ sử dụng các ảnh chụp X-quang ngực hậu – trước (PA) hoặc chính diện có định dạng DICOM hoặc JPEG chất lượng gốc.

Độ phân giải: Ảnh phải đạt tối thiểu 1024×1024 pixel (hoặc cao hơn), đảm bảo độ sắc nét cần thiết để quan sát rõ cấu trúc phổi, đặc biệt là các nốt mờ nhỏ.

Tư thế đứng: Loại trừ ảnh chụp nghiêng, lệch hoặc có tư thế sai (ví dụ: không hít sâu, che khuất phần phổi).

Loại bỏ nhiễu: Không sử dụng các ảnh bị mờ do chuyển động, quá sáng/tối, có bóng che phủ (từ vật thể ngoài cơ thể bệnh nhân hoặc máy móc), hoặc có nhiều ảnh hưởng đến khả năng trích xuất đặc trưng

(3) Tiêu chí lựa chọn phân lớp bệnh

Nhóm bệnh bụi phổi silic: Bao gồm các trường hợp đã được chẩn đoán xác định là mắc bệnh bụi phổi silic, không phân biệt theo mức độ nặng nhẹ do hạn chế về nhãn chi tiết trong dữ liệu gốc. Việc gộp các mức độ tổn thương vào một nhãn duy nhất giúp đảm bảo tính nhất quán trong phân loại và phù hợp với mục tiêu của các mô hình học sâu nhị phân hoặc đa lớp trong nghiên cứu.

Nhóm viêm phổi: Bao gồm các ca viêm phổi đã được xác định từ nguồn dữ liệu công khai có nhãn đáng tin cậy. Các trường hợp không xác định căn nguyên hoặc có biểu hiện chồng lấn với các bệnh phổi khác được loại trừ để tránh tăng độ nhiễu trong huấn luyện.

Nhóm bình thường (đối chứng): Bao gồm các ảnh X-quang ngực không có biểu hiện bệnh lý rõ ràng, được xác nhận là không mắc bệnh bụi phổi silic hoặc các bệnh lý phổi khác. Đây là nhóm đóng vai trò nền để mô hình học sâu phân biệt được đặc trưng của lớp bệnh.

(4) Tiêu chí dữ liệu nhân khẩu học

Tuổi lao động: Chỉ lựa chọn các bệnh nhân trong độ tuổi 18—65, là nhóm có nguy cơ nghề nghiệp cao nhất. Thông tin đủ: Các mẫu chỉ được đưa vào tập dữ liệu chính thức khi đi kèm với đầy đủ thông tin nhân khẩu học cơ bản (giới tính, địa phương, số năm làm việc, nghề nghiệp).

Việc áp dụng đồng bộ các tiêu chuẩn trên giúp đảm bảo rằng tập dữ liệu không chỉ đạt độ tin cậy cao về mặt lâm sàng, mà còn có tính nhất quán về kỹ thuật và phù hợp với yêu cầu của các mô hình học sâu.

Mặc dù dữ liệu đã trải qua kiểm định ban đầu từ nhóm nghiên cứu thuộc đề tài cấp Nhà nước, nhóm nghiên cứu tiếp tục thực hiện quy trình kiểm định thứ cấp nhằm đảm bảo dữ liệu phù hợp với mục tiêu nghiên cứu, đáp ứng các tiêu chuẩn cần thiết để

Bảng 2.1 Tiêu chí lựa chọn và vai trò các nhóm bệnh trong phân lớp dữ liệu

Loại bệnh	Tiêu chí lựa chọn	Vai trò trong mô hình học sâu
Bụi phổi silic	- Chẩn đoán xác định theo tiêu chuẩn ILO. - Không phân mức độ do thiếu nhãn chi tiết.	Lớp chính cần nhận diện; sử dụng huấn luyện mô hình nhị phân hoặc đa lớp.
Viêm phổi do virus	- Nhãn bệnh được xác thực từ nguồn dữ liệu công khai.	Lớp nhiễu đối trọng; giúp mô hình học được đặc trưng phân biệt giữa viêm phổi và bụi phổi silic.
Viêm phổi do vi khuẩn	- Tương tự như trên: nguyên nhân xác định, nhãn rõ ràng.	Lớp nhiễu đối trọng; giúp mô hình học được đặc trưng phân biệt giữa viêm phổi và bụi phổi silic.
Bình thường (đối chứng)	- Không có dấu hiệu bệnh lý phổi. - Được xác nhận âm tính bởi chuyên gia.	Nhóm đối chứng; dùng để phân biệt rõ các ca không bệnh với ca bệnh bụi phổi silic.

huấn luyện mô hình học sâu. Quá trình kiểm định này bao gồm ba bước chính:

Bước 1. Rà soát chất lượng hình ảnh X-quang

Mục tiêu: Đảm bảo chỉ những hình ảnh đáp ứng tiêu chuẩn về độ phân giải, độ tương phản và mức độ thể hiện tổn thương mới được sử dụng trong nghiên cứu.

Phương pháp: Nhóm nghiên cứu phối hợp với bác sĩ chuyên khoa chẩn đoán hình ảnh và bác sĩ bệnh nghề nghiệp để đánh giá toàn bộ tập dữ liệu. Hình ảnh được kiểm tra theo các tiêu chí loại bỏ dữ liệu, bao gồm: Hình ảnh bị mờ hoặc nhiễu quá mức, làm giảm khả năng nhận diện tổn thương. Hình ảnh có tổn thương không rõ ràng hoặc không đồng nhất với nhãn gán ban đầu, có thể dẫn đến sai lệch trong huấn luyện mô hình. Hình ảnh không đáp ứng tiêu chuẩn về độ phân giải, ảnh hưởng đến quá trình tiền xử lý và khả năng trích xuất đặc trưng.

Bước 2. Kiểm tra tính nhất quán của nhãn dữ liệu

Mục tiêu: Đảm bảo nhãn dữ liệu chính xác, thống nhất và phù hợp với các tiêu chí chẩn đoán của mô hình học sâu.

Phương pháp: Các hình ảnh đã được chọn sẽ tiếp tục được soát xét lại nhãn gán bởi hội đồng chuyên gia y tế, bao gồm: Bác sĩ chuyên khoa chẩn đoán hình ảnh đánh giá lại tổn thương trên phim X-quang. Bác sĩ bệnh nghề nghiệp kiểm tra mức độ phù hợp với tiêu chí bệnh bụi phổi silic. Các nhãn gán sẽ được đối chiếu với hồ sơ bệnh án, kết quả kiểm tra lâm sàng và các tiêu chuẩn chẩn đoán của ILO 2000.

Bước 3. Đánh giá mức độ phù hợp của dữ liệu với mô hình học sâu

Mục tiêu: Đảm bảo tập dữ liệu đầu vào phù hợp với đặc trưng của mô hình học sâu, giúp tối ưu hóa quá trình huấn luyện và tăng độ chính xác trong chẩn đoán.

Phương pháp: Kiểm tra thử nghiệm dữ liệu trên mô hình học sâu bằng cách: Huấn luyện thử nghiệm trên một tập con của dữ liệu để đánh giá hiệu suất ban đầu. Xác định các lỗi phân loại phổ biến và điều chỉnh dữ liệu nếu cần. Nếu phát hiện tỷ lệ lỗi cao hoặc sai lệch trong phân loại, nhóm nghiên cứu sẽ rà soát lại dữ liệu đầu vào, xác minh nguyên nhân và thực hiện điều chỉnh.

Kết luận về quy trình kiểm định thứ cấp: Hai cấp kiểm định giúp đảm bảo chất lượng cơ sở dữ liệu. Kiểm định ban đầu do nhóm nghiên cứu thuộc đề tài cấp Nhà nước thực hiện để sàng lọc dữ liệu đúng tiêu chuẩn y khoa. Kiểm định thứ cấp do nhóm nghiên cứu trong luận án thực hiện nhằm tối ưu hóa tính chính xác và đảm bảo dữ liệu phù hợp với huấn luyện mô hình học sâu. Chỉ những hình ảnh đạt chuẩn qua tất cả các giai đoạn kiểm định mới được sử dụng trong nghiên cứu, đảm bảo dữ liệu có độ tin cậy cao nhất, giảm thiểu sai lệch và tối ưu hóa khả năng ứng dụng của mô hình AI trong chẩn đoán bệnh bụi phổi silic.

Sau khi hoàn thành cả hai giai đoạn kiểm định, chỉ những hình ảnh đáp ứng đầy đủ các tiêu chuẩn về chất lượng, tính chính xác trong chẩn đoán và tính nhất quán về nhãn dữ liệu mới được đưa vào sử dụng trong nghiên cứu này. Việc thực hiện kiểm định hai cấp giúp đảm bảo độ tin cậy cao của dữ liệu, giảm thiểu sai lệch và tối ưu hóa hiệu suất của mô hình học sâu trong chẩn đoán bệnh bụi phổi silic.

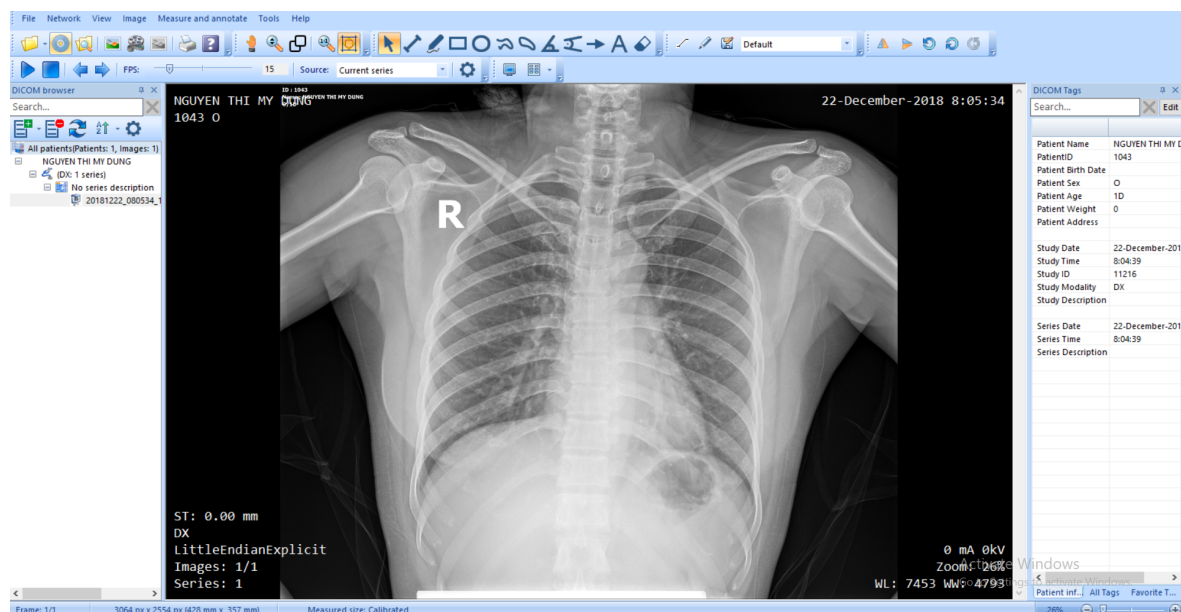
Về tiêu chuẩn loại trừ: Trong quá trình tiếp nhận và xử lý dữ liệu X-quang ngực, một số trường hợp hình ảnh không đáp ứng tiêu chuẩn chẩn đoán và không đủ giá trị sử dụng cho huấn luyện mô hình học sâu nên buộc phải loại trừ. Các tiêu chí loại trừ được xác định như sau: Loại bỏ các trường hợp có hình ảnh X-quang ngực bị mờ nặng, nhiễu cao, hoặc không đạt tiêu chuẩn chẩn đoán (mất một phần khung phổi, bị che bởi dụng cụ y tế). Loại bỏ ảnh thiếu nhãn chẩn đoán xác thực từ chuyên gia. Việc loại bỏ các ảnh không đạt chuẩn giúp đảm bảo tính đồng nhất và độ tin cậy của dữ liệu đầu

vào, đồng thời giảm thiểu nguy cơ mô hình học các đặc trưng sai lệch do ảnh hưởng từ dữ liệu kém chất lượng.

2.2.4 Tiền xử lý dữ liệu

Quá trình tiền xử lý và tăng cường dữ liệu đóng vai trò thiết yếu trong huấn luyện các mô hình học sâu, đặc biệt là trong lĩnh vực thị giác máy tính với dữ liệu hình ảnh y tế. Mục tiêu chính của quy trình này là cải thiện chất lượng dữ liệu đầu vào và tăng cường tính đa dạng của mẫu huấn luyện, từ đó giúp mô hình học được các đặc trưng mang tính khái quát tốt hơn, tránh bị phụ thuộc vào một số biểu hiện cố định trong tập dữ liệu ban đầu. Các bước chính trong quá trình tiền xử lý bao gồm:

(1) Chuyển đổi định dạng dữ liệu: Các ảnh X-quang thu thập được ban đầu được lưu dưới định dạng DICOM (.dcm) – một tiêu chuẩn quốc tế dành riêng cho lĩnh vực hình ảnh y khoa. Định dạng DICOM không chỉ lưu trữ ảnh chất lượng cao mà còn tích hợp nhiều dữ liệu quan trọng liên quan đến bệnh nhân (thông tin cá nhân, thời gian chụp, thiết bị sử dụng, điều kiện kỹ thuật...). DICOM được thiết kế để đảm bảo độ tương thích cao giữa các thiết bị y tế và hỗ trợ tiêu chuẩn hóa việc truyền tải, lưu trữ và xử lý dữ liệu hình ảnh y tế.



Hình 2.1 Hình minh họa ảnh X-quang ngực được lưu ở định dạng DICOM

Tuy nhiên, để phục vụ cho quá trình huấn luyện mô hình học sâu, việc xử lý trực tiếp các tệp DICOM gặp nhiều trở ngại do định dạng này không được hỗ trợ trực tiếp bởi các thư viện học sâu phổ biến như TensorFlow hoặc PyTorch. Ngoài ra, các mô

hình học sâu hiện nay thường được thiết kế để nhận đầu vào là ảnh định dạng JPEG hoặc PNG có kích thước và kênh ảnh cố định. Vì vậy, việc chuyển đổi ảnh X-quang từ định dạng DICOM sang JPEG hoặc PNG là cần thiết để đảm bảo tương thích với quá trình huấn luyện và chuẩn hóa đầu vào. Quá trình chuyển đổi này giúp loại bỏ phần siêu dữ liệu không cần thiết, giảm dung lượng tệp, đồng thời vẫn giữ nguyên độ phân giải cao nhằm bảo toàn các chi tiết bệnh lý.

(2) Chuẩn hóa dữ liệu: Trong thực tế, các ảnh X-quang được thu thập có thể có nhiều kích thước và tỉ lệ khác nhau, tùy thuộc vào thiết bị chụp và quy trình kỹ thuật tại từng cơ sở y tế. Tuy nhiên, các mô hình học sâu, đặc biệt là mạng nơ-ron tích chập yêu cầu đầu vào có kích thước cố định để có thể xử lý đồng bộ trong các tầng tích chập và giảm thiểu lỗi tính toán. Vì vậy, tất cả ảnh đều được chuẩn hóa về cùng một kích thước chuẩn (224×224 pixel), trước khi đưa vào huấn luyện. Kích thước này không chỉ tương thích với nhiều mô hình học sâu được tiền huấn luyện (pre-trained) trên các tập dữ liệu lớn như ImageNet, mà còn đảm bảo tính nhất quán giữa các mẫu dữ liệu.

(3) Tăng cường dữ liệu (Data Augmentation): Trong bối cảnh số lượng ảnh X-quang ngực mắc bệnh bụi phổi silic nghề nghiệp còn hạn chế, việc tăng cường dữ liệu nhằm mở rộng không gian mẫu huấn luyện mà không cần thu thập thêm dữ liệu thực tế. Mục tiêu của các kỹ thuật này là mô phỏng các biến thể có thể gặp trong thực tế lâm sàng (chẳng hạn sự thay đổi tư thế bệnh nhân, điều kiện chiếu sáng khác nhau, vị trí tổn thương thay đổi nhẹ...), từ đó giúp mô hình học sâu tăng khả năng khái quát hóa và tránh hiện tượng học thuộc (overfitting). Các kỹ thuật chính bao gồm:

Xoay nhẹ ảnh (Rotation $\pm 10^\circ$): Mô phỏng sự thay đổi nhỏ trong tư thế bệnh nhân khi chụp X-quang. Phạm vi xoay $\pm 10^\circ$ đảm bảo không làm biến dạng cấu trúc giải phẫu mà vẫn tăng tính đa dạng. Kỹ thuật này giúp mô hình học được các đặc trưng bệnh lý bất biến với góc chụp – đặc biệt hữu ích trong việc nhận diện nốt mờ hoặc xơ hóa không phụ thuộc hướng chiếu.

Lật ngang ảnh (Horizontal Flip): Dù hai bên phổi có một số khác biệt về giải phẫu (tim nằm lệch trái), việc lật ảnh vẫn có giá trị trong việc tăng cường mẫu và buộc mô hình nhận diện đặc trưng bệnh lý một cách đối xứng. Điều này đặc biệt quan trọng khi một bên phổi xuất hiện nhiều mẫu bệnh hơn bên còn lại, gây mất cân bằng.

Dịch chuyển ảnh nhẹ (Translation ± 5 pixels): Phản ánh các sai lệch nhỏ về căn chỉnh thiết bị hoặc chuyển động bệnh nhân. Dịch ảnh ± 5 không làm mất thông tin chẩn đoán, nhưng giúp mô hình không phụ thuộc vào vị trí tuyệt đối của tổn thương – điều đặc biệt hữu ích trong chẩn đoán hình ảnh, nơi cùng một tổn thương có thể xuất hiện ở các vùng khác nhau do góc chụp.

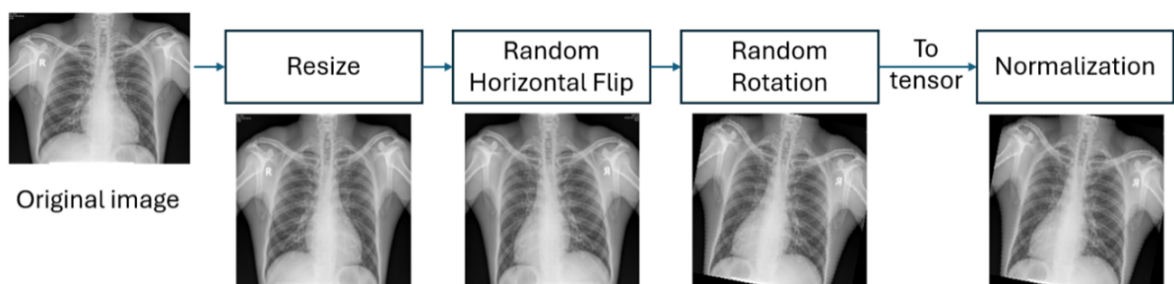
Thay đổi độ sáng (Brightness Adjustment $\pm 20\%$): Mô phỏng các điều kiện chiếu sáng thay đổi trong quá trình thu nhận ảnh hoặc xử lý hậu kỳ. Điều chỉnh độ sáng giúp mô hình duy trì khả năng nhận diện đặc trưng bệnh lý (ví dụ: nốt silic, mảng xơ hóa) bất kể ảnh sáng hơn hoặc tối hơn mức tiêu chuẩn. Tuy nhiên, việc điều chỉnh cần giới hạn ở mức hợp lý để không làm mất thông tin.

Tất cả các phép biến đổi nêu trên đều được thực hiện một cách có kiểm soát để đảm bảo không làm thay đổi bản chất tổn thương hoặc cấu trúc giải phẫu quan trọng.

(4) Cân bằng dữ liệu và giải quyết mất cân bằng lớp: Tập dữ liệu có sự mất cân đối giữa số lượng bệnh nhân mắc bệnh và không mắc bệnh. Nếu không điều chỉnh, mô hình sẽ thiên lệch về lớp chiếm đa số. Để khắc phục, kỹ thuật lấy mẫu có trọng số (weighted sampling) được áp dụng. Theo đó, các mẫu thuộc lớp thiểu số (mắc bệnh) được ưu tiên chọn vào các mini-batch trong quá trình huấn luyện. Nhờ vậy, mô hình có khả năng học được đặc trưng của lớp bệnh lý một cách đầy đủ và cải thiện độ nhạy (sensitivity) trong bài toán phát hiện sớm bệnh bụi phổi silic.

(5) Lọc nhiễu và cải thiện chất lượng ảnh: Một số ảnh vẫn còn nhiễu nhẹ hoặc thiếu tương phản sau bước sàng lọc ban đầu. Do đó, bộ lọc Gaussian được sử dụng để làm mịn ảnh, loại bỏ nhiễu tần số cao mà không ảnh hưởng đến chi tiết quan trọng. Kỹ thuật CLAHE (Contrast Limited Adaptive Histogram Equalization) cũng được áp dụng để tăng độ tương phản cục bộ, hỗ trợ mô hình nhận diện các tổn thương nhỏ như nốt mờ, xơ hóa hoặc dày màng phổi.

(6) Chuẩn hóa dữ liệu đầu vào: Để đảm bảo sự ổn định trong quá trình tối ưu hóa, ảnh được chuẩn hóa bằng Min-Max Scaling (đưa giá trị pixel về khoảng $[0, 1]$).



Hình 2.2 Hình minh họa tiền xử lý dữ liệu

Trong hình 2.2 minh họa chuỗi các phép biến đổi như chuẩn hóa (resize), lật ngang ngẫu nhiên (random horizontal flip), xoay ngẫu nhiên (random rotation), và chuẩn hoá (normalization) được áp dụng tuần tự lên ảnh gốc nhằm biến đổi không gian đầu vào theo các cách hợp lý nhưng không làm thay đổi bản chất nội dung hình ảnh. Các biến đổi này mô phỏng sự thay đổi có thể xảy ra trong thực tế như hướng chụp, vị trí bệnh

nhân hoặc điều kiện thu nhận ảnh, từ đó giúp mô hình trở nên bền vững hơn trước các sai lệch, biến thể trong môi trường triển khai thực tế.

Mặc dù quá trình này được gọi là tăng cường dữ liệu (data augmentation), nhưng không làm tăng số lượng mẫu thực tế trong tập huấn luyện. Thay vào đó, tại mỗi epoch hoặc mỗi lần truy xuất ảnh trong quá trình huấn luyện, mô hình sẽ thấy một phiên bản biến đổi khác nhau của ảnh gốc, nhờ đó có thể học được những đặc trưng bất biến quan trọng hơn. Đây là một chiến lược ngẫu nhiên và động, khác với việc nhân đôi dữ liệu một cách tĩnh. Điều này không chỉ giúp mở rộng không gian đặc trưng mà còn làm giảm nguy cơ overfitting (hiện tượng phổ biến khi mô hình học quá kỹ trên một tập dữ liệu có quy mô hạn chế hoặc ít biến động).

2.2.5 Cấu trúc và đặc điểm tập dữ liệu

Sau quy trình sàng lọc và kiểm định chất lượng nghiêm ngặt trình bày tại Mục 2.2.3, trên cơ sở dữ liệu thu thập và xây dựng, Luận án chia dữ liệu thành ba tập dữ liệu nhỏ để sử dụng ở ba phần nội dung với các ba mục đích nhau, cụ thể: LungDiasePTIT, DNNBPSi, và SVBCX với cấu trúc lớp bệnh được thể hiện trong Bảng 2.2.

Bảng 2.2 Số lượng ảnh X-quang giữa các lớp bệnh trong ba tập dữ liệu nghiên cứu

Lớp bệnh	Tên tập dữ liệu		
	LungDiseasePTIT	DNNBPSi	SVBCX
BPSi	445	535	535
Bình thường	509	1554	1554
Virus	520	0	1493
Vi khuẩn	531	0	2772
Tổng số ảnh	2005	2089	6354
Tổng số lớp	4	2	4
Mục đích sử dụng	Phân loại đa lớp với số mẫu ít	Phân loại nhị phân	Phân loại đa lớp

Ba tập dữ liệu trên được xây dựng với cấu trúc lớp và mục tiêu khác nhau, phục vụ cho từng giai đoạn cụ thể trong thiết kế và đánh giá mô hình:

Tập dữ liệu **LungDiseasePTIT**: là kết quả của quá trình xây dựng và chuẩn hóa dữ liệu do NCS và các cộng sự tại Học viện Công nghệ Bưu chính Viễn thông (PTIT)

thực hiện, nhằm phục vụ cho các nghiên cứu học sâu ít mẫu được trình bày tại Chương 3. Dữ liệu này được tổng hợp từ ba nguồn chính đã trình bày tại Mục 2.2.1, bao gồm dữ liệu được thu thập, tổng hợp từ trong và ngoài nước. Tất cả ảnh đều được sàng lọc, xử lý tiền xử lý và gán nhãn lại dưới sự hướng dẫn và thẩm định của các chuyên gia y tế. Tập dữ liệu LungDiseasePTIT được tổ chức thành bốn lớp: bụi phổi silic, bình thường, viêm phổi do virus, và viêm phổi do vi khuẩn, với tổng cộng hơn 2005 ảnh. Tập dữ liệu này được sử dụng trong môi trường huấn luyện hạn chế dữ liệu, đóng vai trò kiểm chứng khả năng thích nghi và tổng quát hóa của mô hình trong kịch bản Few-Shot Learning.

Tập dữ liệu **DNNBPSi**: Tập dữ liệu phân loại nhị phân, bao gồm hai lớp: ảnh bình thường và ảnh bệnh bụi phổi silic. Việc loại trừ lớp viêm phổi do virus, viêm phổi do vi khuẩn giúp tập trung vào việc đánh giá khả năng phân biệt đặc trưng hình thái của mô hình học sâu trong môi trường ít nhiễu. Trong luận án, DNNBPSi được sử dụng tại Mục 4.1 – Chương 4, để huấn luyện và đánh giá hiệu năng của các kiến trúc học sâu tiêu chuẩn như MobileNetV2, DenseNet201 và Swin Transformer.

Tập dữ liệu **SVBCX** (Silicosis – Viral – Bacterial – Chest X-ray): là tập dữ liệu phân loại đa lớp (multi-class classification), được xây dựng với quy mô lớn nhất và cấu trúc phức tạp nhất trong luận án. Tập dữ liệu này gồm bốn lớp bệnh: bình thường, viêm phổi do virus, viêm phổi do vi khuẩn và bụi phổi silic. Việc bổ sung hai lớp viêm phổi có mục đích tăng mức độ khó và gần với tình huống thực tế, nơi tổn thương bụi phổi silic dễ bị nhầm lẫn với các dạng viêm nhiễm đường hô hấp.

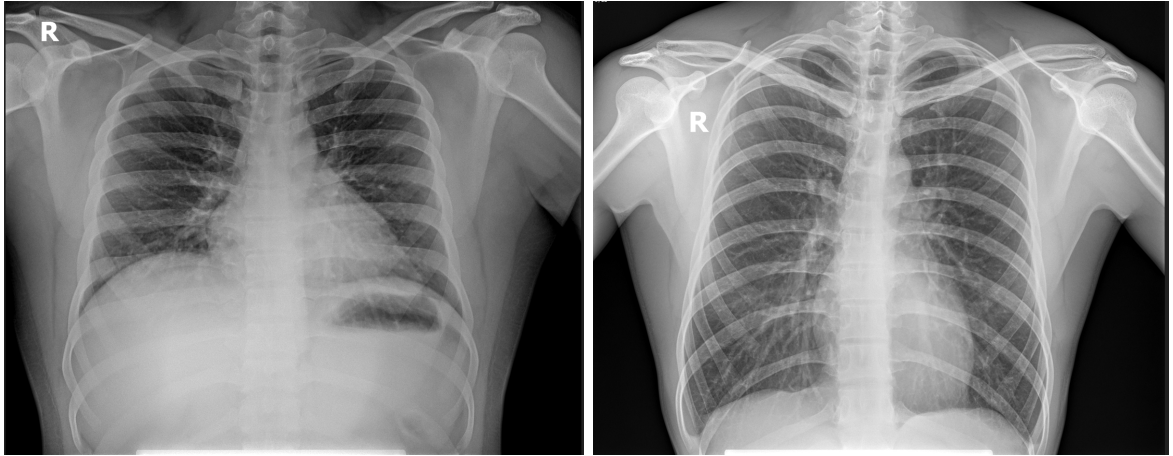
SVBCX đóng vai trò trung tâm trong các thực nghiệm tại Mục 4.2 – Chương 4, nơi mô hình Graph Transformer Post-hoc (GTP) được đề xuất và đánh giá. Với hơn 6354 ảnh, SVBCX cung cấp nền tảng đủ lớn để kiểm định khả năng học quan hệ ngữ cảnh và phân biệt hình thái phức tạp của mô hình đồ thị, đồng thời cho phép đánh giá mô hình trong kịch bản nhiều lớp bệnh.

Hình 2.3 và Hình 2.4 minh họa hình ảnh X-quang ngực của người bình thường và người mắc bệnh bụi phổi silic nghề nghiệp, giúp làm rõ sự khác biệt về hình thái học giữa các lớp bệnh được đưa vào mô hình học sâu.

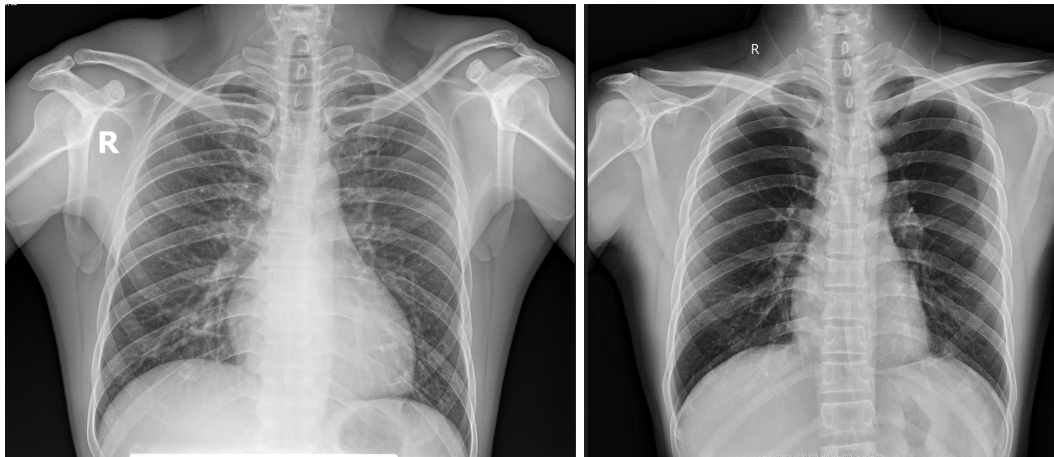
Bộ dữ liệu này có ý nghĩa quan trọng trong việc đánh giá hiệu suất của các mô hình học sâu trong nhiệm vụ phát hiện và phân loại bệnh bụi phổi silic. Việc xây dựng mô hình dựa trên tập dữ liệu này không chỉ hỗ trợ quá trình chẩn đoán sớm hơn mà còn góp phần nâng cao hiệu quả giám sát sức khỏe nghề nghiệp, đặc biệt đối với nhóm lao động có nguy cơ cao tiếp xúc với bụi silic.

Bộ dữ liệu đã được xử lý và chuẩn hóa để đảm bảo:

- (i) Tính đại diện, phản ánh các cấp độ tổn thương phổi trong bệnh bụi phổi silic.



Hình 2.3 Hình ảnh X-quang ngực của người không mắc bệnh bụi phổi silic nghề nghiệp

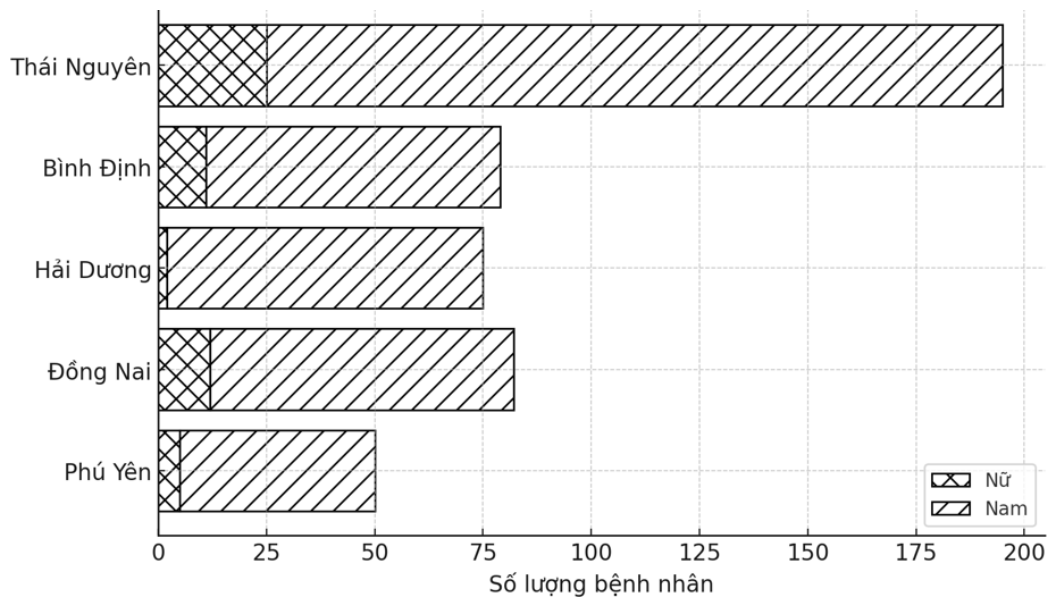


Hình 2.4 Hình ảnh X-quang ngực của người mắc bệnh bụi phổi silic nghề nghiệp

- (ii) Tính cân bằng, giúp mô hình học được sự khác biệt giữa tình trạng bình thường và bệnh lý.
- (iii) Độ chính xác, nhờ quy trình kiểm định nghiêm ngặt, giúp tăng độ tin cậy trong thực nghiệm mô hình.

Bằng cách sử dụng bộ dữ liệu này, nghiên cứu có thể kiểm tra hiệu suất của các mô hình trong nhiệm vụ chẩn đoán bệnh bụi phổi silic, hướng đến ứng dụng thực tế trong hỗ trợ bác sĩ và hệ thống chẩn đoán lâm sàng.

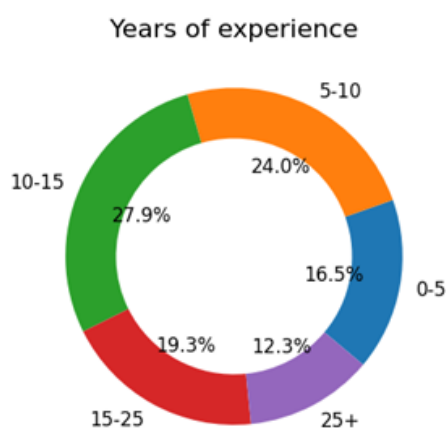
Hình 2.5 trình bày phân bố số lượng người được chẩn đoán mắc bệnh bụi phổi silic theo giới tính và địa phương trong tập dữ liệu SVBCX.



Hình 2.5 Thống kê ca bệnh bụi phổi silic nghề nghiệp theo giới tính

Kết quả thống kê cho thấy số ca bệnh tập trung tại năm tỉnh: Thái Nguyên, Bình Định, Hải Dương, Đồng Nai và Phú Yên – đây đều là các khu vực có hoạt động khai thác khoáng sản, sản xuất vật liệu xây dựng hoặc công nghiệp nặng. Thái Nguyên là địa phương có số lượng ca mắc cao nhất, vượt trội so với các tỉnh còn lại, phản ánh đúng thực trạng phân bố nghề nghiệp và môi trường lao động có nguy cơ cao trong khu vực này. Tỷ lệ giới tính cho thấy nam giới chiếm ưu thế tuyệt đối trong hầu hết các tỉnh. Một số địa phương như Hải Dương và Đồng Nai có ghi nhận một lượng nhỏ ca bệnh ở nữ giới, tuy nhiên con số này vẫn ở mức rất thấp. Sự chênh lệch này phù hợp với đặc thù lao động nặng nhọc trong các ngành có nguy cơ cao phơi nhiễm bụi silic, vốn chủ yếu sử dụng lao động nam giới.

Hình 2.6 trình bày phân bố tỷ lệ người lao động mắc bệnh bụi phổi silic theo nhóm số năm làm việc tại môi trường có nguy cơ phơi nhiễm bụi silic trong tập dữ liệu SVBCX. Dữ liệu được thể hiện dưới dạng biểu đồ tròn với năm phân nhóm: dưới 5 năm, từ 5 đến 10 năm, từ 10 đến 15 năm, từ 15 đến 25 năm và trên 25 năm. Phân tích biểu đồ cho thấy phần lớn các ca bệnh tập trung vào nhóm lao động có thời gian làm việc từ 10 - 15 năm (27.9%) và 5 - 10 năm (24.0%), cho thấy nguy cơ mắc bệnh tăng đáng kể sau khoảng 5 năm tiếp xúc. Nhóm 15 - 25 năm chiếm tỷ lệ 19.3%, tiếp theo là nhóm 0 - 5 năm với 16.5%, phản ánh khả năng xuất hiện bệnh lý ngay cả khi thời gian làm việc chưa dài. Nhóm trên 25 năm chiếm 12.3%, phần nào thể hiện xu hướng nghỉ hưu hoặc chuyển đổi nghề nghiệp ở lao động lớn tuổi.



Hình 2.6 Thống kê ca bệnh bụi phổi silic nghề nghiệp theo số năm kinh nghiệm

2.3 Kết luận chương 2

Trong chương này, luận án đã trình bày quá trình xây dựng, xử lý và chuẩn hóa tập dữ liệu X-quang ngực phục vụ cho bài toán hỗ trợ chẩn đoán bệnh bụi phổi silic nghề nghiệp. Đây là giai đoạn nền tảng, có ảnh hưởng đến chất lượng huấn luyện và hiệu quả của các mô hình. Các nội dung được trình bày trong chương này bao gồm:

- Nguồn gốc dữ liệu được thu thập và xây dựng từ các nguồn trong và ngoài nước.
- Quy trình chuẩn hóa định dạng ảnh, tiền xử lý dữ liệu.
- Phân tích cấu trúc, quy mô, sự phân bố các lớp bệnh trong tập dữ liệu. Phân tích đặc điểm, thách thức của tập dữ liệu.

Những phân tích và xử lý trong chương này không chỉ đảm bảo một nền tảng dữ liệu đáng tin cậy cho huấn luyện các mô hình học sâu, mà còn góp phần định hướng thiết kế kiến trúc mô hình và chiến lược huấn luyện được trình bày ở Chương 3 và Chương 4. Một phần nội dung trình bày trong chương này đã được NCS công bố trong các công trình CT1, CT4, CT8.

CHƯƠNG 3: ĐỀ XUẤT MÔ HÌNH HỌC SÂU ÍT MẪU HỖ TRỢ PHÁT HIỆN VÀ KHOANH VÙNG BỆNH BỤI PHỔI SILIC NGHỀ NGHIỆP

3.1 Đặt vấn đề và động lực nghiên cứu

Chẩn đoán bệnh bụi phổi silic từ ảnh X-quang ngực là một nhiệm vụ phức tạp do đặc điểm hình ảnh tổn thương thường rất mờ nhạt, phân bố không đồng đều và dễ bị nhầm lẫn với các bệnh lý phổi khác. Hơn nữa, số lượng mẫu có nhãn trong lĩnh vực này còn hạn chế, do chi phí gán nhãn cao và yêu cầu trình độ chuyên môn sâu từ bác sĩ chuyên khoa. Các mô hình học sâu truyền thống đòi hỏi lượng dữ liệu huấn luyện lớn, dễ dẫn đến quá khớp và kém hiệu quả trong tình huống khan hiếm dữ liệu. Trong bối cảnh đó, việc phát triển các mô hình học sâu ít mẫu (Few-Shot Learning – FSL) trở thành một hướng tiếp cận khả thi nhằm tận dụng hiệu quả các mẫu bệnh hiếm, đồng thời đảm bảo khả năng tổng quát hóa tốt.

Few-Shot Learning (FSL) là một nhánh đặc biệt của lĩnh vực học máy, được phát triển để giải quyết bài toán phân loại dữ liệu khi số lượng mẫu được gán nhãn rất ít [92,93]. Trong FSL, mô hình học máy sẽ được huấn luyện dựa trên những tập dữ liệu rất nhỏ (thậm chí chỉ một vài mẫu mỗi lớp), nhưng vẫn có khả năng đưa ra dự đoán chính xác cho những dữ liệu mới chưa từng thấy trước đây.

Few-Shot Learning bao gồm các hướng tiếp cận chính như sau:

One-Shot Learning [94]: Mô hình được huấn luyện với duy nhất một mẫu cho mỗi lớp. Một trong những kiến trúc phổ biến trong One-Shot Learning là mạng Siamese, sử dụng cặp ảnh đầu vào để đánh giá sự tương đồng giữa chúng thông qua một hàm đo khoảng cách như Euclid hoặc cosine.

N -Way K -Shot Learning [95]: Mô hình được huấn luyện theo từng *episode*, mỗi *episode* gồm N lớp với K ảnh gán nhãn (support set) và một số ảnh chưa gán nhãn (query set). Mục tiêu là phân loại đúng ảnh query dựa trên thông tin từ support. Các mô hình điển hình thuộc nhóm này gồm Prototypical Networks, Matching Networks và Model-Agnostic Meta-Learning (MAML).

Meta-Learning [96]: Tập trung vào việc “học cách học”, cho phép mô hình thích ứng nhanh với các nhiệm vụ mới chỉ với vài mẫu. Meta-learning thường giúp nâng cao khả năng khái quát hóa khi huấn luyện trên nhiều tác vụ nhỏ và đa dạng.

Trong lĩnh vực chẩn đoán ảnh X-quang ngực – nơi dữ liệu có nhãn chất lượng cao

thường rất khan hiếm – nhiều nghiên cứu gần đây đã khai thác thành công FSL với các mô hình và chiến lược khác nhau. Galán-Cuenca và cộng sự đã áp dụng mạng Siamese kết hợp hàm mất mát có trọng số để phân loại COVID-19 trong điều kiện dữ liệu mất cân bằng nghiêm trọng [97]. Chen và cộng sự sử dụng meta-learning để khai thác mối quan hệ phân cấp giữa các dạng viêm phổi, qua đó tăng cường khả năng khái quát hóa trong thiết lập few-shot [98]. Pramana và cộng sự đã đạt độ chính xác gần 99% khi phân loại lao phổi với mô hình Prototypical Networks [99]. Ngoài ra, các nghiên cứu của Jadon [100] và Angshuman [101] cho thấy tiềm năng ứng dụng FSL trong các bài toán chẩn đoán bệnh phổi hiếm gặp bằng cách kết hợp học phân biệt, autoencoder và học chuyển giao.

Từ những phân tích trên, việc nghiên cứu và đề xuất mô hình học sâu ít mẫu phù hợp là hướng tiếp cận tiềm năng nhằm nâng cao hiệu quả chẩn đoán bệnh bụi phổi silic trong điều kiện dữ liệu khan hiếm.

Những đóng góp chính của chương này có thể được tóm tắt như sau:

- Thiết kế và áp dụng mô hình học sâu ít mẫu (Few-Shot Learning) chuyên biệt cho bài toán chẩn đoán bệnh bụi phổi silic từ ảnh X-quang ngực, giúp cải thiện hiệu suất của mô hình so với một số kiến trúc học ít mẫu khác.
- Xây dựng chiến lược huấn luyện và tinh chỉnh mô hình phù hợp với đặc thù của bộ dữ liệu ảnh X-quang bệnh bụi phổi silic để tăng cường khả năng học và tổng quát hóa.
- Thực hiện đánh giá và so sánh hiệu quả giữa các mô hình học sâu khác nhau, qua đó chứng minh tính phù hợp và ưu thế của mô hình học sâu ít mẫu được đề xuất trong việc chẩn đoán bệnh bụi phổi silic trong điều kiện dữ liệu hạn chế.

3.2 Phát biểu bài toán trong ngữ cảnh dữ liệu y tế hạn chế

Trong chương này, nghiên cứu sử dụng tập dữ liệu LungDiseasePTIT gồm tổng cộng 2005 ảnh X-quang ngực. Mỗi ảnh được gán nhãn theo một trong sáu lớp, bao gồm hai lớp kỹ thuật (Background và Bone) và bốn lớp liên quan đến tình trạng phổi, bao gồm bình thường, viêm phổi do vi khuẩn, viêm phổi do virus, và bụi phổi silic nghề nghiệp. Lớp Background (vùng nền ảnh) đại diện cho khu vực bên ngoài phổi – thường là các vùng tối không chứa thông tin chẩn đoán có giá trị. Lớp Bone (vùng xương) thể hiện các cấu trúc xương bao quanh phổi như xương sườn và cột sống. Hai lớp này xuất hiện trong toàn bộ ảnh và đóng vai trò hỗ trợ trong các tác vụ phân đoạn, giúp mô hình học sâu nhận biết và loại trừ các vùng không liên quan đến tổn thương phổi. Ngược lại, bốn lớp còn lại phản ánh các tình trạng bệnh lý thực thể và được sử dụng trong các thiết lập phân loại và phát hiện bệnh.

Bảng 3.1 trình bày mã lớp, tên lớp, vai trò và số lượng ảnh tương ứng trong tập dữ liệu LungDiseasePTIT. Trong đó, hai lớp *Background* và *Bone* phục vụ cho bài toán phân đoạn kỹ thuật, trong khi bốn lớp còn lại được sử dụng làm mục tiêu trong các bài toán phân loại bệnh phổi dựa trên học sâu.

Bảng 3.1 Phân loại lớp kỹ thuật và lớp bệnh lý trong tập dữ liệu

ID	Tên lớp	Loại lớp	Số lượng	Vai trò
0	Vùng nền ảnh (Background)	Kỹ thuật (Technical)	2005	Vùng nền ngoài phổi, không chứa mô tổn thương.
1	Vùng xương (Bone)	Kỹ thuật (Technical)	2005	Xương sườn, xương sống, hỗ trợ phân đoạn vùng phổi.
2	Phổi bình thường (Normal)	Bệnh lý	509	Lớp không bệnh, làm baseline trong phân loại.
3	Bụi phổi sili (Silicosis)	Bệnh lý	445	Lớp bệnh hiếm, dùng trong kiểm tra mô hình Few-Shot.
4	Viêm phổi do virus (Viral)	Bệnh lý	520	Lớp bệnh được huấn luyện trong tập cơ sở.
5	Viêm phổi do vi khuẩn (Bacterial)	Bệnh lý	531	Lớp bệnh được huấn luyện trong tập cơ sở.

Tập cơ sở $\mathcal{D}^T$ (Base set): Là tập dữ liệu dùng để huấn luyện mô hình, gồm các ảnh X-quang ngực thuộc các lớp đã biết, có số lượng ảnh dán nhãn đầy đủ.

Tập mới $\mathcal{D}^E$ (Novel set): Là tập dữ liệu dùng để kiểm tra mô hình, gồm các ảnh X-quang ngực thuộc lớp bệnh mới, ít mẫu, chỉ xuất hiện trong gia đoạn đánh giá nhằm kiểm tra khả năng tổng quát hóa của mô hình.

$\mathcal{C}^T$ là danh sách/lớp các loại bệnh phổi có trong $\mathcal{D}^T$.

$\mathcal{C}^E$ là danh sách các loại bệnh phổi có trong $\mathcal{D}^E$.

Dữ liệu LungDiasePTIT được chia chiến lược thành hai tập con: tập huấn luyện ($\mathcal{D}^T$) và tập kiểm tra ($\mathcal{D}^E$). Tập huấn luyện bao gồm nhiều lớp cơ sở, trong khi tập kiểm tra chứa các lớp bệnh mới mà mô hình chưa từng gặp trong quá trình huấn luyện. Ký hiệu tập các lớp tương ứng với $\mathcal{D}^T$, $\mathcal{D}^E$, $\mathcal{C}^T$ và $\mathcal{C}^E$ với mỗi tập con có một tập lớp bệnh

phổi riêng biệt, trong đó $|\mathcal{C}^T| + |\mathcal{C}^E| = 6$,

Bài toán như sau:

Đầu vào: một tập ảnh X-quang chưa gán nhãn (*query set*) cùng với một số rất ít ảnh có nhãn thuộc cùng lớp bệnh (*support set*).

Đầu ra: nhãn phân loại (ví dụ: có/không mắc bụi phổi silic) và/hoặc bản đồ phân đoạn vùng tổn thương trên ảnh truy vấn. Để phát hiện bệnh phổi bằng mô hình Few-Shot Learning, mô hình được huấn luyện theo cơ chế episodic learning, trong đó mỗi episode mô phỏng một tác vụ học nhỏ theo thiết lập N -way K -shot

N : Số lớp cần phân biệt trong episode,

K : số mẫu có nhãn mỗi lớp trong support set.

Phương pháp này tạo ra nhiều tập huấn luyện Few-Shot trong quá trình huấn luyện, trong đó mỗi tác vụ chọn một tập nhỏ dữ liệu từ tập huấn luyện $\mathcal{D}^T$ và chia nó thành tập hỗ trợ $\mathcal{S}^T$ và tập truy vấn $\mathcal{Q}^T$.

Tập hỗ trợ $\mathcal{S}^T$ chứa các hình ảnh có gán nhãn về các loại bệnh phổi khác nhau, giúp mô hình học được những thông tin cần thiết. Tập truy vấn $\mathcal{Q}^T$, bao gồm các hình ảnh chưa được gán nhãn, thách thức mô hình trong việc phân loại và phân đoạn các vùng bệnh dựa trên kiến thức đã học từ tập hỗ trợ. Bằng cách mô phỏng môi trường Few-Shot Learning, phương pháp này giúp mô hình thích nghi với các trường hợp mới với lượng dữ liệu tối thiểu, giảm sự phụ thuộc vào các bộ dữ liệu lớn.

Trong khung học theo tập hợp này (episodic learning framework), mỗi tập huấn luyện (episode) được xác định bởi hai tập dữ liệu chính: Tập hỗ trợ

$$\mathcal{S}^* = \{(x_i^s, a_i^s) \mid x_i^s \in \mathbb{R}^{H \times W \times C}, a_i^s \in \mathbb{R}^{H \times W}\}_{i=1}^{NK}, \quad (3.1)$$

trong đó

- x_i^s : hình ảnh hỗ trợ có kích thước $H \times W \times C$, trong đó H là chiều cao, W là chiều rộng, C là số kênh màu,
- a_i^s : nhãn tương ứng có kích thước $H \times W$, biểu thị khu vực bệnh lý trên hình ảnh, nhãn trong tập hỗ trợ,
- N : số lượng lớp, số loại bệnh lý cần phân loại,
- K : số lượng mẫu trên mỗi lớp (shot),
- chỉ số $*$ biểu thị giai đoạn, có thể là huấn luyện T , hoặc kiểm tra E .

Dấu $*$ biểu thị tập dữ liệu thuộc giai đoạn huấn luyện (T) hoặc kiểm tra (E), N là số lượng lớp cần phân loại, và K là số lượng mẫu được gán nhãn trên mỗi lớp, điều

này tạo thành bài toán N -way K -shot. H , W và C lần lượt là chiều cao, chiều rộng và số kênh màu của ảnh.

Tập hỗ trợ chứa các cặp (hình ảnh, nhãn) được sử dụng để cung cấp thông tin cho mô hình về các lớp bệnh phổi. Trong bài toán N -way K -shot, tập $\mathcal{S}^*$ có tổng cộng $N \times K$ mẫu, giúp mô hình học cách nhận diện các đặc trưng của từng lớp với số lượng mẫu hạn chế

$$\mathcal{Q}^* = \{x_j^q \mid x_j^q \in \mathbb{R}^{H \times W \times C}\}_{j=1}^M, \quad (3.2)$$

trong đó, x_j^q là một hình ảnh trong tập truy vấn, và M là số lượng hình ảnh trong tập truy vấn. Chỉ số $*$ biểu thị giai đoạn huấn luyện hoặc kiểm tra, chỉ số j đánh dấu từng mẫu trong tập truy vấn.

Tập truy vấn $\mathcal{Q}^*$ chứa các hình ảnh cần được phân loại và phân đoạn bởi mô hình, dựa trên thông tin học được từ tập hỗ trợ. Mô hình sẽ dự đoán nhãn và mặt nạ phân đoạn cho mỗi hình ảnh trong $\mathcal{Q}^*$.

Trong bài toán, mục tiêu là xác định vector xuất hiện đa nhãn $\mathbf{y}_C \in \mathbb{R}^N$ bằng cách sử dụng hàm phân loại f_C . Đồng thời, trong bài toán phân đoạn cần tạo một mặt nạ phân đoạn $\mathbf{Y}_S \in \mathbb{R}^{H \times W}$ để xác định vùng tổn thương trên ảnh X-quang, thông qua hàm phân đoạn f_S .

Các phương trình toán học mô tả hai nhiệm vụ này như sau:

$$\begin{aligned} \mathbf{y}_C &= f_C(x_j^q, \mathcal{S}^*; \theta_C), \\ \mathbf{Y}_S &= f_S(x_j^q, \mathcal{S}^*; \theta_S). \end{aligned} \quad (3.3)$$

f_C nhận đầu vào là hình ảnh truy vấn x_j^q và tập hỗ trợ $\mathcal{S}^*$, sau đó dự đoán các lớp bệnh phổi xuất hiện trong hình ảnh. θ_C là tham số học của mô hình phân loại.

f_S là hàm phân đoạn, nhận đầu vào là hình ảnh truy vấn x_j^q và tập hỗ trợ $\mathcal{S}^*$, sau đó tạo ra mặt nạ phân đoạn để xác định vùng bệnh lý trên hình ảnh X-quang ngược. θ_S là tham số học của mô hình phân đoạn.

3.3 Giải quyết bài toán với Few-Shot Learning

3.3.1 Mô hình hóa bài toán

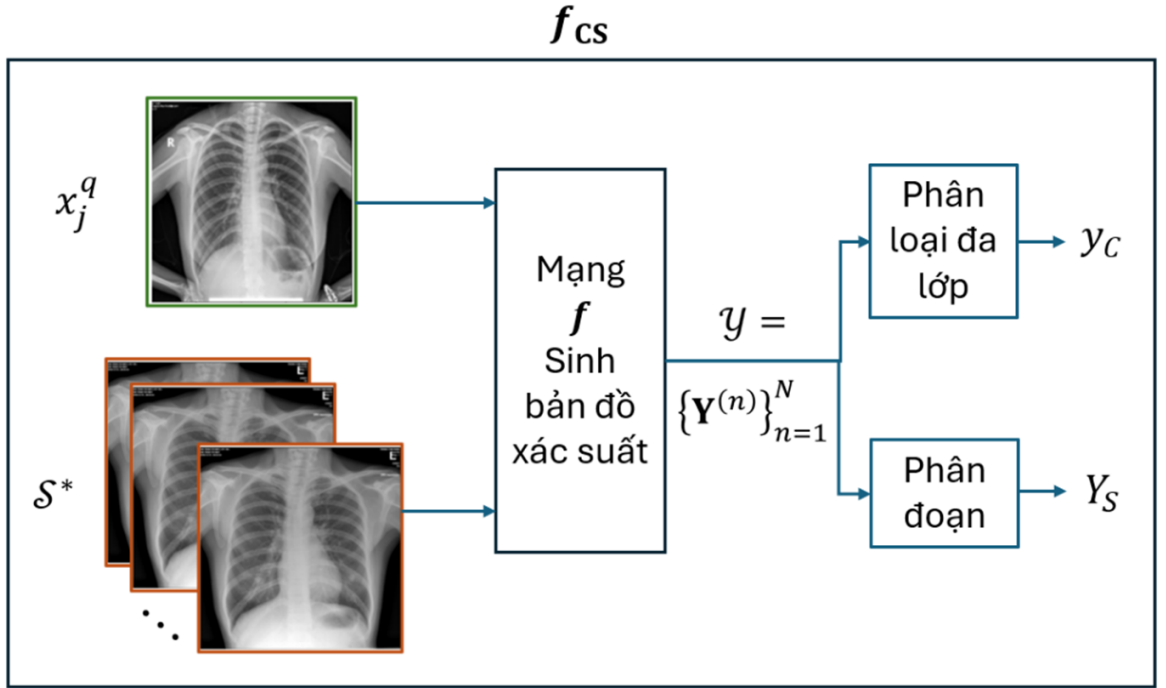
Trong học sâu (deep learning), một mạng nơ-ron có thể được coi như một hàm ánh xạ từ dữ liệu đầu vào (ảnh, chuỗi, tín hiệu...) sang đầu ra (nhãn lớp, mặt nạ, giá trị số). Đối với bài toán chẩn đoán bệnh bụi phổi silic nghề nghiệp trong điều kiện dữ liệu hạn chế, luận án mô hình hóa theo khung học ít mẫu (Few-Shot Learning – FSL). Cụ thể:

Thay vì tối ưu hóa riêng biệt các hàm f_C và f_S , nghiên cứu tập trung vào một phương pháp tiếp cận hợp nhất, nhằm tối ưu hóa chung một hàm duy nhất f_{CS} kết hợp cả hai nhiệm vụ phân loại và phân đoạn Few-Shot (FS-CS).

$$\{y_C, Y_S\} = f_{CS}(x_j^q, \mathcal{S}^*; \theta_{CS}) \quad (3.4)$$

Trong đó:

- θ_{CS} : tập tham số học được tối ưu trong quá trình huấn luyện.
- y_C : vector phân loại (cho biết lớp nào có trong ảnh).
- Y_S : mặt nạ phân đoạn (chỉ rõ vị trí từng lớp trong ảnh).



Hình 3.1 Luồng tổng quan bài toán

Sinh bản đồ xác suất (Probability Maps)

Để tạo ra hai đầu ra trên, mô hình trước tiên sinh ra một tập bản đồ xác suất $\mathcal{Y}$, mỗi bản đồ ứng với một lớp trong tập hỗ trợ:

$$\mathcal{Y} = f(x_j^q, \mathcal{S}^*; \theta) = \{Y^{(n)}\}_{n=1}^N \quad (3.5)$$

Trong đó:

- $Y^{(n)} \in \mathbb{R}^{H \times W}$: bản đồ xác suất cho lớp n , kích thước bằng ảnh gốc.

- Giá trị tại mỗi pixel $\mathbf{Y}^{(n)}(p)$ thể hiện xác suất pixel p thuộc về lớp n .

Nói cách khác, thay vì dự đoán trực tiếp nhãn, mô hình trước tiên phân bố xác suất trên từng điểm ảnh, rồi từ đó suy ra phân loại và phân đoạn.

Suy luận chi tiết (Inference)

1. Phân loại đa lớp (Class-wise Occurrence Prediction)

Từ mỗi bản đồ $\mathbf{Y}^{(n)}$, mô hình quyết định lớp n có xuất hiện trong ảnh hay không. Quy tắc như sau:

$$\hat{\mathbf{y}}_C^{(n)} = \begin{cases} 1, & \text{nếu } \max_{\mathbf{p} \in [H] \times [W]} \mathbf{Y}^{(n)}(\mathbf{p}) \geq \delta \\ 0, & \text{ngược lại} \end{cases} \quad (3.6)$$

- δ : ngưỡng quyết định (decision threshold), trong thực nghiệm đặt mặc định giá trị này là 0.5.
- Lý do dùng max pooling thay vì average pooling: ngay cả khi vùng bệnh rất nhỏ, chỉ cần một số pixel có xác suất cao là đủ để mô hình phát hiện.

2. Phân đoạn (Segmentation)

Để đảm bảo rằng mỗi pixel đều được gán nhãn, mô hình bổ sung thêm một lớp nền (background). Xác suất nền được tính như:

$$\mathbf{Y}_{\text{bg}} = \frac{1}{N} \sum_{n=1}^N (1 - \mathbf{Y}^{(n)}) \quad (3.7)$$

Sau đó ghép lớp nền với các bản đồ xác suất để tạo thành tensor phân đoạn:

$$\mathbf{Y}_S = [\mathbf{Y} || \mathbf{Y}_{\text{bg}}] \in \mathbb{R}^{H \times W \times (N+1)} \quad (3.8)$$

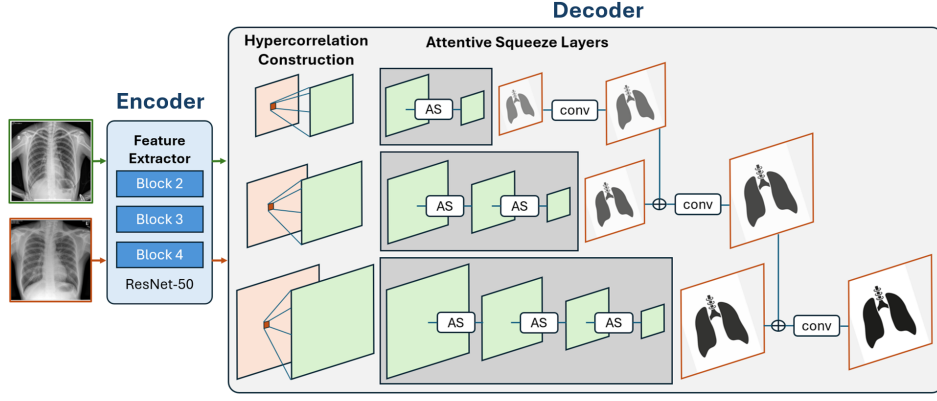
Mặt nạ phân đoạn cuối cùng được lấy bằng cách chọn lớp có xác suất cao nhất tại mỗi pixel:

$$\hat{\mathbf{Y}}_S = \arg \max_{n \in [N+1]} \mathbf{Y}_S \quad (3.9)$$

Nghĩa là, với mỗi điểm ảnh, mô hình sẽ gán nhãn theo lớp có xác suất cực đại, bao gồm cả background.

3.3.2 Kiến trúc mô hình

Trong bài toán phân loại bệnh phổi silic nghề nghiệp với học ít mẫu, luận án kế thừa và phát triển từ các công trình nghiên cứu trước đây như các kiến trúc ASNet, HSNet... Các thành phần quan trọng được giới thiệu trong các nghiên cứu này, bao gồm Hypercorrelation Construction và Attentive Squeeze Layers, đã được tích hợp và điều chỉnh trong phương pháp đề xuất.



Hình 3.2 Kiến trúc mạng và luồng xử lý dữ liệu của mô hình

Quá trình thiết kế và thực nghiệm cho thấy việc sắp xếp và cấu trúc lại các lớp mạng đóng vai trò quan trọng trong việc nâng cao hiệu năng của mô hình. Đặc biệt, trong kiến trúc ASNet gốc, các siêu tương quan (hypercorrelation features) ở tầng thứ 2 và 3 được xử lý qua cùng một chuỗi biến đổi, dẫn đến khả năng khai thác thông tin chưa tối ưu. Từ quan sát này, nghiên cứu đề xuất một kiến trúc theo dạng kim tự tháp (pyramid-like structure), trong đó các siêu tương quan có kích thước lớn hơn sẽ được đưa qua nhiều phép biến đổi hơn. Cách tổ chức này cho phép mô hình trích xuất được nhiều thông tin hơn từ các đặc trưng có độ phân giải cao, đồng thời làm giàu và cô đọng biểu diễn ở các tầng sâu hơn. Các kết quả thực nghiệm sau đó đã chứng minh hiệu quả của kiến trúc đề xuất, cho thấy sự vượt trội đáng kể so với các phương pháp tham chiếu.

Kiến trúc đề xuất gồm bốn khối chính: (i) Feature Extractor, (ii) Hypercorrelation Construction, (iii) Attentive Squeeze Layers, (iv) Decoder. Quá trình lan truyền dữ liệu được mô tả như sau:

(i) Feature Extractor - Encoder

Ảnh truy vấn x_q và ảnh hỗ trợ x_s được đưa qua backbone ResNet-50 để thu được đặc trưng ở nhiều mức độ:

$$F_q^l = \phi^l(x_q), F_s^l = \phi^l(x_s), l \in \{2, 3, 4\}. \quad (3.10)$$

Trong đó $\phi^l(\cdot)$ biểu thị phép trích xuất đặc trưng tại tầng l . Kết quả thu được là ba cặp đặc trưng song song $(F_q^2, F_s^2), (F_q^3, F_s^3), (F_q^4, F_s^4)$.

Việc khai thác đặc trưng ở nhiều tầng đảm bảo rằng mô hình không chỉ nắm bắt được các tín hiệu cục bộ như biên và kết cấu (tầng thấp), mà còn duy trì được thông tin ngữ nghĩa toàn cục (tầng cao).

(ii) Hypercorrelation Construction

Từ các đặc trưng đã trích xuất, mô hình xây dựng không gian tương quan bằng

phép ánh xạ:

$$\mathcal{H}^l = \text{Corr}(F_q^l, F_s^l), l \in \{2, 3, 4\}. \quad (3.11)$$

Trong đó $\text{Corr}(\cdot)$ là phép tính tương quan điểm-điểm (pixel-to-pixel correlation). Bước này tạo ra một biểu diễn chung phản ánh mối quan hệ giữa ảnh truy vấn và ảnh hỗ trợ. Không gian tương quan cho phép mô hình định vị các vùng trong ảnh truy vấn có mức độ tương đồng cao với vùng đã gán nhãn trong ảnh hỗ trợ. Đây chính là cơ chế giúp Few-Shot Learning vượt qua hạn chế về dữ liệu gán nhãn ít.

(iii) Attentive Squeeze Layers

Mỗi không gian tương quan $\mathcal{H}^l$ được tinh chỉnh qua chuỗi lớp Attentive Squeeze (AS):

$$\hat{\mathcal{H}}^l = AS(\mathcal{H}^l), l \in \{2, 3, 4\}. \quad (3.12)$$

Ứng với mỗi lớp, số khối attentive squeeze là khác nhau để biến đổi các đầu vào với kích cỡ khác nhau, tăng dần theo l từ 2 đến 4. Kết quả thu được sẽ được biến đổi thành các ma trận với các kích thước tăng dần, là ba luồng đặc trưng $\hat{\mathcal{H}}^2, \hat{\mathcal{H}}^3, \hat{\mathcal{H}}^4$. Sau đó, đặc trưng ở tầng 2 sẽ được đi qua lớp tích chập để tăng kích thước lên sao cho tương đồng với kích thước của $\hat{\mathcal{H}}^3$. Tiếp theo, chúng sẽ được kết hợp với nhau qua phép cộng từng phần tử. Đầu ra ở pha này sau đó lại được tiếp tục đi qua lớp biến đổi conv để tăng kích thước và tổng hợp với $\hat{\mathcal{H}}^4$. Các bản đồ đặc trưng sẽ được sinh ra sau khi qua bước biến đổi cuối cùng.

(iv) Decoder: Các đặc trưng hợp nhất cuối cùng được đưa qua các lớp conv để sinh bản đồ phân vùng tổn thương, đồng thời tổng hợp thành đầu ra phân loại.

Lớp AS biến đổi các đặc trưng tương quan, đồng thời áp dụng cơ chế chú ý nhằm nhấn mạnh các vùng tương quan quan trọng, đồng thời loại bỏ nhiễu từ các vùng nền. Việc kết hợp thông tin từ nhiều mức độ đảm bảo mô hình vừa nhạy với chi tiết nhỏ, vừa duy trì bối cảnh ngữ nghĩa tổng thể.

Việc thiết kế kiến trúc mạng như trên thể hiện sự phối hợp chặt chẽ giữa trích xuất đặc trưng, xây dựng tương quan và cơ chế chú ý đa tầng. Dòng chảy dữ liệu qua các khối được tổ chức hợp lý, cho phép mô hình tận dụng tối đa thông tin từ số ít ảnh hỗ trợ.

Trong quá trình huấn luyện mô hình, việc lựa chọn hàm mất mát đóng vai trò quan trọng để định hướng tối ưu hóa. Luận án sử dụng hai loại hàm mất mát chính, tương ứng với hai nhiệm vụ khác nhau: mất mát phân loại (classification loss) và mất mát phân đoạn (segmentation loss).

Hàm mất mát phân loại

Hàm mất mát phân loại được xây dựng dựa trên binary cross-entropy, tính trên vector nhãn đa lớp (multihot labels). Đầu tiên, các bản đồ xác suất $\mathbf{Y}^{(n)}$ được tổng hợp bằng average pooling theo không gian, sau đó so sánh với nhãn thực tế. Công thức được biểu diễn như sau:

$$\mathcal{L}_C = -\frac{1}{N} \sum_{n=1}^N \mathbf{y}_{gt}^{(n)} \log \frac{1}{HW} \sum_{\mathbf{p} \in [H] \times [W]} \mathbf{Y}^{(n)}(\mathbf{p}) \quad (3.13)$$

trong đó:

N : số lớp quan tâm,

$\mathbf{y}_{gt}^{(n)}$: nhãn thực tế, được biểu diễn dưới dạng multi-hot encoded vector,

$\mathbf{Y}^{(n)}(\mathbf{p})$: xác suất tại vị trí pixel $\mathbf{p}$ thuộc lớp n .

Như vậy, $\mathcal{L}_C$ phản ánh mức độ khớp giữa nhãn ảnh thực tế và dự đoán sau khi gộp không gian.

Hàm mất mát phân đoạn

Đối với nhiệm vụ phân đoạn, hàm mất mát được xây dựng dưới dạng cross-entropy trung bình giữa phân bố xác suất tại từng điểm ảnh và mặt nạ nhãn thực tế. Hàm mất mát này được định nghĩa:

$$\mathcal{L}_S = -\frac{1}{(N+1)HW} \sum_{n=1}^{N+1} \sum_{\mathbf{p} \in [H] \times [W]} \mathbf{Y}_{gt}^{(n)}(\mathbf{p}) \log \mathbf{Y}_S^{(n)}(\mathbf{p}) \quad (3.14)$$

trong đó:

$\mathbf{Y}_{gt}^{(n)}(\mathbf{p})$: giá trị nhãn thực tế tại vị trí $\mathbf{p}$,

$\mathbf{Y}_S^{(n)}(\mathbf{p})$: xác suất dự đoán tại vị trí $\mathbf{p}$,

$N+1$: số lớp, bao gồm cả lớp nền (background).

$\mathcal{L}_S$ đánh giá trực tiếp khả năng phân loại chính xác của mô hình trên từng pixel, từ đó hướng dẫn mạng học được cấu trúc không gian và ranh giới vùng bệnh.

Cả hai hàm mất mát $\mathcal{L}_C$ và $\mathcal{L}_S$ đều chia sẻ mục tiêu phân loại, nhưng khác nhau ở cấp độ giám sát:

$\mathcal{L}_C$: phân loại trên toàn ảnh, phù hợp khi chỉ có nhãn ảnh mức cao (weak supervision).

$\mathcal{L}_S$: phân loại từng điểm ảnh, phù hợp khi có mặt nạ nhãn đầy đủ (strong supervision).

Trong quá trình huấn luyện, việc lựa chọn giữa hai hàm mất mát phụ thuộc vào mức độ giám sát dữ liệu sẵn có.

3.4 Thiết kế thực nghiệm và cấu hình huấn luyện

3.4.1 Cấu hình thực nghiệm

Trong nghiên cứu này đánh giá phương pháp đề xuất bằng cách so sánh nó với bốn thuật toán Few-Shot tiêu biểu trong phân đoạn và phát hiện đối tượng.

Mô hình IG-FSL được so sánh với các phương pháp cơ sở nổi bật trong lĩnh vực phân đoạn và phát hiện đối tượng sau:

- Path Aggregation Network (PANet) [102]: đạt giải cao trong cuộc thi COCO 2017 Challenge, đạt hạng nhất trong phân đoạn đối tượng và hạng nhì trong phát hiện đối tượng, nhờ khả năng tích hợp hiệu quả các đặc trưng trên nhiều cấp độ.
- Prior Guided Feature Enrichment Network (PFENet) [103]: Một phương pháp mạnh trong Few-Shot Learning (FSL), giúp tổng quát hóa tốt trên tập dữ liệu hạn chế bằng cách tận dụng thông tin từ các đặc trưng đã có trước đó.
- Hypercorrelation Squeeze Networks (HSNet) [104]: Đề xuất một mô hình mới cho Few-Shot Classification và Segmentation (FS-CS), sử dụng tích chập 4D nhẹ để cân bằng giữa hiệu suất và độ chính xác.
- Attentive Squeeze Network (ASNet) [105]: Đạt hiệu quả cao nhờ cơ chế Attentive Squeeze, tăng khả năng phân biệt đặc trưng giữa các lớp trong FSL.

Hai chỉ số đánh giá

Hai nhiệm vụ chính (phân loại và phân đoạn) được đánh giá bằng các chỉ số riêng biệt.

Phân loại đa nhãn (Multi-label classification) dùng chỉ số Exact Ratio (ER) đo lường độ chính xác của vecto phân loại đa nhãn

$$ER = 1[y_C = Y_{gt}] \quad (3.15)$$

trong đó: Y_{gt} Vector nhãn thực tế của dữ liệu, biểu thị các lớp bệnh có mặt trong ảnh y_C Vector nhãn dự đoán bởi mô hình.

Phân đoạn hình ảnh (Segmentation evaluation) Sử dụng chỉ số trung bình IoU (Mean Intersection over Union - mIoU)

$$mIoU = \frac{1}{N} \sum_n IoU_n \quad (3.16)$$

trong đó: IoU_n biểu diễn giá trị Intersection over Union cho lớp bệnh thứ n , đo lường độ trùng lặp giữa mặt nạ dự đoán và mặt nạ thực tế cho lớp n .

3.4.2 Cấu hình huấn luyện

Quá trình tối ưu hóa siêu tham số được thực hiện bằng phương pháp tìm kiếm theo lưới (grid search) trên một tập validation được tách từ dữ liệu huấn luyện. Các tham số quan trọng được khảo sát bao gồm

Learning rate: $\{10^{-3}, 10^{-4}, 5 \times 10^{-5}\}$

Batch size: $\{4, 8, 16, 24\}$

Số epoch: $\{50, 75, 100, 125, 150\}$

Ngưỡng phân loại: $\delta \in [0.3, 0.7]$

Kết quả cho thấy IG-FSL đạt hiệu suất tối ưu khi sử dụng Adam với learning rate 10^{-4} , batch size = 8, và số epoch = 100. Các giá trị tối ưu này, cùng với siêu tham số của các mô hình so sánh, được tổng hợp trong Bảng 3.2.

Trong huấn luyện meta, mô hình được tối ưu theo phương pháp huấn luyện theo tập (episodic training). Mỗi episode bao gồm một tập hỗ trợ (support set) với N lớp và K mẫu mỗi lớp, cùng một tập truy vấn (query set). Với bài toán 1-way 1-shot, mỗi task chỉ bao gồm một mẫu hỗ trợ, trong khi bài toán 2-way 1-shot chứa hai lớp, mỗi lớp một mẫu hỗ trợ.

Trong quá trình huấn luyện, chia dữ liệu thành các episodes cho training, validation và testing theo tỷ lệ 8:1:1. Tập training được sử dụng để cập nhật tham số mô hình, validation để lựa chọn siêu tham số và điều chỉnh chiến lược học, còn test set để đánh giá hiệu năng tổng quát.

3.5 Kết quả thực nghiệm

3.5.1 So sánh kết quả với các phương pháp nền tảng

Kết quả trong Bảng 3.3: Kết quả hiệu suất của các phương pháp khác nhau trong hai thiết lập thử nghiệm 1-way 1-shot và 2-way 1-shot, được đánh giá trên các hiệu số phân loại (0/1 ER) và phân đoạn (mIoU) cho thấy tất cả các mô hình đều hoạt động hiệu quả trong cả nhiệm vụ phân loại và phân đoạn, phù hợp với hiểu biết chung về Few-Shot Learning.

Trong thiết lập 1-way 1-shot: IG-FSL (84.40% ER) và ASNet (83.13% ER) đạt hiệu suất phân loại cao nhất, vượt trội hơn các phương pháp khác. PFENet và PANet có hiệu suất thấp hơn (77.57% và 74.70%) trong phân loại. Về phân đoạn, kết quả mIoU của IG-FSL (46.00%) là tốt nhất, theo sau là ASNet (44.16%), trong khi PANet và PFENet có kết quả thấp hơn (40.10% và 40.82%).

Bảng 3.2 Cấu hình huấn luyện và siêu tham số

Tham số	Giá trị
Kích thước ảnh đầu vào	224×224 pixels
Backbone	ResNet-50 (tiền huấn luyện trên ImageNet)
Optimizer	Adam [106]
Learning rate	10^{-4}
Batch size	8
Số epoch	100
Ngưỡng phân loại δ	0.5
Chia dữ liệu	Train : Val : Test = 8 : 1 : 1
Thiết lập thử nghiệm	1-way 1-shot; 2-way 1-shot; 1-way (1–4 shot)
Cấu hình phần cứng	GPU NVIDIA Tesla T4, 16GB VRAM

Bảng 3.3 Kết quả hiệu suất của các phương pháp khác nhau trong hai thiết lập thử nghiệm

Kết quả hiệu suất				
Phương pháp	1-way 1-shot (%)		2-way 1-shot (%)	
metric	cls. 0/1 ER	seg. mIoU	cls. 0/1 ER	seg. mIoU
PANet	74,70	40,10	62,62	39,78
PFENet	77,57	40,82	62,40	40,22
HSNet	82,11	43,78	63,79	41,52
ASNet	83,13	44,16	63,40	42,35
IG-FSL	84,40	46,00	65,00	43,05

Trong thiết lập 2-way 1-shot: Tất cả các mô hình đều giảm hiệu suất do số lượng lớp cần dự đoán tăng lên. IG-FSL vẫn đạt hiệu suất cao nhất (65.00% ER), tiếp theo là ASNet (64.40%). Phân đoạn (mIoU): IG-FSL dẫn đầu (43.05%), ASNet theo sau (42.35%). PANet và PFENet có sự suy giảm mạnh nhất, đặc biệt trong phân đoạn, với mIoU của PFENet giảm xuống 38.22%.

Trong bài toán 1-way, mô hình chỉ cần xác định liệu ảnh truy vấn có thuộc vào một lớp bệnh đã biết hay không, tức là một bài toán phân loại nhị phân đơn giản. Tuy

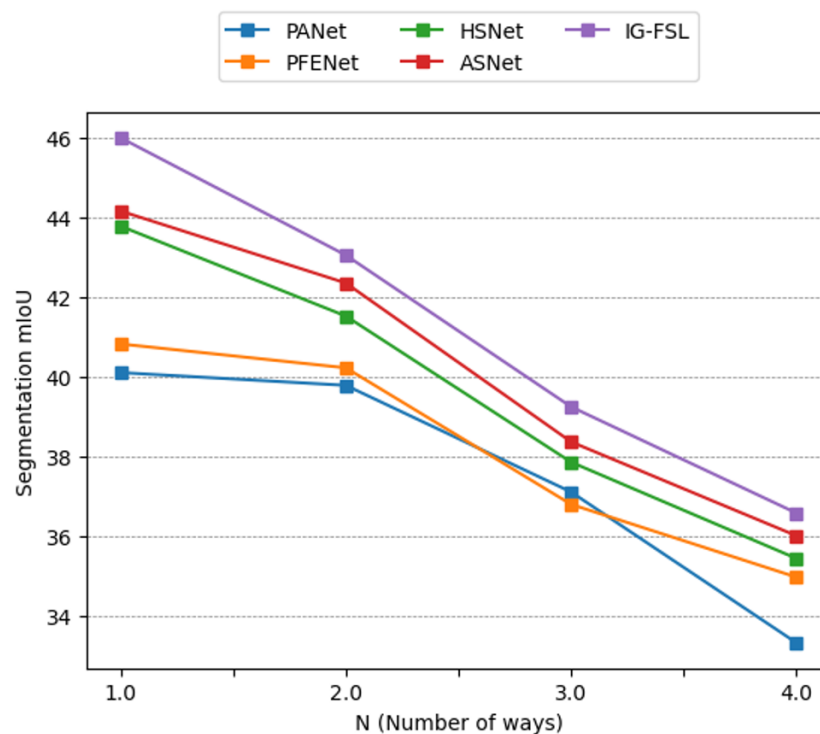
nhiên, trong bài toán 2-way, mô hình phải phân biệt giữa hai lớp khác nhau, làm tăng mức độ phức tạp và kéo theo sự suy giảm độ chính xác.

Xu hướng này cũng thể hiện rõ trong nhiệm vụ phân đoạn. Trong bài toán 1-way, mô hình chỉ cần phát hiện một vùng tổn thương duy nhất, trong khi bài toán 2-way yêu cầu mô hình phải phân biệt hai vùng tổn thương khác nhau trên ảnh. Điều này dẫn đến sự suy giảm đáng kể của chỉ số mIoU, do mô hình phải đối mặt với sự chồng chéo giữa các vùng tổn thương, làm tăng độ khó trong việc xác định ranh giới chính xác.

Nhìn chung, IG-FSL và ASNet thể hiện hiệu suất vượt trội trong cả hai kịch bản thử nghiệm, chứng minh khả năng tổng quát hóa ngay cả khi điều kiện dữ liệu hạn chế. Trong khi đó, PFENet và PANet kém hiệu quả hơn, đặc biệt trong bài toán 2-way 1-shot, khi yêu cầu phân biệt nhiều lớp bệnh hơn.

3.5.2 So sánh kết quả dựa trên số lượng lớp

Phần này tập trung vào việc đánh giá ảnh hưởng của số lượng lớp (ways) đến hiệu suất của mô hình IG-FSL trong bài toán phát hiện và khoanh vùng bệnh bụi phổi silic nghề nghiệp. Các thí nghiệm được thực hiện bằng cách thay đổi số lượng lớp từ 1 đến 4, với mục tiêu phân tích khả năng tổng quát hóa của mô hình khi độ phức tạp của bài toán tăng lên.

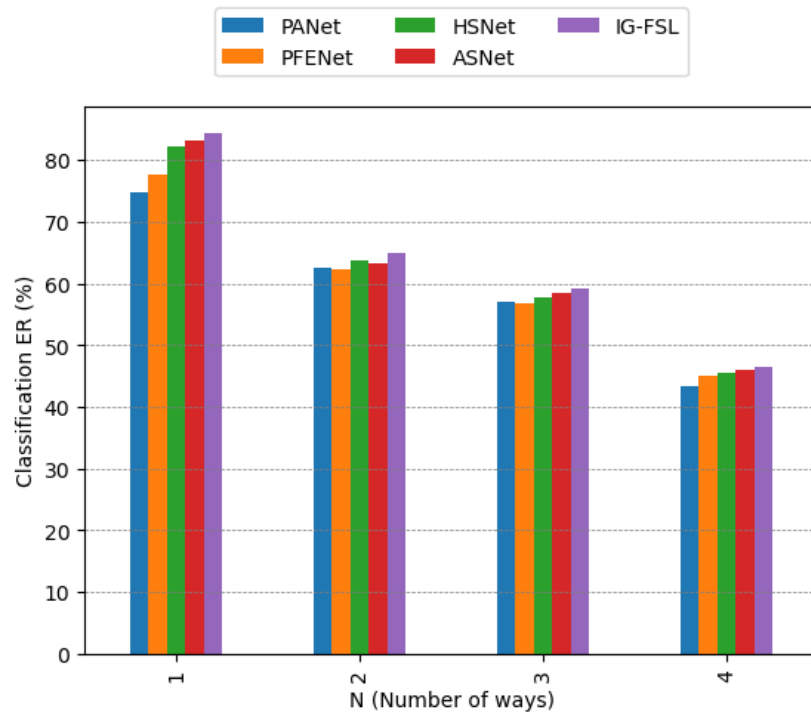


Hình 3.3 Kết quả phân đoạn dựa trên số lượng lớp của FSL

Trong nghiên cứu này, chúng tôi để đánh giá tác động của số lượng lớp bệnh cần phân loại N -way đến với hiệu suất của các mô hình. Thí nghiệm được thực hiện với $N = \{1, 2, 3, 4\}$.

Đối với thiết lập 1-way, lớp bệnh bụi phổi silic được chọn làm mục tiêu cho tập kiểm tra. Khi số lượng lớp tăng lên ($N > 1$), phân chia tập huấn luyện thành $N - 1$ lớp bệnh phổi, giữ lại các lớp còn lại để làm tập kiểm tra (đảm bảo tính không trùng lặp giữa tập huấn luyện và kiểm tra).

Hình 3.3 thể hiện rõ xu hướng suy giảm hiệu suất phân đoạn (mIoU) khi số lượng lớp (ways) tăng lên. Tuy nhiên, IG-FSL duy trì vị trí dẫn đầu ổn định, với chỉ số mIoU cao nhất ở tất cả các thiết lập N . Sự chênh lệch hiệu suất giữa IG-FSL và các phương pháp khác như PANet, PFENet, HSNet và ASNet càng trở nên rõ rệt hơn khi N tăng, cho thấy khả năng khái quát hóa mạnh mẽ của IG-FSL trong điều kiện độ phức tạp cao.



Hình 3.4 Kết quả phân loại dựa trên số lượng lớp của FSL

Hình 3.4 cho thấy xu hướng tương tự đối với nhiệm vụ phân loại: tỷ lệ lỗi (error rate) tăng dần khi số lớp (N) tăng từ 1 lên 4. Tuy nhiên, IG-FSL và ASNet vẫn duy trì được hiệu suất tương đối tốt và ổn định hơn so với các phương pháp còn lại, đặc biệt khi $N = 4$ trong khi PANet và PFENet có xu hướng sụt giảm mạnh hơn.

Điều này nhấn mạnh một điểm quan trọng: bài toán Few-Shot trở nên thách thức hơn khi số lớp mục tiêu tăng, do không gian quyết định phân loại mở rộng, từ đó đòi hỏi

mô hình phải phân biệt được nhiều lớp hơn với số mẫu vẫn giữ nguyên. Việc IG-FSL vẫn đạt hiệu quả cao trong bối cảnh này chứng tỏ tính linh hoạt và khả năng mở rộng (scalability) của mô hình đối với các tình huống thực tiễn phức tạp, như khi phải phân loại nhiều dạng tổn thương phổi khác nhau từ ảnh X-quang.

3.5.3 So sánh hiệu suất giữa các mô hình theo số lượng mẫu

Phần này tập trung vào việc đánh giá ảnh hưởng của số lượng mẫu (shots) đến hiệu suất của mô hình IG-FSL và các phương pháp nền tảng trong bài toán phát hiện và khoanh vùng bệnh bụi phổi silic nghề nghiệp. Các thí nghiệm được thực hiện bằng cách thay đổi số lượng mẫu từ 1-shot đến 4-shot, với mục tiêu phân tích khả năng cải thiện hiệu suất khi mô hình được cung cấp thêm dữ liệu huấn luyện.

Bảng 3.4 So sánh các phương pháp khác nhau dựa trên số liệu phân loại (0/1 ER) và phân đoạn (mIoU)

metric	cls. 0/1 ER (%)				seg. mIoU (%)			
methods	1-shot	2-shot	3-shot	4-shot	1-shot	2-shot	3-shot	4-shot
PANet	74,70	74,91	76,42	78,53	40,10	40,98	42,72	44,84
PFENet	77,57	78,42	79,21	79,97	40,82	41,17	43,34	45,11
HSNet	82,11	83,78	85,34	86,74	43,78	44,44	45,02	46,29
ASNet	83,13	84,22	86,37	88,36	44,16	45,74	46,33	47,55
IG-FSL	84,46	86,09	87,66	89,52	46,00	46,27	47,32	47,89

Kết quả thực nghiệm trong Bảng 3.4 các kịch bản từ 1-shot đến 4-shot, được đánh giá qua hai tiêu chí: Exact Ratio (ER) cho phân loại và Mean Intersection over Union (mIoU) cho phân đoạn.

Nhìn chung, hiệu suất của tất cả các mô hình đều tăng lên rõ rệt khi số lượng mẫu hỗ trợ (shots) tăng, cho thấy lợi ích của việc cung cấp thêm ví dụ học trong bối cảnh học ít mẫu. Cụ thể: PANet cải thiện nhẹ: tỷ lệ lỗi phân loại giảm từ 74.70% đến 78.53%, mIoU tăng từ 40.10% đến 44.84%. PFENet cũng thể hiện sự tiến bộ: phân loại tăng từ 77.57% đến 79.97%, phân đoạn từ 40.82% đến 45.11%. Các mô hình mạnh hơn như HSNet và ASNet cho thấy mức cải thiện rõ rệt hơn, đặc biệt là ở các thiết lập 3-shot và 4-shot. ASNet đạt 88.36% chính xác phân loại và 47.55% mIoU ở 4-shot – tiệm cận với mô hình đề xuất.

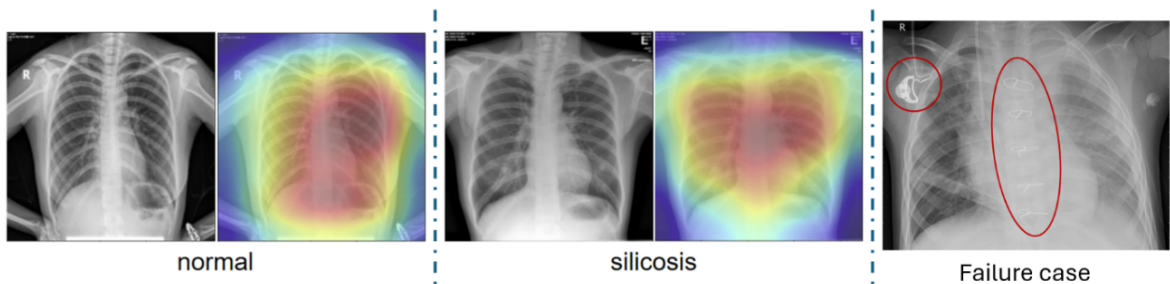
Hiệu năng vượt trội của IG-FSL: Mô hình IG-FSL tiếp tục khẳng định vị trí dẫn đầu ở cả phân loại và phân đoạn. Phân loại: đạt 89.52% chính xác trong 4-shot – cao nhất trong tất cả các mô hình. Phân đoạn: đạt 47.89% mIoU, cũng là giá trị cao nhất được ghi nhận.

Các kết quả trình bày trong Bảng 3.4 cho thấy IG-FSL không chỉ hoạt động hiệu quả trong điều kiện cực kỳ hạn chế về dữ liệu (1-shot) mà còn tăng ổn định và mạnh mẽ khi được cung cấp thêm dữ liệu huấn luyện. Điều này khẳng định tính linh hoạt trong khả năng mở rộng (scalability) và tính thực tiễn của mô hình trong bối cảnh hạn chế dữ liệu y tế.

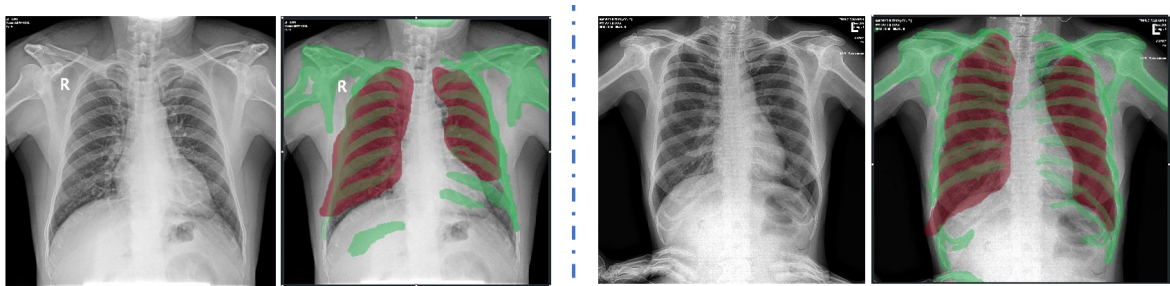
3.5.4 Trực quan hóa kết quả

Trực quan hóa vùng chú ý của mô hình bằng Grad-CAM

Hình 3.5 trình bày kết quả trực quan hóa bằng phương pháp Grad-CAM, nhằm minh họa vùng tập trung chú ý của mô hình trong ba tình huống: (1) ảnh X-quang bình thường, (2) ảnh mắc bệnh bụi phổi silic, và (3) một trường hợp mô hình dự đoán sai.



Hình 3.5 Bản đồ Grad-CAM minh họa vùng chú ý của IG-FSL



Hình 3.6 Kết quả phân đoạn trong trường hợp xét hai nhãn, trong đó vùng xanh lá cây thể hiện phần xương, vùng đỏ thể hiện vùng phổi.

Kết quả cho thấy mô hình IG-FSL có xu hướng tập trung vào vùng phổi, đặc biệt

là các vùng tổn thương. Điều này phản ánh khả năng học đúng đặc trưng có giá trị chẩn đoán, dù mô hình chỉ được huấn luyện với số lượng mẫu rất hạn chế. Ở các trường hợp dự đoán sai, bản đồ Grad-CAM cho thấy mô hình bị nhiễu bởi các yếu tố không liên quan như thiết bị y tế cấy ghép hoặc bất thường giải phẫu, cho thấy độ nhạy của mô hình với các đặc trưng ngoại lai. Việc sử dụng Grad-CAM góp phần nâng cao khả năng diễn giải và độ tin cậy của mô hình IG-FSL.

Trực quan hóa kết quả phân đoạn

Hình 3.6 trình bày kết quả phân đoạn của mô hình IG-FSL trong bài toán mở rộng có nhãn cấu trúc. Kết quả cho thấy mô hình tái tạo khá chính xác biên dạng của hai lá phổi, đặc biệt ở vùng trung tâm và đỉnh phổi. Một số sai lệch nhỏ xuất hiện ở vùng đáy phổi và rìa gần cột sống, do tương phản thấp hoặc bóng tim che phủ. Việc xác định rõ vùng phổi và cấu trúc xương cũng hỗ trợ loại bỏ nhiễu nền, là tiền đề để phát triển mô hình chẩn đoán tích hợp (classification-segmentation) trong các nghiên cứu tiếp theo.

Ma trận nhầm lẫn

Bảng 3.5 cho thấy mô hình IG-FSL đạt hiệu quả phân loại tốt trên bốn lớp bệnh phổi. Các lớp viral pneumonia và bacterial pneumonia được nhận diện chính xác cao nhất (trên 75%), trong khi lớp silicosis có độ chính xác thấp hơn (66%), chủ yếu bị nhầm với lớp normal (9 mẫu). Nguyên nhân là do đặc trưng nốt mờ nhỏ, phân bố thưa của bệnh bụi phổi silic giai đoạn sớm dễ bị mô hình nhầm lẫn với ảnh phổi bình thường. Mô hình vẫn duy trì khả năng phân biệt tốt các lớp bệnh chính, đặc biệt trong điều kiện học ít mẫu.

Bảng 3.5 Ma trận nhầm lẫn kết quả phân loại

True \ Pred	BPSi	Bình thường	Vi khuẩn	Virus
BPSi	30	9	2	4
Bình thường	7	40	2	2
Vi khuẩn	2	5	39	6
Virus	3	1	7	42

3.6 Kết luận chương 3

Chương 3 đã trình bày chi tiết quá trình đề xuất và đánh giá mô hình học sâu ít mẫu IG-FSL nhằm hỗ trợ phát hiện và khoanh vùng bệnh bụi phổi silic nghề nghiệp từ hình ảnh X-quang ngực. Với đặc thù của bài toán y tế có dữ liệu hạn chế, IG-FSL được

phát triển dựa trên framework FS-CS, tích hợp hiệu quả hai nhiệm vụ phân loại đa nhãn và phân đoạn trong thiết lập N -way K -shot, sử dụng kiến trúc ResNet-50 làm bộ trích xuất đặc trưng.

Kết quả thực nghiệm cho thấy IG-FSL đạt hiệu suất vượt trội trong việc phân biệt các lớp bệnh, đặc biệt với bụi phổi silic, dù chỉ sử dụng số lượng ảnh gán nhãn rất ít, cụ thể đạt 84,4% ER và 46,0% mIoU trong kịch bản 1-shot, và 89,52% ER cùng 47,89% mIoU trong kịch bản 4-shot. Việc trực quan hóa vùng chú ý bằng Grad-CAM đã xác nhận rằng mô hình tập trung đúng vào các đặc trưng tổn thương có ý nghĩa chẩn đoán, như các nốt trắng nhỏ trong silicosis, nhưng đồng thời phát hiện một số yếu tố nhiễu từ thiết bị y tế làm ảnh hưởng đến độ tin cậy của dự đoán.

Chương 3 khẳng định tính khả thi và hiệu quả của việc áp dụng Few-Shot Learning khi dữ liệu khan hiếm, đồng thời mở ra một hướng tiếp cận thực tiễn cho các hệ thống hỗ trợ chẩn đoán hình ảnh tại các cơ sở y tế có điều kiện hạn chế, với các hướng nghiên cứu tương lai bao gồm cải tiến kiến trúc mô hình và tích hợp dữ liệu lâm sàng, một phần nội dung này đã được NCS công bố trong công trình CT1.

CHƯƠNG 4: ĐỀ XUẤT MÔ HÌNH HỌC SÂU HỖ TRỢ PHÂN LOẠI BỆNH BỤI PHỔI SILIC NGHỀ NGHIỆP

Để giải quyết bài toán phát hiện và phân loại bệnh bụi phổi silic từ ảnh X-quang ngực, luận án tập trung vào hai hướng tiếp cận chính:

Bài toán phân loại nhị phân: Nhằm xác định sự hiện diện của bệnh bụi phổi silic, bằng cách phân biệt giữa hai nhóm: *bình thường* và *bụi phổi silic nghề nghiệp*.

Bài toán phân loại đa lớp: Nhằm nhận diện và phân biệt nhiều loại bệnh phổi trên ảnh X-quang, bao gồm: *bụi phổi silic*, *viêm phổi do vi khuẩn*, *viêm phổi do virus* và *bình thường*. Bài toán này phản ánh tình huống thực tế tại cơ sở y tế, nơi bác sĩ cần phân biệt tổn thương phổi có nhiều nguyên nhân khác nhau.

Từ hai bài toán cốt lõi trên, luận án xác lập hệ thống sáu câu hỏi nghiên cứu (Research Questions – RQ), đóng vai trò định hướng cho việc xây dựng mô hình, thiết kế thực nghiệm và phân tích kết quả:

RQ1: Mức độ cải thiện của mô hình Graph Transformer Post-hoc (GTP) so với các mô hình học sâu truyền thống trong phân loại bệnh bụi phổi silic là như thế nào?

RQ2: Việc sử dụng hàm mất mát cân bằng Balanced Cross-Entropy (BalCE) có giúp nâng cao hiệu quả phân loại đối với các lớp bệnh có ít dữ liệu hay không?

RQ3: Kỹ thuật học tổng hợp (ensemble learning) có giúp cải thiện độ chính xác và độ ổn định của mô hình trong các kịch bản thực nghiệm khác nhau hay không?

RQ4: Hiệu quả phân loại của mô hình được phản ánh như thế nào thông qua các chỉ số đánh giá như độ nhạy, độ đặc hiệu, AUC và ma trận nhầm lẫn?

RQ5: Mức độ diễn giải mô hình đối với bệnh bụi phổi silic có giúp làm nổi bật được các vùng tổn thương đặc trưng trên ảnh X-quang hay không?

RQ6: Hiệu suất mô hình trí tuệ nhân tạo được đề xuất có thể so sánh như thế nào với chẩn đoán của chuyên gia lâm sàng trong một số trường hợp cụ thể?

4.1 Ứng dụng các mô hình học sâu trong hỗ trợ phân loại bệnh bụi phổi silic nghề nghiệp

4.1.1 Phát biểu bài toán

Input: Tập dữ liệu ảnh X-quang ngực DNNBPSi.

Output: Nhãn nhị phân tương ứng với mỗi ảnh đầu vào, cho biết ảnh đó thuộc về

bệnh nhân bình thường (0) hoặc mắc bệnh bụi phổi silic (1).

Cho một tập dữ liệu huấn luyện $D = \{(x_i, y_i)\}_{i=1}^N$, trong đó:

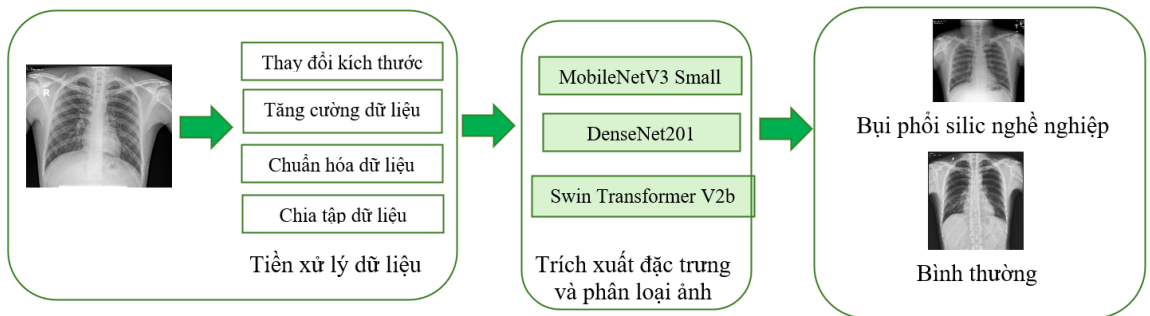
- $x_i \in \mathbb{R}^{H \times W \times C}$ là ảnh X-quang ngực đã được tiền xử lý với H là chiều cao, W là chiều rộng, C là số kênh màu của ảnh.
- $y_i \in \{0, 1\}$ là nhãn nhị phân tương ứng với tình trạng bệnh của ảnh, trong đó $y_i = 0$ là bệnh nhân không mắc bệnh bụi phổi silic nghề nghiệp, $y_i = 1$ là bệnh nhân mắc bệnh bụi phổi silic nghề nghiệp.

Mục tiêu là tìm một hàm ánh xạ tối ưu $f_\theta : \mathbb{R}^{H \times W \times C} \rightarrow [0, 1]$ sao cho $\hat{y}_i = f_\theta(x_i) \approx y_i$, trong đó $\hat{y}_i$ là xác suất dự đoán rằng ảnh x_i thuộc lớp dương tính (bệnh nhân có khả năng mắc bệnh bụi phổi silic nghề nghiệp).

4.1.2 Kiến trúc chung của mô hình học sâu áp dụng trong chẩn đoán ảnh X-quang

Như đã đề cập trong Chương 1, nhiều mô hình học sâu đã được phát triển để hỗ trợ chẩn đoán viêm phổi từ ảnh X-quang ngực, với các chỉ số hiệu suất đa dạng nhằm đánh giá tính hiệu quả. Tuy nhiên, một thách thức lớn là xác định mô hình đạt được sự cân bằng tối ưu giữa hiệu năng phân loại và chi phí tính toán, đồng thời phù hợp với các điều kiện triển khai thực tế trong lâm sàng, đặc biệt đối với các bệnh lý có tổn thương mờ nhạt và phân bố không đồng đều như bệnh bụi phổi silic nghề nghiệp.

Trên cơ sở đó, Luận án lựa chọn tiếp cận bài toán phân loại ảnh X-quang ngực thành hai lớp bằng các kiến trúc học sâu có đặc điểm bộ trợ lẫn nhau: MobileNetV3-small, DenseNet201 và Swin Transformer. Tổng quan về kiến trúc của hệ thống học sâu được trình bày trong Hình 4.1.



Hình 4.1 Kiến trúc khung mô hình học sâu áp dụng trong phân loại bệnh bụi phổi silic nghề nghiệp

Khung học sâu được thiết kế gồm hai tầng chính:

Tầng thứ nhất – tiền xử lý ảnh: Bao gồm các bước cơ bản như chuyển đổi định dạng ảnh từ DICOM sang PNG hoặc JPEG, chuẩn hóa kích thước ảnh về $224 \times 224 \times 3$, và áp dụng các kỹ thuật tăng cường dữ liệu. Dữ liệu được chia thành hai tập: tập huấn luyện và tập kiểm tra, phục vụ cho việc huấn luyện và đánh giá mô hình.

Tầng thứ hai – mô hình học sâu: Ảnh sau tiền xử lý được đưa vào các kiến trúc học sâu (MobileNetV3-small, DenseNet201, Swin Transformer) để trích xuất đặc trưng và thực hiện phân loại.

Sau khi trích xuất đặc trưng, ảnh được đưa qua các tầng phân loại và hàm sigmoid để dự đoán xác suất thuộc lớp “bình thường” hoặc “mắc bệnh”.

4.1.3 Kết quả thực nghiệm

4.1.3.1 Thiết lập thực nghiệm

Các thí nghiệm được triển khai trên nền tảng phần cứng gồm GPU NVIDIA Tesla T4 (RAM 15GB), sử dụng thư viện học sâu PyTorch để xây dựng, huấn luyện và đánh giá mô hình. Mục tiêu là so sánh hiệu quả phân loại giữa ba kiến trúc học sâu: MobileNetV3_small, DenseNet201, và Swin Transformer, trong cùng một điều kiện huấn luyện và cùng một tập dữ liệu, nhằm đảm bảo tính công bằng và nhất quán trong đánh giá.

Tập dữ liệu DNNBPSi bao gồm các ảnh X-quang ngực được gán nhãn nhị phân. Dữ liệu được chia theo tỷ lệ 80:20 thành tập huấn luyện và tập kiểm tra, chi tiết ở Bảng 4.1.

Bảng 4.1 Cấu trúc phân bố dữ liệu DNNBPSi

Dữ liệu	Bụi phổi Silic	Bình thường	Tổng số
Tập huấn luyện	428	1244	1672
Tập kiểm tra	107	310	417
Tổng số	535	1554	2089

Các mô hình được huấn luyện với cùng số epoch, batch size, phương pháp tăng cường dữ liệu và kỹ thuật dừng sớm (early stopping). Bảng 4.2 trình bày chi tiết các siêu tham số được áp dụng cho từng mô hình.

Bảng 4.2 Cấu hình huấn luyện của các mô hình học sâu

Tham số	MobileNetV3 Small	DenseNet201	Swin Transformer
Kích thước ảnh đầu vào	$224 \times 224 \times 3$	$224 \times 224 \times 3$	$224 \times 224 \times 3$
Batch size	32	32	32
Số epoch	50	50	50
Hàm mất mát	Cross-Entropy	Cross-Entropy	Cross-Entropy
Bộ tối ưu hóa	Adam	Adam	Adam
Learning rate ban đầu	$1 \cdot 10^{-4}$	$1 \cdot 10^{-4}$	$1 \cdot 10^{-4}$
Weight Decay	0.0001	0.0001	0.0001
Early Stopping	Patience = 10	Patience = 10	Patience = 10
Tăng cường dữ liệu	Có	Có	Có

4.1.3.2 Kết quả và nhận xét

Độ chính xác và hiệu suất

Bảng 4.3 trình bày kết quả phân loại của ba mô hình học sâu: MobileNetV3-small, DenseNet201 và Swin Transformer trên cùng tập dữ liệu DNNBPSi. Trong số đó, Swin Transformer thể hiện hiệu suất vượt trội, với độ chính xác cao nhất trên cả tập huấn luyện (95,36%) và tập kiểm tra (86,13%). Đồng thời, mô hình này đạt chỉ số F1-macro lên tới 82,96, phản ánh khả năng cân bằng tốt giữa độ nhạy và độ đặc hiệu trên các lớp bệnh, đặc biệt quan trọng trong bối cảnh dữ liệu không đồng đều và tồn tại hiện tượng mất cân bằng lớp nghiêm trọng. Điều này cho thấy Swin Transformer không chỉ học tốt trên tập huấn luyện mà còn có năng lực khái quát hóa cao trên dữ liệu thực tế, vượt trội hơn so với các mô hình còn lại.

Phân tích ma trận nhầm lẫn

Ma trận nhầm lẫn (Bảng 4.4) của mô hình Swin Transformer cho thấy khả năng nhận diện chính xác cả hai lớp. Tuy nhiên, số lượng trường hợp bỏ sót bệnh (FN = 33) vẫn còn đáng lưu ý, ảnh hưởng đến độ nhạy. Mặc dù vậy, độ chính xác tổng thể và khả năng phân biệt lớp “bình thường” rất tốt (TN = 277), khẳng định tính ổn định của mô hình.

Ma trận nhầm lẫn cho Swin Transformer:

Bảng 4.3 Độ chính xác và hiệu suất giữa các mô hình

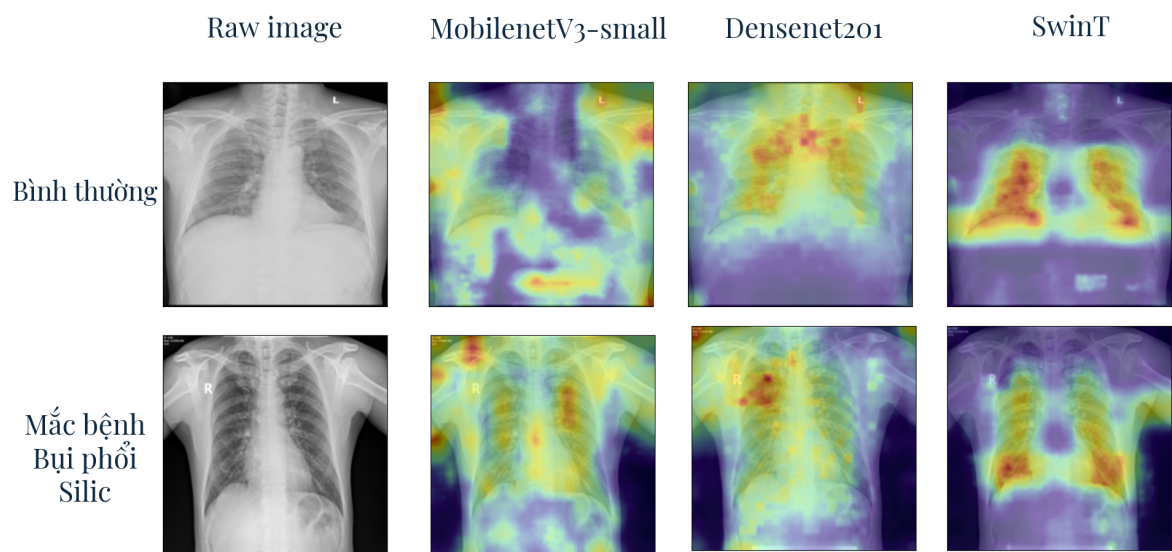
Mô hình	Accuracy		F1-macro
	Huấn luyện	Kiểm tra	
MobileNetV3-small	73.29	74.34	42.64
Densenet201	78.88	76.98	57.32
Swin Transformer	95.36	86.13	82.96

Bảng 4.4 Kết quả ma trận nhầm lẫn

		Thực tế	
		Bụi phổi silic	Bình thường
Dự đoán	Bụi phổi silic	89	18
	Bình thường	33	277

Phân tích Grad-Cam

Phân tích Grad-CAM trong Hình 4.2 cho thấy Swin Transformer tập trung mạnh vào vùng phổi, đặc biệt tại các vị trí tổn thương dạng nốt hoặc xơ hóa, giúp mô hình đưa ra quyết định chính xác. DenseNet201 và MobileNetV3-small có xu hướng phân tán sự chú ý ra khỏi vùng tổn thương, khiến chất lượng phân loại giảm.



Hình 4.2 Kết quả Grad-CAM của các mô hình học sâu

Trên các ảnh bình thường, Swin Transformer vẫn duy trì sự tập trung chủ yếu vào cấu trúc phổi, tránh nhận diện sai vào các vùng không liên quan (như xương sườn, mô mềm ngoài phổi. . .). Những kết quả này chứng minh rằng cơ chế attention trong Swin Transformer không chỉ mang lại hiệu suất tốt về định lượng, mà còn đảm bảo tính hợp lý về mặt ngữ nghĩa và cấu trúc ảnh y học.

4.2 Đề xuất mô hình học sâu trên đồ thị

4.2.1 Phát biểu bài toán

Phát biểu bài toán

Input: Tập dữ liệu X-quang ngực

Tập dữ liệu gồm n mẫu, với mỗi mẫu (x_i, y_i) bao gồm một ảnh đầu vào x_i và nhãn đầu ra tương ứng y_i được lấy từ $X \times Y$, trong đó:

$X \subset \mathbb{R}^{3 \times H \times W}$ là không gian đầu vào, mỗi ảnh được biểu diễn dưới dạng một ma trận kích thước $H \times W$, đại diện cho chiều cao và chiều rộng của ảnh.

$Y \subset \{1, 0\}^4$ là không gian nhãn đầu ra, với mỗi nhãn y_i là một vector one-hot biểu diễn xác suất ảnh thuộc về một trong 4 lớp: “bình thường”, “bụi phổi silic”, “viêm phổi do vi khuẩn”, “viêm phổi do virus”.

Output: Vector one-hot $y_i \in \{1, 0\}^4$ dự đoán lớp của ảnh x_i .

Mục tiêu: Huấn luyện một mô hình $f_\theta(x_i)$, trong đó $f_\theta : X \times Y$, ánh xạ từ x_i sang y_i và tối ưu hóa hàm mất mát để giảm thiểu sai số giữa nhãn thực tế và dự đoán.

Mô hình bao gồm hai phần chính:

Bộ mã hóa (encoder) $v : \mathcal{F}(x \mid \theta_w) \in \mathbb{R}^d$ chịu trách nhiệm trích xuất các vector đặc trưng kích thước d từ ảnh X-quang.

Bộ phân loại (classifier) $z = \mathcal{W}(v \mid \theta_w) \in \mathbb{R}^4$ tạo ra các giá trị logit z , dự đoán xác suất ảnh thuộc về các lớp bệnh.

Cuối cùng, mô hình được tối ưu hóa dựa trên các giá trị logit z và cải thiện hiệu suất phân loại thông qua việc giảm thiểu hàm mất mát.

4.2.2 Đề xuất kiến trúc Graph Transformer Post-hoc

Kiến trúc GTP được xây dựng dựa trên ý tưởng của Graph Transformer và cơ chế attention toàn cục của Graph Attention Networks. Mặc dù được gọi là “post-hoc”, toàn bộ pipeline vẫn được huấn luyện theo cơ chế end-to-end, với các backbone (DenseNet201, Swin Transformer, MobileNetV3) và GTP được tối ưu đồng thời dựa

trên cùng một hàm mất mát phân loại. So với cách tiếp cận truyền thống chỉ sử dụng bộ phân loại tuyến tính sau encoder, việc tích hợp GTP mang lại ưu điểm để giải quyết bài toán dữ liệu y tế "tinh vi", khó phân biệt và có sự chênh lệch số lượng giữa các lớp bệnh.

Trong luận án, GTP được thiết kế như một module dạng plugin có thể tích hợp linh hoạt phía sau bất kỳ backbone nào. Thiết kế này giúp mô hình nâng cao độ chính xác phân loại, trong khi vẫn duy trì tốc độ suy luận (inference time) và mức sử dụng bộ nhớ (memory usage), đảm bảo hiệu quả đồng thời về hiệu năng và tài nguyên tính toán.

Ngoài ra, GTP có khả năng tận dụng thông tin về mối liên hệ giữa các ảnh bụi phổi, được biểu diễn dưới dạng thuộc tính cạnh (edge attributes) nếu dữ liệu này được cung cấp. Khi có thông tin này, mô hình có thể học và duy trì trật tự cũng như tính nhất quán của dữ liệu ảnh ban đầu, ngay cả khi kích thước batch giảm trong giai đoạn suy luận. Trong trường hợp không có thông tin liên kết, các thuộc tính cạnh được khởi tạo rỗng và mô hình vẫn hoạt động bình thường. Cơ chế này giúp GTP duy trì độ ổn định và độ chính xác cao hơn trong quá trình hoạt động, đồng thời mở rộng khả năng biểu diễn cấu trúc dữ liệu mà không làm tăng đáng kể chi phí tính toán.

Kiến trúc GTP gồm hai thành phần chính:

- (i) bộ trích xuất đặc trưng ảnh.
- (ii) bộ lan truyền, tổng hợp thông tin trên đồ thị bằng cơ chế attention.

4.2.2.1 Trích xuất đặc trưng ảnh

DNN encoder: trích xuất đặc trưng ảnh

Cho một tập huấn luyện gồm b ảnh X-quang ngược $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^b$, mỗi ảnh được đưa qua backbone học sâu để tạo ra vector đặc trưng $\mathbf{c}_i \in \mathbb{R}^d$.

Tập đặc trưng $F = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_b\}$ đóng vai trò làm nút trong đồ thị

4.2.2.2 Lan truyền và tổng hợp thông tin trên đồ thị

GTN encoder: lan truyền, tổng hợp thông tin trên đồ thị bằng cơ chế attention

Từ tập đặc trưng, xây dựng đồ thị đầy đủ $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{F})$ [109].

Xây dựng một đồ thị đầy đủ từ các vector đặc trưng của các hình ảnh trong cùng một batch.

$\mathcal{V}$ là tập hợp các hình ảnh trong mỗi batch, với $|\mathcal{V}| = b$.

$\mathcal{E} = \{\mathbf{e}_{ij}\}_{i,j=\overline{1,b}}$ là tập hợp các cạnh kết nối các hình ảnh.

$\mathcal{F} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_b\}$ là tập hợp các đặc trưng của các nút trong đồ thị.

Các trọng số cạnh $\mathbf{e}_{ij}$ là các tham số có thể học được, được điều chỉnh động theo dữ liệu huấn luyện.

Cơ chế hoạt động của GTN Encoder: GTN sử dụng cơ chế chú ý để xử lý thông tin trong đồ thị, giúp mô hình học được cách các nút (đặc trưng của hình ảnh) và cạnh (mối quan hệ giữa các đặc trưng) tương tác với nhau.

Với mỗi nút nguồn i , GTN tính toán việc tổng hợp thông tin từ các nút khác j trong đồ thị thông qua công thức:

Tổng hợp thông tin từ các nút láng giềng

Với mỗi nút nguồn i , vector đặc trưng tổng hợp được tính bởi công thức:

$$\hat{\mathbf{c}}_i = \sum_{j \in \mathcal{N}(i)} \alpha_{i,j} (\mathbf{W}_v \mathbf{c}_j + \mathbf{W}_e \mathbf{e}_{ij}). \quad (4.1)$$

$\mathcal{N}(i)$ là tập hợp tất cả các nút láng giềng của nút i ,

$\hat{\mathbf{c}}_i$ là vector đặc trưng cập nhật cho nút i ,

$\alpha_{i,j}$ là trọng số chú ý, dùng để xác định mức độ quan trọng (trọng số) của nút j khi tổng hợp thông tin cho nút i ,

$\mathbf{W}_v$ là bộ tham số của mô hình dùng để biến đổi vector đặc trưng $\mathbf{c}_j$ của nút j ,

$\mathbf{W}_e$ là bộ tham số của mô hình dùng để điều chỉnh trọng số cạnh $\mathbf{e}_{ij}$ giữa nút i, j .

GTN encoder tạo ra sự tương tác giữa các hình ảnh trong batch, giúp mô hình phân biệt được các đặc trưng khác nhau và xử lý tốt hơn các vấn đề phức tạp như tương đồng cao giữa các hình ảnh hoặc mất cân bằng dữ liệu.

Công thức trên cho phép mô hình tổng hợp thông tin từ các nút láng giềng, với trọng số chú ý $\alpha_{i,j}$ được tính toán để xác định mức độ ảnh hưởng của từng nút láng giềng đối với nút trung tâm. Trọng số chú ý này giúp GTN nhận diện các mối quan hệ quan trọng giữa các vùng trong ảnh X-quang, từ đó cải thiện độ chính xác trong quá trình phân loại. Các ma trận trọng số $\mathbf{W}_v$ và $\mathbf{W}_e$ giúp điều chỉnh và học được các đặc trưng từ nút và cạnh, làm cho việc tổng hợp thông tin trở nên hiệu quả hơn.

Tính hệ số chú ý (Attention Coefficient) $\alpha_{i,j}$

$$\alpha_{i,j} = \text{softmax} \left(\frac{(\mathbf{W}_q \mathbf{c}_i)^\top (\mathbf{W}_k \mathbf{c}_j + \mathbf{W}_e \mathbf{e}_{ij})}{\sqrt{d}} \right), \quad (4.2)$$

$\alpha_{i,j}$ xác định mức độ quan trọng của nút j đối với nút i . Hệ số này được tính thông qua “dot-product attention” (chú ý dựa trên tích chấm), sau đó áp dụng hàm softmax để chuẩn hóa. Giá trị này cho biết mức độ ảnh hưởng của nút j đối với nút i .

W_q, W_k : là các bộ tham số được dùng để tính toán giá trị ‘key’ và ‘query’ trong module attention.

$W_q e_i$ là phép biến đổi đặc trưng e_i của nút i thông qua ma trận tham số W_q .

$W_k e_j$ là phép biến đổi đặc trưng e_j của nút j thông qua ma trận tham số W_k . Nó tạo ra vector key cho nút j .

$W_e e_{ij}$ là phép biến đổi trọng số cạnh e_{ij} của nút j thông qua ma trận tham số W_k . Nó tạo ra vector key cho nút j .

$\sqrt{d}$ là giá trị chuẩn hóa, với d là số chiều của các vector đặc trưng, được dùng để tránh giá trị dot product quá lớn, dẫn đến việc làm mất cân bằng trong quá trình tính attention, giúp ổn định quá trình huấn luyện.

Softmax: Hàm softmax được áp dụng lên tất cả các hệ số chú ý $\alpha_{i,j}$ để chuẩn hóa các trọng số attention, đảm bảo tổng các giá trị của $\alpha_{i,j}$ cho tất cả các nút láng giềng $j \in \mathcal{N}(i)$ là sẽ bằng 1.

Việc tính toán hệ số chú ý giúp mô hình học được nút nào đóng vai trò quan trọng hơn trong quá trình tổng hợp thông tin.

Kết hợp thông tin residual và thông tin tổng hợp

Sau khi tính toán hệ số chú ý $\alpha_{i,j}$ ở công thức (4.2), mô hình đã xác định được mức độ ảnh hưởng của các nút láng giềng j đối với nút i . Quá trình này giúp mô hình thu thập thông tin từ các nút lân cận dựa trên tầm quan trọng của chúng. Tuy nhiên, việc chỉ dựa vào thông tin từ các nút láng giềng có thể làm mất đi các đặc trưng quan trọng của nút i ban đầu. Để khắc phục vấn đề này, mô hình áp dụng một cơ chế residual nhằm kết hợp cả thông tin gốc từ nút i và thông tin tổng hợp từ các nút láng giềng.

Tiếp theo, trong công thức (4.3), đặc trưng e_i của nút i được tính bằng cách kết hợp giữa thông tin ban đầu r_i và thông tin mới $\hat{e}_i$ đã tổng hợp từ các nút lân cận, thông qua một hệ số điều khiển β_i . Cụ thể, quá trình kết hợp này được điều chỉnh dựa trên mức độ ảnh hưởng của từng loại thông tin, giúp mô hình duy trì sự cân bằng giữa đặc trưng ban đầu và thông tin thu thập được từ các nút khác.

$$c'_i = \beta_i r_i + (1 - \beta_i) \hat{e}_i \quad (4.3)$$

Công thức (4.3) nhằm mục tiêu duy trì thông tin quan trọng từ nút ban đầu i , đồng

thời kết hợp với thông tin đã được tổng hợp từ các nút láng giềng. Quá trình này giúp mô hình cân bằng giữa việc giữ lại đặc trưng gốc và học từ các đặc trưng không gian phức tạp, hỗ trợ mô hình trong việc tránh mất đi các đặc điểm quan trọng trong quá trình tổng hợp thông tin từ các nút khác, từ đó có thể giảm thiểu hiện tượng over-smoothing.

Công thức tính hệ số β_i để kết hợp thông tin

$$r_i = \mathbf{W}_r \mathbf{c}_i,$$

$$\beta_i = \text{sigmoid}(\mathbf{W}_g[\hat{\mathbf{c}}_i; \mathbf{r}_i; \hat{\mathbf{c}}_i - \mathbf{r}_i]). \quad (4.4)$$

Hệ số β_i được tính bằng hàm sigmoid, giúp giá trị nằm trong khoảng từ 0 đến 1. Nó giúp mô hình điều chỉnh tỉ trọng giữa thông tin residual r_i và thông tin tổng hợp $\hat{\mathbf{c}}_i$.

$\mathbf{W}_g$ là trọng số học được của mô hình, dùng để biến đổi đặc trưng đầu vào.

Hệ số này đảm bảo rằng sự kết hợp giữa hai nguồn thông tin không bị quá mức thiên lệch về một phía, giúp mô hình tránh hiện tượng over-smoothing.

Biến đổi và chuẩn hóa đặc trưng

Sau khi kết hợp thông tin từ residual và các nút láng giềng, mô hình sẽ biến đổi đặc trưng qua lớp tuyến tính, sau đó chuẩn hóa đặc trưng để đảm bảo các đặc trưng đầu ra có dạng hợp lý, ổn định hơn và hỗ trợ quá trình huấn luyện hiệu quả. Công thức (4.5) mô tả quá trình này:

$$\mathbf{v}_i = \text{BatchNorm}(\text{Linear}(\mathbf{c}'_i)). \quad (4.5)$$

Vector đặc trưng $\mathbf{v}_i$ của nút i được chuẩn hóa thông qua hai bước:

$\text{Linear}(\mathbf{c}'_i)$: Biến đổi vector đặc trưng $\mathbf{c}'_i$ thông qua một lớp tuyến tính. BatchNorm : Áp dụng chuẩn hóa trên vector kết quả, giúp giảm thiểu hiện tượng lệch phân phối và tăng tốc độ huấn luyện.

Luồng xử lý tổng quan của mô hình đề xuất

Trong thiết kế đề xuất, embedding của mỗi nút được trích xuất từ DNN encoder được ký hiệu là $\mathbf{c}_i \in \mathbb{R}^{d_c}$, trong đó d_c biểu thị số chiều đặc trưng của mỗi mẫu ảnh. Mỗi cạnh nối một cặp nút (i, j) trong đồ thị đầy đủ được biểu diễn bằng một vector embedding có thể học được $\mathbf{e}_{ij} \in \mathbb{R}^{d_e}$, với d_e là số chiều embedding của cạnh. Để bảo đảm cả thông tin nút và cạnh đều được chiếu vào một không gian tiềm ẩn chung, ta đưa vào một tập các ma trận trọng số có thể huấn luyện. Cụ thể, ma trận biến đổi nút được định nghĩa là $\mathbf{W}_v \in \mathbb{R}^{m \times d_c}$, trong khi ma trận biến đổi cạnh là $\mathbf{W}_e \in \mathbb{R}^{m \times d_e}$, cả hai cùng ánh xạ đầu vào của chúng vào không gian tiềm ẩn có số chiều m .

Đối với cơ chế attention, các phép chiếu query và key lần lượt được tham số hóa bởi $\mathbf{W}_q \in \mathbb{R}^{m \times d_c}$ và $\mathbf{W}_k \in \mathbb{R}^{m \times d_c}$. Nhờ vậy, hệ số attention $\alpha_{i,j}$ được tính như một vô hướng thông qua tích vô hướng giữa query đã biến đổi $\mathbf{W}_q \mathbf{c}_i$ và tổ hợp giữa key cùng embedding cạnh, tức $\mathbf{W}_k \mathbf{c}_j + \mathbf{W}_e \mathbf{e}_{ij}$. Nhánh residual được tham số hóa bởi $\mathbf{W}_r \in \mathbb{R}^{m \times d_c}$, chiếu embedding gốc của nút vào cùng không gian tiềm ẩn. Một kết nối residual có cổng (gated residual connection) được bổ sung thông qua công thức $\beta_i = \sigma(\mathbf{W}_g [\hat{\mathbf{c}}_i; \mathbf{r}_i; \hat{\mathbf{c}}_i - \mathbf{r}_i])$, trong đó $\mathbf{W}_g \in \mathbb{R}^{1 \times 3m}$ và $\sigma(\cdot)$ là hàm sigmoid, đảm bảo $\beta_i \in (0, 1)$.

Cuối cùng, biểu diễn nút đã cập nhật $\mathbf{c}'_i \in \mathbb{R}^m$ được ánh xạ qua một lớp fully connected với tham số $\mathbf{W}_{\text{lin}} \in \mathbb{R}^{d \times m}$ và bias $\mathbf{b}_{\text{lin}} \in \mathbb{R}^d$, sau đó được chuẩn hóa bằng BatchNorm, để thu được embedding cuối cùng $\mathbf{v}_i \in \mathbb{R}^d$. Như vậy, đặc trưng nút, embedding cạnh và hệ số attention đều được xử lý nhất quán trong không gian tiềm ẩn $\mathbb{R}^m$, trước khi được chiếu sang không gian đặc trưng đầu ra $\mathbb{R}^d$.

4.2.2.1 Thuật toán lan truyền tiến

Để hệ thống hóa toàn bộ quy trình xử lý, Thuật toán 4.1 mô tả các bước lan truyền tiến (forward pass) của mô hình Graph Transformer Post-hoc.

4.2.2.2 Cấu hình chi tiết của kiến trúc GTP trong thực nghiệm

Để phục vụ cho quá trình thực nghiệm, cấu hình chi tiết của kiến trúc GTP được chi tiết trong bảng 4.5.

Bảng 4.5 mô tả chi tiết kích thước đầu vào, đầu ra và các tham số chính của từng tầng, từ lớp trích xuất đặc trưng đến lớp phân loại cuối cùng. Cấu hình này cho phép mô hình tận dụng đồng thời thông tin hình ảnh và mối quan hệ liên mẫu (inter-sample relationships), từ đó tăng cường khả năng học biểu diễn, cải thiện hiệu quả phân loại.

4.2.3 Tổng quan quy trình xử lý và khả năng tích hợp mô hình Graph Transformer hậu xử lý

Kiến trúc tổng thể của mô hình Graph Transformer Post-hoc (GTP) được triển khai thông qua một chuỗi các bước xử lý cụ thể như sau:

- (1) Tiếp nhận dữ liệu đầu vào và trích xuất đặc trưng ảnh

Các ảnh X-quang ngực có kích thước $H \times W \times 3$, được chia thành các lô huấn luyện với kích thước cố định.

Thuật toán 4.1: Lan truyền tiền của Graph Transformer Post-hoc

Input: Lô huấn luyện gồm các mẫu $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^b$

Output: Biểu diễn mức ảnh cuối cùng $\{\mathbf{v}_i\}_{i=1}^b$

```
1 Bước 1: Bộ mã hoá DNN;
2 for  $i \leftarrow 1$  to  $b$  do
3    $\mathbf{c}_i \leftarrow \text{DNNEncoder}(x_i);$            // Trích xuất vector đặc
   trung ảnh
4 Bước 2: Xây dựng đồ thị đầy đủ;
5  $\mathcal{V} \leftarrow \{\mathbf{c}_i\}_{i=1}^b;$            // Nút đặc trưng từ DNN encoder
6  $\mathcal{E} \leftarrow \{\mathbf{e}_{ij}\}_{i,j=1}^b;$            // Trọng số cạnh được học
7  $\mathcal{G} \leftarrow (\mathcal{V}, \mathcal{E}, \mathcal{F} = \mathcal{V});$            // Xây dựng đồ thị
8 Bước 3: Mã hoá đồ thị bằng GTN;
9 for  $i \leftarrow 1$  to  $b$  do
10    $\hat{\mathbf{c}}_i \leftarrow \sum_{j \in \mathcal{N}(i)} \alpha_{i,j} (\mathbf{W}_v \mathbf{c}_j + \mathbf{W}_e \mathbf{e}_{ij});$    // Tổng hợp thông
   tin từ các nút lân cận
11    $\mathbf{r}_i \leftarrow \mathbf{W}_r \mathbf{c}_i;$            // Kết nối dư
12    $\beta_i \leftarrow \sigma(\mathbf{W}_g [\mathbf{c}_i; \mathbf{r}_i; \hat{\mathbf{c}}_i - \mathbf{r}_i]);$    // Hệ số cộng hợp nhất
13    $\mathbf{c}'_i \leftarrow \beta_i \cdot \mathbf{r}_i + (1 - \beta_i) \cdot \hat{\mathbf{c}}_i;$  // Đặc trưng nút cuối cùng
14 Bước 4: Chuẩn hoá đầu ra;
15 for  $i \leftarrow 1$  to  $b$  do
16    $\mathbf{v}_i \leftarrow \text{BatchNorm}(\text{Linear}(\mathbf{c}'_i));$            // Chuẩn hoá đầu ra
17 return  $\{\mathbf{v}_i\}_{i=1}^b;$ 
```

Mỗi ảnh trong lô huấn luyện được đưa qua một backbone học sâu để trích xuất, sinh vector đặc trưng $\mathbf{e}_i \in \mathbb{R}^m$. phản ánh các thông tin có giá trị chẩn đoán. Tập các vector đặc trưng $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_b\}$ tạo thành đầu ra của khối mã hóa, đóng vai trò là đầu vào cho bước xử lý đồ thị.

(2) Tổ chức đồ thị:

Đồ thị đầy đủ $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ được khởi tạo. Trong đó: mỗi nút biểu $\mathbf{v} \in \mathcal{V}$ diễn vector đặc trưng ảnh $\mathbf{c}_i$, các cạnh $\mathbf{e} \in \mathcal{E}$ giữa hai nút $\mathbf{v}_i$ và $\mathbf{v}_j$ trong không gian đặc trưng. Khác với các phương pháp khởi tạo tĩnh dựa trên độ tương đồng, các đặc trưng cạnh được khởi tạo như những tham số có thể học được và được cập nhật động trong quá trình huấn

Bảng 4.5 Cấu hình chi tiết của kiến trúc GTP trong thực nghiệm

Layer Label	Input Layer Dimension	Output Dimension
Input Layer		$B \times 224 \times 224 \times 3$
Feature Extractor	$B \times 224 \times 224 \times 3$	$B \times 1024$
TransformerConv	Input tensor: $B \times 1024$ Edge index: $2 \times B^2$ Edge attribution: $B^2 \times 1$	$B \times (1920 \times n_{head})$
Linear	$B \times (1920 \times n_{head})$, $n_{head} = 3$	$B \times 1024$
Relu	$B \times 1024$	$B \times 1024$
FC Layer		$B \times 4$

luyện, giúp mô hình học và điều chỉnh thích ứng các quan hệ ngữ cảnh giữa các mẫu

(3) Lan truyền thông tin trên đồ thị:

Các vector đặc trưng nút được cập nhật thông qua các tầng Graph Transformer, sử dụng cơ chế truyền thông điệp (message passing) và attention đa đầu (multi-head attention) để khai thác sâu các mối quan hệ phụ thuộc và tương tác ngữ cảnh giữa các nút trong đồ thị.

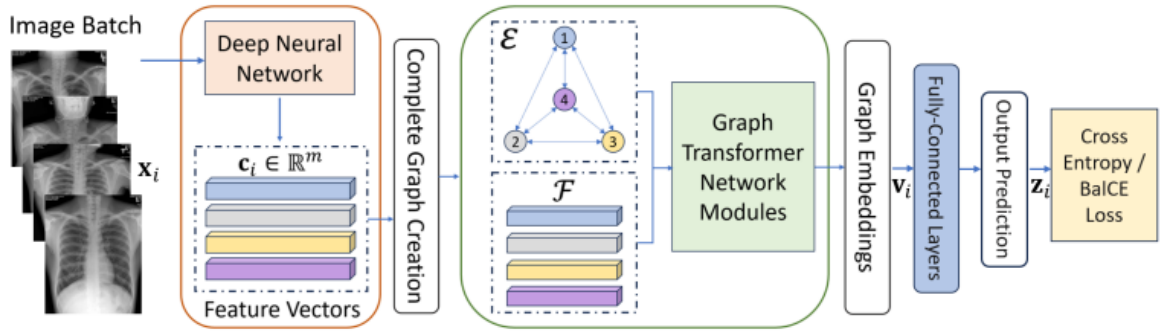
(4) Tạo embedding ngữ cảnh hóa:

Kết quả của quá trình lan truyền trên đồ thị là các vector embedding mới, trong đó mỗi embedding không chỉ giữ lại đặc trưng ban đầu của ảnh mà còn tích hợp thêm thông tin từ các ảnh khác trong lô. Nhờ đó, embedding này phản ánh tốt hơn mối quan hệ ngữ cảnh giữa các ảnh, giúp phân biệt rõ ràng hơn những đặc trưng tinh vi và giảm thiểu nhầm lẫn giữa các lớp có đặc điểm tương đồng

(5) Phân loại ảnh: Các embedding được đưa vào các lớp fully connected để dự đoán xác suất phân lớp tương ứng với các loại bệnh.

(6) Tối ưu hóa mô hình: Mô hình được huấn luyện dựa trên hàm mất mát phân loại Cross-Entropy hoặc Balanced Cross-Entropy.

Hình 4.3 minh họa kiến trúc và quy trình hoạt động tổng thể của mô hình Graph Transformer Post-hoc (GTP) được đề xuất. Một batch ảnh X-quang ngực $\mathbf{x}_i$ trước hết được đưa vào bộ mã hoá DNN nhằm trích xuất các vector đặc trưng $\mathbf{c}_i \in \mathbb{R}^{d_c}$. Tập vector này được sử dụng để xây dựng một đồ thị đầy đủ, trong đó các đỉnh biểu diễn ảnh trong batch và các cạnh được tham số hoá bởi embedding có thể học được $\mathbf{e}_{ij} \in \mathbb{R}^{d_e}$. Đồ thị



Hình 4.3 Kiến trúc tổng thể của mô hình đề xuất:

$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{F})$ sau đó được xử lý bởi mô-đun Graph Transformer Network (GTN) để tạo ra các biểu diễn nút tinh chỉnh (graph embeddings) $\mathbf{v}_i \in \mathbb{R}^d$. Các embedding này tiếp tục được đưa qua các lớp fully connected để sinh ra logit $\mathbf{z}_i$. Cuối cùng, dự đoán đầu ra được tính bằng lớp softmax và mô hình được tối ưu bằng hàm mất mát CE hoặc BalCE.

Khác với các phương pháp chỉ dựa trên đặc trưng của từng ảnh riêng lẻ, kiến trúc GTP mở rộng khả năng học sang không gian tương tác toàn cục, cho phép mô hình khai thác và tích hợp các mối quan hệ ngữ cảnh giữa các ảnh trong cùng tập dữ liệu. Nhờ vậy, GTP có khả năng nhận diện tốt hơn các đặc trưng tinh vi, tăng cường khả năng phân biệt giữa các lớp bệnh có biểu hiện tương đồng và cải thiện độ chính xác trong chẩn đoán bệnh bụi phổi silic.

4.3 Lựa chọn hàm mất mát

Trong các mô hình học sâu, hàm mất mát (loss function) đóng vai trò quan trọng trong việc đo lường sai lệch giữa đầu ra dự đoán và nhãn mục tiêu, từ đó định hướng quá trình tối ưu hóa trọng số thông qua thuật toán lan truyền ngược [119]. Việc lựa chọn hàm mất mát phù hợp có ảnh hưởng đáng kể đến khả năng hội tụ và hiệu quả phân biệt giữa các lớp – đặc biệt trong bối cảnh dữ liệu huấn luyện không cân bằng.

Đối với bài toán phân loại ảnh X-quang ngực nhằm hỗ trợ chẩn đoán bệnh bụi phổi silic, chúng tôi nhận thấy rằng việc sử dụng các hàm mất mát tiêu chuẩn như Cross-Entropy (CE) có thể dẫn đến sự thiên lệch của mô hình về phía các lớp phổ biến, chẳng hạn như lớp ảnh phổi bình thường. Hệ quả là mô hình đạt độ chính xác tổng thể cao nhưng lại suy giảm hiệu quả phát hiện các lớp thiểu số – trong đó có bụi phổi silic.

Để giải quyết vấn đề này, Luận án áp dụng và so sánh hai loại hàm mất mát: Cross-Entropy Loss truyền thống và Balanced Cross-Entropy Loss (BalCE) [121] – một biến thể có điều chỉnh trọng số theo phân bố lớp.

4.3.1 Hàm mất mát tiêu chuẩn

Hàm mất mát tiêu chuẩn (Cross-Entropy) có nguồn gốc từ lý thuyết thông tin của Shannon [122], và được sử dụng rộng rãi trong học máy như là hệ quả trực tiếp của nguyên lý cực đại hóa khả năng hợp lý [123]. Đây là lựa chọn tiêu chuẩn cho các bài toán phân loại nhiều lớp kết hợp với hàm softmax.

Với đầu ra logit $\mathbf{z}_k$ tương ứng với lớp k , hàm CE được định nghĩa như sau:

$$\begin{aligned} \mathcal{L}_{CE}(\mathcal{W}(\mathcal{F}(\mathbf{x} \mid \theta_f) \mid \theta_w), k) \\ = -\log(p(k \mid \mathbf{x}; \theta_f, \theta_w)) = -\log \left[\frac{e^{\mathbf{z}_k}}{\sum_{j=1}^4 e^{\mathbf{z}_j}} \right]. \end{aligned} \quad (4.6)$$

trong đó, $\mathcal{F}(\mathbf{x} \mid \theta_f)$ là hàm trích xuất đặc trưng từ ảnh X-quang $\mathbf{x}$ với tham số θ_f , thường là một mạng nơ-ron sâu CNN chuyển đầu vào x thành vecto đặc trưng.

$\mathcal{W}(\cdot \mid \theta_w)$ là lớp phân loại áp dụng lên vecto đặc trưng để sinh ra logits $\mathbf{z}_k$. Giá trị $\mathbf{z}_k$ là logit (đầu ra chưa chuẩn hóa) tương ứng với lớp k , phản ánh mức độ mô hình “tin tưởng” vào lớp đó trước khi chuẩn hóa softmax.

$\frac{e^{\mathbf{z}_k}}{\sum_{j=1}^4 e^{\mathbf{z}_j}}$ là xác suất dự đoán cho lớp k , được tính qua hàm softmax. Hàm softmax chuẩn hóa các logits thành phân bố xác suất, với tổng xác suất của tất cả các lớp bằng 1.

$-\log$ là hàm mất mát Cross-Entropy, đo lường độ khác biệt giữa phân bố xác suất dự đoán và nhãn thực tế. $\mathcal{L}_{CE}$ là giá trị mất mát Cross-Entropy cho một mẫu dữ liệu. Giá trị này càng lớn nếu xác suất dự đoán càng nhỏ.

Hàm mất mát trong công thức (4.6) giả định mọi lớp có phân bố đồng đều, không xét đến sự mất cân bằng trong phân bố mẫu huấn luyện. Do đó, trong bài toán phân loại ảnh X-quang ngực, hàm mất mát CE có thể dẫn đến thiên lệch về lớp phổ biến (như ảnh phổi bình thường), làm giảm hiệu quả phát hiện lớp thiểu số (như bụi phổi silic).

4.3.2 Hàm mất mát cân bằng

Hàm mất mát cân bằng Balanced Cross-Entropy (BalCE) là một biến thể của hàm Cross-Entropy truyền thống, được thiết kế để giải quyết các bài toán phân loại trong điều kiện dữ liệu phân bố mất cân bằng. Hàm BalCE điều chỉnh đóng góp của từng lớp vào hàm mất mát thông qua trọng số tỷ lệ nghịch với tần suất lớp trong từng lô huấn luyện (*mini-batch*). Cách tiếp cận này có nền tảng từ các phương pháp có trọng số (*cost-sensitive learning*) trong học sâu [124], và được định hình rõ ràng hơn thông qua khái niệm “số lượng mẫu hiệu dụng” (*effective number of samples*) [125].

Hàm mất mát cân bằng BalCE (Balanced Cross-Entropy Loss) là một hàm mất mát được thiết kế để xử lý dữ liệu phân lớp mất cân bằng bằng cách điều chỉnh trọng số cho từng lớp dựa trên tần suất xuất hiện trong lô huấn luyện [121].

Khi xét đến số lượng mẫu của từng lớp là n_k , hàm Bal-CE được định nghĩa lại như sau:

$$\mathcal{L}_{\text{Bal-CE}}\left(\mathcal{W}\left(\mathcal{F}(\mathbf{x} \mid \theta_f) \mid \theta_w\right), k\right) = -\log \left[\frac{n_k e^{\mathbf{z}_k}}{\sum_{j=1}^4 n_j e^{\mathbf{z}_j}} \right]. \quad (4.7)$$

Các hệ số n_k giúp hiệu chỉnh xác suất softmax theo tỷ lệ xuất hiện của từng lớp trong tập huấn luyện. Nhờ đó, ảnh hưởng của các lớp chiếm ưu thế (nhiều mẫu) được giảm xuống, trong khi vai trò của các lớp thiểu số được tăng cường, giúp mô hình học được phân biệt tốt hơn trong môi trường dữ liệu mất cân bằng.

4.4 Kỹ thuật kết hợp mô hình

Kỹ thuật kết hợp mô hình (*ensemble learning*) là một hướng tiếp cận trong học máy nhằm cải thiện hiệu suất và độ ổn định của hệ thống dự đoán bằng cách tổng hợp đầu ra từ nhiều mô hình học thành phần [124]. Thay vì chỉ dựa vào một mô hình đơn lẻ, học kết hợp tận dụng sự đa dạng giữa các mô hình để tạo nên một dự đoán hợp nhất, có tính tổng quát cao hơn. Các mô hình thành phần có thể được huấn luyện trên các tập dữ liệu khác nhau, sử dụng các đặc trưng hoặc kiến trúc mạng khác nhau. Dự đoán cuối cùng được tổng hợp thông qua các chiến lược phổ biến như bỏ phiếu đa số (majority voting) hoặc trung bình có trọng số (weighted averaging). Những kỹ thuật này vừa đơn giản, dễ triển khai, vừa cho thấy hiệu quả rõ rệt trong các bài toán phân loại phức tạp.

Trong các bài toán phân loại ảnh y tế, đặc biệt là chẩn đoán từ ảnh X-quang ngực, mô hình học sâu thường gặp khó khăn do đặc trưng hình ảnh giữa các lớp bệnh lý có mức độ tương đồng cao. Đặc biệt, hiện tượng mất cân bằng lớp nghiêm trọng – khi các lớp bệnh hiếm xuất hiện rất ít – khiến mô hình dễ nghiêng về lớp chiếm ưu thế và bỏ sót các dấu hiệu quan trọng. Trong nghiên cứu này, lớp bụi phổi silic vừa có tần suất xuất hiện thấp, vừa dễ bị nhầm lẫn với các dạng viêm phổi (vi khuẩn, virus). Điều này dẫn đến hiệu suất mô hình đơn lẻ thường không đồng đều và thiếu ổn định.

Để khắc phục những hạn chế trên, luận án triển khai kỹ thuật học kết hợp bằng cách tích hợp đầu ra từ sáu mô hình học sâu khác nhau. Cả sáu mô hình được huấn luyện trên cùng tập dữ liệu, sau đó kết hợp kết quả theo hai chiến lược với nguyên lý kết hợp mô hình như sau:

Giả sử tồn tại m mô hình học sâu riêng biệt $f^1, f^2, \dots, f^m$, tất cả đều được huấn luyện trên cùng một tập dữ liệu. Trong nghiên cứu này, hai kỹ thuật kết hợp đơn giản

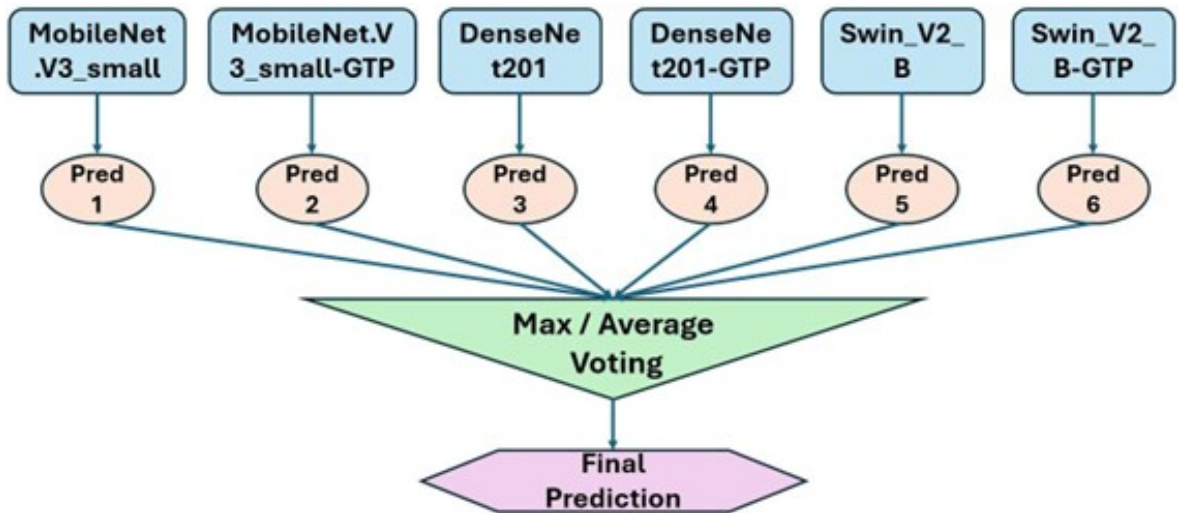
nhưng hiệu quả được áp dụng:

(1) Bỏ phiếu đa số (Max Voting): Mỗi mô hình đưa ra một dự đoán rời rạc, sau đó hệ thống chọn nhãn xuất hiện nhiều nhất làm đầu ra cuối cùng. Cách tiếp cận này đơn giản và dễ triển khai, phù hợp khi các mô hình có độ chính xác tương đương.

(2) Trung bình trọng số (Weighted Averaging): Thay vì dự đoán nhãn rời rạc, mỗi mô hình đưa ra một phân phối xác suất trên các lớp, sau đó tiến hành tính trung bình có trọng số để tạo ra đầu ra cuối cùng. Với $\mathbf{x}$ là mẫu đầu vào, công thức tổng hợp được biểu diễn như sau:

$$pred_{final} = \sum_{i=1}^m f^i(\mathbf{x}) * w^i \quad (4.8)$$

trong đó w^i là trọng số phản ánh mức độ tin cậy hoặc hiệu suất của mô hình thứ i . Trong trường hợp các trọng số W_i bằng nhau, phương pháp này trở thành phép trung bình đơn giản. Cách tiếp cận này đặc biệt hiệu quả khi các mô hình có khả năng bổ trợ lẫn nhau về mặt dự đoán.



Hình 4.4 Tổng quan kỹ thuật kết hợp các mô hình

Hình 4.4 minh họa tổng quan kỹ thuật kết hợp sáu mô hình học sâu được huấn luyện trên cùng tập dữ liệu, bao gồm: MobileNetV3-small, MobileNetV3-small-GTP, DenseNet201, DenseNet201-GTP, Swin-V2-B và Swin-V2-B-GTP. Mỗi mô hình tạo ra một dự đoán riêng (Pred 1 đến Pred 6). Các dự đoán này sau đó được tổng hợp theo hai chiến lược: bỏ phiếu đa số (Max Voting) hoặc trung bình có trọng số (Weighted Averaging). Kết quả cuối cùng là một nhãn phân loại duy nhất (Final Prediction).

Việc tích hợp nhiều mô hình học sâu theo cách tổng hợp mang lại nhiều lợi ích: Việc áp dụng kỹ thuật kết hợp mô hình mang lại nhiều lợi ích: Tăng độ ổn định. Tận

dụng ưu điểm mô hình thành phần. Cải thiện hiệu suất ở lớp hiểm: tăng độ nhạy ở lớp bụi phổi silic, là lớp vốn dễ bị bỏ sót khi sử dụng mô hình đơn.

4.5 Thiết lập thực nghiệm và đánh giá

4.5.1 Thiết lập thực nghiệm

Thiết lập cấu hình thực nghiệm

- Sử dụng GPU: NVIDIA Tesla T4, RAM 15GB.
- Số epoch: 50
- Kích thức batch: 32
- Bộ tối ưu: Rectified Adam ($\epsilon = 1.10^{-8}$)
- Tốc độ học (learning rate) $= 1.10^{-5}$

Các mô hình sử dụng:

Ba kiến trúc học sâu phổ biến được chọn làm mô hình nền tảng:

- MobileNet [126]
- DenseNet201 [110, 127]
- SwinTransformer [109]

Mỗi kiến trúc được triển khai dưới hai phiên bản

- Phiên bản tiêu chuẩn: Sử dụng nguyên bản mô hình gốc.
- Phiên bản tích hợp Graph Transformer Post-hoc (GTP): tương ứng với mỗi kiến trúc nguyên bản, thêm khối GTP phía sau bộ trích xuất đặc trưng.

Kỹ thuật học tổng hợp (ensemble):

Hai cơ chế tích hợp mô hình được áp dụng để nâng cao hiệu quả phân loại: Max Voting, Averaging

4.5.2 Kết quả thực nghiệm

4.5.2.1 Ảnh hưởng của hàm mất mát CE và Bal-CE đối với hiệu suất mô hình

Trong thực nghiệm này, đánh giá độ chính xác phân loại của hai mô hình MobileNet_V3_small và MobileNet_V3_small-GTP trên bốn lớp bệnh phổi khác nhau (bụi phổi silic nghề nghiệp, bình thường, viêm phổi do vi khuẩn và viêm phổi do virus).

Cả hai mô hình được đều sử dụng hai hàm mất mát: Cross-Entropy (LCE) và Balanced Cross-Entropy (LBal-CE). Kết quả so sánh của bốn mô hình được trình bày tại bảng 4.6.

Bảng 4.6 Kết quả tác động của hai hàm mất mát lên hiệu suất mô hình

Model	Silicosis Bụi phổi Silic	Normal Bình thường	Bacterial Vi khuẩn	Viral Virus
$\mathcal{L}_{CE}$				
MobileNet V3 small	71,96	97,10	85,74	71,48
MobileNet V3 small-GTP	74,77	93,87	84,84	72,15
$\mathcal{L}_{Bal-CE}$				
MobileNet V3 small	72,03	95,16	76,17	83,89
MobileNet V3 small-GTP	79,44	84,83	78,45	85,77

Từ kết quả ở bảng 4.6, nhận thấy:

(1) Tích hợp khối Graph Transformer Post-hoc (GTP) có xu hướng cải thiện hiệu năng tổng thể của mô hình, đặc biệt trên các lớp thiểu số, khó phân loại.

(2) Sử dụng hàm mất mát Bal-CE cho phép tái cân bằng đóng góp giữa các lớp bằng cách điều chỉnh trọng số trong quá trình tối ưu, từ đó cải thiện khả năng nhận diện các lớp ít mẫu.

(3) Sự kết hợp giữa GTP và Bal-CE cho thấy tiềm năng trong việc xử lý dữ liệu mất cân bằng, tạo nền tảng để cải thiện hơn nữa hiệu quả phân loại trên các lớp bệnh hiếm (số lượng dữ liệu hạn chế).

Ảnh hưởng của việc tích hợp GTP (trong cùng hàm mất mát CE): So với mô hình gốc, MobileNet_V3_small-GTP cải thiện độ chính xác phân loại ở lớp bụi phổi silic nghề nghiệp (tăng từ 71,96% lên 74,77%) và lớp virus (từ 71,48% lên 72,15%). Tuy nhiên, hiệu suất ở lớp bình thường lại giảm nhẹ (từ 97,10% xuống 93,87%), cho thấy việc bổ sung GTP giúp mô hình học được các đặc trưng phân biệt hữu ích cho các lớp ít dữ liệu.

Ảnh hưởng của hàm mất mát Bal-CE: Khi giữ nguyên kiến trúc và thay đổi hàm mất mát từ CE sang Bal-CE, hiệu suất phân loại lớp bụi phổi silic nghề nghiệp và virus được cải thiện ở cả hai mô hình. Cụ thể: MobileNetV3_small: bụi phổi silic tăng

từ 71,96% lên 72,03%, virus tăng mạnh từ 71,48% lên 83,89%. MobileNetV3_small-GTP: bụi phổi silic tăng từ 74,77% lên 79,44%, virus tăng từ 72,15% lên 85,77%. Trong khi đó, độ chính xác ở lớp bình thường và vi khuẩn giảm nhẹ.

Kết quả trên cho thấy thứ tự phân cấp hiệu năng
Trên các lớp thiếu số (BPSi, virus) như sau:

$$ACC_{DNN-CE} < ACC_{GTP-CR} < ACC_{GTP-BalCE}$$

Trên các lớp đa số (bình thường, vi khuẩn) như sau:

$$ACC_{DNN-CE} > ACC_{GTP-CR} > ACC_{GTP-BalCE}$$

4.5.2.2 Hiệu quả của kiến trúc GTP so với DNN

Bảng 4.7 trình bày kết quả thực nghiệm của việc sử dụng các mô hình DNN truyền thống với hàm mất mát Cross-Entropy (CE) và các mô hình GTP sử dụng hàm mất mát Balanced Cross-Entropy (Bal-CE). Các kết quả thực nghiệm được đánh giá thông qua hai chỉ số: Macro-F1 và AUC ROC, cho từng lớp bệnh lý.

Kết quả trong Bảng 4.7 cho thấy rằng việc bổ sung GTP vào các mô hình DNN không phải lúc nào cũng mang lại điểm Macro-F1 cao hơn. Cụ thể, mô hình MobileNet_V3_small-GTP có Macro-F1 thấp hơn 0.02 so với phiên bản gốc không tích hợp GTP. Giá trị Macro-F1 cao hơn cho thấy mô hình DNN với CE duy trì được sự cân bằng tốt hơn giữa độ chính xác (precision) và độ bao phủ (recall) trên tất cả các lớp, mặc dù về mặt lý thuyết hàm Bal-CE được kỳ vọng có vai trò tái cân bằng này.

Đối với chỉ số AUC ROC, tất cả các mô hình đều đạt trên 0.9, phản ánh năng lực phân biệt xuất sắc trong nhiệm vụ phân loại. Tuy nhiên, khi so sánh trực tiếp các mô hình có và không có “-GTP”, kết quả có sự dao động. Cụ thể, MobileNet-V3.small-GTP cho AUC ROC thấp hơn một chút so với phiên bản không GTP; trong khi đó, DenseNet201 và DenseNet201-GTP lại không có sự khác biệt đáng kể.

Về thời gian suy luận, việc tích hợp GTP chỉ làm tăng nhẹ độ trễ: MobileNetV3_small từ 0,013s lên 0,016s; DenseNet201 từ 0,28s lên 0,29s. Mức tăng này được đánh giá là không đáng kể, cho phép GTP có thể được triển khai trong các ứng dụng yêu cầu gần thời gian thực nếu được tối ưu hợp lý.

Quan sát tổng thể cho thấy, mỗi loại mô hình đều có điểm mạnh riêng, có khả năng bổ sung cho nhau; do đó, việc tích hợp chúng có thể tạo ra hiệu quả vượt trội hơn. Để đánh giá thêm về tính ổn định và khả năng khái quát, tiến hành kiểm chứng chéo 5-fold cross-validation.

Bảng 4.7 Bảng tóm tắt kết quả của các mô hình với hai chỉ số đánh giá Macro-F1 và AUC-ROC

Model	Thời gian suy luận	Dung lượng RAM GPU	Macro-F1	AUC-ROC			
				BPSi	Bình thường	Vi khuẩn	Virus
MobileNetV3_small	0,013	3.45 GB	0,8210	0,9768	0,9903	0,9342	0,9103
MobileNetV3_small-GTP	0,016	3.82 GB	0,8015	0,9505	0,9840	0,9455	0,9257
DenseNet201	0,28	8.63 GB	0,8368	0,9557	0,9891	0,9385	0,9122
DenseNet201-GTP	0,29	9.24 GB	0,8317	0,9512	0,9902	0,9493	0,9283
Swin_V2.b	0,82	14.82 GB	0,8410	0,9722	0,9884	0,9633	0,9499
Swin_V2.b-GTP	0,84	15.48 GB	0,8406	0,9749	0,9897	0,9576	0,9352
Max voting 3 DNNs	1,113	—	0,8808	0,9936	0,9973	0,9837	0,9777
Averaging 3 DNNs	1,113	—	0,9306	0,9985	0,9994	0,9928	0,9902
Max voting 3 GTPs	1,146	—	0,9164	0,9910	0,9968	0,9850	0,9793
Averaging 3 GTPs	1,146	—	0,9214	0,9903	0,9962	0,9869	0,9821
Max voting 6 models	2,259	—	0,9617	0,9969	0,9987	0,9950	0,9932
Averaging 6 models	2,259	—	0,9749	0,9973	0,9989	0,9973	0,9964

Bảng 4.8 trình bày kết quả độ chính xác phân loại của các mô hình qua 5 lần gập, kèm giá trị trung bình và độ lệch chuẩn. Kết quả cho thấy:

Các mô hình tích hợp GTP có ưu thế hơn so với phiên bản gốc.

Trung bình qua các lần gập, mô hình có GTP đạt độ chính xác cao hơn. Độ lệch chuẩn rất nhỏ (chỉ 0.7 ở MobileNet-V3.small-GTP), cho thấy các mô hình này ổn định và có khả năng tái lập kết quả trên các phân chia dữ liệu khác nhau.

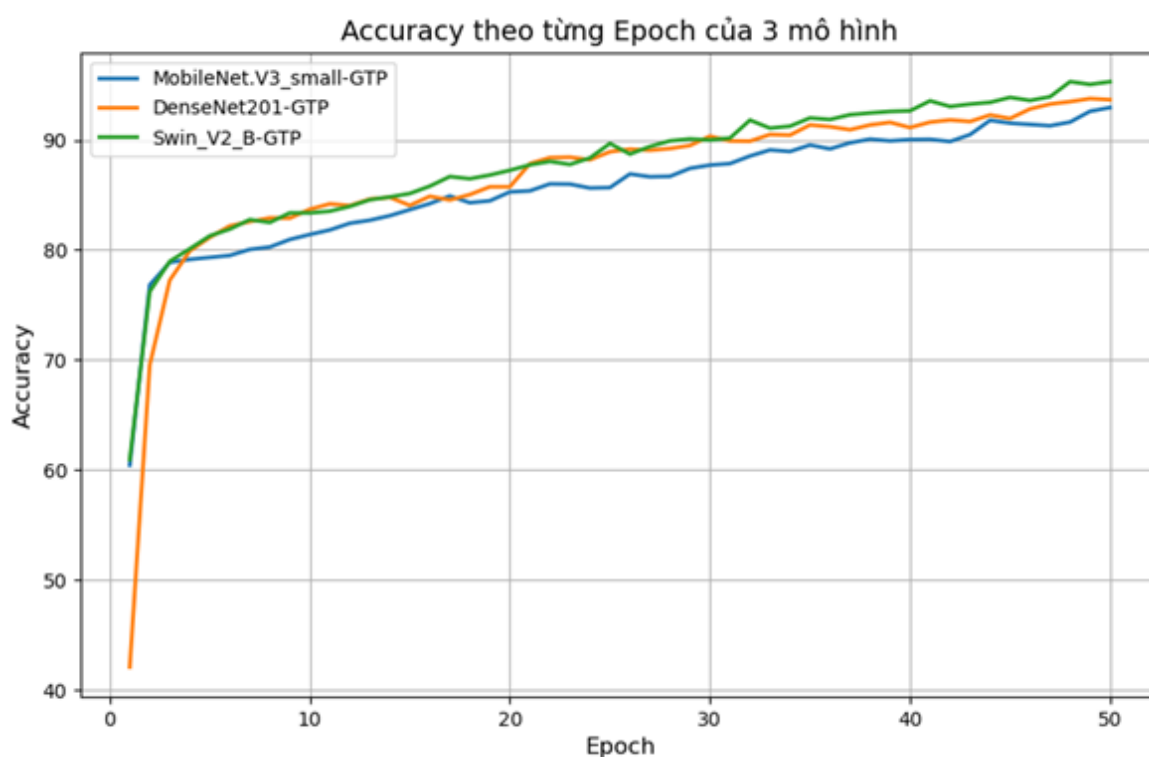
Điều này chứng tỏ quá trình huấn luyện không quá nhạy cảm với thay đổi trong thành phần dữ liệu, một tín hiệu tích cực về tính bền vững (robustness) và khả năng khái quát (generalization).

Các kết quả kiểm chứng chéo đã phản ánh hiệu năng tổng thể của mô hình. Tuy nhiên, để đánh giá sâu hơn về động thái huấn luyện (training dynamics), luận án bổ

Bảng 4.8 Kết quả độ chính xác phân loại (%) của các mô hình qua 5 lần kiểm chứng chéo (5-fold cross-validation)

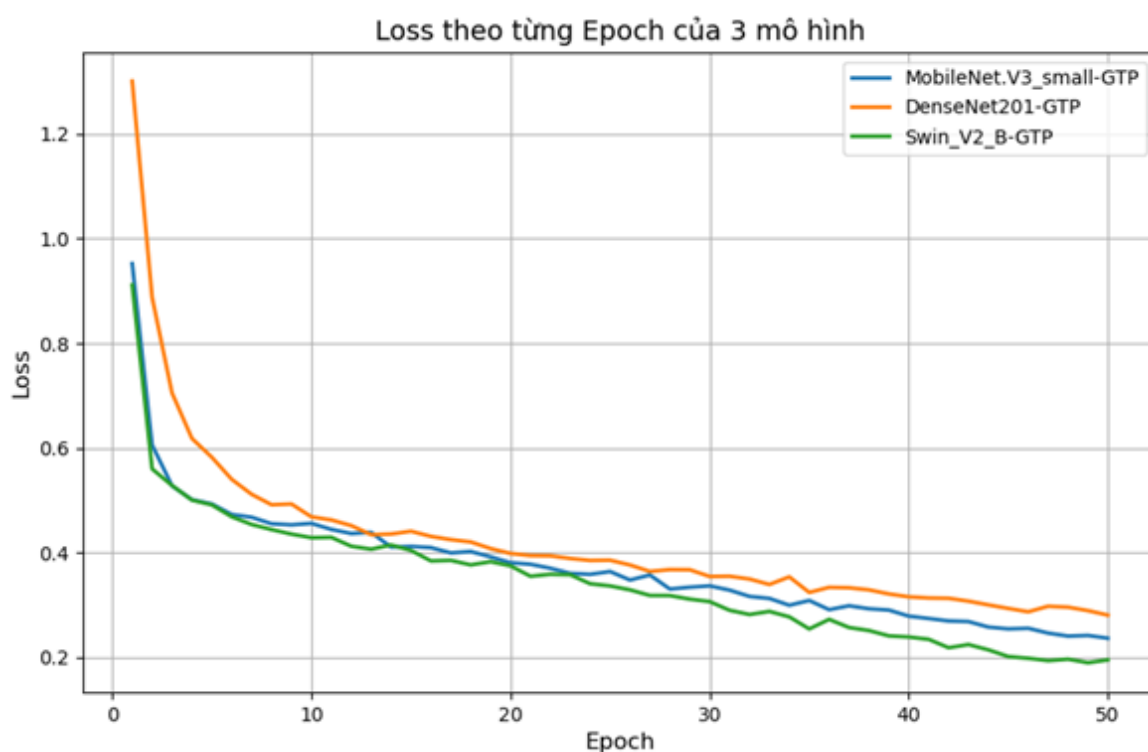
Mô hình	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean $\pm$ Std
MobileNet._3_small	79,95	80,72	80,10	80,65	81,13	80,51 $\pm$ 0,43
MobileNetV3_small-GTP	81,09	82,80	82,12	82,67	83,05	82,35 $\pm$ 0,70
DenseNet201	79,14	80,30	80,76	80,45	80,32	80,19 $\pm$ 0,55
DenseNet201-GTP	82,93	82,71	81,93	82,19	81,98	82,35 $\pm$ 0,40
Swin_V2.b	82,22	82,84	82,55	83,02	82,59	82,64 $\pm$ 0,27
Swin_V2.b-GTP	84,10	84,40	84,49	84,25	84,50	84,35 $\pm$ 0,15

sung phân tích diễn tiến của hàm mất mát và độ chính xác theo từng epoch. Phần này tập trung vào ba mô hình GTP tiêu biểu, nhằm làm rõ tốc độ hội tụ cũng như tính ổn định trong quá trình huấn luyện



Hình 4.5 Accuracy theo từng Epoch của 3 mô hình

Quan sát Hình 4.5 cho thấy cả ba mô hình đều có xu hướng giảm loss nhanh chóng trong những epoch đầu và dần tiến tới trạng thái hội tụ. Đáng chú ý, DenseNet201-GTP khởi đầu với giá trị loss cao hơn so với hai mô hình còn lại nhưng nhanh chóng thu



Hình 4.6 Loss theo từng Epoch của 3 mô hình

hẹp khoảng cách và đạt trạng thái ổn định từ khoảng epoch thứ 20 trở đi. Trong khi đó, Swin_V2.b-GTP duy trì mức loss thấp nhất trong phần lớn quá trình huấn luyện, phản ánh khả năng tối ưu hóa và năng lực khái quát hóa tốt hơn.

Về độ chính xác (Hình 4.6), cả ba mô hình đều đạt mức cải thiện rõ rệt trong những epoch đầu tiên và tiếp tục gia tăng dần trong suốt quá trình huấn luyện. DenseNet201-GTP có khởi đầu chậm hơn nhưng tốc độ cải thiện nhanh, trong khi MobileNet_V3_small-GTP duy trì sự ổn định cao song thường đạt kết quả thấp hơn hai mô hình còn lại ở giai đoạn sau. Swin_V2.b-GTP thể hiện ưu thế nổi bật, đạt độ chính xác cao nhất vào cuối quá trình huấn luyện, qua đó khẳng định hiệu quả vượt trội trong bài toán phân loại bụi phổi silic nghề nghiệp

Như vậy, mặc dù hiệu quả của GTP so với các mô hình DNN truyền thống có sự khác biệt theo từng chỉ số đánh giá và từng lớp bệnh, nhưng các kết quả từ kiểm chứng chéo cũng như phân tích diễn tiến huấn luyện đã cho thấy rõ ràng rằng GTP góp phần nâng cao tính ổn định, khả năng khái quát và cải thiện hiệu năng trên các lớp thiểu số, qua đó khẳng định tiềm năng ứng dụng của kiến trúc này.

4.5.2.3 Tác dụng của việc kết hợp các mô hình đối với hiệu suất phân loại

Dựa trên Bảng 4.7, việc kết hợp các mô hình học sâu bằng kỹ thuật học tổng hợp các mô hình giúp nâng cao hiệu suất phân loại, đặc biệt trong bối cảnh dữ liệu mất cân bằng.

Có hai phương pháp chính được áp dụng để kết hợp các mô hình: max voting và averaging. Cả hai phương pháp đều cho thấy những kết quả cải thiện đáng kể trong các chỉ số đánh giá so với các mô hình đơn lẻ.

Hiệu suất tổng thể:

Kỹ thuật Averaging với ba mô hình DNN đạt Macro-F1 là 0,9306, còn với ba mô hình GTP là 0,9214. Đáng chú ý, khi kết hợp toàn bộ 6 mô hình (3 DNN + 3 GTP), Macro-F1 tăng lên 0,9749, vượt trội so với mọi mô hình đơn lẻ (cao nhất chỉ 0,8410 với Swin_V2.b). Điều này cho thấy ensemble có khả năng tận dụng ưu điểm về độ ổn định của DNN và khả năng phân biệt lớp thiểu số của GTP.

AUC-ROC theo lớp:

Ensemble 6 mô hình (Averaging) đạt AUC-ROC lên tới 0,9964 với lớp virus và 0,9973 với lớp bụi phổi silic, cho thấy khả năng phân biệt các lớp bệnh hiếm được cải thiện đáng kể khi kết hợp nhiều mô hình với đặc trưng học khác nhau.

Thời gian suy luận và mức sử dụng tài nguyên:

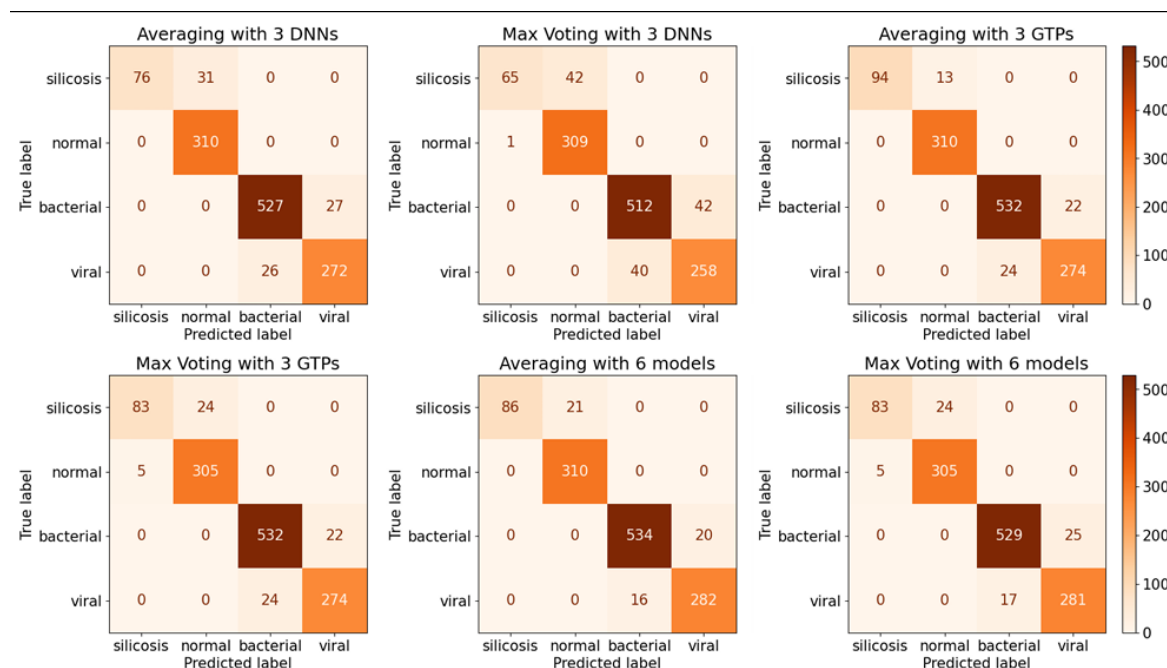
Việc hợp nhất sáu mô hình làm tăng thời gian suy luận lên 2,259 giây, cao hơn so với các mô hình đơn lẻ (từ 0,013 đến 0,84 giây). Tuy nhiên, mức tăng này hoàn toàn chấp nhận được trong các hệ thống ứng dụng thực tế, nơi thời gian xử lý tính theo giây vẫn đảm bảo yêu cầu lâm sàng.

Về mặt tài nguyên tính toán, kết quả trong Bảng 4.7 cho thấy dung lượng RAM GPU sử dụng tăng nhẹ khi tích hợp GTP: MobileNetV3_small tăng từ 3.45 GB lên 3.82 GB, DenseNet201 tăng từ 8.63 GB lên 9.24 GB, và Swin_V2.b tăng từ 14.82 GB lên 15.48 GB. Mức tăng trung bình chỉ khoảng 7% cho thấy chi phí bộ nhớ là không đáng kể so với lợi ích về hiệu suất.

Tóm lại, Phương pháp học tổng hợp, đặc biệt là khi kết hợp giữa DNN và GTP, đã chứng minh hiệu quả trong việc cải thiện hiệu suất phân loại tổng thể và nâng cao khả năng nhận diện các lớp bệnh thiểu số. Mặc dù chi phí tính toán và thời gian suy luận có tăng, song mức tăng này nằm trong giới hạn chấp nhận được và hoàn toàn khả thi khi triển khai thực tế.

4.5.2.4 Đánh giá mô hình thông qua ma trận nhầm lẫn

Ma trận nhầm lẫn là một công cụ trong đánh giá mô hình phân loại, đặc biệt trong các bài toán y học, nơi độ chính xác tổng thể là chưa đủ để phản ánh chất lượng chẩn đoán.



Hình 4.7 Ma trận nhầm lẫn

Hình 4.7 trình bày ma trận nhầm lẫn tương ứng với sáu kịch bản tích hợp mô hình, bao gồm: ba mô hình học sâu DNN, ba mô hình GTP, và tổ hợp cả sáu mô hình. Mỗi mô hình được đánh giá với hai chiến lược tổng hợp: Averaging và Max Voting.

Nhận diện lớp bụi phổi silic: Phương pháp Averaging vượt trội hơn Max Voting trong việc nhận diện lớp hiếm gặp này. “Averaging với 3 GTPs” đạt hiệu quả cao nhất với 94/107 mẫu đúng, tiếp theo là “Averaging 6 models” với 86 mẫu đúng. Kết quả này chứng minh rằng tích hợp GTP với kỹ thuật trung bình hóa cải thiện đáng kể khả năng chẩn đoán Silicosis, lớp bệnh dễ bị bỏ sót do mất cân bằng dữ liệu.

Nhận diện lớp viêm phổi do virus: Lớp thiểu số này đạt độ chính xác cao nhất với “Averaging 6 models” (282 mẫu đúng), khẳng định rằng kết hợp đa mô hình, bao gồm cả DNNs và GTPs, nâng cao hiệu quả phân loại các bệnh phổi có biểu hiện tương đồng trên ảnh X-quang.

4.5.2.5 Phân tích độ nhạy, độ đặc hiệu theo từng lớp bệnh

Trong các bài toán y tế, đặc biệt là đối với bệnh bụi phổi silic – vốn thường tiến triển âm thầm và dễ bị bỏ sót – hai chỉ số độ nhạy (sensitivity) và độ đặc hiệu (specificity) được trích xuất từ ma trận nhầm lẫn giúp hiểu rõ hơn khả năng phát hiện đúng bệnh (tránh bỏ sót) và khả năng loại trừ chính xác các ca âm tính (tránh nhầm lẫn).

Bảng 4.9 Độ nhạy và độ đặc hiệu theo từng lớp bệnh đối với các mô hình

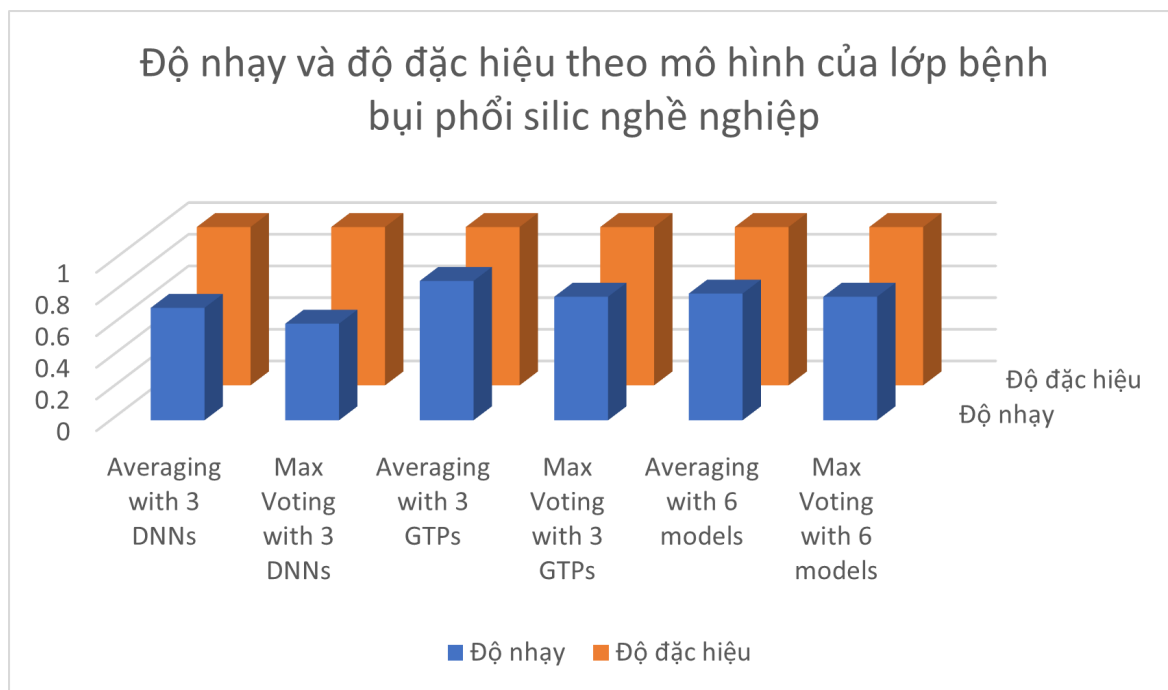
Mô hình	BPSi		Bình thường		Vi khuẩn		Virus	
	Độ nhạy	Độ đặc hiệu	Độ nhạy	Độ đặc hiệu	Độ nhạy	Độ đặc hiệu	Độ nhạy	Độ đặc hiệu
Averaging 3 DNN	0,71	1	1	0,97	0,95	0,96	0,91	0,97
Max Voting 3 DNN	0,61	1	1	0,96	0,92	0,94	0,87	0,96
Averaging 3 GTP	0,88	1	1	0,99	0,96	0,97	0,92	0,98
Max Voting 3 GTP	0,78	1	0,98	0,97	0,96	0,97	0,92	0,98
Averaging 3 mô hình	0,8	1	1	0,98	0,96	0,98	0,95	0,98
Max Voting 6 mô hình	0,78	1	0,98	0,97	0,95	0,98	0,94	0,97

Bảng 4.9 trình bày chi tiết độ nhạy và độ đặc hiệu của sáu mô hình tích hợp khác nhau, được đánh giá trên bốn lớp bệnh.

Kết quả cho thấy: Tất cả các mô hình đều đạt độ đặc hiệu rất cao (gần hoặc bằng 1), cho thấy khả năng loại bỏ đúng các ca không mắc bệnh là rất tốt. Trong khi đó, độ nhạy giữa các mô hình có sự khác biệt rõ rệt, phản ánh khả năng phát hiện đúng ca bệnh khác nhau.

Mô hình Averaging 3 GTP đạt độ nhạy cao nhất (0.88) cho lớp bụi phổi silic, vượt trội so với các mô hình khác. Các mô hình kết hợp 6 thành phần (3 DNN + 3 GTP) cũng duy trì độ nhạy và đặc hiệu ổn định, cho thấy hiệu quả của kỹ thuật hội tụ mô hình (ensemble)

Kết quả phân tích tại Hình 4.8. Độ nhạy độ đặc hiệu của các mô hình đối với lớp bệnh bụi phổi silic cho thấy mô hình Averaging_3 GTPs là lựa chọn tối ưu cho lớp bụi phổi silic (0,88) và duy trì đặc hiệu tuyệt đối. Với chỉ số độ nhạy vẫn còn khiêm tốn, phản ánh thách thức của bài toán: các tổn thương thường mờ, nhỏ, lan tỏa và dễ bị nhầm lẫn với các bệnh lý về phổi khác. Bên cạnh đó, hiện tượng dịch chuyển miền dữ liệu (domain shift) giữa dữ liệu tự thu thập và dữ liệu công khai có thể làm giảm tính ổn



Hình 4.8 Độ nhạy và độ đặc hiệu theo mô hình

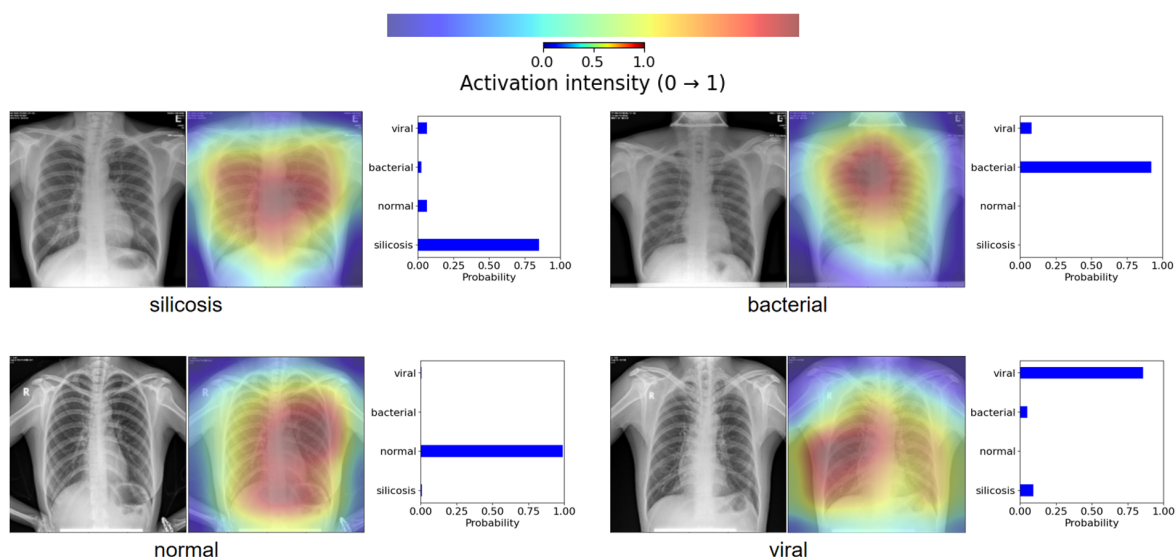
định của mô hình. Trong các nghiên cứu tiếp theo, cần xem xét những hướng tiếp cận nâng cao hơn như: mô hình đa phương thức (multi-modal) kết hợp ảnh X-quang ngực với metadata lâm sàng (tuổi, giới, tiền sử nghề nghiệp, thời gian phơi nhiễm bụi...).

Từ góc nhìn ứng dụng, việc duy trì độ nhạy cao trong khi vẫn đảm bảo độ đặc hiệu ở mức gần như tuyệt đối có ý nghĩa quan trọng trong hỗ trợ chẩn đoán bệnh bụi phổi silic nghề nghiệp. Mặc dù cần thận trọng khi đánh giá kết luận tổng quát từ một tập dữ liệu cụ thể song kết quả bước đầu cho thấy mô hình tích hợp GTP với Averaging là lựa chọn phù hợp với yêu cầu vừa phát hiện chính xác vừa giảm thiểu nguy cơ nhầm lẫn trong phân loại.

4.5.2.6 Đánh giá mô hình qua các đặc trưng học được và phân tích trực quan

Hình 4.9 trình bày bản đồ nhiệt Grad-CAM từ mô hình DenseNet201-GTP, áp dụng trên ảnh X-quang cho các lớp bụi phổi silic, vi khuẩn, bình thường và virus. Các bản đồ nhiệt (màu vàng-đỏ) tập trung chính xác vào vùng phổi, phản ánh khả năng mô hình học các đặc trưng bệnh lý cụ thể:

Lớp bụi phổi silic: Tập trung ở vùng giữa phổi với xác suất cao (0,75), chỉ ra nhận diện nốt hoặc xơ hóa đặc trưng. Lớp vi khuẩn: Khu vực tổn thương lan tỏa ở trung tâm



Hình 4.9 Bản đồ nhiệt Grad-CAM của mô hình DenseNet201-GTP cho 4 lớp phân loại

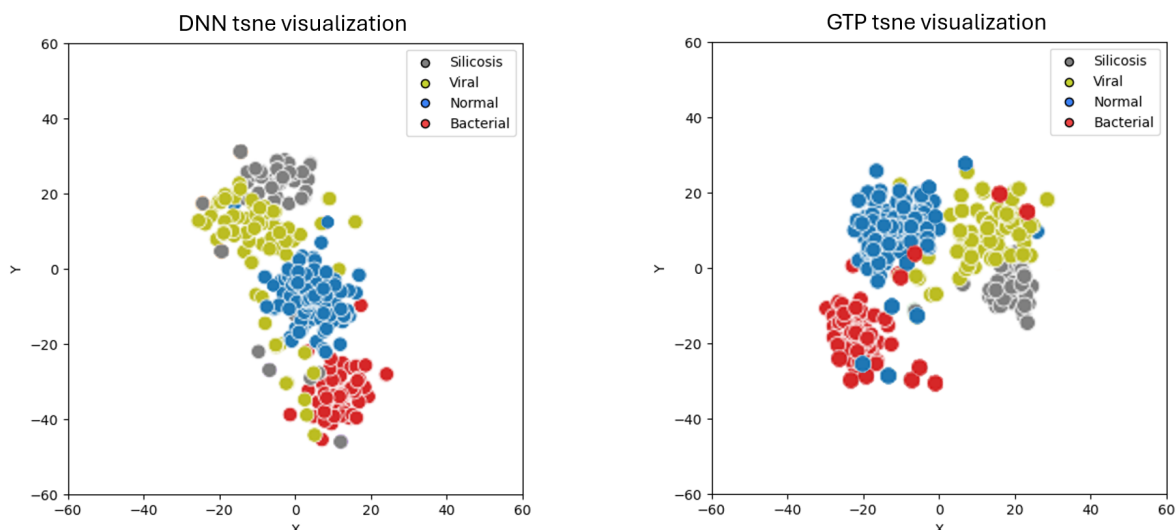
phổi, với xác suất 0,75, phù hợp với viêm phổi vi khuẩn. Lớp bình thường: Tập trung thấp (0,25) ở trung tâm, xác suất cao nhất cho Normal (0,75), phản ánh phổi khỏe mạnh. Lớp virus: Vùng tổn thương rõ ở trung tâm với xác suất gần 1,00, chỉ ra mô hình tự tin nhận diện viêm phổi virus. Màu sắc và xác suất đi kèm minh họa độ tin cậy. Mặc dù Grad-CAM chưa thể phản ánh đầy đủ cơ chế học bên trong của mô hình, song các bản đồ nhiệt chứng minh rằng mô hình đã học được thông tin cốt lõi từ dữ liệu, biết tập trung vào vùng phổi, nơi chứa các đặc trưng bệnh lý.

Ngoài ra, để phân tích cách mô hình biểu diễn đặc trưng trong không gian tiềm ẩn, luận án áp dụng kỹ thuật t-SNE. Trên hình 4.10, đối với DNN t-SNE, lớp bụi phổi silic và lớp virus phân tán, có sự chồng lấn. Ngược lại, với GTP t-SNE, lớp bụi phổi silic và vi khuẩn gom cụm chặt chẽ hơn, phân bố tập trung và tách biệt rõ hơn.

Những kết quả này giúp lý giải tại sao trong các bảng 4.6 và 4.7, mô hình GTP thường đạt độ chính xác cao hơn ở các lớp bệnh thiếu số như bụi phổi silic và lớp virus. Sự cải thiện này có thể đến từ khả năng học đặc trưng phân biệt tốt hơn, cho phép các cụm dữ liệu trong không gian nhúng trở nên rời rạc và dễ tách lớp.

4.5.2.7 So sánh hiệu suất mô hình với chuyên gia

Để đánh giá khả năng ứng dụng trong thực tiễn lâm sàng, mô hình được kiểm thử trên một tập dữ liệu độc lập gồm 417 ảnh X-quang ngực, trong đó có 107 trường hợp được xác định mắc bệnh bụi phổi silic và 310 trường hợp bình thường. Tập dữ liệu này



Hình 4.10 Trực quan hóa t-SNE của các đặc trưng được học bởi mô hình: DenseNet201 (Hình bên trái) và DenseNet201-GTP (Hình bên phải).

được Hội đồng chuyên gia thuộc Hội Y học Lao động Thái Nguyên thẩm định, dưới sự chủ trì của GS.TS Đỗ Văn Hàm – một chuyên gia lâu năm trong lĩnh vực bệnh nghề nghiệp.

Quá trình đánh giá được thực hiện độc lập, tuân thủ nguyên tắc ẩn danh và đối chiếu chéo, nhằm đảm bảo tính khách quan tối đa. Các bác sĩ tiến hành kiểm tra kỹ thuật ảnh chụp và phân tích hình ảnh từng trường hợp mà không có thông tin định danh bệnh nhân.

Kết quả đánh giá của chuyên gia được trình bày trong Bảng 4.10, cho thấy có 40/107 ca bụi phổi silic được phân loại đúng, trong khi 67 ca còn lại bị đánh giá nhầm là bình thường. Ngoài ra, có 4/310 ca bình thường bị chẩn đoán nhầm là bệnh lý, chiếm tỷ lệ sai số thấp nhưng vẫn đáng lưu ý trong công tác sàng lọc.

Bảng 4.10 Ma trận nhầm lẫn từ kết quả chẩn đoán của chuyên gia

	Dự đoán BPSi	Dự đoán bình thường
Thực tế BPSi	40	67
Thực tế bình thường	4	306

Những kết quả này phản ánh thách thức trong việc chẩn đoán bụi phổi silic, đặc biệt ở những trường hợp biểu hiện không điển hình hoặc tổn thương mờ nhạt trên phim X-quang. Đây là cơ sở quan trọng để so sánh với kết quả từ hệ thống đề xuất, nhằm đánh giá mức độ hỗ trợ mà mô hình có thể mang lại trong thực hành lâm sàng.

4.6 Kết luận chương 4

Chương 4 đã trình bày toàn bộ quá trình thiết lập thực nghiệm, đánh giá và phân tích hiệu quả của các mô hình đề xuất đối với bài toán hỗ trợ chẩn đoán bệnh bụi phổi silic nghề nghiệp từ ảnh X-quang ngực. Các kết quả chính có thể tóm lược như sau:

(1) Ảnh hưởng của hàm mất mát: Việc sử dụng hàm mất mát Balanced Cross-Entropy (Bal-CE) đã cải thiện đáng kể khả năng nhận diện các lớp thiểu số như bụi phổi silic và virus, trong khi Cross-Entropy (CE) cho kết quả cân bằng hơn trên các lớp đa số. Sự kết hợp giữa GTP và Bal-CE mang lại hiệu quả vượt trội trong xử lý dữ liệu mất cân bằng.

(2) So sánh GTP với DNN truyền thống: Kết quả cho thấy mô hình GTP tuy không phải lúc nào cũng đạt Macro-F1 cao hơn DNN, song GTP cho hiệu quả vượt trội ở các lớp khó phân loại (Silicosis, Viral). Phân tích kiểm chứng chéo 5-fold chứng minh tính ổn định và khả năng tái lập kết quả của các mô hình.

(3) Hiệu quả của kỹ thuật học tổng hợp: Cả hai phương pháp max voting và averaging đều cải thiện hiệu suất so với các mô hình đơn lẻ. Đặc biệt, averaging cho kết quả tốt nhất, trong đó kết hợp đồng thời 6 mô hình (3 DNN + 3 GTP) đạt Macro-F1 và AUC ROC cao nhất, khẳng định sức mạnh của học tổng hợp.

(4) Phân tích theo độ nhạy, độ đặc hiệu và ma trận nhầm lẫn: Các mô hình ensemble giúp nâng cao khả năng phân biệt, đặc biệt ở lớp bệnh bụi phổi silic. Kết quả confusion matrix chỉ ra sự nhầm lẫn chủ yếu giữa các lớp có biểu hiện lâm sàng tương đồng, trong khi Silicosis được cải thiện đáng kể khi sử dụng GTP.

(5) Khả năng giải thích và trực quan hóa: Phân tích Grad-CAM cho thấy các mô hình tập trung vào vùng phổi trên X-quang, phù hợp với kiến thức y khoa. Biểu diễn t-SNE minh chứng rằng GTP giúp đặc trưng của lớp bụi phổi silic được phân bố tập trung và tách biệt rõ ràng hơn so với DNN, giải thích cho kết quả phân loại vượt trội ở các lớp này.

(6) Đánh giá trên tập dữ liệu kiểm định độc lập, kết hợp với chẩn đoán từ Hội đồng chuyên gia, chỉ ra rằng hệ thống có thể hỗ trợ hiệu quả cho bác sĩ trong việc phát hiện bụi phổi silic, đặc biệt ở những trường hợp biểu hiện mờ nhạt hoặc không điển hình.

Các thực nghiệm đã chứng minh rằng kiến trúc Graph Transformer Post-hoc (GTP), kết hợp với hàm mất mát Bal-CE và kỹ thuật học tổng hợp, mang lại sự cải thiện đáng kể về hiệu năng, đặc biệt trong việc nhận diện lớp bệnh thiếu số. Tuy nhiên, một số chỉ số (độ nhạy của lớp bụi phổi silic) vẫn còn khiêm tốn, và luận án chưa khai thác được dữ liệu bệnh bụi phổi silic nghề nghiệp từ nguồn quốc tế do hạn chế về điều kiện truy cập. Đây sẽ là những hướng nghiên cứu mở rộng nhằm tiếp tục nâng cao độ tin cậy và khả năng ứng dụng thực tiễn của hệ thống.

Tổng kết lại, Chương 4 đã trình bày các đóng góp chính của luận án trong việc phát triển mô hình học sâu hỗ trợ phân loại bệnh bụi phổi silic nghề nghiệp. Cụ thể, luận án đã:

(i) Đề xuất mô hình Graph Transformer Post-hoc (GTP) cho phép khai thác mối quan hệ liên ảnh, giúp tăng cường khả năng phân biệt giữa các lớp bệnh có đặc trưng hình thái tương đồng.

(ii) Sử dụng hàm mất mát cân bằng (Balanced Cross-Entropy) nhằm khắc phục vấn đề mất cân bằng dữ liệu giữa các nhóm bệnh.

(iii) Tích hợp cơ chế học tổng hợp (ensemble learning) giữa các mô hình học sâu khác nhau để nâng cao độ chính xác, độ ổn định và khả năng khái quát.

Những kết quả trình bày đã được công bố trong các công trình khoa học liên quan (CT2, CT3, CT4, CT5 và CT6) góp phần minh chứng cho tính khả thi, tiềm năng ứng dụng của hướng nghiên cứu.

KẾT LUẬN

1. Kết quả nghiên cứu và đóng góp chính của Luận án

Xuất phát từ mục tiêu hỗ trợ chẩn đoán bệnh bụi phổi silic nghề nghiệp từ ảnh X-quang ngực trong điều kiện dữ liệu hạn chế và không cân bằng, luận án đã đề xuất, triển khai và kiểm chứng các hướng tiếp cận chính, với những đóng góp như sau:

Đóng góp thứ nhất: Đề xuất, tăng cường và hệ thống hóa được bộ dữ liệu.

Luận án đã kế thừa bộ dữ liệu từ đề tài KC.10.33/16-20, đồng thời bổ sung thêm nguồn dữ liệu công khai, hình thành một tập dữ liệu có cấu trúc, được tiền xử lý, gán nhãn và kiểm định chất lượng bởi chuyên gia y tế. Bộ dữ liệu này được chuẩn hóa về định dạng, độ phân giải và thông tin nhãn, đảm bảo tính tin cậy và sẵn sàng cho các nghiên cứu học máy và học sâu tiếp theo trong lĩnh vực chẩn đoán hình ảnh.

Đóng góp thứ hai: Đề xuất cải tiến mô hình phân vùng và phân loại tổn thương bệnh bụi phổi silic nghề nghiệp từ ảnh X-quang ngực dựa trên hai hướng tiếp cận:

(1) Điều chỉnh và ứng dụng phương pháp học ít mẫu cho phép đồng thời phân đoạn và phân loại.

(2) Phát triển kiến trúc Graph Transformer Post-hoc (GTP) kết hợp hàm mất mát cân bằng (Balanced Loss), kết hợp kỹ thuật học tổng hợp (Ensemble Learning) nhằm khai thác mối quan hệ không gian giữa các vùng đặc trưng, cải thiện độ ổn định, nâng cao hiệu quả mô hình phân loại.

2. Đề xuất các vấn đề cần tiếp tục nghiên cứu

Mở rộng bộ dữ liệu và tích hợp dữ liệu lâm sàng: Dù tập dữ liệu đã được xây dựng và triển khai, nhưng việc mở rộng bộ dữ liệu này bằng cách kết hợp với các nguồn dữ liệu khác, như hình ảnh từ CT scan, dữ liệu lâm sàng hoặc thông tin về tiền sử bệnh nhân, sẽ giúp tăng cường khả năng phân loại chính xác hơn nữa. Hơn nữa, sự đa dạng hóa dữ liệu huấn luyện sẽ giúp các mô hình học sâu phát triển khả năng tổng quát tốt hơn khi đối diện với các tình huống thực tế.

Tăng cường khả năng tổng quát của mô hình: Mặc dù các mô hình hiện tại đã đạt được kết quả khả quan trong điều kiện huấn luyện có kiểm soát, song cần tiếp tục nghiên cứu các hướng tiếp cận mới như học sâu đa tầng (multi-level learning), học đa nhiệm (multi-task learning) hoặc học không giám sát (unsupervised learning). Những hướng này có tiềm năng cải thiện độ bền vững và khả năng thích ứng của mô hình.

Ứng dụng mô hình trong thực tế lâm sàng: Triển khai các mô hình đã phát triển vào thực tế lâm sàng là một thách thức lớn nhưng cũng là cơ hội để chứng minh sự hiệu quả. Cần tiến hành các thử nghiệm lâm sàng với sự tham gia của bác sĩ và chuyên gia y tế để đánh giá độ tin cậy, hiệu năng và khả năng mở rộng của mô hình trong hỗ trợ chẩn đoán bệnh bụi phổi silic cũng như các bệnh lý phổi, bệnh lý nghề nghiệp khác.

3. Hạn chế và hướng khắc phục

Mặc dù đạt được những kết quả tích cực, luận án vẫn tồn tại một số hạn chế:

Hạn chế về dữ liệu: Tập dữ liệu còn hạn chế về quy mô và sự cân bằng lớp, đặc biệt đối với bệnh bụi phổi silic nghề nghiệp. Hướng khắc phục: Mở rộng quy mô dữ liệu, kết hợp thêm dữ liệu công khai và dữ liệu đa phương thức để cải thiện tính khái quát.

Hạn chế về mô hình: Các mô hình học sâu yêu cầu tài nguyên tính toán lớn, đồng thời khả năng diễn giải còn hạn chế. Hướng khắc phục: Tối ưu hóa mô hình theo hướng nhẹ hơn, kết hợp thêm cơ chế trực quan hóa để tăng khả năng giải thích.

Hạn chế về ứng dụng thực tiễn: Kết quả hiện tại chủ yếu mới dừng lại ở mức nghiên cứu thử nghiệm, chưa có điều kiện triển khai trong môi trường lâm sàng thực tế. Hướng khắc phục: Từng bước triển khai thử nghiệm lâm sàng, thu thập phản hồi từ bác sĩ chuyên khoa, và điều chỉnh mô hình cho phù hợp với nhu cầu sử dụng thực tế.

DANH MỤC CÔNG TRÌNH KHOA HỌC

I. Tạp chí

- [1] **Nguyễn Thị Tân Tiên**, Bui Quoc Bao, Pham Van Cuong (2025), “A data-constrained approach for occupational silicosis detection on chest X-rays with few-shot learning”, Journal of Computer Science and Cybernetics, 41(1), 1–16. doi: 10.15625/1813-9663/21668.
- [2] **Nguyễn Thị Tân Tiên** (2025), *Phát hiện bệnh bụi phổi silic nghề nghiệp từ hình ảnh X-quang ngực bằng kiến trúc DenseNet*, Tạp chí Khoa học Công nghệ Thông tin và Truyền thông, số 01 (CS.01), Trang 44–51.
- [3] **Nguyễn Thị Tân Tiên**, Nguyễn Văn Tảo, Ngô Hữu Huy (2025), *Ứng dụng mô hình Swin Transformer trong hỗ trợ chẩn đoán bệnh bụi phổi silic nghề nghiệp*, Tạp chí Khoa học Công nghệ Thông tin và Truyền thông. (số 02 (CS.02), Trang 32–29
- [4] **Nguyễn Thị Tân Tiên**, Bui Quoc Bao, Pham Van Cuong, Tran Tien Cong, Le Minh Duy, Nguyen Van Tao (2025), “FSGCN: Feature-Connected graph neural network for filne-grained silicosis classification”, Journal of Computer Science and Cybernetics. (Đã được chấp nhận đăng)
- [5] **Tien Nguyen-Tan** , Bao Quoc-Bui, Duy Minh-Le, Cong Tran, Cuong Pham, *Innovative Silicosis and Pneumonia Classification: Leveraging Graph Transformer Post-hoc Modeling and Ensemble Techniques*, BMC Medical Imaging (Q2 SCIE, đang được phản biện và đã nộp lại sau nhận xét của phản biện) .
- [6] Duy Minh Le, **Tien Nguyen-Tan**, Huyen Nguyen, Ngoc Anh Nguyen, Cong Tran, Cuong Pham, *SiCLIP: An Explainable Multimodal Framework with CLIP-ViT for Occupational Silicosis Diagnosis*, Artificial Intelligence in Medicine (Q1 SICE, đang được phản biện và đã nộp lại sau nhận xét của phản biện).

II. Hội thảo

- [7] Huu Quang Hoa Nguyen, Quy Dat Tran, Tran Quoc Bao Le, Cong Tran, Cuong Pham, Hang Phuong Nguyen and **Tien Thi Tan Nguyen** (2025), A Deep Learning-based Web Application for CT reconstruction from X-ray image, Accepted for presentation at International Conference on Advances in Information and Communication Technology (SCOPUS, Q4, ISSN 2367-3370)
- [8] Cong Tran, Khanh Nguyen-Trong, Cuong Pham, Dat Tran-Anh, and **Tien Nguyen-Thi-Tan** (2022), *Improving text recognition by combining visual and lin-*

guistic features of text, Proceedings of the 11th International Symposium on Information and Communication Technology (SoICT '22). ACM, pp. 329–335. doi: 10.1145/3568562.3568624. (SCOPUS indexed).

TÀI LIỆU THAM KHẢO

- [1] Bộ Y tế, “7.000 người lao động được chẩn đoán mắc bệnh nghề nghiệp mỗi năm,” *Cổng thông tin điện tử Bộ Y tế*, 2024. [Online]. Available: <https://moh.gov.vn/>. Accessed: Apr. 7, 2025.
- [2] V. Nguyễn, “Đặc điểm dịch tễ học và dịch tễ học phân tử của bệnh bụi phổi silic ở người lao động tại một số tỉnh Việt Nam,” Trường Đại học Y Hà Nội, Hà Nội, Việt Nam, 2023.
- [3] H. V. Đỗ, “Vệ sinh lao động và bệnh nghề nghiệp,” pp. 69–82, 2007.
- [4] Bộ Khoa học và Công nghệ, “Kết quả nghiên cứu bệnh bụi phổi silic và những cảnh báo đối với người lao động trong môi trường có bụi silic tự do,” *Cổng thông tin điện tử Bộ Khoa học và Công nghệ*, 2021. <https://www.most.gov.vn/vn/Pages/chitiettin.aspx?IDNews=20266>. Accessed: Apr. 7, 2025.
- [5] E. K. Austin, C. James, and J. Tessier, “Early detection methods for silicosis in Australia and internationally: A review of the literature,” *Int. J. Environ. Res. Public Health*, vol. 18, no. 15, Art. no. 8123, 2021. DOI: doi:10.3390/ijerph18158123
- [6] A. Durairaj, P. Howlett, and J. Feary, “S17 The diagnostic accuracy of chest X-ray for the diagnosis of silicosis and how this relates to silica exposure,” *Thorax*, vol. 79, Suppl. 2, pp. A19.2–A20, 2024, DOI: 10.1136/thorax-2024-BTSabstracts.23.
- [7] B. Priego-Torres, D. Sánchez-Morillo, E. Khalili, M. Á. Conde-Sánchez, A. García-Gámez, and A. León-Jiménez, “Automated engineered-stone silicosis screening and staging using deep learning with X-rays,” *Computers in Biology and Medicine*, vol. 191, Art. no. 110153, 2025, DOI: 10.1016/j.combiomed.2025.110153.
- [8] J. Yi, K. Tapia, J. W. Robinson, and Z. Gandomkar, “Radiologists’ performance in diagnosing silicosis on high-resolution computed tomography (HRCT) scans: An online platform,” in *Medical Imaging 2024: Image Perception, Observer Performance, and Technology Assessment*, vol. 12929, Paper 1292908, SPIE, 2024, DOI: 10.1117/12.3006114.
- [9] G. J. Williams, P. Macaskill, M. Kerr, D. A. Fitzgerald, D. Isaacs, M. Codarini, M. McCaskill, K. Prelog, and J. C. Craig, “Variability and accuracy in interpretation of consolidation on chest radiography for diagnosing pneumonia in children

- under 5 years of age,” *Pediatric Pulmonology*, vol. 48, no. 12, pp. 1195–1200, 2013, DOI: 10.1002/ppul.22806.
- [10] N. H. Nguyen, H. Q. Nguyen, N. T. Nguyen, T. V. Nguyen, H. H. Pham, and T. N.-M. Nguyen, “A clinical validation of VinDr-CXR, an AI system for detecting abnormal chest radiographs,” *arXiv preprint*, arXiv:2104.02256, 2021, DOI: 10.48550/arXiv.2104.02256.
- [11] S. Sharma and K. Guleria, “A systematic literature review on deep learning approaches for pneumonia detection using chest X-ray images,” *Multimedia Tools and Applications*, vol. 83, no. 8, pp. 24101–24151, 2024, DOI: 10.1007/s11042-023-16419-1.
- [12] B. C. trị, “Nghị quyết số 57-NQ/TW ngày 22/12/2024 về phát triển khoa học, công nghệ, đổi mới sáng tạo và chuyển đổi số quốc gia,” 22/12/2024.
- [13] B. Y. tế, “Thông tư số 02/2023/TT-BYT sửa đổi, bổ sung một số điều của Thông tư số 15/2016/TT-BYT quy định về bệnh nghề nghiệp được hưởng bảo hiểm xã hội,” Hà Nội, 2023.
- [14] B. Yang, X. Liu, C. Peng, X. Meng, and Q. Jia, “Silicosis: from pathogenesis to therapeutics,” *Frontiers in Pharmacology*, vol. 16, Art. no. 1516200, 2025, DOI: 10.3389/fphar.2025.1516200.
- [15] S. Chen, M. Liu, and F. Xie, “Global and national burden and trends of mortality and disability-adjusted life years for silicosis, from 1990 to 2019: results from the Global Burden of Disease study 2019,” *BMC Pulmonary Medicine*, vol. 22, Art. no. 240, 2022, DOI: 10.1186/s12890-022-02040-9.
- [16] C. C. Leung, I. T. S. Yu, and W. Chen, “Silicosis,” *The Lancet*, vol. 379, no. 9830, pp. 2008–2018, 2012, DOI: 10.1016/S0140-6736(12)60235-9.
- [17] X.-M. Qi, Y. Luo, M.-Y. Song, Y. Liu, T. Shu, Y. Liu, J.-L. Pang, J. Wang, and C. Wang, “Pneumoconiosis: current status and future prospects,” *Chinese Medical Journal*, vol. 134, no. 8, pp. 898–907, 2021, DOI: 10.1097/CM9.0000000000001461.
- [18] International Labour Organization, “Nearly 3 million people die of work-related accidents and diseases each year,” Nov. 26, 2023. <https://www.ilo.org/resource/news/nearly-3-million-people-die-work-related-accidents-and-diseases>.
- [19] C. C. Leung, I. T. S. Yu, and W. Chen, “Silicosis,” *The Lancet*, vol. 379, no. 9830, pp. 2008–2018, 2012, DOI: 10.1016/S0140-6736(12)60235-9.

- [20] N. T. T. Huyền, L. M. Anh, and T. A. J. T. c. Y. h. D. p. Thành, “Thực trạng bệnh nghề nghiệp tại Việt Nam giai đoạn 2016–2020,” vol. 31, no. 9 Phụ bản, pp. 303–311, 2021. DOI: 10.51403/0868-2836/2021/468
- [21] T. T. K. Nhung *et al.*, “Thực trạng mắc bệnh bụi phổi silic của người lao động tại hai Công ty và một số yếu tố liên quan ở Đồng Nai năm 2020,” *Tạp chí Nghiên cứu Y học*, vol. 144, no. 8, pp. 401–409, Aug. 2021., DOI: 10.52852/tc-ncyh.v144i8.451
- [22] T. T. X. Lê *et al.*, “Đặc điểm các hình ảnh bất thường trên phim X-quang phổi và một số yếu tố liên quan của người lao động tiếp xúc với bụi silic tại Phú Yên năm 2020,” *Tạp chí Nghiên cứu Y học*, vol. 144, no. 8, pp. 300–307, 2021, DOI: 10.52852/tcncyh.v144i8.445.
- [23] International Labour Office. Guidelines for the use of the ILO International Classification of Radiographs of Pneumoconioses. Geneva: International Labour Office; 2011. (Occupational Safety and Health Series, No. 22).
- [24] V. Schlünssen *et al.*, “The prevalences and levels of occupational exposure to dusts and/or fibres (silica, asbestos and coal): A systematic review and meta-analysis from the WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury,” *Environment International*, vol. 178, Art. no. 107980, 2023, DOI: 10.1016/j.envint.2023.107980.
- [25] International Labour Office, “Guidelines for the use of the ILO International Classification of Radiographs of Pneumoconioses (Occupational Safety and Health Series No. 22, Rev. 2000),” Geneva: International Labour Office, 2002.
- [26] Y. Liu, H. Li, T. Luo, C. Zhang, Z. Xiao, Y. Wei, Y. Gao, F. Shi, F. Shan, and D. Shen, “Structural Attention Graph Neural Network for Diagnosis and Prediction of COVID-19 Severity,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 2, pp. 557–567, 2023, DOI: 10.1109/TMI.2022.3226575.
- [27] G. Panagopoulos, G. Nikolentzos, and M. Vazirgiannis, “Transfer Graph Neural Networks for Pandemic Forecasting,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 6, pp. 4838–4845, 2021. DOI: <https://ojs.aaai.org/index.php/AAAI/article/view/16616>.
- [28] D. R. Sarvamangala and R. V. Kulkarni, “Convolutional neural networks in medical image understanding: a survey,” *Evolutionary Intelligence*, vol. 15, no. 1, pp. 1–22, 2022, DOI: 10.1007/s12065-020-00540-3.
- [29] E. Çallı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen, and K. Murphy, “Deep learning for chest X-ray analysis: A survey,” *Medical Image Analysis*, vol. 72, Art. no. 102125, 2021, DOI: 10.1016/j.media.2021.102125.

- [30] N. Việt, N. N. Anh, P. T. Quân, L. T. Hương, and L. T. T. Xuân, “Mối liên quan giữa một số đặc điểm dịch tễ học, gen TNF-A với nguy cơ mắc bệnh bụi phổi Silic,” *Tạp chí Nghiên cứu Y học*, vol. 150, no. 2, pp. 202–209, Feb. 2022, DOI: 10.52852/tencyh.v150i2.734.
- [31] A. Davri, E. Birbas, T. Kanavos, G. Ntritsos, N. Giannakeas, A. T. Tzallas, and A. Batistatou, “Deep learning for lung cancer diagnosis, prognosis and prediction using histological and cytological images: A systematic review,” *Cancers*, vol. 15, no. 15, Art. no. 3981, 2023, DOI: 10.3390/cancers15153981.
- [32] M. A. Balcı, L. M. Batrancea, Ö. Akgüller, and A. Nichita, “A series-based deep learning approach to lung nodule image classification,” *Cancers*, vol. 15, no. 3, Art. no. 843, 2023, DOI: 10.3390/cancers15030843.
- [33] Y. H. Bhosale and K. S. Patnaik, “Application of deep learning techniques in diagnosis of COVID-19 (coronavirus): A systematic review,” *Neural Processing Letters*, vol. 55, no. 3, pp. 3551–3603, 2023, DOI: 10.1007/s11063-022-11023-0.
- [34] R. Kundu, R. Das, Z. W. Geem, G.-T. Han, and R. Sarkar, “Pneumonia detection in chest X-ray images using an ensemble of deep learning models,” *PLOS ONE*, vol. 16, no. 9, Art. no. e0256630, 2021, DOI: 10.1371/journal.pone.0256630.
- [35] S. Albahli, H. T. Rauf, A. Algosaiibi, and V. E. Balas, “AI-driven deep CNN approach for multi-label pathology classification using chest X-rays,” *PeerJ Computer Science*, vol. 7, Art. no. e495, 2021, DOI: 10.7717/peerj-cs.495.
- [36] C. X. Cai, B. Y. Zhu, and H. Chen, “Computer-aided diagnosis for pneumoconiosis based on texture analysis on digital chest radiographs,” *Applied Mechanics and Materials*, vols. 241–244, pp. 244–247, 2013, DOI: 10.4028/www.scientific.net/AMM.241-244.244.
- [37] M. Masud, A. K. Bairagi, Abdullah-Al Nahid, N. Sikder, S. Rubaiee, A. Ahmed, and D. Anand, “A pneumonia diagnosis scheme based on hybrid features extracted from chest radiographs using an ensemble learning algorithm,” *Journal of Healthcare Engineering*, vol. 2021, Art. no. 8862089, 2021, DOI: 10.1155/2021/8862089.
- [38] T. B. Chandra and K. Verma, “Pneumonia detection on chest X-ray using machine learning paradigm,” in *Proceedings of the 3rd International Conference on Computer Vision and Image Processing (CVIP 2018)*, Singapore: Springer, 2020, pp. 21–33, DOI: 10.1007/978-981-32-9088-4_3.
- [39] K. M. Kuo, P. C. Talley, C. H. Huang, and L. C. Cheng, “Predicting hospital-acquired pneumonia among schizophrenic patients: A machine learning ap-

- proach,” *BMC Medical Informatics and Decision Making*, vol. 19, Art. no. 42, 2019, DOI: 10.1186/s12911-019-0792-1.
- [40] H. Yue, Q. Yu, Y. Liu, J. Liu, Y. Huang, W. Chen, X. Li, and X. Yang, “Machine learning-based CT radiomics method for predicting hospital stay in patients with pneumonia associated with SARS-CoV-2 infection: A multicenter study,” *Annals of Translational Medicine*, vol. 8, no. 14, Art. no. 859, 2020, DOI: 10.21037/atm-20-3026.
- [41] P. Harber, J. M. McCoy, K. Howard, D. Greer, and J. Luo, “Artificial intelligence-assisted occupational lung disease diagnosis,” *Chest*, vol. 100, no. 2, pp. 340–346, 1991, DOI: 10.1378/chest.100.2.340.
- [42] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.
- [43] D. Shen, G. Wu, and H.-I. Suk, “Deep learning in medical image analysis,” *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, 2017, DOI: 10.1146/annurev-bioeng-071516-044442.
- [44] G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017, DOI: 10.1016/j.media.2017.07.005.
- [45] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, et al., “CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning,” *arXiv*, 2017, DOI: 10.48550/arXiv.1711.05225.
- [46] S. Okeke, M. Sain, U. J. Maduh, and D.-U. Jeong, “An efficient deep learning approach to pneumonia classification in healthcare,” *Journal of Healthcare Engineering*, vol. 2019, Art. no. 4180949, 7 pp., 2019, DOI: 10.1155/2019/4180949.
- [47] R. E. Al Mamlook, S. Chen, and H. F. Bzizi, “Investigation of the performance of machine learning classifiers for pneumonia detection in chest X-ray images,” in *Proc. 2020 IEEE Int. Conf. Electro Information Technology (EIT)*, Chicago, IL, USA, 2020, pp. 98–104, DOI: 10.1109/EIT48999.2020.9208232.
- [48] Zhang, J., Xie, Y., Pang, G., Liao, Z., Verjans, J., Li, W., Sun, Z., He, J., Li, Y., Shen, C., & Xia, Y. (2021). Viral Pneumonia Screening on Chest X-Rays Using Confidence-Aware Anomaly Detection. *IEEE transactions on medical imaging*, 40(3), 879–890. <https://doi.org/10.1109/TMI.2020.3040950>
- [49] T. Rahman, M. E. H. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Kadir, and S. Kashem, “Transfer learning with deep convo-

- lutional neural network (CNN) for pneumonia detection using chest X-ray,” *Applied Sciences*, vol. 10, no. 9, Art. no. 3233, 2020, DOI: 10.3390/app10093233.
- [50] G. Liang and L. Zheng, “A transfer learning method with deep residual network for pediatric pneumonia diagnosis,” *Computer Methods and Programs in Biomedicine*, vol. 187, Art. no. 104964, 2020, DOI: 10.1016/j.cmpb.2019.06.023.
- [51] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, “Automated detection of COVID-19 cases using deep neural networks with X-ray images,” *Computers in Biology and Medicine*, vol. 121, Art. no. 103792, 2020, DOI: 10.1016/j.combiomed.2020.103792.
- [52] A. J. DeGrave, J. D. Janizek, and S.-I. Lee, “AI for radiographic COVID-19 detection selects shortcuts over signal,” *Nature Machine Intelligence*, vol. 3, pp. 610–619, 2021, DOI: 10.1038/s42256-021-00338-7.
- [53] Turker Tuncer, Fatih Ozyurt, Sengul Dogan, Abdulhamit Subasi, A novel Covid-19 and pneumonia classification method based on F-transform, *Chemometrics and Intelligent Laboratory Systems*, Volume 210, 2021, 104256, ISSN 0169-7439, j.chemolab.2021.104256
- [54] G. K. Sharma, P. Harjule, B. Agarwal, and R. Kumar, “Silicosis detection using extended transfer learning model,” in *Recent Trends in Image Processing and Pattern Recognition (RTIP2R 2023)*, Commun. Comput. Inf. Sci., vol. 2027, Cham, Switzerland: Springer, 2024, DOI: 10.1007/978-3-031-53085-2_10.
- [55] D. Ahmedt-Aristizabal, M. A. Armin, S. Denman, C. Fookes, and L. Petersson, “Graph-based deep learning for medical diagnosis and analysis: Past, present and future,” *Sensors*, vol. 21, no. 14, p. 4758, 2021, DOI: 10.3390/s21144758.
- [56] A. Sekuboyina, D. O noro-Rubio, J. Kleesiek, and B. Malone, “A relational-learning perspective to multi-label chest X-ray classification,” in *Proc. 2021 IEEE 18th Int. Symp. Biomedical Imaging (ISBI)*, Nice, France, 2021, pp. 1618–1622, DOI: 10.1109/ISBI48211.2021.9433786.
- [57] L. Zhang, Y. Zhao, T. Che, S. Li, and X. Wang, “Graph neural networks for image-guided disease diagnosis: A review,” *iRadiology*, vol. 1, no. 2, pp. 151–166, 2023, DOI: 10.1002/ird3.20.
- [58] D. Raghu, H. K. Tripathy, and R. Borreo, “A novel graph convolutional neural networks (GCNNs)-based framework to enhance the detection of COVID-19 from X-ray and CT scan images,” *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 4, 2024, DOI: 10.14569/IJACSA.2024.0150473.

- [59] B. Wang, H. Pan, A. Aboah, Z. Zhang, E. Keles, D. Torigian, B. Turkbey, E. Krupinski, J. Udupa, and U. Bagci, “GazeGNN: A gaze-guided graph neural network for chest X-ray classification,” in *Proc. 2024 IEEE/CVF Winter Conf. Applications of Computer Vision (WACV)*, Waikoloa, HI, USA, 2024, pp. 2183–2192, DOI: 10.1109/WACV57701.2024.00219.
- [60] J. Sultana, R. Qin, and Z. Yin, “Seeing through expert’s eyes: Leveraging radiologist eye gaze and speech report with graph neural networks for chest X-ray image classification,” in *Computer Vision – ACCV 2024*, Lecture Notes in Computer Science, vol. 15473, Singapore: Springer, 2025, DOI: 10.1007/978-981-96-0901-7_9.
- [61] A. Paul, Y. X. Tang, T. C. Shen, and R. M. Summers, “Discriminative ensemble learning for few-shot chest X-ray diagnosis,” *Medical Image Analysis*, vol. 68, Art. no. 101911, 2021, DOI: 10.1016/j.media.2020.101911.
- [62] Junwen Chen, Tong Liu, Yangguang Cui, Xiaoqiang Li, Weiqin Tong, A meta-learning based method for few-shot pneumonia identification using chest X-ray images, *Biomedical Signal Processing and Control*, Volume 95, Part B, 2024, 106433, ISSN 1746-8094, j.bspc.2024.106433
- [63] D. Atukunda, R. Mwangi, and P. Muriithi, “Few-shot pneumonia detection: Integrating Siamese networks with transfer learning for improved chest X-ray image classification,” *Authorea*, Aug. 27, 2024.<https://www.authorea.com/users/820956/articles/1219219-few-shot-pneumonia-detection-integrating-siamese-networks>
- [64] M. J. Kim, S. G. Chae, S. J. Bae, et al., “Unsupervised few-shot learning architecture for diagnosis of periodontal disease in dental panoramic radiographs,” *Scientific Reports*, vol. 14, Art. no. 23237, 2024, DOI: 10.1038/s41598-024-73665-5.
- [65] T. T. T. Nguyen, Q. B. Bui, and C. Pham, “A data-constrained approach for occupational silicosis detection on chest X-rays with few-shot learning,” *Journal of Computer Science and Cybernetics*, vol. 41, no. 1, pp. 1–16, Mar. 2025, DOI: 10.15625/1813-9663/21668.
- [66] D. Atukunda, W. Mwangi, and P. Muriithi, “An ensemble learning approach using self-supervised and meta-learning for few-shot pneumonia detection in chest X-ray images,” *Research Square*, 2025, DOI: 10.21203/rs.3.rs-7063509/v1.
- [67] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, DOI: 10.1038/nature14539.

- [68] J. Wang, S. Wang, and Y. Zhang, “Deep learning on medical image analysis,” *CAAI Transactions on Intelligence Technology*, vol. 10, no. 1, pp. 1–35, 2025, DOI: 10.1049/cit2.12356.
- [69] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A survey of convolutional neural networks: Analysis, applications, and prospects,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 12, pp. 6999–7019, 2021, DOI: 10.1109/TNNLS.2021.3084827.
- [70] C. Chen, N. A. Mat Isa, and X. Liu, “A review of convolutional neural network based methods for medical image classification,” *Computers in Biology and Medicine*, vol. 185, Art. no. 109507, 2025, DOI: 10.1016/j.compbiomed.2024.109507.
- [71] M. A. K. Raiaan, S. Sakib, N. M. Fahad, A. A. Mamun, M. A. Rahman, S. Shatabda, and M. S. H. Mukta, “A systematic review of hyperparameter optimization techniques in Convolutional Neural Networks,” *Decision Analytics Journal*, vol. 11, Art. no. 100470, 2024, DOI: 10.1016/j.dajour.2024.100470.
- [72] G. K. Sharma, P. Harjule, B. Agarwal, and R. Kumar, “Silicosis Detection Using Extended Transfer Learning Model,” in *Recent Trends in Image Processing and Pattern Recognition: 6th International Conference, RTIP2R 2023, Derby, UK, December 7–8, 2023, Revised Selected Papers, Part II* (Communications in Computer and Information Science, vol. 2027), Cham: Springer, pp. 111–126, 2024, DOI: 10.1007/978-3-031-53085-2_10.
- [73] O. Stephen, M. Sain, U. J. Maduh, and D.-U. Jeong, “An efficient deep learning approach to pneumonia classification in healthcare,” *Journal of Healthcare Engineering*, vol. 2019, Art. no. 4180949, 2019, DOI: 10.1155/2019/4180949.
- [74] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. R. Acharya, “Automated detection of COVID-19 cases using deep neural networks with X-ray images,” *Computers in Biology and Medicine*, vol. 121, Art. no. 103792, 2020, DOI: 10.1016/j.compbiomed.2020.103792.
- [75] R. E. Al Mamlook, S. Chen, and H. F. Bzizi, “Investigation of the performance of machine learning classifiers for pneumonia detection in chest X-ray images,” in *Proc. 2020 IEEE Int. Conf. Electro Information Technology (EIT)*, pp. 98–104, 2020, DOI: 10.1109/EIT48999.2020.9208232.
- [76] J. D. Janizek, G. Erion, A. J. DeGrave, and S.-I. Lee, “An adversarial approach for the robust classification of pneumonia from chest radiographs,” in *Proc. ACM Conf. Health, Inference, and Learning (CHIL '20)*, pp. 69–79, 2020, DOI: 10.1145/3368555.3384458.

- [77] J. Zhang, Y. Xie, G. Pang, Z. Liao, J. Verjans, W. Li, Z. Sun, J. He, Y. Li, C. Shen, and Y. Xia, “Viral pneumonia screening on chest X-rays using confidence-aware anomaly detection,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 879–890, 2021, DOI: 10.1109/TMI.2020.3040950.
- [78] T. Tuncer, F. Ozyurt, S. Dogan, and A. Subasi, “A novel Covid-19 and pneumonia classification method based on F-transform,” *Chemometrics and Intelligent Laboratory Systems*, vol. 210, Art. no. 104256, 2021, DOI: 10.1016/j.chemolab.2021.104256.
- [79] T. Rahman, M. E. H. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Kadir, and S. B. A. Kashem, “Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray,” *Applied Sciences*, vol. 10, no. 9, Art. no. 3233, 2020, DOI: 10.3390/app10093233.
- [80] G. Liang and L. Zheng, “A transfer learning method with deep residual network for pediatric pneumonia diagnosis,” *Computer Methods and Programs in Biomedicine*, vol. 187, Art. no. 104964, 2020, DOI: 10.1016/j.cmpb.2019.06.023.
- [81] A. U. Ibrahim, M. Ozsoz, S. Serte, F. M. Al-Turjman, and P. S. Yakoi, “Pneumonia classification using deep learning from chest X-ray images during COVID-19,” *Cognitive Computation*, vol. 16, no. 4, pp. 1589–1601, 2024, DOI: 10.1007/s12559-020-09787-5.
- [82] U. Shah, A. Abd-Alrazeq, T. Alam, M. Househ, and Z. Shah, “An efficient method to predict pneumonia from chest X-rays using deep learning approach,” in *The Importance of Health Informatics in Public Health during a Pandemic* (Studies in Health Technology and Informatics, vol. 272), IOS Press, pp. 457–460, 2020, DOI: 10.3233/SHTI200594.
- [83] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 2097–2106, 2017, DOI: 10.1109/CVPR.2017.369.
- [84] A. E. W. Johnson, T. J. Pollard, S. J. Berkowitz, N. R. Greenbaum, M. P. Lungren, C.-y. Deng, R. G. Mark, S. Horng, and B. Moody, “MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports,” *Scientific Data*, vol. 6, no. 1, Art. no. 317, 2019, DOI: 10.1038/s41597-019-0322-0.
- [85] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A.

- Kazerooni, H. MacMahon, E. J. R. van Beek, D. Yankelevitz, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M. Engelmann, G. E. Laderach, ... L. P. Clarke, “The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans,” *Medical Physics*, vol. 38, no. 2, pp. 915–931, 2011, DOI: 10.1118/1.3528204.
- [86] Y. Akhter, R. Ranjan, M. Vatsa, R. Singh, S. Chaudhury, A. Agrawal, P. Rao, S. Aggarwal, and A. Kalyanpur, “Silicodata: A Chest X-ray Dataset for Silicosis Detection,” *Dataset*, figshare, 2025. Available: https://figshare.com/articles/dataset/Silicodata_A_benchmark_chest_X-ray_dataset_for_Silicosis_detection_with_radiologist_s_annotations/25383697.
- [87] H. Q. Nguyen *et al.*, “VinDr-CXR: An open dataset of chest X-rays with radiologist’s annotations,” *Scientific Data*, vol. 9, no. 1, Art. no. 429, 2022, DOI: 10.1038/s41597-022-01498-w.
- [88] L. T. Hương, *Báo cáo tổng hợp kết quả thực hiện nhiệm vụ: “Nghiên cứu đặc điểm dịch tễ học phân tử, yếu tố nguy cơ và ứng dụng kỹ thuật tiên tiến trong chẩn đoán sớm bệnh bụi phổi silic tại Việt Nam”*, Bộ Khoa học và Công nghệ, 2021.
- [89] D. S. Kermany, K. Zhang, and M. Goldbaum, “Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification,” *Mendeley Data*, V3, 2018, DOI: 10.17632/rscbjbr9sj.3.
- [90] C. I. Patel, A. Patel, and D. Patel, “Optical character recognition by open source OCR tool Tesseract: A case study,” *International Journal of Computer Applications*, vol. 55, no. 10, pp. 50–56, 2012, DOI: 10.5120/8794-2784.
- [91] R. Smith, “An overview of the Tesseract OCR engine,” in *Proc. 9th Int. Conf. Document Analysis and Recognition (ICDAR 2007)*, vol. 2, pp. 629–633, 2007, DOI: 10.1109/ICDAR.2007.4376991.
- [92] A. Parnami and M. Lee, “Learning from few examples: A summary of approaches to few-shot learning,” *arXiv preprint*, arXiv:2203.04291, 2022, DOI: 10.48550/arXiv.2203.04291.
- [93] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, “Generalizing from a few examples: A survey on few-shot learning,” *ACM Computing Surveys*, vol. 53, no. 3, Art. no. 63, pp. 1–34, 2020, DOI: 10.1145/3386252.
- [94] G. Koch, R. Zemel, and R. Salakhutdinov, “Siamese neural networks for one-shot image recognition,” in *Proceedings of the 32nd International Conference on Ma-*

- chine Learning (ICML)*, JMLR: W&CP, vol. 37, Lille, France, 2015. Available: <https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf>
- [95] J. Snell, K. Swersky, and R. S. Zemel, “Prototypical networks for few-shot learning,” *arXiv preprint*, arXiv:1703.05175, 2017, DOI: 10.48550/arXiv.1703.05175
 - [96] T. M. Hospedales, A. Antoniou, P. Micaelli, and A. J. Storkey, “Meta-Learning in Neural Networks: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 9, pp. 5149–5169, 2022, DOI: 10.1109/TPAMI.2021.3079209
 - [97] A. Galán-Cuenca, A. J. Gallego, M. Saval-Calvo, and A. Pertusa, “Few-shot learning for COVID-19 chest X-ray classification with imbalanced data: An inter vs. intra domain study,” *Pattern Analysis and Applications*, vol. 27, no. 3, Art. no. 69, 2024, DOI: 10.1007/s10044-024-01285-w.
 - [98] J. Chen, T. Liu, Y. Cui, X. Li, and W. Tong, “A meta-learning based method for few-shot pneumonia identification using chest X-ray images,” *Biomedical Signal Processing and Control*, vol. 95, Art. no. 106433, 2024, DOI: 10.1016/j.bspc.2024.106433.
 - [99] M. F. Maulana, A. Y. Pramana, F. I. Permana, and D. R. Fudholi, “Few-Shot Learning Approach on Tuberculosis Classification Based on Chest X-Ray Images,” in *Proc. 2024 IEEE Int. Conf. Future Machine Learning and Data Science (FMLDS)*, pp. 57–62, 2024. (Preprint: arXiv:2409.11644, DOI: 10.48550/arXiv.2409.11644).
 - [100] S. Jadon, “COVID-19 detection from scarce chest X-ray image data using few-shot deep learning approach,” in *Medical Imaging 2021: Imaging Informatics for Healthcare, Research, and Applications*, vol. 11601, Paper 116010X, SPIE, 2021, DOI: 10.1117/12.2581496.
 - [101] A. Paul, Y.-X. Tang, T. C. Shen, and R. M. Summers, “Discriminative ensemble learning for few-shot chest X-ray diagnosis,” *Medical Image Analysis*, vol. 68, Art. no. 101911, 2021, DOI: 10.1016/j.media.2020.101911.
 - [102] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path aggregation network for instance segmentation,” in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 8759–8768, 2018, DOI: 10.1109/CVPR.2018.00913.
 - [103] Z. Tian, H. Zhao, M. Shu, Z. Yang, R. Li, and J. Jia, “Prior guided feature enrichment network for few-shot segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 2, pp. 1050–1065, 2022, DOI: 10.1109/TPAMI.2020.3013717.

- [104] J. Min, D. Kang, and M. Cho, “Hypercorrelation squeeze for few-shot segmentation,” in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, pp. 6941–6952, 2021, DOI: 10.1109/ICCV48922.2021.00686.
- [105] D. Kang and M. Cho, “Integrative few-shot learning for classification and segmentation,” in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 9979–9990, 2022, DOI: 10.1109/CVPR52688.2022.00974.
- [106] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint*, arXiv:1412.6980, 2014, DOI: 10.48550/arXiv.1412.6980.
- [107] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint*, arXiv:1704.04861, 2017, DOI: 10.48550/arXiv.1704.04861.
- [108] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, “Searching for MobileNetV3,” in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, pp. 1314–1324, 2019, DOI: 10.1109/ICCV.2019.00140.
- [109] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, pp. 10012–10022, 2021, DOI: 10.1109/ICCV48922.2021.00986.
- [110] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 4700–4708, 2017, DOI: 10.1109/CVPR.2017.243.
- [111] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint*, arXiv:1609.02907, 2016, DOI: 10.48550/arXiv.1609.02907.
- [112] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, “Graph neural networks: A review of methods and applications,” *AI Open*, vol. 1, pp. 57–81, 2020, DOI: 10.1016/j.aiopen.2021.01.001.
- [113] B. Khemani, S. Patil, K. Kotecha, and S. Tanwar, “A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions,” *Journal of Big Data*, vol. 11, Art. no. 18, 2024, DOI: 10.1186/s40537-023-00876-4.
- [114] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, “Neural message passing for quantum chemistry,” in *Proc. 34th Int. Conf. Ma-*

- chine Learning (ICML 2017)*, PMLR, vol. 70, pp. 1263–1272, 2017. (Preprint: arXiv:1704.01212). DOI: 10.48550/arXiv.1704.01212.
- [115] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph attention networks,” *arXiv preprint*, arXiv:1710.10903, 2017, DOI: 10.48550/arXiv.1710.10903.
 - [116] W. L. Hamilton, R. Ying, and J. Leskovec, “Inductive representation learning on large graphs,” in *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, 31st Conf. Neural Information Processing Systems, Long Beach, CA, USA, 2017. <https://papers.nips.cc/paper/6703-inductive-representation-learning-on-large-graphs>.
 - [117] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, “How powerful are graph neural networks?,” *arXiv preprint*, arXiv:1810.00826, 2018, DOI: 10.48550/arXiv.1810.00826.
 - [118] Ashish Vaswani, et al. *Attention Is All You Need. In Advances in Neural Information Processing Systems (NIPS), 2017.* arXiv:1706.03762.
 - [119] J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” *Journal of Big Data*, vol. 6, Art. no. 27, 2019, DOI: 10.1186/s40537-019-0192-5.
 - [120] 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," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.
 - [121] Q. Jodelet, X. Liu, and T. Murata, “Balanced softmax cross-entropy for incremental learning with and without memory,” *Computer Vision and Image Understanding*, vol. 225, Art. no. 103582, 2022, DOI: 10.1016/j.cviu.2022.103582.
 - [122] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948, DOI: 10.1002/j.1538-7305.1948.tb01338.x.
 - [123] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: <https://www.deeplearningbook.org>
 - [124] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, “Cost-sensitive learning of deep feature representations from imbalanced data,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3573–3587, 2018, DOI: 10.1109/TNNLS.2017.2732482.

- [125] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, “Class-balanced loss based on effective number of samples,” in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 9268–9277, 2019, DOI: 10.1109/CVPR.2019.00949.
- [126] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint*, arXiv:1704.04861, 2017, DOI: 10.48550/arXiv.1704.04861.
- [127] A. Alhudhaif, K. Polat, and O. Karaman, “Determination of COVID-19 pneumonia based on generalized convolutional neural network model from chest X-ray images,” *Expert Systems with Applications*, vol. 180, Art. no. 115141, 2021, DOI: 10.1016/j.eswa.2021.115141