



BỘ LAO ĐỘNG - THƯƠNG BINH VÀ XÃ HỘI
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT VĨNH LONG

KHÓA LUẬN TỐT NGHIỆP
ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

PHÁT HIỆN BỆNH PHỔI DỰA VÀO ẢNH X-RAY
LÒNG NGỰC BẰNG PHƯƠNG PHÁP MÁY HỌC

CBHD:	TS. Phan Anh Cang
Sinh viên:	Nguyễn Hồng Hải
Mã số sinh viên:	19004053

Vĩnh Long - Năm 2023



BỘ LAO ĐỘNG - THƯƠNG BINH VÀ XÃ HỘI
TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT VĨNH LONG

KHÓA LUẬN TỐT NGHIỆP
ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

PHÁT HIỆN BỆNH PHỔI DỰA VÀO ẢNH X-RAY
LÒNG NGỰC BẰNG PHƯƠNG PHÁP MÁY HỌC

CBHD:	TS. Phan Anh Cang
Sinh viên:	Nguyễn Hồng Hải
Mã số sinh viên:	19004053

Vĩnh Long - Năm 2023

LỜI CAM ĐOAN

Tôi xin cam đoan kết quả đạt được trong luận văn là sản phẩm của riêng cá nhân, không sao chép lại của người khác. Trong toàn bộ nội dung của luận văn, những điều được trình bày hoặc là của cá nhân hoặc là được tổng hợp từ nhiều nguồn tài liệu. Tất cả các tài liệu tham khảo đều có xuất xứ rõ ràng và được trích dẫn hợp pháp. Đây là công trình nghiên cứu của cá nhân tôi và được sự hướng dẫn thực hiện của Tiến sĩ Phan Anh Cang. Các nội dung nghiên cứu trong đề tài **“Phát hiện bệnh phổi dựa vào ảnh X-Ray lồng ngực bằng phương pháp máy học”** của tôi là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây và dựa trên một số tài liệu, đề tài trước. Tôi xin hoàn toàn chịu trách nhiệm và chịu mọi hình thức kỷ luật theo quy định cho lời cam đoan của mình.

Vĩnh Long, ngày tháng năm 2023

Người thực hiện đề tài

Nguyễn Hồng Hải

LỜI CẢM ƠN

Trước hết, em xin bày tỏ lòng biết ơn sâu sắc đến tất cả các giảng viên thuộc Khoa Công nghệ thông tin - Trường Đại học Sư Phạm Kỹ Thuật Vĩnh Long, đặc biệt là thầy Phan Anh Cang. Em không thể đánh giá đúng bằng lời văn những đóng góp to lớn của các thầy cô, những người đã dành cho em rất nhiều thời gian và tận tâm hướng dẫn, hỗ trợ, và góp ý trong quá trình thực hiện khóa luận tốt nghiệp này.

Em muốn gửi lời cảm ơn chân thành đến tất cả các giảng viên đã truyền đạt những kiến thức quý báu về học tập và cuộc sống trong suốt thời gian em học tập tại trường. Các thầy cô đã dành thời gian để đọc và phản biện bài báo cáo của em, giúp em hiểu rõ hơn về những vấn đề mà em nghiên cứu.

Dù em đã cố gắng hết sức trong quá trình học tập và thực hiện khóa luận, em không thể tránh khỏi những thiếu sót. Em rất mong nhận được sự góp ý quý báu từ tất cả các thầy cô cũng như các bạn để làm cho kết quả của em trở nên hoàn thiện hơn từng ngày.

Cuối cùng, em xin kính chúc mừng tất cả các thầy cô với sức khỏe dồi dào và thành công trong sự nghiệp cao quý của mình.

Một lần nữa, em xin bày tỏ lòng biết ơn chân thành và sâu sắc!

Vĩnh Long, ngày tháng năm 2023

Người thực hiện đề tài

Nguyễn Hồng Hải

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
DANH SÁCH CÁC TỪ VIẾT TẮT	vi
DANH MỤC HÌNH ẢNH.....	vii
DANH SÁCH BẢNG	viii
CHƯƠNG 1 - CƠ SỞ KHOA HỌC CỦA ĐỀ TÀI.....	1
1.1 Giới thiệu.....	1
1.2 Lý do chọn đề tài.....	2
1.3 Cơ sở lý luận	3
1.4 Cơ sở thực tiễn	4
1.5 Những nghiên cứu liên quan	4
1.6 Vấn đề đặt ra	7
1.7 Hướng giải quyết của đề tài	7
CHƯƠNG 2 - CƠ SỞ LÝ THUYẾT	8
2.1 Bệnh tràn khí màng phổi	8
2.1.1 Nguyên nhân	8
2.1.2 Dấu hiệu và triệu chứng	9
2.1.3 Biến chứng và mức độ nguy hiểm của bệnh tràn khí màng phổi.....	10
2.2 Đặc điểm nhận dạng tràn khí màng phổi trên ảnh X-Ray.....	11
2.3 Tổng quan về trí tuệ nhân tạo.....	12
2.3.1 Khái niệm trí tuệ nhân tạo.....	12
2.3.2 Một số công nghệ Trí tuệ nhân tạo phổ biến hiện nay.....	12
2.4 Machine Learning	15
2.4.1 Khái niệm Machine Learning	15
2.4.2 Machine Learning Workflow.....	16

2.4.3 Ứng dụng của Machine Learning.....	16
2.5 Deep Learning	17
2.5.1 Khái niệm Deep Learning	17
2.5.2 Cách thức hoạt động của Deep Learning	18
2.5.3 Các ưu và nhược điểm của Deep Learning	19
2.5.4 Ứng dụng của Deep Learning	20
2.5.5 Mạng nơ-ron tích chập CNN.....	21
2.5.6 Kiến trúc mạng Densenet121	23
2.5.7 Kiến trúc mạng Xception	25
2.5.8 Kiến trúc mạng Resnet50	26
2.5.9 Kiến trúc mạng Unet	28
2.5.10 Kiến trúc mạng DeepLabV3+	29
2.5.11 Kiến trúc mạng Vision Transformer	30
2.5.12 Phương pháp đánh giá mô hình mạng học sâu	32
CHƯƠNG 3 - PHƯƠNG PHÁP NGHIÊN CỨU	36
3.1 Tập dữ liệu	36
3.2 Môi trường thực nghiệm	36
3.2 Mô hình kiến trúc đề xuất	37
3.4 Kịch bản và các tham số liên quan.....	38
CHƯƠNG 4 - KẾT QUẢ.....	39
4.1 Kết quả huấn luyện.....	39
4.4.1 Kết quả huấn luyện pha phân loại	39
4.4.2 Kết quả huấn luyện pha phân đoạn	46
4.2 Kết quả kiểm thử.....	51
4.2.1 Phân loại tràn khí màng phổi	51
4.2.2 Phân vùng tràn khí màng phổi	51

CHƯƠNG 5: KẾT LUẬN	53
5.1 Kết quả đạt được	53
5.2 Hạn chế của đề tài	53
5.2 Hướng phát triển	53
TÀI LIỆU KHAM KHẢO.....	54

DANH SÁCH CÁC TỪ VIẾT TẮT

STT	Kí hiệu	Nội dung đầy đủ
1	AI	Artificial Intelligent
2	ARDS	Acute Respiratory Distress Syndrome
3	CAD	Computer Aided Design
4	COPD	Bệnh phổi tắc nghẽn mạn tính
5	CDC	Centers for Disease Control and Prevention
6	CPU	Central Processing Unit
7	CLAHE	Contrast Limited Adaptive Histogram Equalization
8	CNN	Convolutional Neural Networks
9	CV	Computer Vision
10	DNN	Deep Neural Network
11	DL	Deep Learning
12	FC	Fully Connected
13	GPU	Graphics Processing Unit
14	mAP	Mean Average Precision
15	ML	Machine Learning
16	MLW	Machine Learning Workflow
17	RT-PCR	Real time Polemerase Chain Reaction
18	RSNA	Radiological Society of North America
19	TPU	Tensor Processing Unit
20	SIRM	Società Italiana di Radiologia Medica e Interventistica
21	IOT	Internet Of Things
22	WHO	World Health Organization

DANH MỤC HÌNH ẢNH

Hình 2.1. Ảnh chụp X-Ray tràn khí phổi [3]	11
Hình 2.2. Mô hình mạng CNN[20]	22
Hình 2.3. Thông số kiến trúc DenseNet121[27]	23
Hình 2.4. Kiến trúc mạng Xception[28]	25
Hình 2.5. Kiến trúc depthwise separable convolution[28]	25
Hình 2.6. Kiến trúc Resnet50[29]	26
Hình 2.7. Kiến trúc mạng Unet[30]	28
Hình 2.8. Kiến trúc mạng DeeplabV3+ [31].....	29
Hình 2.9. Mô tả Atrous Spatial Pyramid Pooling [31]	30
Hình 2.10. Kiến trúc Vision Transformer[32]	31
Hình 2.11. Sự thay đổi của TPR và FPR[33].....	33
Hình 2.12. Confusion matrix[34]	34
Hình 3.1. Kiến trúc đề xuất mô hình phân loại và phân đoạn.....	37
Hình 4.1. Biểu đồ so sánh độ chính xác các mô hình phân loại	40
Hình 4.2. Biểu đồ so sánh độ mất mát các mô hình phân loại.....	41
Hình 4.3. Biểu đồ so sánh AUC các mô hình phân loại	43
Hình 4.4. Biểu đồ so sánh Confusion matrix-Accuracy phân loại.....	44
Hình 4.5. Biểu đồ so sánh Confusion matrix-Precision phân loại	44
Hình 4.6. Biểu đồ so sánh Confusion matrix-Recall phân loại.....	45
Hình 4.7. Biểu đồ so sánh Confusion matrix-F1-score phân loại.....	45
Hình 4.8. Biểu đồ so sánh thời gian huấn luyện các mô hình phân loại	46
Hình 4.9. Biểu đồ so sánh độ chính xác các mô hình phân đoạn.....	47
Hình 4.10. Biểu đồ so sánh độ mất mát các mô hình phân đoạn	49
Hình 4.11. Biểu đồ so sánh Dice coefficient các mô hình phân đoạn	50
Hình 4.12. Biểu đồ so sánh thời gian huấn luyện các mô hình phân đoạn	50

DANH SÁCH BẢNG

Bảng 4.1. Thông số môi trường huấn luyện.....	36
Bảng 4.2. Thông số kích bản huấn luyện.....	38
Bảng 4.3. Độ chính xác các mô hình pha phân loại.....	39
Bảng 4.4. Độ mất mát các mô hình pha phân loại	40
Bảng 4.5. ROC-AUC các mô hình pha phân loại	42
Bảng 4.6. Kết quả Confusion matrix các mô hình phân loại	43
Bảng 4.7. Độ chính xác các mô hình pha phân đoạn.....	46
Bảng 4.8. Độ mất mát các mô hình phân đoạn	48
Bảng 4.9. Dice coefficient các mô hình phân đoạn	49
Bảng 4.10. Kết quả kiểm thử các mô hình phân loại	51
Bảng 4.11. Kết quả kiểm thử các mô hình phân đoạn	51

CHƯƠNG 1 - CƠ SỞ KHOA HỌC CỦA ĐỀ TÀI

1.1 Giới thiệu

Tràn khí màng phổi nói đến một tình trạng xẹp của một hoặc cả hai lá phổi do thoát khí vào trong khoang màng phổi. Khoang màng phổi là khoảng không gian nằm giữa các lá màng phổi, hai lớp màng bao bọc các lá phổi của bạn. Trên lâm sàng, có thể gặp các trường hợp tràn khí màng phổi như tự phát, thứ phát và tràn khí màng phổi do chấn thương[1].

Tràn khí màng phổi tự phát

Tràn khí màng phổi tự phát là hiện tượng khí từ nhu mô phổi thoát ra và ứ đọng trong khoang màng phổi, từ đó làm tổn thương phế nang.

Nguyên nhân của tràn khí màng phổi tự phát không phải do những chấn thương từ bên ngoài hoặc do một vết thương nào đó. Ở trạng thái bình thường, khoang màng phổi không chứa khí và áp lực sẽ là âm làm cho phổi dễ dàng trao đổi khí, ngược lại khi tràn khí màng phổi tự phát diễn ra, phổi sẽ bị xẹp lại làm cho nó không thể thực hiện trao đổi khí được nữa.

Tràn khí màng phổi tự phát là bệnh lý diễn ra trên người bình thường trước đó, thường là nam giới còn trẻ (tuổi từ 15 đến 34). Cơ chế là do vỡ những bóng khí trên bề mặt phổi, có thể là do bẩm sinh hoặc viêm tiểu phế quản. Những trường hợp này thường xảy ra đối với người gầy và cao nhiều hơn vì áp lực đỉnh phổi của họ thấp nên dễ gây nên hiện tượng vỡ bóng khí.

Hút thuốc lá và/hoặc cần sa làm tăng đáng kể nguy cơ xảy ra tràn khí màng phổi tự phát, cũng như tăng khả năng bị xẹp phổi lần nữa.

Tràn khí màng phổi tự phát vẫn có khả năng bị tái phát sau đó với tỉ lệ là 30%.

Tràn khí màng phổi thứ phát

Là bệnh lý tràn khí màng phổi diễn ra trên một bệnh nhân đã bị bệnh phổi trước đó, trường hợp này sẽ có tiên lượng xấu hơn tràn khí màng phổi tự phát. Tràn khí màng phổi thứ phát thường gặp ở những đối tượng trên 30 tuổi.

Theo thống kê từ Bộ Y tế thì có tới hơn 30% tỷ lệ người bệnh đã từng bị tràn khí màng phổi sẽ bị tái phát bệnh ít nhất một lần. Nguyên nhân dẫn tới tình trạng tái phát tràn khí màng phổi có thể bắt nguồn từ những yếu tố sau đây:

- Người bệnh đang mắc phải các bệnh lý mạn tính có liên quan đến hệ hô hấp như bệnh phổi tắc nghẽn mạn tính (COPD) hoặc một số căn bệnh về phổi khác như: viêm phổi mạn, bệnh lao phổi, xơ nang, bệnh sarcoid, xơ hóa phổi vô căn, ung thư phổi.
- Trường hợp bệnh nhân đã được điều trị tràn khí màng phổi nhưng chưa thực hiện đầy đủ các bước dự phòng tái phát tràn khí màng phổi.
- Tình trạng bệnh nhân bị tràn khí màng phổi nguyên phát nhưng không được điều trị mà bệnh tự khỏi hầu hết sẽ bị tái phát bệnh bởi màng phổi đã bị tổn thương nhưng không được xử lý.
- Một số trường hợp tái phát tràn khí màng phổi có thể đến từ yếu tố công việc như làm việc trong môi trường bị ô nhiễm gây ảnh hưởng đường hô hấp, làm việc trong môi trường có áp suất cao như hàng không hoặc thợ lặn,...

Tràn khí màng phổi do chấn thương

Chấn thương ở ngực hay vết thương do bệnh lý, vi trùng gây ra đều có thể dẫn đến tràn khí màng phổi là xẹp phổi. Chấn thương thường gặp là vết thương dao, vết thương do đạn bắn hoặc đòn đánh lực mạnh vào ngực. Bất kỳ vết thương cùn hoặc xuyên thấu nào đối với ngực của bạn đều có thể gây xẹp phổi.

Một số thương tích có thể xảy ra khi bị hành hung hoặc đụng xe, trong khi những thương tích khác có thể vô tình xảy ra trong các thủ tục y tế liên quan liên quan như: sinh thiết phổi, sinh thiết gan, hồi sức tim phổi, đặt ống lồng ngực,...

1.2 Lý do chọn đề tài

Tràn khí màng phổi (Pneumothorax) là một tình trạng y tế khẩn cấp có thể gây nguy hiểm đến tính mạng. Khi khí bất thường tích tụ trong không gian giữa màng phổi và màng ngoài, áp lực trong khí quản và phổi bị suy giảm, ảnh hưởng đến sự thông khí và chức năng hô hấp của bệnh nhân. Tràn khí màng phổi có thể xảy ra do nhiều nguyên nhân, bao gồm chấn thương, bệnh lý phổi, hoặc các thủ thuật y tế. Điều quan trọng là phát hiện và chẩn đoán tràn khí màng phổi một cách chính xác và nhanh chóng để đảm bảo các biện pháp điều trị kịp thời và hiệu quả.

Tuy nhiên, việc chẩn đoán tràn khí màng phổi từ ảnh chụp X-ray đòi hỏi sự kỹ thuật và kinh nghiệm của các chuyên gia y tế. Sử dụng máy học để phát hiện tràn khí màng phổi từ ảnh chụp X-ray có thể cải thiện khả năng chẩn đoán và giúp tiết kiệm

thời gian và công sức của các chuyên gia y tế. Máy học có khả năng tự động hóa quá trình phân tích ảnh chụp X-ray và xác định các đặc trưng hình ảnh quan trọng, như khu vực tràn khí, kích thước, hình dạng, và vị trí. Bằng cách áp dụng các thuật toán máy học và kỹ thuật học sâu, mô hình có thể học từ các bộ dữ liệu X-ray đã được gán nhãn để nhận biết tràn khí màng phổi với độ chính xác cao.

Điều này không chỉ giúp cung cấp kết quả chẩn đoán nhanh chóng mà còn giảm bớt sự phụ thuộc vào nhân lực y tế chuyên gia. Việc phát hiện tràn khí màng phổi bằng máy học có thể giúp nâng cao khả năng tiếp cận chẩn đoán và điều trị cho người dân ở các khu vực xa xôi, nơi không có sẵn các chuyên gia y tế có kỹ năng chẩn đoán tràn khí màng phổi. Ngoài ra, máy học còn có thể tạo ra các hệ thống thông minh hỗ trợ quyết định cho các bác sĩ, giúp họ đưa ra quyết định chẩn đoán chính xác và nhanh chóng.

Việc áp dụng máy học để phát hiện tràn khí màng phổi từ ảnh chụp X-ray là một xu hướng tiềm năng đang phát triển rộng rãi và mang lại những tiềm năng đáng kể trong lĩnh vực y tế. Nghiên cứu về **"Phát hiện bệnh phổi dựa vào ảnh X-Ray lồng ngực bằng phương pháp máy học"** không chỉ đóng vai trò quan trọng trong việc nâng cao chất lượng chẩn đoán, mà còn có khả năng giảm thiểu thời gian chờ đợi và tối ưu hóa quy trình chăm sóc sức khỏe cho bệnh nhân.

1.3 Cơ sở lý luận

Deep Learning là một phương pháp học máy mạnh mẽ dựa trên khái niệm của não bộ, giúp hiểu rõ hơn về các dạng dữ liệu thông qua việc xây dựng nhiều tầng biểu diễn, bao gồm cả khái niệm cụ thể và trừu tượng. Hiện nay, Deep Learning đã được ứng dụng rộng rãi trong nhiều lĩnh vực khác nhau, mang lại nhiều đột phá vượt trội.

Thuật toán Deep Learning được hình thành từ các Mạng Nơ-ron Nhân tạo với các lớp ẩn, dựa trên cấu trúc của mạng thần kinh sinh học. Deep Neural Networks có khả năng mô phỏng các mối quan hệ phức tạp phi tuyến trong cả bài toán giám sát (dự đoán kết quả) và bài toán không giám sát (tìm kiếm mẫu và dữ liệu ngoại lệ). Deep Learning có độ chính xác cao, khả năng học sâu và rộng, và đạt được kết quả vượt trội. Mặc dù là một lĩnh vực mới, các kiến trúc Deep Learning có khả năng tự động mô hình hoá mối quan hệ phức tạp phi tuyến trong dữ liệu. Do đó, các kỹ thuật Deep Learning có thể làm việc trực tiếp với dữ liệu thô mà không cần thông qua các bước trích xuất đặc trưng phức tạp, chỉ phù hợp với từng vấn đề cụ thể. Có nhiều kiến

trúc Deep Learning như Convolutional Neural Network (mạng nơ-ron tích chập), Deep Belief Network, và Vision Transformer. Các kiến trúc này đã mang lại nhiều thành công đột phá trong nhiều lĩnh vực như thị giác máy tính và xử lý ngôn ngữ tự nhiên.

Trong phạm vi đề tài này, chúng ta sẽ tìm hiểu về kiến trúc và thực hiện huấn luyện mô hình mạng trên tập dữ liệu ảnh y tế, được lấy từ Kaggle và các nguồn tài liệu liên quan. Mục tiêu là phân loại và phân đoạn bệnh tràn khí màng phổi trong môi trường Google Colab, sử dụng các mô hình mạng như Densenet121, Xception, Vision Tranformer, Resnet50 cho bài toán phân loại, và mạng phân vùng Unet, DeeplabV3+ cho bài toán phân đoạn.

1.4 Cơ sở thực tiễn

Chụp X-Ray và chẩn đoán bệnh dựa vào ảnh chụp đã và đang được áp dụng rộng rãi trong việc phát hiện một số bệnh lý, nhưng rất khó để phân biệt bệnh tràn khí màng phổi với một số căn nguyên khác. Việc lý giải các hình ảnh thường gặp nhiều sai lầm do kinh nghiệm, cảm giác hoặc do nhận thức cá nhân. Vì thế khi phân tích, các bác sĩ phải luôn kết hợp với đặc điểm dịch tễ và biểu hiện lâm sàng để đưa ra chẩn đoán chính xác. Ngoài ra, Deep Learning ứng dụng trong lĩnh vực y tế đã ghi nhận được nhiều nghiên cứu mang đến những kết quả khả quan. Đó là tiền đề để thực hiện nghiên cứu phát hiện bệnh tràn khí màng phổi dựa trên dấu hiệu tràn khí trên phổi.

1.5 Những nghiên cứu liên quan

Trong bài báo “Deep Learning for Diagnosis and Segmentation of Pneumothorax: The Results on the Kaggle Competition and Validation Against Radiologists” [12] năm 2020 của các tác giả Alexey Tolkachev, Ilyas Sirazitdinov, Maksym Kholiavchenko, Tamerlan Mustafaev, Bulat Ibragimov. Trong nghiên cứu này, nhóm tác giả đã thực hiện việc kết hợp mạng thần kinh tích chập Unet với nhiều mạng trích xuất đặc trưng như ResNet34, SE-ResNext50, và DenseNet121 và đạt được hiệu suất tính rất cao với Dice coefficient lần lượt đạt 0.8574, 0.8432, 0.8632 ứng với từng mạng trích xuất.

Nghiên cứu tiếp theo được đăng trong bài báo “Deep Learning for Pneumothorax Detection and Localization in Chest Radiographs” [13] năm 2019 của các tác giả André Gooßen, Hrishikesh Deshpande, Tim Harder, Evan Schwab, Ivo

Baltruschat, Thusitha Mabotuwana, Nathan Cross, Axel Saalbach. Trong bài báo này các tác giả đã thực hiện nghiên cứu trên các kỹ thuật học sâu khác nhau như: CNN, MIL, FCNs đã đạt được các hiệu suất AUC lần lượt là 0.96, 0.93, 0.92 cho ba kỹ thuật trên nhằm so sánh và tìm ra kỹ thuật tối ưu nhất cho việc phát hiện tràn khí màng phổi trên ảnh X-Ray.

Trong bài báo “Automatic Segmentation of Pneumothorax in Chest Radiographs Based on a Two-Stage Deep Learning Method” [14] năm 2020 của nhóm các tác giả Xiao Han, Jing Zhang, Minghui Wang, Jianhui He, Yuqi Fang, Jun Lan, Sen Yang, Xiyue Wang đã phát triển một mô hình phân vùng với sự kết hợp của kiến trúc Unet và DeepLabV3 đã đạt được hiệu suất mô hình rất cao với AUC 0.9795 và Dice score 0.8883.

Nghiên cứu tiếp theo được đăng trong bài báo “CheXLocNet: Automatic localization of pneumothorax in chest radiographs using deep convolutional neural networks” [15] năm 2020 của nhóm tác giả Hongyu Wang, Hong Gu, Pan Qin, Jia Wang. Trong bài báo này, các tác giả đã phát triển một mạng lưới thần kinh tích chập được gọi là CheXLocNet để phân chia các tổn thương tràn khí màng phổi đã đạt được hiệu suất mô hình rất tốt trong cả phân loại và phân đoạn. Hơn nữa, kết quả phân đoạn có thể cung cấp thông tin hình học toàn diện về các tổn thương, có thể mang lại lợi ích cho việc theo dõi sự phát triển tuần tự của các tổn thương với độ chính xác cao với IoU đạt 0.81 và Dice score là 0.82.

Nghiên cứu tiếp theo được đăng trong bài báo “Deep learning for pneumothorax detection and localization using networks fine-tuned with multiple institutional datasets” [16] năm 2020 của nhóm tác giả Crosby, Jennie, Rhines, Thomas, Li, Feng, MacMahon, Heber, Giger, Maryellen. Trong nghiên cứu này, ảnh chụp phổi được chia thành hai ảnh đỉnh (phần trên của phổi) và sau đó các mảnh 256 x 256 được trích xuất từ ảnh đỉnh. Mạng neural VGG19 được điều chỉnh tinh chỉnh cho nhiệm vụ phân biệt giữa các hình ảnh có và không có pneumothorax. Một mạng được điều chỉnh tinh chỉnh với các ảnh đỉnh (giảm kích thước xuống 256 x 256) và một mạng khác được điều chỉnh tinh chỉnh với các mảnh 256 x 256 trong các ảnh đỉnh. Những mạng điều chỉnh tinh chỉnh này đã được kiểm tra trên một tập dữ liệu kiểm tra độc lập và phân tích ROC được thực hiện. Mạng dựa trên ảnh đỉnh đã cho một AUC là 0,80 (khoảng tin cậy 95%: 0,79, 0,81) và mạng dựa trên mảnh đã cho một AUC là 0,73 (khoảng tin

cây 95%: 0,71, 0,74) trong nhiệm vụ phân biệt giữa các hình ảnh có và không có pneumothorax. Khi các đầu ra từ hai mạng được kết hợp thông qua bỏ phiếu mềm, ta quan sát được một sự tăng đáng kể về hiệu suất so với mỗi mạng đơn lẻ ($AUC=0,83$, $p<0,001$).

Nghiên cứu tiếp theo được đăng trong bài báo “Detection of the location of pneumothorax in chest X-rays using small artificial neural networks and a simple training process” [17] năm 2021 của nhóm tác giả Yongil Cho, Jong Soo Kim, Tae Ho Lim, Inhye Lee, Jongbong Choi. Mục đích của nghiên cứu này là đánh giá hiệu suất chẩn đoán đạt được bằng cách sử dụng mạng neural nhân tạo kết nối đầy đủ nhỏ (ANNs) và quá trình huấn luyện đơn giản, thuật toán Kim-Monte Carlo, để phát hiện vị trí của pneumothorax trên ảnh chụp X-Ray phổi. Tổng cộng 1.000 hình ảnh X-Ray phổi với pneumothorax được lấy ngẫu nhiên từ cơ sở dữ liệu hình ảnh công cộng của NIH (Viện Quốc gia Sức khỏe) và được sử dụng làm tập huấn luyện và tập kiểm tra. Mỗi hình ảnh X-Ray với pneumothorax được chia thành 49 hộp để xác định vị trí pneumothorax. Đối với mỗi hộp trong các hình ảnh X-Ray phổi chứa trong tập kiểm tra, diện tích dưới đường cong đặc trưng nhận biết (ROC) (AUC) là 0,882, và độ nhạy và độ đặc hiệu lần lượt là 80,6% và 83,0%. Ngoài ra, một phương pháp học sâu thông thường hiện đang được sử dụng cho nhận dạng hình ảnh, mạng neural tích chập (CNN), cũng được áp dụng vào cùng tập dữ liệu này để so sánh. Hiệu suất của ANN kết nối đầy đủ nhỏ tốt hơn so với CNN. Đối với hiệu suất chẩn đoán của CNN với các hàm kích hoạt khác nhau, CNN với hàm kích hoạt sigmoid cho các nút ẩn kết nối đầy đủ tốt hơn so với CNN với hàm kích hoạt đơn vị tuyến tính (RELU). Nghiên cứu này cho thấy phương pháp có thể phát hiện vị trí của pneumothorax trên ảnh chụp X-Ray phổi một cách chính xác, giảm đáng kể thời gian trễ khi chẩn đoán các bệnh khẩn cấp như pneumothorax và tăng tính hiệu quả của thực hành lâm sàng và chăm sóc bệnh nhân.

Nghiên cứu tiếp theo được đăng trong bài báo “Deep multi-instance transfer learning for pneumothorax classification in chest X-ray images” [18] năm 2021 của nhóm tác giả Yuchi Tian , Jiawei Wang , Wenjie Yang , Jun Wang , Dahong Qian. Trong bài báo này, các tác giả đã phát triển một mạng lưới thần kinh tích chập sử dụng mạng nơ-ron sâu dựa trên Residual Neural Network (ResNet) để nhận diện tràn khí màng phổi: chiến lược học đặc trưng cục bộ (LFL) kết hợp với học đa thể hiện

toàn cục (GMIL). Hầu hết các vùng không có bất thường trên hình ảnh được loại bỏ để học các đặc trưng phân biệt. Hai tập dữ liệu được sử dụng để đánh giá quy mô lớn: một tập dữ liệu riêng (27.955 hình ảnh X-quang ngực mặt trước) và một tập dữ liệu công cộng (National Institutes of Health [NIH] ChestX-ray14; 112.120 hình ảnh X-quang ngực mặt trước). Hiệu suất của mô hình trong việc nhận diện được đánh giá bằng các độ đo như độ chính xác, độ chính xác dương tính, độ nhạy, độ đặc hiệu, F1-score, đường cong hoạt động nhận dạng (ROC), và diện tích dưới đường cong ROC (AUC). Kết quả thực nghiệm chứng minh rằng chiến lược học được đề xuất có thể đạt hiệu suất tiên tiến trên tập dữ liệu NIH với độ chính xác, ROC (AUC), Precision, Recall, độ đặc hiệu và F1-score lần lượt là 94,4%, 97,3%, 94,2%, 94,6% , 94,2% và 94,4%.

1.6 Vấn đề đặt ra

Nghiên cứu và phát triển đề tài **“Phát hiện bệnh phổi dựa vào ảnh X-Ray lồng ngực bằng phương pháp máy học”** nhằm mục đích:

- Phát hiện tràn khí màng phổi dựa vào ảnh X-Ray phổi, khoang vùng phổi bị tràn khí.
- Tinh chỉnh mô hình đạt độ chính xác cao.

1.7 Hướng giải quyết của đề tài

Để thực hiện các mục đích được đề ra ở phần vấn đề đặt ra cần tiến hành:

- Tìm hiểu về bệnh tràn khí màng phổi.
- Tìm hiểu về những nghiên cứu liên quan về tràn khí màng phổi.
- Thu thập tập dữ liệu về X-Ray phổi.
- Xây dựng mô hình Deep Learning để phát hiện tràn khí màng phổi

CHƯƠNG 2 - CƠ SỞ LÝ THUYẾT

2.1 Bệnh tràn khí màng phổi

Tràn khí màng phổi là sự tích tụ không khí bất thường trong khoang màng phổi giữa phổi và thành ngực. Các triệu chứng thường bao gồm đột ngột xuất hiện cơn đau nhói một bên ngực và khó thở. Trong một số ít trường hợp, van một chiều được hình thành bởi một vùng mô bị tổn thương và lượng không khí trong khoảng trống giữa thành ngực và phổi tăng lên, đây được gọi là tràn khí màng phổi tự phát. Điều này có thể gây ra tình trạng thiếu oxy ngày càng trầm trọng và huyết áp thấp. Dẫn đến một loại cú sốc gọi là sốc tắc nghẽn, có thể gây tử vong.[1]

Theo một thống kê y tế tại Anh vào năm 2020 [2], trung bình cứ 10.000 người trưởng thành lại có khoảng 15 người mắc chứng tràn khí màng phổi tự phát. Trong đó, nam giới dễ mắc bệnh hơn nữ giới, lứa tuổi mắc bệnh chủ yếu là dưới 40 tuổi. Những người hút thuốc lá có nguy cơ cao mắc bệnh hơn, nguyên nhân là hóa chất trong khói thuốc khiến cho thành màng phổi mỏng hơn, dễ rách hơn.

Ngoài ra còn 1 dạng tràn khí màng phổi xảy ra ở phụ nữ độ tuổi sinh sản là tràn khí màng phổi theo chu kỳ kinh nguyệt. Đặc điểm bệnh khá giống với trường hợp tự phát thông thường, song khởi phát bệnh liên quan đến các chu kỳ kinh nguyệt. Nguyên nhân là do trong chu kỳ, một lượng nhỏ mô tử cung di chuyển lên làm tổ ở màng phổi, gây rách, tổn thương và tràn dịch.

Những người từng bị tràn dịch màng phổi dù được điều trị tích cực thì nguy cơ tái phát khá cao, khoảng 30%. Hầu hết trường hợp tái phát xảy ra ở cùng bên phổi, trong vòng 3 năm đầu kể từ lần phát bệnh trước.

2.1.1 Nguyên nhân

Tràn khí màng phổi tự phát xảy ra ở những bệnh nhân không có bệnh phổi, ở nam giới trẻ tuổi, cao, gầy ở độ tuổi thiếu niên và tuổi 20. Đây được cho là do sự vỡ tự phát của các kén khí hoặc bóng khí dưới màng phổi vùng đỉnh do thuốc lá hoặc di truyền. Bệnh thường xảy ra lúc nghỉ ngơi, mặc dù một số trường hợp xảy ra trong các hoạt động liên quan đến gắng sức hoặc kéo dài. tràn khí màng phổi tự phát cũng xảy ra khi lặn và bay ở độ cao lớn.

Tràn khí màng phổi tự phát thứ phát xảy ra ở bệnh nhân bị bệnh phổi. Bệnh thường là kết quả của vỡ của một bóng hoặc kén hoặc ở bệnh nhân bị bệnh phổi tắc nghẽn mạn tính nặng (thể tích thở ra gắng sức trong một giây [FEV1 < 1 L], Nhiễm

Pneumocystis jirovecii liên quan đến HIV, xơ nang, hoặc bất kỳ bệnh nhu phổi nền nào (Xem bảng Nguyên nhân tràn khí màng phổi tự phát thứ phát). Tràn khí màng phổi tự phát thứ phát nghiêm trọng hơn tràn khí màng phổi tự phát nguyên phát vì bệnh xảy ra ở những bệnh nhân có bệnh phổi làm giảm thể tích dự trữ phổi của họ.

Tràn khí màng phổi liên quan đến kinh nguyệt là một dạng tràn khí màng phổi tự phát hiếm xảy ra trong vòng 48 giờ khi bắt đầu có kinh nguyệt ở phụ nữ tiền mãn kinh và đôi khi ở phụ nữ sau mãn kinh đang dùng estrogen. Nguyên nhân là nội mạc tử cung trong lồng ngực, có thể là do di căn của nội mạc tử cung qua cơ hoành hoặc thuyên tắc qua các tĩnh mạch vùng chậu.

Tràn khí màng phổi sau can thiệp là do các biện pháp can thiệp y khoa, bao gồm sinh thiết xuyên thành, chọc dịch màng phổi, đặt catheter tĩnh mạch trung tâm, thông khí nhân tạo và hồi sức tim phổi.

Tràn khí màng phổi do chấn thương có thể do chấn thương cùn hoặc chấn thương xuyên thấu vào thành ngực. Cơ chế phổ biến nhất là sự xâm nhập của các điểm xương sắc nhọn tại chỗ gãy xương sườn mới, gây tổn thương mô phổi. Tràn khí màng phổi do chấn thương cũng có thể được quan sát thấy ở những người tiếp xúc với vụ nổ, ngay cả khi không có vết thương rõ ràng ở ngực.[3]

2.1.2 Dấu hiệu và triệu chứng

Triệu chứng cơ năng của tràn khí màng phổi thường rõ ràng bao gồm[2]:

- Đau ngực đột ngột là triệu chứng thường gặp nhất ở bệnh nhân khi xảy ra thông khí màng phổi. Đôi khi đau tăng khi hít thở. Một số bệnh nhân, không phải tất cả, bị khó thở.
- Một số bệnh nhân còn bị ho, một số bị đau vai, hay đau nhói giữa hai bả vai. Ở phần lớn những bệnh nhân, các triệu chứng này không kéo dài lâu và rất hiếm khi nặng hơn.
- Bệnh nhân có thể bị choáng (tái xanh người, vã mồ hôi, mạch đập nhanh, nông, huyết áp tụt, tay chân lạnh, vã mồ hôi, tinh thần hốt hoảng, lo âu).
- Ngoài ra, bệnh nhân có cảm giác khó thở, người bệnh thường ở trong tình trạng sốc. Nếu tràn khí màng phổi không cấp cứu kịp thời có thể dẫn đến nguy kịch do xẹp phổi, suy hô hấp.

Triệu chứng thực thể của tràn khí màng phổi như như:

- Lồng ngực không cân đối, bên có tràn khí màng phổi thường sẽ bị phồng to hơn và kém di động hơn bên còn lại.
- Khi sờ phổi thì thấy rung thanh có thể bị mất đi.
- Vùng khí có xuất hiện tràn khí thì gõ nghe vang, có thể có những vùng thấp xuất hiện tràn dịch, máu thì có thể gõ nghe đục.
- Rì rào phế nang có thể giảm hay thậm chí mất hẳn.[4]

Triệu chứng bệnh có thể nhiều hoặc ít, kéo dài hoặc âm thầm trong thời gian ngắn phụ thuộc vào lượng không khí tồn tại ở khoang màng phổi nhiều hay ít. Tuy nhiên, một số ít bệnh nhân dù bệnh tiến triển nặng, màng khí khiến phổi bị xẹp lại nhưng triệu chứng chỉ là cơn đau ngực và khó thở nhẹ. [2]

Người bệnh nên đi khám bác sĩ ngay khi xuất hiện triệu chứng đau ngực, khó thở đột ngột. Đôi khi nguyên nhân gây ra tình trạng này có thể không phải là tràn khí màng phổi song chẩn đoán xác định nguyên nhân và điều trị sớm vẫn là cần thiết.

Nếu cơn đau ngực trở nên nặng hơn, khiến bạn không thể thở nổi hoặc thở khó khăn thì cần cấp cứu y tế ngay lập tức.

2.1.3 Biến chứng và mức độ nguy hiểm của bệnh tràn khí màng phổi

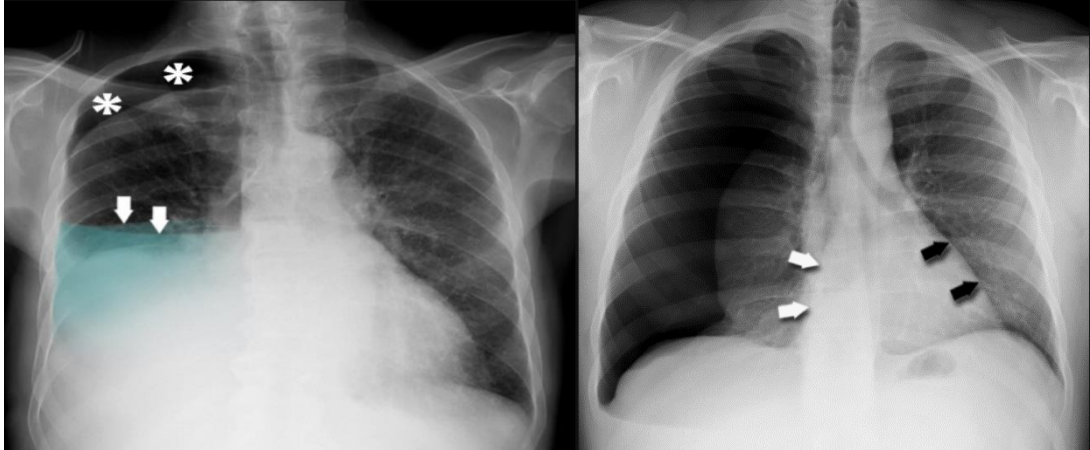
Hội chứng tràn khí màng phổi cần được cấp cứu y tế và dẫn lưu khí, nhất là các trường hợp khí tích tụ lớn hơn 20% thể tích phổi. Biến chứng nguy hiểm nhất là xẹp phổi dẫn đến suy hô hấp, suy tuần hoàn khiến bệnh nhân tử vong nhanh chóng. Ngoài ra, bệnh nhân có thể phải đối mặt với các biến chứng tràn khí màng phổi như:

- Rò rỉ không khí: khiến tràn khí màng phổi tái phát liên tục, cần phẫu thuật đóng lỗ rò rỉ.
- Nồng độ oxy thấp: ảnh hưởng tới hoạt động của nhiều cơ quan trong cơ thể, thậm chí đe dọa đến tính mạng.
- Chèn ép tim: Khí tích tụ có thể làm tăng áp lực đẩy tim và mạch máu, cản trở lưu thông máu. Biến chứng chèn ép tim có thể gây tử vong nhanh chóng.
- Suy hô hấp: Khi oxy trong máu quá thấp, suy hô hấp có thể khiến bệnh nhân hôn mê, mất ý thức, rối loạn nhịp tim dẫn đến tử vong.
- Shock: Shock là tình trạng nguy kịch khi tràn dịch màng phổi xảy ra làm thiếu hụt oxy trầm trọng, huyết áp giảm thấp.

Hội chứng tràn khí màng phổi là biến chứng hô hấp nghiêm trọng, cần sớm điều trị và cấp cứu để phòng ngừa bệnh tiến triển nặng và nguy hiểm.[1]

2.2 Đặc điểm nhận dạng tràn khí màng phổi trên ảnh X-Ray

Thông thường, chụp X-Ray tràn khí màng phổi là xét nghiệm đầu tiên được chỉ định để đánh giá tràn khí màng phổi, bởi vì nó đơn giản, chi phí thấp, nhanh chóng và không xâm lấn.[3]



Hình 2.1. Ảnh chụp X-Ray tràn khí phổi [3]

Ảnh phim tràn khí màng phổi X-Ray là hình ảnh tăng sáng không có vân phổi, phổi ép lại và nhìn thấy được đường viền lá tạng rõ nét. Đôi khi hình ảnh bị thu nhỏ lại thành một khối giống như khối u ở rốn phổi, gian sườn bị giãn, đẩy tim và trung thất, vòm hoành hạ thấp. Phim chụp X-Ray của bệnh nhân tràn khí màng phổi sẽ thường xuất hiện các đặc điểm sau [3]:

- Phổi bị xẹp: Đây là sự tích tụ khí trong khoảng trống giữa 2 lớp: lớp lót của thành ngực bên trong và lót bên ngoài của phổi. Tràn khí màng phổi sẽ tác động và làm mô phổi kéo vào bên trong, từ đó sẽ đẩy không khí ra khỏi phế nang.
- Hình ảnh tăng sáng và vùng ngoại vi không xuất hiện vân phổi
- Đường viền của màng phổi tạng rõ nét: trên phim X-Ray có thể nhìn thấy mép phổi (màng phổi tạng) và có màu đen xung quanh mép này thì nên nghi ngờ có tràn khí màng phổi. Điều này sẽ dẫn đến việc đánh giá ngay khí quản và trung thất của bệnh nhân.
- Trung thất có hiện tượng bị đè đẩy: Phát hiện này có tầm quan trọng to lớn vì đây chính là dấu hiệu lâm sàng của một bệnh lý nghiêm trọng hơn ở phổi, cần được chẩn đoán và xử trí khẩn cấp.

2.3 Tổng quan về trí tuệ nhân tạo

2.3.1 Khái niệm trí tuệ nhân tạo

Trí tuệ nhân tạo tiếng anh là Artificial Intelligence - viết tắt là AI. Đây là một ngành thuộc lĩnh vực khoa học máy tính. Được hình thành do con người lập trình trí tuệ nhân tạo với mục tiêu giúp máy tính có thể tự động hóa các hành vi thông minh như con người. Những đặc điểm của trí tuệ nhân tạo AI:

- Ứng dụng các hệ thống học máy để giúp mô phỏng trí tuệ của con người trong những xử lý mà con người sẽ làm tốt hơn cả máy tính.
- Trí tuệ nhân tạo giúp cho máy tính có được những trí tuệ của con người: khả năng biết suy nghĩ, biết lập luận để có thể giải quyết được vấn đề, biết giao tiếp do hiểu ngôn ngữ, tiếng nói, biết học và tự thích nghi...
- Trí thông minh nhân tạo là một trong những ngành trọng yếu của tin học, liên quan đến cách cư xử, sự học hỏi và khả năng thích ứng thông minh của máy móc mà chúng ta không ngờ.

2.3.2 Một số công nghệ Trí tuệ nhân tạo phổ biến hiện nay

a. Sản sinh ngôn ngữ tự nhiên (*Natural language generation*)

Máy móc xử lý và giao tiếp theo một cách khác với bộ não con người. Tạo ngôn ngữ tự nhiên là một công nghệ phổ biến giúp chuyển đổi dữ liệu có cấu trúc thành ngôn ngữ bản địa. Máy móc được lập trình với các thuật toán để chuyển đổi dữ liệu sang định dạng mong muốn của người dùng.

Ngôn ngữ tự nhiên là một nhánh của trí tuệ nhân tạo giúp các nhà phát triển nội dung tự động hóa nội dung và phân phối ở định dạng mong muốn. Các nhà phát triển nội dung có thể sử dụng nội dung được tự động hóa để quảng cáo trên các nền tảng mạng xã hội khác nhau và các nền tảng truyền thông khác để tiếp cận đối tượng mục tiêu.

Sự can thiệp của con người sẽ giảm đáng kể vì dữ liệu sẽ được chuyển đổi thành các định dạng mong muốn. Dữ liệu có thể được hiển thị dưới dạng biểu đồ, đồ thị...

b. Nhận dạng giọng nói (*Speech recognition*)

Nhận dạng giọng nói là một nhánh quan trọng khác của trí tuệ nhân tạo, chuyển đổi giọng nói của con người thành một định dạng hữu ích và có thể hiểu được bằng các ứng dụng máy tính. Công nghệ này là cầu nối tương tác giữa máy tính và con người. Ứng dụng Siri của iPhone là một ví dụ điển hình về nhận dạng giọng nói.

c. Trợ lý ảo (Virtual agent)

Trợ lý ảo (virtual agent) đã trở thành công cụ rất hữu ích đối với các nhà thiết kế hướng dẫn. Trợ lý ảo là một ứng dụng máy tính tương tác với con người. Các ứng dụng web và di động cung cấp tính năng chatbot hoạt động như các trợ lý dịch vụ khách hàng để tương tác với con người và trả lời các truy vấn của họ.

Trợ lý ảo Google (Google Assistant) giúp tổ chức các cuộc họp, trong khi Alexia trợ lý ảo của Amazon giúp việc mua sắm của bạn trở nên dễ dàng hơn. Trợ lý ảo cũng hoạt động giống như một trợ lý ngôn ngữ, tùy theo sự lựa chọn và sở thích của bạn. Trợ lý ảo Watson của IBM có thể hiểu được các truy vấn dịch vụ khách hàng điển hình được hỏi theo nhiều cách khác nhau.

Bên cạnh đó, các trợ lý ảo cũng hoạt động như một phần mềm dưới dạng dịch vụ (SaaS) - mô hình phân phối dịch vụ ứng dụng phần mềm, trong đó nhà cung cấp không bán sản phẩm phần mềm mà bán dịch vụ dựa trên phần mềm đó.

d. Quản lý quyết định (Decision management)

Các công ty hiện nay đang triển khai các hệ thống quản lý quyết định để chuyển đổi và phân tích dữ liệu thành các mô hình dự đoán. Hệ thống quản lý quyết định được sử dụng để nhận thông tin cập nhật, sau đó tiến hành phân tích dữ liệu kinh doanh nhằm hỗ trợ quá trình ra quyết định của doanh nghiệp.

Quản lý quyết định giúp đưa ra quyết định nhanh chóng, tránh được những rủi ro, và tự động hóa quy trình ra quyết định. Hệ thống quản lý quyết định được triển khai rộng rãi trong lĩnh vực tài chính, y tế, thương mại, bảo hiểm, thương mại điện tử...

e. Sinh trắc học (Biometrics)

Sinh trắc học là phép đo và phân tích thống kê các đặc điểm sinh học và hành vi độc đáo của con người. Công nghệ này chủ yếu được sử dụng để nhận dạng và kiểm soát truy cập, hoặc để xác định các cá nhân thuộc diện bị giám sát.

Xác thực sinh trắc học (biometric authentication) ngày càng trở nên phổ biến trong các hệ thống an ninh công cộng và doanh nghiệp, điện tử tiêu dùng và các ứng dụng điểm bán lẻ POS. Ngoài vấn đề bảo mật, động lực thúc đẩy xác thực sinh trắc học là sự tiện lợi vì công nghệ này không yêu cầu phải nhớ mật khẩu hay mang theo mã thông báo bảo mật (security tokens).

Xác thực sinh trắc học thực hiện thông qua các dấu hiệu nhận diện như: dấu vân tay, dải tai, võng mạc, hình dạng bàn tay, khuôn mặt hoặc chữ ký bằng văn bản. Trong đó, dấu vân tay là loại xác thực sinh trắc học được sử dụng phổ biến và lâu đời nhất.

f. Học máy (Machine learning)

Học máy là một lĩnh vực của trí tuệ nhân tạo cho phép máy móc hiểu được các tập dữ liệu mà không cần được lập trình. Kỹ thuật học máy giúp các doanh nghiệp đưa ra những quyết định sáng suốt với các phân tích dữ liệu được thực hiện bằng thuật toán và mô hình thống kê.

Học máy được ứng dụng trong nhiều lĩnh vực khác nhau. Lĩnh vực chăm sóc sức khỏe và chuyên môn y tế cần các kỹ thuật học máy để phân tích dữ liệu bệnh nhân nhằm dự đoán bệnh và điều trị hiệu quả. Ngành tài chính ngân hàng cần học máy để phân tích dữ liệu khách hàng nhằm xác định và đề xuất các lựa chọn đầu tư cho khách hàng cũng như phòng ngừa rủi ro và lừa đảo. Các nhà bán lẻ sử dụng học máy để phân tích dữ liệu khách hàng, qua đó dự đoán được những thay đổi trong sở thích và hành vi của khách hàng.

g. Tự động hóa quy trình bằng robot (Robotic process automation)

Tự động hóa quy trình bằng robot (RPA) là một ứng dụng của trí tuệ nhân tạo giúp định cấu hình robot (ứng dụng phần mềm) để diễn giải, truyền đạt và phân tích dữ liệu. Công cụ AI này giúp tự động hóa một phần hoặc toàn bộ các hoạt động thủ công lặp đi lặp lại và theo quy luật.

h. Mạng ngang hàng (Peer-to-peer network)

Mạng ngang hàng giúp kết nối các hệ thống và máy tính khác nhau để chia sẻ dữ liệu mà không cần truyền dữ liệu qua máy chủ. Mạng ngang hàng có khả năng giải quyết các vấn đề phức tạp nhất. Công nghệ này được sử dụng trong tiền điện tử. Việc triển khai mạng ngang hàng giúp tiết kiệm chi phí vì các máy trạm (máy tính có hiệu năng cao hơn và cấu hình vượt trội hơn máy tính thông thường) riêng lẻ được kết nối mà không cần cài đặt các máy chủ.

i. Các nền tảng học sâu (Deep learning platforms)

Học sâu là một nhánh khác của trí tuệ nhân tạo hoạt động dựa trên các mạng nơ-ron nhân tạo (neural network), lấy cảm hứng từ bộ não con người. Học sâu giúp giải quyết những vấn đề phức tạp như cách mà bộ não con người vẫn làm, thông qua việc sử dụng nhiều thuật toán khác nhau.

Thuật ngữ “sâu” được đặt ra bởi vì tồn tại nhiều lớp ẩn trong các mạng nơ-ron. Thông thường, một mạng nơ-ron có 2-3 lớp ẩn và có thể có tối đa 150 lớp ẩn. Học sâu được dùng để đào tạo một mô hình và một đơn vị xử lý đồ họa dựa trên lượng dữ liệu khổng lồ. Các thuật toán hoạt động theo hệ thống phân cấp để tự động hóa các phân tích dự đoán.

Học sâu được ứng dụng rộng rãi trong nhiều lĩnh vực như hàng không vũ trụ và quân sự để phát hiện các vật thể thông qua vệ tinh, cải thiện an toàn lao động cho công nhân bằng cách xác định các sự cố rủi ro khi công nhân đến gần thiết bị máy móc, giúp phát hiện tế bào ung thư...

j. Phần cứng tối ưu hóa cho AI (AI-optimised hardware)

Giới kinh doanh hiện có nhu cầu cao về phần mềm trí tuệ nhân tạo. Khi sự chú ý dành cho phần mềm tăng lên, nhu cầu về phần cứng để hỗ trợ phần mềm cũng nảy sinh. Một con chip thông thường không thể hỗ trợ các mô hình trí tuệ nhân tạo, do đó một thế hệ chip trí tuệ nhân tạo mới đang được phát triển nhằm phục vụ các mạng nơ-ron, học sâu và thị giác máy tính (computer vision).

Phần cứng cho trí tuệ nhân tạo bao gồm CPU để xử lý khối lượng công việc ngày càng tăng, silicon tích hợp cho mục đích đặc biệt dành cho mạng nơ-ron, hay chip mô phỏng hệ thần kinh người... Các công ty như Nvidia, Qualcomm và AMD đang tạo ra những con chip có thể thực hiện các phép tính AI phức tạp. Chăm sóc sức khỏe và công nghiệp ô-tô có thể là những ngành sẽ được hưởng lợi từ loại chip này.

2.4 Machine Learning

2.4.1 Khái niệm Machine Learning

Machine Learning (ML) hay máy học là một nhánh của trí tuệ nhân tạo (AI), nó là một lĩnh vực nghiên cứu cho phép máy tính có khả năng cải thiện chính bản thân chúng dựa trên dữ liệu mẫu (training data) hoặc dựa vào kinh nghiệm (những gì đã được học). Machine learning có thể tự dự đoán hoặc đưa ra quyết định mà không cần được lập trình cụ thể.

Bài toán machine learning thường được chia làm hai loại là dự đoán (prediction) và phân loại (classification). Các bài toán dự đoán như dự đoán giá nhà, giá xe... Các bài toán phân loại như nhận diện chữ viết tay, nhận diện đồ vật...

2.4.2 Machine Learning Workflow

MLW sẽ cho thấy quy trình để làm việc với machine learning như thế nào. Cụ thể từng bước trong machine learning workflow như sau như sau:

- Data collection - thu thập dữ liệu: để máy tính có thể học được bạn cần có một bộ dữ liệu (dataset), bạn có thể tự thu thập chúng hoặc lấy các bộ dữ liệu đã được công bố trước đó. Lưu ý là bạn phải thu thập từ nguồn chính thống, có như vậy dữ liệu mới chính xác và máy có thể học một cách đúng đắn và đạt hiệu quả cao hơn.
- Preprocessing - tiền xử lý: bước này dùng để chuẩn hóa dữ liệu, loại bỏ các thuộc tính không cần thiết, gán nhãn dữ liệu, mã hóa một số đặc trưng, trích xuất đặc trưng, rút gọn dữ liệu nhưng vẫn đảm bảo kết quả các kỹ thuật tiền xử lý dữ liệu phổ biến hiện nay bao gồm: xử lý dữ liệu bị khuyết (missing data), mã hóa các biến nhóm (encoding categorical variables), chuẩn hóa dữ liệu (standardizing data), co giãn dữ liệu (scaling data),... Bước này tốn thời gian nhất tỉ lệ thuận với số lượng dữ liệu bạn có. Bước 1 và 2 thường chiếm hơn 70% tổng thời gian thực hiện.
- Training model - huấn luyện mô hình: bước này là bước bạn huấn luyện cho mô hình hay chính là cho nó học trên dữ liệu bạn đã thu thập và xử lý ở hai bước đầu.
- Evaluating model - đánh giá mô hình: sau khi đã huấn luyện mô hình xong, chúng ta cần dùng các độ đo để đánh giá mô hình, tùy vào từng độ đo khác nhau mà mô hình cũng được đánh giá tốt hay không khác nhau. Độ chính xác của mô hình đạt trên 80% được cho là tốt.
- Improve - cải thiện: sau khi đã đánh giá mô hình, các mô hình đạt độ chính xác không tốt thì cần được train lại, chúng ta sẽ lặp lại từ bước 3, cho đến khi đạt độ chính xác như kỳ vọng. Tổng thời gian của 3 bước cuối rơi vào khoảng 30% tổng thời gian thực hiện.

2.4.3 Ứng dụng của Machine Learning

Một số ứng dụng thực tế của Machine Learning như:

- Speech recognition: Dùng để nhận dạng giọng nói tự động (ASR), nhận dạng giọng nói máy tính hoặc chuyển giọng nói thành văn bản. Đây là một

khả năng sử dụng xử lý ngôn ngữ tự nhiên (NLP) để dịch giọng nói của con người sang định dạng viết.

- Customer service: Chatbots trực tuyến đang thay thế các tác nhân con người trong hành trình của khách hàng, thay đổi cách chúng ta nghĩ về sự tương tác của khách hàng trên website và nền tảng xã hội.
- Computer vision: Công nghệ AI này cho phép máy tính lấy thông tin có ý nghĩa từ video, hình ảnh kỹ thuật số và các đầu vào trực quan khác, sau đó thực thi hành động thích hợp.
- Recommendation engines: Sử dụng dữ liệu hành vi tiêu dùng trong quá khứ, các thuật toán AI learning có thể giúp khám phá các xu hướng dữ liệu có thể được sử dụng để phát triển các chiến lược cross-sell hiệu quả hơn.
- Automated stock trading: Được thiết kế để tối ưu hóa danh mục đầu tư chứng khoán, các nền tảng giao dịch tần suất cao do AI điều khiển để hàng triệu giao dịch mỗi ngày mà không cần đến sự can thiệp của con người.
- Fraud detection: Các ngân hàng và các tổ chức tài chính có thể sử dụng máy học để phát hiện các giao dịch đáng ngờ.

Ví dụ đơn giản như dự báo thời tiết, người ta sẽ dùng các phép tính và những quan sát, ghi nhận về thời tiết trong quá khứ để dự báo về thời tiết của những ngày kế tiếp. Tuy nhiên sẽ thế nào nếu như có cực kỳ nhiều quan sát được thực hiện, có thể lên đến hàng triệu, hàng tỉ quan sát, lúc đó con người không thể nào thực hiện được việc tính toán trên dữ liệu lớn như vậy. Hơn nữa, việc tính toán với dữ liệu lớn như vậy có thể gặp sai sót và dẫn đến kết quả dự đoán bị sai.

Khi này, việc áp dụng machine learning vào để cho máy tính học các quan sát được ghi nhận trong quá khứ, chúng có thể dự đoán được thời tiết trong tương lai với độ chính xác cao hơn rất nhiều so với con người dự đoán.

2.5 Deep Learning

2.5.1 Khái niệm Deep Learning

Deep Learning là tập hợp con của Machine Learning và nó có tác dụng hỗ trợ cho máy tính tự huấn luyện chính nó để có thể thực hiện tác vụ tương tự như con người. Điều này chính là giúp máy tính bắt chước con người cách học hỏi và suy nghĩ. Deep Learning hoạt động dựa vào mạng lưới thần kinh nhân tạo và nó bao gồm nhiều lớp chứa các dữ liệu mô phỏng cách thức vận hành của não người. Các hệ thống

của Deep Learning có khả năng cải thiện được những hiệu suất của chúng với quyền truy cập vào dữ liệu sẽ được nhiều hơn.

Thông thường, phiên bản máy sẽ có nhiều kinh nghiệm hơn; những máy móc đã có đủ kinh nghiệm thì sẽ được mang đi phục vụ cho những công việc như: phát hiện chẩn đoán trong y khoa, lái xe, các hệ thống trợ lý ảo,...

Deep Learning có hỗ trợ cho việc dịch ngôn ngữ, phân loại các hình ảnh, nhận dạng giọng nói. Chính vì thế, nó có thể được ứng dụng để giải quyết mọi nhu cầu cần nhận dạng mẫu mà không cần đến sự can thiệp của con người.

Bộ phận mạnh thần kinh nhân tạo này giống với não người bao gồm: Các node (đơn vị thần kinh trong mạng thần kinh nhân tạo) là những neuron thần kinh. Hầu hết các node bản thân thường chỉ có khả năng trả lời cho những câu hỏi đơn giản và cơ bản nhất; với những tác vụ khó thì chúng sẽ tiến hành liên kết với nhau để trả lời.

Có thể dạy hoặc có thể huấn luyện chúng bằng những thuật toán cụ thể. Những node giải đáp các câu hỏi phức tạp thì sẽ được gọi là mạng lưới thần kinh sâu (DNN), nó được định nghĩa như sau: Mạng lưới thần kinh sâu có khả năng thực hiện các hoạt động phức tạp như biểu diễn, trừu tượng có mang các ý nghĩa về âm thanh, văn bản và hình ảnh. Chúng được đánh giá là lĩnh vực phát triển nhất trong Machine Learning.

2.5.2 Cách thức hoạt động của Deep Learning

Deep Learning được xem là một phương pháp của học máy theo phương pháp: lập trình AI sẽ được huấn luyện để dự đoán đầu ra dựa vào một tập đầu vào. Ví dụ cụ thể: Dự đoán hành động của con mèo khi gặp chuột rồi tiến hành huấn luyện nó bằng phương pháp học có giám sát.

Khi ta muốn dự đoán được hành động của nó vào các đầu vào sẽ được thực hiện như sau:

- Lựa chọn con mèo phù hợp.
- Lúc này: các bộ phận trên cơ thể của con mèo như mắt, móng vuốt, tai,.. của con mèo sẽ trở nên rất nhanh nhạy.
- Con chuột sẽ xuất hiện ở đâu.

Về cơ bản thì Deep Learning sẽ không khác với học máy thông thường. Tuy nhiên, với ví dụ trên thì thời gian cần dành khá nhiều mới có thể thiết kế được các tính năng đại diện cho con mèo. Những gì cần làm là cung cấp cho hệ thống một số lượng hình ảnh con mèo, video con mèo bắt chuột thì hệ thống có thể tự học được

những tính năng đại diện dành cho 1 con mèo. Với những tác vụ như thị giác máy tính, nhận dạng giọng nói, robot hoặc dịch máy thì hiệu suất của Deep Learning có thể vượt xa những hệ thống máy khác. Tuy nhiên, việc xây dựng được hệ thống Deep Learning lại không mấy dễ dàng so với hệ thống máy móc thông thường.

Cách thức hoạt động của thuật toán Deep Learning diễn ra như sau: Các dòng thông tin sẽ được trải qua nhiều lớp cho đến lớp sau cùng. Lấy quy trình học của con người làm ví dụ cụ thể. Qua các lớp đầu tiên sẽ tập trung vào việc học các khái niệm cụ thể hơn trong khi các lớp sâu hơn sẽ sử dụng thông tin đã học để nghiên cứu và phân tích sâu hơn trong các khái niệm trừu tượng. Quy trình xây dựng biểu diễn dữ liệu này được gọi là trích xuất tính năng. Kiến trúc phức tạp của việc học sâu được cung cấp từ mạng lưới thần kinh sâu với khả năng thực hiện trích xuất tính năng tự động. Ngược lại, trong học máy thông thường còn gọi là học nông, nhiệm vụ này được thực hiện khi truy xuất các thuật toán cụ thể.

2.5.3 Các ưu và nhược điểm của Deep Learning

Deep Learning (Học sâu) có những ưu điểm và nhược điểm riêng. Dưới đây là một số điểm cần lưu ý:

Ưu điểm của Deep Learning:

- Khả năng học tập tự động: Deep Learning có khả năng học tập tự động từ dữ liệu, mà không cần định nghĩa cụ thể các đặc trưng hay quy tắc. Mô hình Deep Learning có khả năng tự tìm hiểu và trích xuất các đặc trưng phức tạp từ dữ liệu đầu vào.
- Xử lý dữ liệu lớn: Deep Learning hiệu quả trong việc xử lý dữ liệu lớn. Nhờ kiến trúc mạng neural sâu, nó có khả năng học từ dữ liệu phức tạp và đạt được kết quả tốt hơn so với các phương pháp truyền thống.
- Độ chính xác cao: Deep Learning thường mang lại độ chính xác cao trong nhiều tác vụ phân loại và nhận dạng, vượt trội so với các phương pháp truyền thống.

Nhược điểm của Deep Learning:

- Đòi hỏi dữ liệu lớn: Mô hình Deep Learning cần một lượng lớn dữ liệu huấn luyện để có thể học tốt. Thiếu dữ liệu có thể gây ra hiện tượng quá khớp (overfitting) hoặc kết quả không ổn định.

- Tính phức tạp và tài nguyên: Mô hình Deep Learning có kiến trúc phức tạp và đòi hỏi tài nguyên tính toán cao. Việc huấn luyện một mạng neural sâu có thể tốn nhiều thời gian và yêu cầu phần cứng mạnh mẽ.
- Khó hiểu và không giải thích được: Deep Learning thường được coi là "hộp đen" vì khó hiểu cách mô hình ra quyết định và không cung cấp giải thích rõ ràng về quá trình ra kết quả.
- Dễ bị tấn công: Mô hình Deep Learning có thể dễ dàng bị tấn công và đánh lừa bằng cách tạo ra dữ liệu đầu vào giả mạo hoặc thay đổi nhằm gây ra lỗi phân loại.
- Tuy Deep Learning mang lại nhiều ưu điểm, nhưng cần phải cân nhắc và đánh giá kỹ lưỡng các nhược điểm để đảm bảo sự áp dụng hiệu quả và an toàn của nó trong các tác vụ thực tế.

2.5.4 Ứng dụng của Deep Learning

Ứng dụng trong ngành công nghệ cao: Ứng dụng vượt bậc của Deep Learning không thể không kể đến đó chính là chế tạo Robot. Hiện nay, các phiên bản Robot giống con người với khả năng cảm nhận, phản ứng với môi trường đang dần được ra đời. Hiện nay, Robot còn có thể hợp tác với các hoạt động của con người và chúng có thể thực hiện được các nhiệm vụ riêng biệt phù hợp với thể mạnh của chúng. Robot đang góp phần thay thế con người trong việc thực hiện các công việc khó khăn hơn. Đây chính là phát minh tuyệt vời nhờ ứng dụng Deep Learning.

Ứng dụng trong nông nghiệp: Hiện nay, nhờ có Deep Learning mà người nông dân có thể triển khai những thiết bị có khả năng phân biệt được cỏ dại với cây trồng. Từ đó, các loại máy móc phun thuốc diệt cỏ có thể chọn lọc phun lên cỏ dại để đảm bảo cho cây trồng không bị ảnh hưởng. Ngoài vai trò loại bỏ cỏ dại bằng thuốc diệt cỏ thì nhờ Deep Learning mà sản lượng nông nghiệp ngày càng được cải thiện. Bên cạnh đó, Deep Learning đang được mở rộng hơn nữa sang các hoạt động như: thu hoạch, tưới tiêu, bón phân, gieo trồng.

Ứng dụng trong điện tử: Deep Learning được ứng dụng cho công việc nhận diện không mặt trong các thiết bị điện tử như điện thoại, các phần mềm như facebook, messenger,... Thêm vào đó, Deep Learning còn được dùng trong nghe dịch và nói tự động mà trang web google dịch đang sử dụng. Ngoài ra, một số thiết bị trợ giúp tại

nhà cũng được ứng dụng Deep Learning để đáp ứng giọng nói và hiểu được sở thích của bản (đây chính là ứng dụng học sâu).

Ứng dụng cho lĩnh vực không gian vũ trụ và quốc phòng: Deep Learning được dùng với các mục đích như: xác định các vật thể như vệ tinh, xác định khu vực trung tâm, xác định khu vực an toàn và không an toàn trong quân đội. AI kết hợp với phân tích không gian địa lý có thể hỗ trợ việc trích xuất thông tin hữu ích từ các thiết bị được liên kết như radar và hệ thống nhận dạng tự trị. Dữ liệu này có thể hỗ trợ việc phát hiện bất kỳ hoạt động bất hợp pháp hoặc đáng ngờ nào, cũng như cảnh báo cho các cơ quan chức năng thích hợp. Việc nhận dạng và phân loại mục tiêu cũng có thể được hỗ trợ bởi rô bốt hỗ trợ AI và thị giác máy tính hỗ trợ IoT.

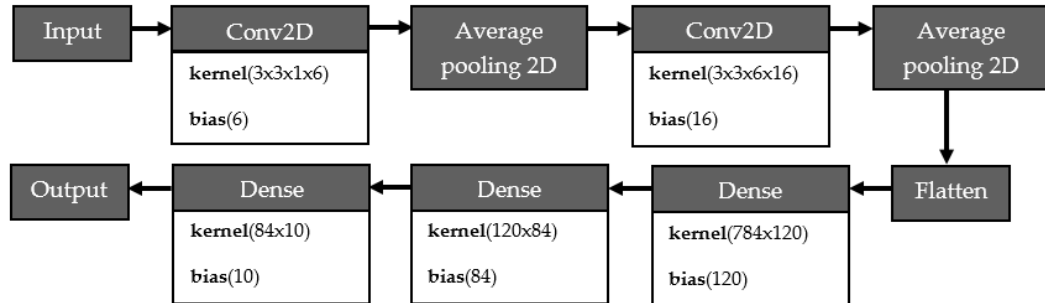
Ứng dụng trong y tế và chăm sóc sức khỏe: Deep Learning phát huy được hiệu quả siêu nổi bật của mình bằng cách hiển thị dữ liệu chất lượng cao và đo lường khả năng của các mạng thần kinh rồi thực hiện phân loại hình ảnh. Đồng thời, các nhà nghiên cứu về căn bệnh ung thư cũng đang ứng dụng DL để có thể tự động phát hiện ra các tế bào phát bệnh, nó được thực hiện như sau: chế tạo kính hiển vi tiên tiến để có thể tạo ra bộ dữ liệu có chiều cao phù hợp, được dùng để đào tạo một ứng dụng học sâu để có thể xác định tế bào ung thư. Ngoài ra, Deep Learning còn có bước tiến quan trọng trong việc cải thiện chất lượng sức khỏe của bệnh nhân bằng sự dự đoán với các sự kiện y tế từ dữ liệu hồ sơ sức khỏe điện tử.

Ứng dụng trong công nghiệp: Deep Learning có khả năng cải thiện sự an toàn cho các công nhân xung máy móc hạng nặng bằng cách: tự động phát hiện nếu như người hoặc vật nằm trong khoảng cách không an toàn đối với máy. Sản xuất thông minh hỗ trợ hệ thống thông tin đã tăng năng suất và chất lượng của các tổ chức công nghiệp. Sử dụng phân tích dữ liệu, mô hình thống kê và thuật toán dự đoán đã tăng lên theo các bước nhảy vọt, vì chất lượng và xu hướng của dữ liệu do máy tạo và con người tạo ra được cải thiện theo thời gian.

2.5.5 Mạng nơ-ron tích chập CNN

Convolutional Neural Network (CNN - Mạng nơ-ron tích chập) là một trong những mô hình Deep Learning tiên tiến. Nó giúp cho chúng ta xây dựng được những hệ thống thông minh với độ chính xác cao như hiện nay. CNN được sử dụng nhiều trong các bài toán nhận dạng các object trong ảnh.

Trong các mạng MLP truyền thống, mỗi nơ-ron trong lớp phía trước sẽ được kết nối đến tất cả các nơ-ron ở lớp phía sau, điều này khiến cho khối lượng tính toán trong mạng tăng mạnh khi tăng độ sâu của mô hình (tăng số lượng lớp) cho mô hình.



Hình 2.2. Mô hình mạng CNN[20]

Sự ra đời của mạng CNN đã giúp giải quyết vấn đề trên bằng cách sử dụng các vùng tiếp nhận cục bộ, tập trọng số chia sẻ và phương pháp lấy tích chập để trích xuất thông tin thay cho phương pháp cổ điển.

Sau đây, chúng tôi sẽ đi vào chi tiết hơn về mạng nơ-ron tích chập và các tham số sẽ được nghiên cứu đánh giá.

2.5.5.1 Lớp tích chập (convolutional layer)

Lớp tích chập là một thành phần cốt lõi của mạng nơ-ron tích chập (CNN), sử dụng để trích xuất các thông tin đặc tính của hình ảnh (feature map) hỗ trợ cho quá trình “học” của mạng CNN [20]. Phương thức hoạt động của lớp này được thực hiện thông qua quá trình trượt và lấy tích chập của bộ lọc (filter/kernel) trên toàn bộ ảnh. Kết quả đầu ra là đặc tính của ảnh tương ứng với bộ lọc đã sử dụng, với càng nhiều bộ lọc được sử dụng, chúng tôi sẽ thu được càng nhiều đặc tính của ảnh tương ứng.

Trong phạm vi bài báo này, chúng tôi sẽ xét đến các tham số có tác động đến lớp tích chập gồm: kích thước bộ lọc, bước trượt (stride) [20].

2.5.5.2 Lớp lấy mẫu xuống (pooling layer)

Lớp lấy mẫu xuống có tác dụng giảm kích thước của dữ liệu hình ảnh từ đó giúp cho mạng có thể học được các thông tin có tính chất khái quát hơn, đồng thời quá trình này giảm số lượng các thông số trong mạng [21].

Các phương pháp lấy mẫu xuống thường được sử dụng là Max Pooling và Average Pooling [22].

2.5.5.3 Lớp dropout

Lớp dropout là một kỹ thuật được sử dụng để hạn chế hiện tượng overfitting (hiện tượng mạng nơ-ron quá bám sát vào tập dữ liệu huấn luyện và không đáp ứng được với các tập dữ liệu mới), thường gặp ở mạng CNN và giúp mô hình tính toán nhanh hơn. Dropout sử dụng phương pháp loại bỏ một số nơ-ron ngẫu nhiên trong mạng với một xác suất cho trước bằng cách thiết lập tất cả trọng số nơ-ron đó bằng 0, đồng nghĩa với các liên kết tới nơ-ron đó đều không có giá trị, khi đó mô hình sẽ phải cố gắng nhận dạng đúng trong khi thiếu thông tin từ các nơ-ron bị loại bỏ. Điều này sẽ giúp tăng tỉ lệ nhận dạng của mô hình nhưng không quá phụ thuộc vào dữ liệu huấn luyện [20].

2.5.5.4 Lớp kết nối đầy đủ (Fully-connected layer - FC)

Đầu vào của lớp kết nối đầy đủ là đầu ra từ lớp lấy mẫu xuống hoặc lớp tích chập cuối cùng, nó được làm phẳng và sau đó được đưa vào lớp kết nối đầy đủ để chuyển tiếp. Lớp FC có nhiệm vụ tổng hợp thông tin đưa ra lớp quyết định (output) cho ra kết quả chính xác nhất.

2.5.6 Kiến trúc mạng Densenet121

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

Hình 2.3. Thông số kiến trúc DenseNet121[27]

Kiến trúc Densenet-121 là một mô hình mạng nơ-ron sâu (deep neural network) được đề xuất bởi Gao Huang, Zhuang Liu, Laurens van der Maaten và Kilian Q. Weinberger vào năm 2016. Nó là một biến thể của mạng Densenet và đã được sử dụng phổ biến trong lĩnh vực nhận dạng hình ảnh và phân loại.[26]

Dưới đây là một tổng quan về mô hình Densenet-121:

Kiến trúc cơ bản: Mô hình Densenet-121 là một mạng sâu với tổng cộng 121 lớp. Nó bao gồm các lớp Convolutional, Batch Normalization và Activation (thường là ReLU) để xử lý thông tin đầu vào.

Đơn vị cơ bản: Đơn vị cơ bản trong Densenet-121 được gọi là Dense Block. Mỗi Dense Block bao gồm một loạt các Residual Unit, trong đó các đầu vào từ các lớp trước được nối kết (concatenated) với nhau trước khi đi qua các lớp Convolutional và Batch Normalization. Điều này cho phép mô hình tích hợp thông tin từ tất cả các lớp trước và truyền nó qua các lớp sau.

Kết nối mật độ (Dense Connection): Mô hình Densenet-121 sử dụng kết nối mật độ giữa các lớp để tạo ra một mạng cực kỳ mật độ. Mỗi lớp trong một Dense Block nhận đầu vào từ tất cả các lớp trước đó trong Dense Block. Điều này giúp truyền ngược hiệu quả đặc trưng từ các lớp trước đó và tạo ra một luồng thông tin rõ ràng và liên tục trong mạng.

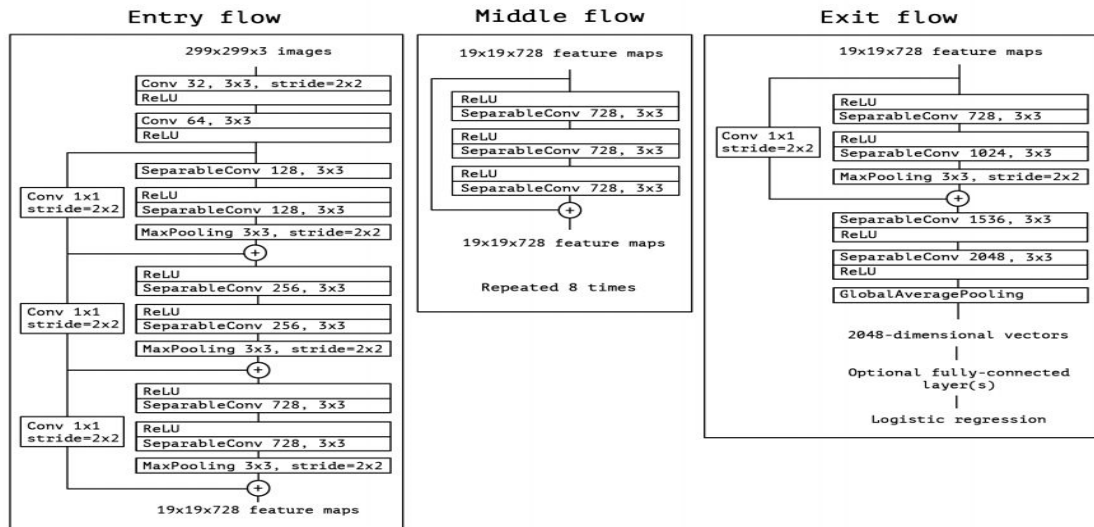
Kiến trúc chính: Mô hình Densenet-121 bao gồm nhiều Dense Block được xếp chồng lên nhau. Kích thước của các Dense Block và số lượng các Residual Unit trong mỗi Dense Block có thể khác nhau.

Lớp Global Average Pooling: Sau khi đi qua các Dense Block, đầu ra được đưa vào một lớp Global Average Pooling để tính trung bình của các đặc trưng trên toàn bộ không gian.

Lớp đầu ra: Cuối cùng, một lớp Fully Connected được sử dụng để tạo ra đầu ra dự đoán, có số lượng nơ-ron tương ứng với số lớp phân loại.

Mô hình Densenet-121 được đánh giá cao vì khả năng học tập mạnh mẽ và khả năng sử dụng thông tin đặc trưng một cách hiệu quả. Nó đã đạt được hiệu suất tốt trong nhiều nhiệm vụ như phân loại hình ảnh và phân đoạn hình ảnh.

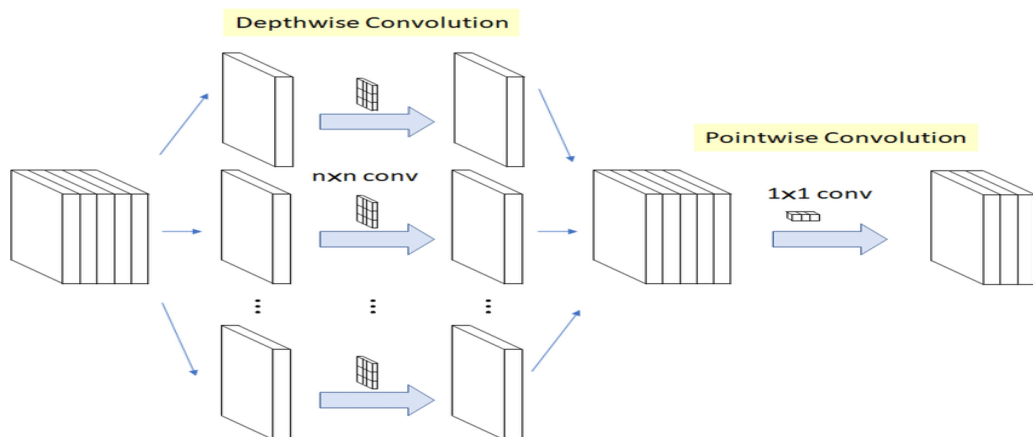
2.5.7 Kiến trúc mạng Xception



Hình 2.4. Kiến trúc mạng Xception[28]

Kiến trúc Xception là một mô hình mạng nơ-ron sâu (deep neural network) được giới thiệu bởi François Chollet vào năm 2016. Tên gọi "Xception" là sự kết hợp của hai từ "Extreme" và "Inception", ám chỉ sự mở rộng và phát triển từ kiến trúc Inception.

Xception là một kiến trúc mạng tích chập (convolutional neural network - CNN) dựa trên ý tưởng của kiến trúc Inception của Google. Tuy nhiên, Xception có một cải tiến quan trọng bằng cách áp dụng kiến trúc "depthwise separable convolution" (convolution tách riêng theo chiều sâu).



Hình 2.5. Kiến trúc depthwise separable convolution[28]

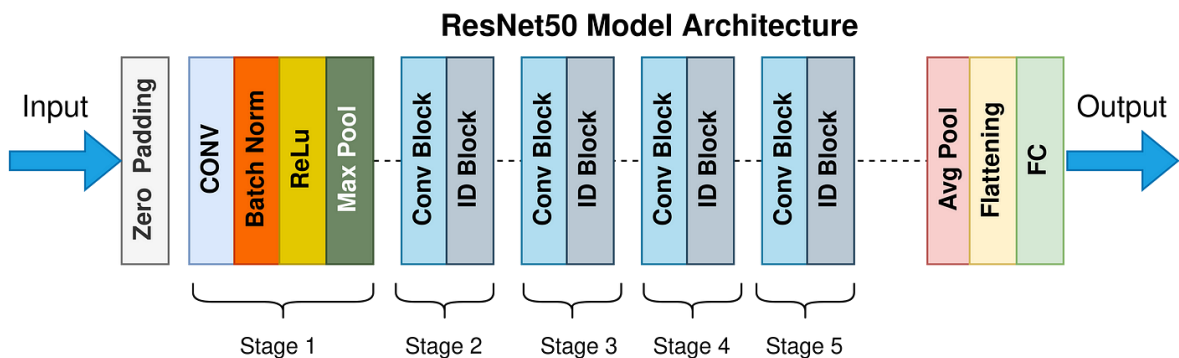
Trong kiến trúc Xception, thay vì sử dụng các bộ lọc (filters) tiêu chuẩn như trong lớp convolution thông thường, nó sử dụng hai lớp convolution riêng biệt: lớp depthwise convolution (convolution theo chiều sâu) và lớp pointwise convolution

(convolution điểm). Lớp depthwise convolution áp dụng một bộ lọc độc lập cho từng kênh đầu vào, trong khi lớp pointwise convolution thực hiện việc kết hợp thông tin từ các kênh độc lập thành kênh đầu ra cuối cùng.

Kiến trúc Xception giúp giảm số lượng tham số và tính toán so với kiến trúc Inception, đồng thời vẫn duy trì khả năng học các đặc trưng phức tạp từ dữ liệu. Nó tận dụng việc tách riêng phân tách theo chiều sâu và không gian trong quá trình tính toán, giúp cải thiện hiệu suất và tăng tốc quá trình huấn luyện mô hình.

Kiến trúc Xception thường được sử dụng cho các nhiệm vụ như phân loại hình ảnh, nhận dạng đối tượng và phân đoạn hình ảnh. Nó đã đạt được hiệu suất rất cao trên nhiều bộ dữ liệu thị giác máy tính và trở thành một trong những kiến trúc quan trọng và phổ biến trong lĩnh vực này.

2.5.8 Kiến trúc mạng Resnet50



Hình 2.6. Kiến trúc Resnet50[29]

Resnet-50 là một mô hình mạng nơ-ron sâu (deep neural network) trong lĩnh vực học sâu và được đề xuất bởi Kaiming He, Xiangyu Zhang, Shaoqing Ren và Jian Sun vào năm 2015. Nó là một biến thể của mạng Residual Network (Resnet) và trở thành một trong những mô hình nổi tiếng và phổ biến trong lĩnh vực nhận dạng hình ảnh và phân loại.

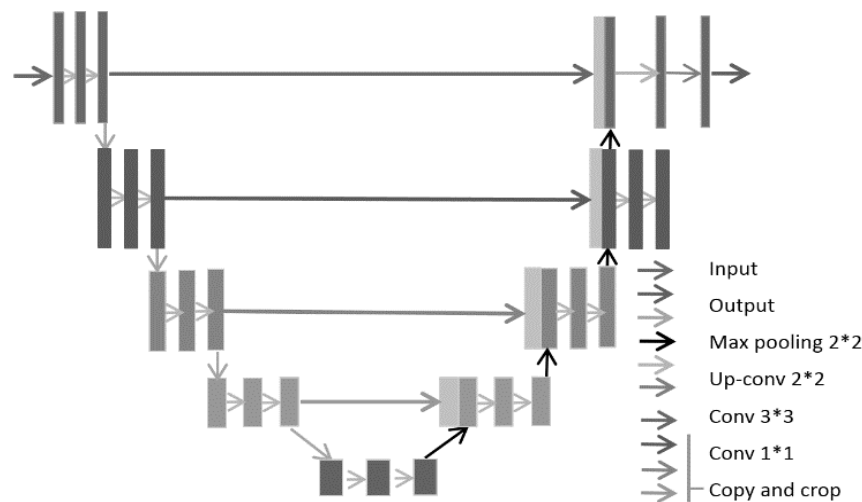
Dưới đây là một tổng quan về mô hình Resnet-50:

- Kiến trúc cơ bản: Mô hình Resnet-50 là một mạng sâu với 50 lớp. Nó bao gồm các lớp Convolutional, Batch Normalization, Activation (thường là ReLU) và lớp Max Pooling để giảm kích thước ảnh.
- Đơn vị cơ bản: Đơn vị cơ bản trong Resnet-50 là Residual Block. Mỗi Residual Block bao gồm hai nhánh:

- Nhánh chính: Bao gồm các lớp Convolutional và Batch Normalization để xử lý thông tin đầu vào.
- Nhánh rút ngắn (Shortcut Connection): Sử dụng một lớp Convolutional 1×1 để ánh xạ thông tin từ đầu vào trực tiếp vào đầu ra của khối. Điều này giúp mô hình có khả năng vượt qua hiện tượng vanishing gradient và cho phép truyền ngược hiệu quả các đặc trưng quan trọng từ các lớp trước.
- Kiến trúc chính: Mô hình Resnet-50 có cấu trúc chính với năm nhóm (groups) của các Residual Blocks, trong đó mỗi nhóm có số lượng khối khác nhau. Các nhóm đầu tiên có các Residual Blocks với kích thước đầu vào và đầu ra giống nhau, trong khi các nhóm sau có kích thước đầu vào và đầu ra khác nhau.
- Lớp Global Average Pooling: Sau khi đi qua các nhóm Residual Blocks, đầu ra được đưa vào một lớp Global Average Pooling để tính trung bình của các đặc trưng trên toàn bộ không gian.
- Lớp đầu ra: Cuối cùng, một lớp Fully Connected được sử dụng để tạo ra đầu ra dự đoán, có số lượng nơ-ron tương ứng với số lớp phân loại.

Resnet-50 được đánh giá cao với khả năng học sâu mà không gặp vấn đề vanishing gradient, cũng như khả năng đạt hiệu suất tốt trong nhiều nhiệm vụ như phân loại hình ảnh, phân đoạn và nhận dạng đối tượng. Kiến trúc Resnet-50 đã trở thành một trong những kiến trúc mạng quan trọng và phổ biến trong cộng đồng học sâu.[24]

2.5.9 Kiến trúc mạng Unet



Hình 2.7. Kiến trúc mạng Unet[30]

Mô hình U-Net là một kiến trúc mạng nơ-ron sâu (deep neural network) được sử dụng phổ biến trong các nhiệm vụ phân đoạn ảnh và xử lý ảnh y tế. Nó đã được đề xuất bởi Olaf Ronneberger, Philipp Fischer và Thomas Brox vào năm 2015, và từ đó đã trở thành một trong những mô hình phổ biến và hiệu quả trong lĩnh vực này.

Kiến trúc của mô hình U-Net có hình dạng giống hình chữ U, do đó được gọi là "U-Net". Nó bao gồm hai phần chính: đường đi thu hẹp (contracting path) và đường đi mở rộng (expanding path).

Contracting path: Phần này giúp mô hình thu thập thông tin về cấu trúc và đặc trưng của ảnh đầu vào. Nó bao gồm một loạt các khối gộp (downsampling) liên tiếp, mỗi khối gộp bao gồm một lớp Convolutional, sau đó là một lớp Activation (thường là ReLU), và cuối cùng là một lớp Max Pooling để giảm kích thước ảnh.

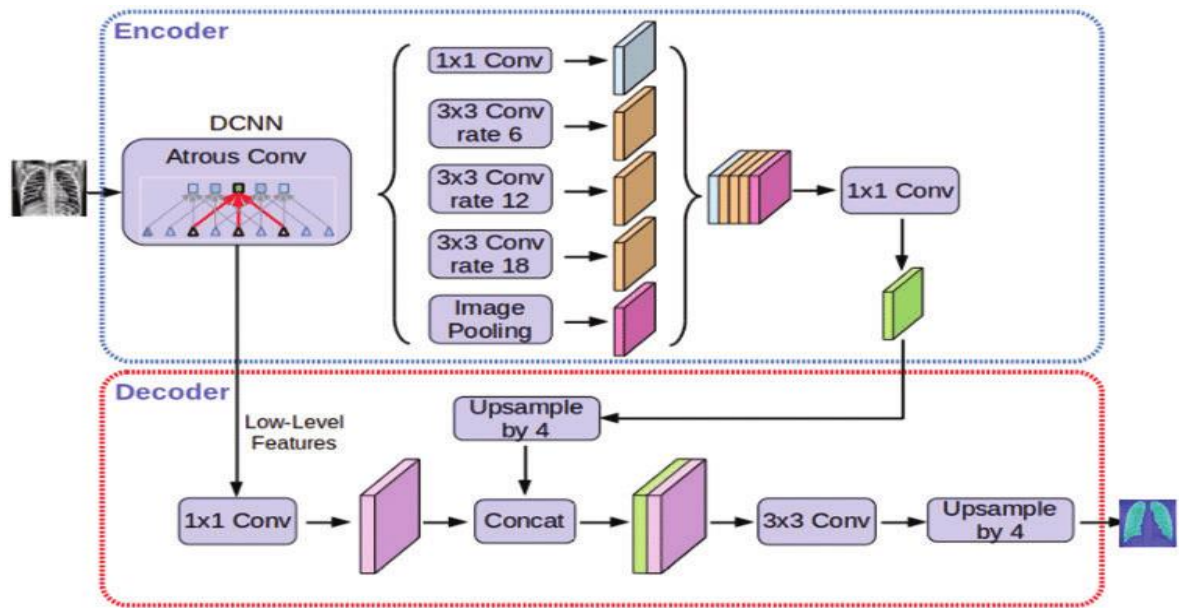
Bottleneck: Sau khi thông tin đã được thu gọn, một lớp Convolutional được sử dụng để duy trì thông tin chi tiết và đặc trưng quan trọng từ phần thu hẹp.

Expanding path: Phần này giúp tái tạo ảnh với độ phân giải tương tự ảnh gốc. Nó bao gồm một loạt các khối mở rộng (upsampling) liên tiếp, mỗi khối mở rộng bao gồm các lớp Upconvolution (hoặc Transposed Convolution), một lớp Concatenation (nối kết) với đầu vào từ đường đi thu hẹp tương ứng, một lớp Convolutional, và sau đó là lớp Activation.

Đầu ra: Một lớp Convolutional cuối cùng được sử dụng để tạo ra đầu ra dự đoán, có kích thước tương tự với ảnh gốc và số lượng kênh tương ứng với số lớp phân đoạn.[23]

Mô hình U-Net có khả năng kết hợp thông tin cục bộ và thông tin ngữ cảnh, từ đó cho phép nó đạt được độ chính xác cao trong việc phân đoạn ảnh. Nó đã được áp dụng thành công trong nhiều nhiệm vụ phân đoạn như phân đoạn cấu trúc y sinh học, tế bào, bướu, và cảnh quan y tế trong hình ảnh y tế.

2.5.10 Kiến trúc mạng DeepLabV3+



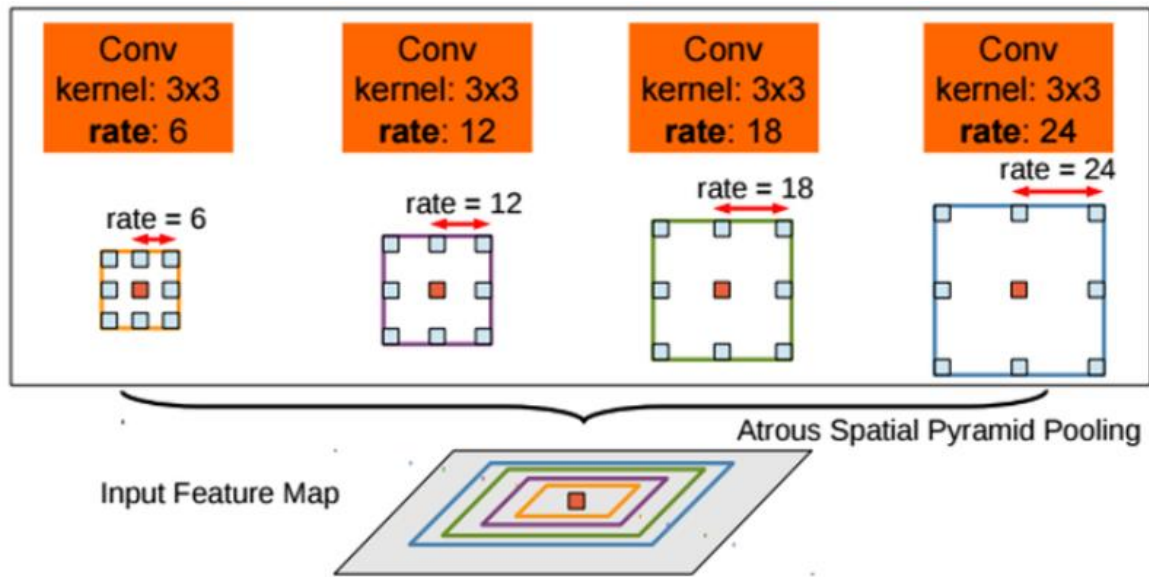
Hình 2.8. Kiến trúc mạng DeeplabV3+ [31]

DeepLabV3+ là một phiên bản nâng cấp của kiến trúc DeepLab được phát triển để giải quyết bài toán phân đoạn hình ảnh (image segmentation) với độ chính xác cao. Nó kết hợp các thành phần quan trọng từ các phiên bản trước đó của DeepLab và áp dụng một số cải tiến để nâng cao hiệu suất và chất lượng phân đoạn.

Dưới đây là mô tả tổng quan về kiến trúc DeepLabV3+:

Backbone Network: DeepLabv3+ sử dụng một mạng nơ-ron tích chập (convolutional neural network - CNN) làm phần cốt lõi (backbone) của mô hình. Thông thường, ResNet hoặc Xception được sử dụng làm backbone network. Mạng này được sử dụng để trích xuất các đặc trưng từ hình ảnh đầu vào và tạo ra các biểu diễn đặc trưng có độ phân giải thấp.

Atrous Spatial Pyramid Pooling (ASPP): DeepLabv3+ sử dụng kiến trúc Atrous Spatial Pyramid Pooling (ASPP) để thu thập thông tin từ các vùng bối cảnh ở nhiều tỷ lệ khác nhau. ASPP sử dụng các lớp atrous convolution ở các tỷ lệ atrous rate khác nhau để tạo ra các biểu diễn đặc trưng với các lĩnh vực nhìn rộng và chi tiết. Điều này



Hình 2.9. Mô tả Atrous Spatial Pyramid Pooling [31]

giúp mô hình có khả năng phân đoạn các đối tượng có kích thước và hình dạng khác nhau.

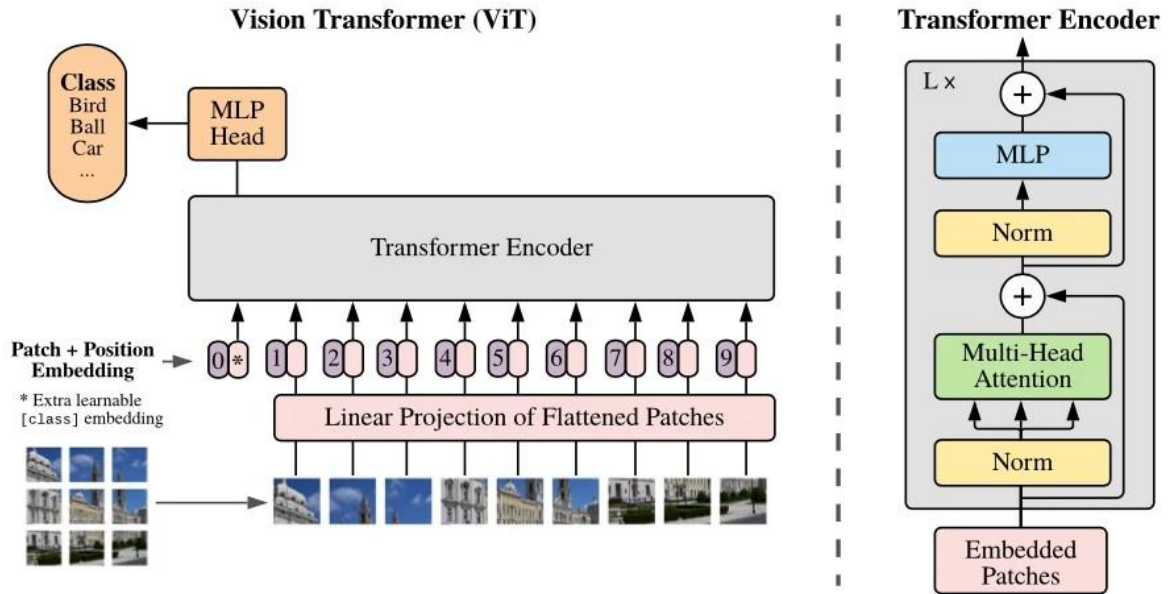
Encoder-Decoder Architecture: DeepLabv3+ sử dụng kiến trúc encoder-decoder để kết hợp thông tin chi tiết từ các biểu diễn đặc trưng có độ phân giải thấp và thông tin bối cảnh rộng từ các biểu diễn đặc trưng có độ phân giải cao. Mô hình sử dụng lớp upsampling và skip connections để kết hợp các đặc trưng từ các lớp phân giải thấp và lớp phân giải cao, từ đó tạo ra kết quả phân đoạn chi tiết và chính xác.

DeepLabv3+ là một kiến trúc mạng mạnh mẽ cho bài toán phân đoạn hình ảnh. Nó kết hợp các thành phần quan trọng như ASPP và kiến trúc encoder-decoder để đạt được kết quả phân đoạn chính xác và chi tiết. Cải tiến refinement cũng cải thiện chất lượng kết quả. DeepLabv3+ đã đạt được kết quả ấn tượng trên nhiều bộ dữ liệu phân đoạn và trở thành một trong những kiến trúc phổ biến trong lĩnh vực này.

2.5.11 Kiến trúc mạng Vision Transformer

Vào năm 2022, Vision Transformer (ViT) nổi lên như một giải pháp thay thế cạnh tranh so với các mạng thần kinh tích chập (Convolutional Neural Network, gọi tắt là CNN) vốn đang là ứng dụng tiên tiến trong thị giác máy tính, và được sử dụng rộng rãi trong các tác vụ nhận dạng hình ảnh khác nhau. Các mô hình ViT được đánh giá là vượt trội hơn so với CNN gần 4 lần về hiệu quả tính toán và độ chính xác.

Các mô hình Transformer từ trước đến nay đã trở thành nền tảng của rất nhiều mô hình khác, tạo ra một bước ngoặt lớn trong lĩnh vực Xử lý ngôn ngữ tự nhiên



Hình 2.10. Kiến trúc Vision Transformer[32]

(NLP). Trong nghiên cứu thị giác máy tính, ngày càng có nhiều sự quan tâm dành cho Vision Transformer (ViT) – một loại mô hình Transformer.

Kiến trúc tổng thể của mô hình ViT được đưa ra như sau theo cách thức từng bước:

- Chia hình ảnh thành các mảng (patch) với kích thước từng mảng cố định
- Làm phẳng các mảng hình ảnh
- Tạo các feature embedding có chiều thấp hơn từ các mảng hình ảnh phẳng này
- Bao gồm thứ tự các mảng
- Chuỗi feature embedding được làm đầu vào cho transformer encoder
- Thực hiện pre-train đối với mô hình ViT với các nhãn hình ảnh, sau đó được giám sát hoàn toàn trên một tập dữ liệu lớn
- Tinh chỉnh model trên bộ dữ liệu riêng của từng bài toán

ViT đạt được kết quả đáng chú ý hơn CNN, trong khi dùng ít tài nguyên tính toán hơn cho pre-training. Nhìn chung, so với CNN, ViT có xu hướng quy nạp yếu hơn, dẫn đến sự phụ thuộc ngày càng nhiều vào việc điều chỉnh mô hình (model regularization) hoặc tăng dữ liệu (AugReg) khi pre-train trên các tập dữ liệu nhỏ hơn.

ViT là một mô hình trực quan dựa trên kiến trúc của một kiến trúc Transformer ban đầu được thiết kế cho các tác vụ dựa trên văn bản. Mô hình ViT chuyển hình ảnh đầu vào thành một loạt các mảng hình ảnh, và dự đoán trực tiếp các nhãn lớp cho

hình ảnh. ViT cho thấy một hiệu suất phi thường khi được huấn luyện trên đủ dữ liệu, vượt xa hiệu suất của một CNN hiện đại tương tự với tài nguyên tính toán ít hơn 4 lần.

Những kiến trúc Transformer này có tỷ lệ thành công cao trong mô hình NLP và hiện cũng được áp dụng cho hình ảnh và các nhiệm vụ nhận dạng hình ảnh. CNN sử dụng mảng pixel, trong khi ViT chia hình ảnh thành các tokens trực quan. Transformer trực quan chia hình ảnh thành các mảng hình ảnh có kích thước cố định, mã hóa từng mảng theo thứ tự làm đầu vào cho Transformer encoder. Hơn nữa, các mô hình ViT vượt trội hơn CNN gần bốn lần về hiệu quả tính toán và độ chính xác.[5]

2.5.12 Phương pháp đánh giá mô hình mạng học sâu

2.5.12.1 Đánh giá độ đo Accuracy:

Trong bài toán phát hiện và phân lớp đối tượng đặc biệt là bài toán phân loại đa lớp áp dụng nhiều mô hình mạng nơ-ron khác nhau, việc lựa chọn một phương pháp phù hợp để đánh giá và so sánh là rất cần thiết. Một trong các phương pháp được sử dụng phổ biến là xác định độ đo Loss và Accuracy.

Accuracy (độ chính xác) chỉ đơn giản đánh giá mô hình thường xuyên dự đoán đúng đến mức nào. Độ chính xác là tỉ lệ giữa số điểm dữ liệu được dự đoán đúng và tổng số điểm dữ liệu.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{Số\ lượng\ mẫu} \quad (1)$$

Trong đó:

- TP (True positive): dương tính thật, mô hình dự đoán là “có” và thực tế là “có”.
- FP (False positive): dương tính giả, mô hình dự đoán là “có” nhưng thực tế là “không”.
- FN (False negative): âm tính giả, mô hình dự đoán “không” nhưng thực tế “có”.
- TN (True negative): là âm tính thật, mô hình dự đoán là “không” và thực tế là “không”.

2.5.12.2 ROC Curve (Receiver Operating Characteristic):

Để vẽ ROC ta sẽ sử dụng 2 khái niệm:

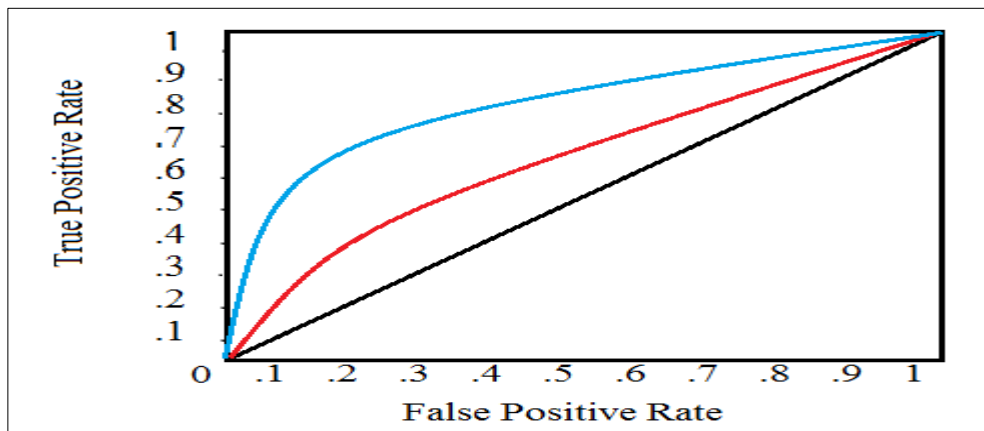
- Đầu tiên là True Positive Rate (TPR) chính là Recall.

- False Positive Rate (FPR) là tỷ lệ cảnh báo sai.

$$\text{TruePositiveRate} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalseNegative}} \quad (2)$$

$$\text{FalsePositiveRate} = \frac{\text{FalsePositive}}{\text{FalsePositive} + \text{TrueNegative}} \quad (3)$$

Thực chất đường ROC để chỉ lên mối quan hệ giữa TPR và FPR khi chúng ta thay đổi ngưỡng threshold của model. Ví dụ ta có thể đặt một ngưỡng threshold=0.5 khi phân loại mail rác, các mail nào có probability trên threshold sẽ được coi là mail rác. Bây giờ ta sẽ thay đổi threshold để xem sự thay đổi của TPR và FPR trên đồ thị.



Hình 2.11. Sự thay đổi của TPR và FPR[33]

Trong lĩnh vực Machine Learning, việc tính toán hiệu suất của mô hình là một điều rất quan trọng. Trong các bài toán phân loại, ngoài phương pháp phổ biến là sử dụng Confusion Matrix thì còn phải nhắc đến một phương pháp rất quan trọng khác, đó là đường cong AUC-ROC. AUC-ROC là một phương pháp tính toán hiệu suất của một mô hình phân loại theo các ngưỡng phân loại khác nhau.

2.5.12.3 Confusion matrix:

Trong những bài toán phân loại lớp, sau khi đã xử lý dữ liệu và đưa vào mô hình học máy, đầu ra của mô hình sẽ là một vector xác suất tương ứng của từng lớp. Ta có thể đánh giá mức độ chính xác của mô hình thông qua chỉ số về accuracy (độ chính xác): là phần trăm các lớp đã phân loại đúng / tổng số dự đoán, thế nhưng với chỉ số các lớp đã phân loại sai thì sẽ ảnh hưởng rất lớn đến việc chẩn đoán. Giả sử với bài toán chẩn đoán ung thư với việc phân loại một người thành 2 lớp: bị ung thư và không bị ung thư. Nếu trong 100 người cần phân loại có 10 người bị ung thư mà mô hình dự đoán cả 100 người đều không bị ung thư thì độ chính xác của mô hình đã là 0.9

khá cao. Nhưng với 10 người bị ung thư đã bị bỏ qua thì rõ ràng mô hình này không đáng tin cậy. Cần phải có một phương pháp đánh giá tốt hơn là độ chính xác trong bài toán phân loại và một trong đó là sử dụng Confusion matrix.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Hình 2.12. Confusion matrix[34]

Confusion matrix là một phương pháp đánh giá kết quả của những bài toán phân loại với việc xem xét cả những chỉ số về độ chính xác và độ bao quát của các dự đoán cho từng lớp. Một confusion matrix gồm 4 chỉ số sau đối với mỗi lớp phân loại:

- TP (True Positive): Số lượng dự đoán chính xác. Là khi mô hình dự đoán đúng một người bị ung thư lành tính.
- TN (True Negative): Số lượng dự đoán chính xác một cách gián tiếp. Là khi mô hình dự đoán đúng một người bị ung thư ác tính, tức là việc không chọn trường hợp bị ung thư lành tính là chính xác.
- FP (False Positive): Số lượng các dự đoán sai lệch. Là khi mô hình dự đoán một người bị ung thư lành tính nhưng người đó bị ung thư ác tính.
- FN (False Negative): Số lượng các dự đoán sai lệch một cách gián tiếp. Là khi mô hình dự đoán một người không bị ung thư lành tính nhưng người đó bị ung thư ác tính, tức là việc không chọn trường hợp bị ung thư lành tính là sai.

Từ 4 chỉ số này, ta có 3 con số để đánh giá mức độ tin cậy của một mô hình:

- Công thức tính precision:

$$Precision = \frac{TP}{TP + FP} = \frac{TP}{all\ detections} \quad (4)$$

- Công thức tính recall:

$$Recall = \frac{TP}{TP + FN} = \frac{TP}{all\ ground\ truth} \quad (5)$$

F1-Score là trung bình hài hòa của Precision và recall và nó thể hiện sự đóng góp của cả hai điểm. Sự đóng góp phụ thuộc vào giá trị beta trong công thức (19) dưới đây:

$$F1 - Score = 2 \frac{(Precision * Recall)}{Precision + Recall} \quad (6)$$

Một mô hình có chỉ số F1-score cao chỉ khi cả 2 chỉ số Precision và Recall đều cao. Một trong 2 chỉ số này thấp đều sẽ kéo điểm F1-score xuống. Trường hợp xấu nhất khi 1 trong hai chỉ số Precision và Recall bằng 0 sẽ kéo điểm F1-score về 0. Trường hợp tốt nhất khi cả điểm chỉ số đều đạt giá trị bằng 1, khi đó điểm F1-score sẽ là 1.

2.5.12.4 Dice coefficient:

Dice coefficient (hệ số Dice) là một phép đo sự tương đồng (similarity measure) thường được sử dụng trong lĩnh vực xử lý hình ảnh và phân đoạn hình ảnh (image segmentation). Nó được sử dụng để đánh giá mức độ khớp (overlap) giữa hai khu vực phân đoạn hoặc đối tượng trên hình ảnh.

Hệ số Dice được tính bằng cách so sánh kích thước khu vực chung giữa hai đối tượng (hoặc vùng phân đoạn) với tổng kích thước của hai đối tượng. Kết quả của Dice coefficient dao động từ 0 đến 1, với 0 cho thấy không có sự khớp giữa hai đối tượng và 1 cho thấy sự khớp hoàn toàn.

Công thức tính Dice coefficient cho hai vùng (A và B) được biểu diễn như sau:

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \quad (7)$$

Trong đó:

- $|A \cap B|$ đại diện cho kích thước (diện tích) khu vực chung giữa A và B.
- $|A|$ đại diện cho kích thước (diện tích) của vùng A.
- $|B|$ đại diện cho kích thước (diện tích) của vùng B.

Dice coefficient là một phép đo thường được sử dụng để đánh giá chất lượng phân đoạn hình ảnh trong các bài toán như phân đoạn tế bào, phân đoạn đối tượng, hay phân đoạn cấu trúc ảnh. Càng gần 1 hệ số Dice, thì sự tương đồng giữa hai vùng phân đoạn càng cao.

CHƯƠNG 3 - PHƯƠNG PHÁP NGHIÊN CỨU

3.1 Tập dữ liệu

Tập dữ liệu SIIM-ACR Pneumothorax Segmentation [9] là một tập dữ liệu y tế được sử dụng để huấn luyện và đánh giá các mô hình học máy và học sâu trong việc phát hiện và phân vùng tràn khí màng phổi trên ảnh chụp X-Ray ngực.

Một số thông tin về tập dữ liệu này:

- Nguồn gốc: Tập dữ liệu SIIM-ACR Pneumothorax Segmentation được tổ chức bởi Hiệp hội Y tế ảnh X-Ray Mỹ (American College of Radiology - ACR) và Tổ chức ảnh hưởng hệ thống thông tin y tế (Society for Imaging Informatics in Medicine - SIIM).
- Số lượng và độ đa dạng: Tập dữ liệu SIIM-ACR Pneumothorax Loaiementation (kích thước 2.08 GB) gồm 12.047 hình ảnh X-Ray ngực từ nhiều bệnh viện khác nhau với độ phân giải 256x256, trong đó có 2669 ảnh phổi bị bệnh và 9378 ảnh phổi bình thường .
- Định dạng: Tập dữ liệu chứa các hình ảnh X-Ray ngực được lưu dưới định dạng DICOM (Digital Imaging and Communications in Medicine), là định dạng phổ biến trong lĩnh vực y tế. Các đánh dấu phân loại được cung cấp dưới dạng các hình ảnh mặt nạ (mask) tương ứng với từng hình ảnh.

3.2 Môi trường thực nghiệm

Môi trường máy tính đơn trên nền tảng cung cấp của Google Colab Pro với CPU Intel Xeon Processor with two cores 2.30 GHz đi kèm với 25 GB Ram cho phép huấn luyện với batch size lớn hơn, với GPU được sử dụng trong việc nghiên cứu là Nvidia Tesla V100 với hiệu năng vượt trội của GPU này sẽ giúp cho quá trình huấn luyện được nhanh và chính xác hơn.

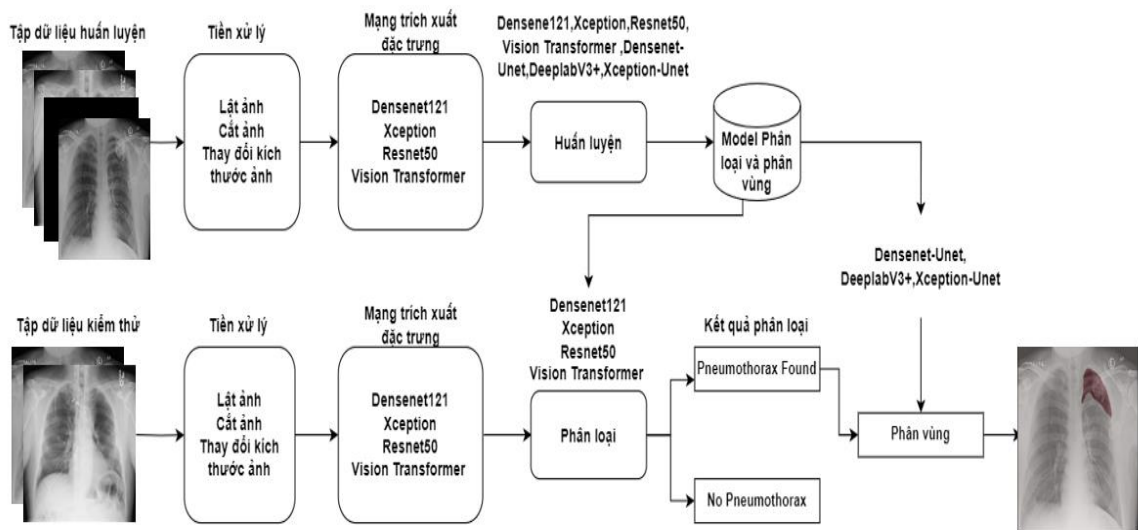
Bảng 4.1. Thông số môi trường huấn luyện

Tên môi trường	Cấu hình	Thư viện
Google Colab Pro	CPU: Intel Xeon Processor GPU: Nvidia Tesla V100 (16GB Ram) RAM: 25GB	Tensorflow 2.12 Keras 2.12

3.2 Mô hình kiến trúc đề xuất

Để xây dựng một hệ thống phát hiện bệnh tràn khí màng phổi, chúng tôi đã áp dụng một loạt các mô hình học sâu hiệu quả. Trong quá trình nghiên cứu và phát triển, chúng tôi đã sử dụng một số mô hình như Densenet121, Resnet50, Xception, Vision Transformer để thực hiện quá trình phân loại sau đó sử dụng Unet và DeeplabV3+ với các mạng trích xuất trên để thực hiện phân đoạn. Các mô hình này đã được chọn vì khả năng chính xác và khả năng học tập sâu. Mô hình đề xuất pha phân loại và phân đoạn như sau:

Hình 3.1. Kiến trúc đề xuất mô hình phân loại và phân đoạn



Mục tiêu chính của chúng tôi là nghiên cứu và tìm ra mô hình tối ưu nhất trong việc phát hiện và chẩn đoán bệnh tràn khí màng phổi. Quá trình này được chia thành hai pha chính: phân loại và phân đoạn. Trong pha phân loại, chúng tôi tập trung vào việc xác định xem liệu bệnh nhân có bị tràn khí màng phổi hay không, đồng thời xác định loại tràn khí màng phổi nếu có. Điều này đòi hỏi một mô hình phân loại chính xác và đáng tin cậy.

Sau đó, trong pha phân đoạn, chúng tôi tập trung vào việc khoanh vùng chính xác vị trí của tràn khí màng phổi trên hình ảnh y tế. Điều này giúp cho việc định vị và đánh dấu rõ ràng vùng bệnh để phục vụ cho quá trình chẩn đoán và điều trị.

Bằng cách kết hợp hai pha này lại với nhau, chúng tôi nhằm tạo ra một hệ thống toàn diện, giúp cải thiện khả năng phát hiện và khoanh vùng bệnh tràn khí màng phổi. Qua quá trình nghiên cứu và tối ưu hóa, chúng tôi hy vọng mô hình này sẽ đem lại

kết quả chính xác và đáng tin cậy, đồng thời hỗ trợ các chuyên gia y tế trong việc chẩn đoán và điều trị bệnh tràn khí màng phổi một cách hiệu quả.

3.4 Kịch bản và các tham số liên quan

Bảng 4.2. Thông số kịch bản huấn luyện

Kịch bản	Rút trích đặc trưng	Mạng phân vùng	Batch size	Learning rate	Số epoch
1	Densenet121	Unet	16	0.0001	150
2	Resnet50	DeeplabV3+	16	0.0001	150
3	Xception	Unet	16	0.0001	150
4	Vision Transformer	—	16	0.0001	150

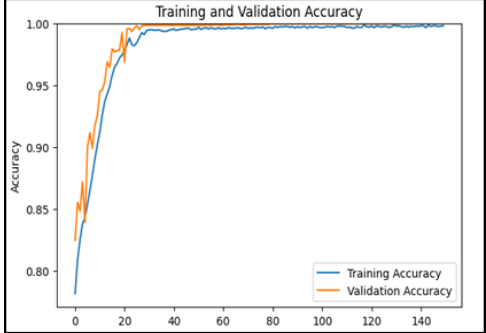
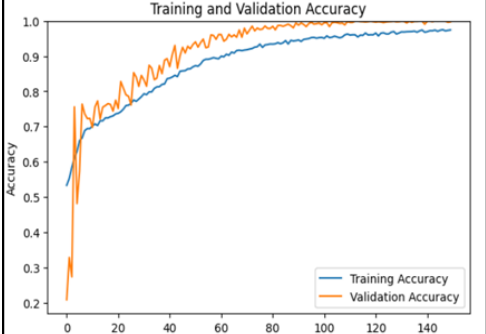
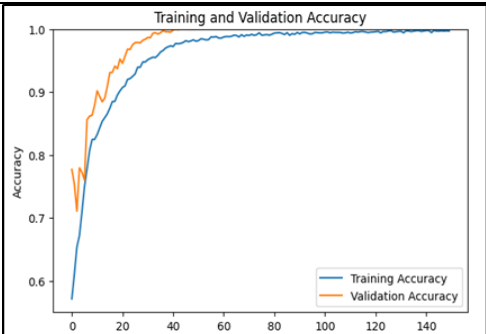

CHƯƠNG 4 - KẾT QUẢ

4.1 Kết quả huấn luyện

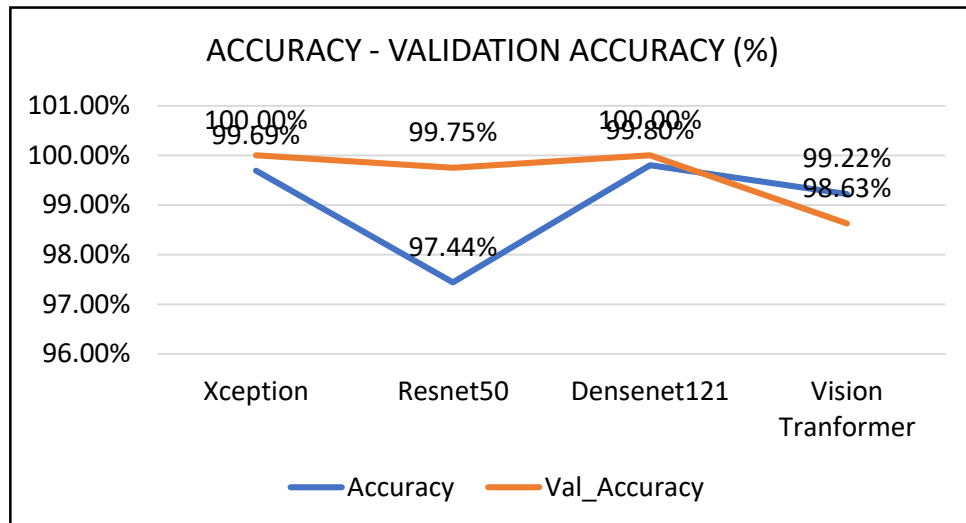
4.4.1 Kết quả huấn luyện pha phân loại

4.4.1.1 Độ chính xác

Bảng 4.3. Độ chính xác các mô hình pha phân loại

Kịch bản	Mô hình mạng	Biểu đồ độ chính xác
1	Densenet121	
2	Resnet50	
3	Xception	
4	Vision Transformer	

Sau khi huấn luyện qua 150 epochs thì ta thu được kết quả sau: Độ chính xác của các mô hình Densenet121, ResNet50, Xception, Vision Trasformer ở tập thăm định lần lượt đạt 100%, 99.75%, 100%, 98.63% . Mô hình Densenet121 và Xception đạt được kết quả tốt nhất với độ chính xác ở tập thăm định đạt 100% so với 2 mô hình còn lại.

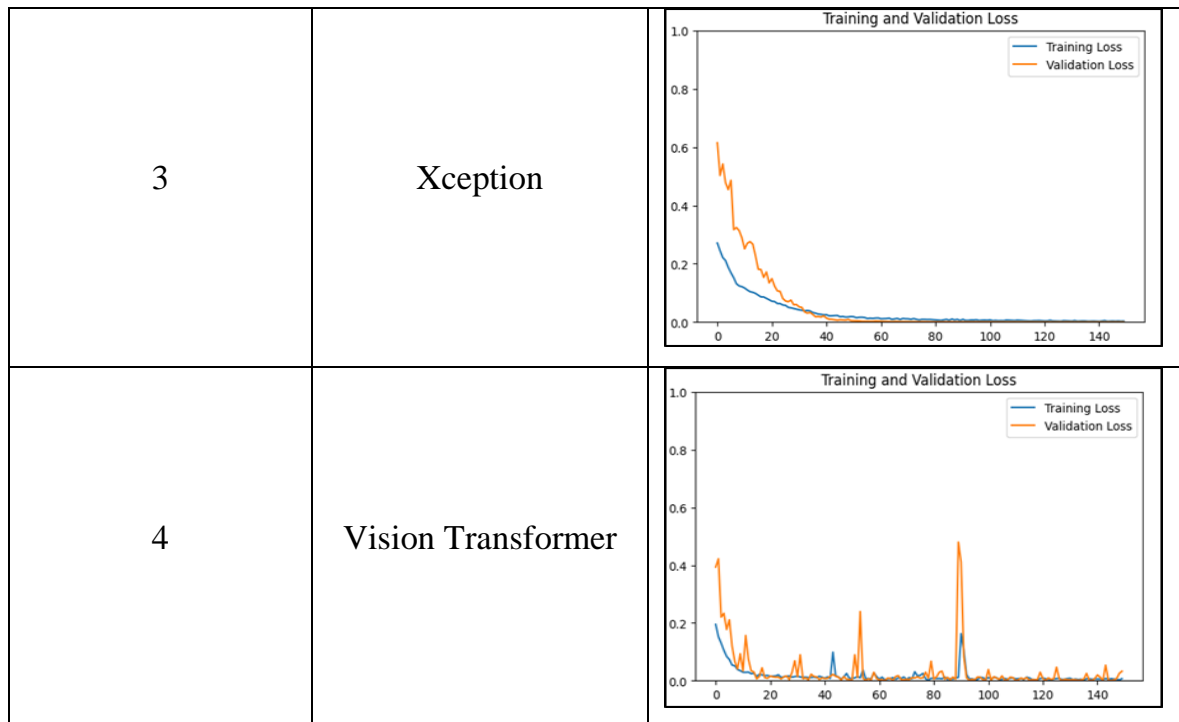


Hình 4.1. Biểu đồ so sánh độ chính xác các mô hình phân loại

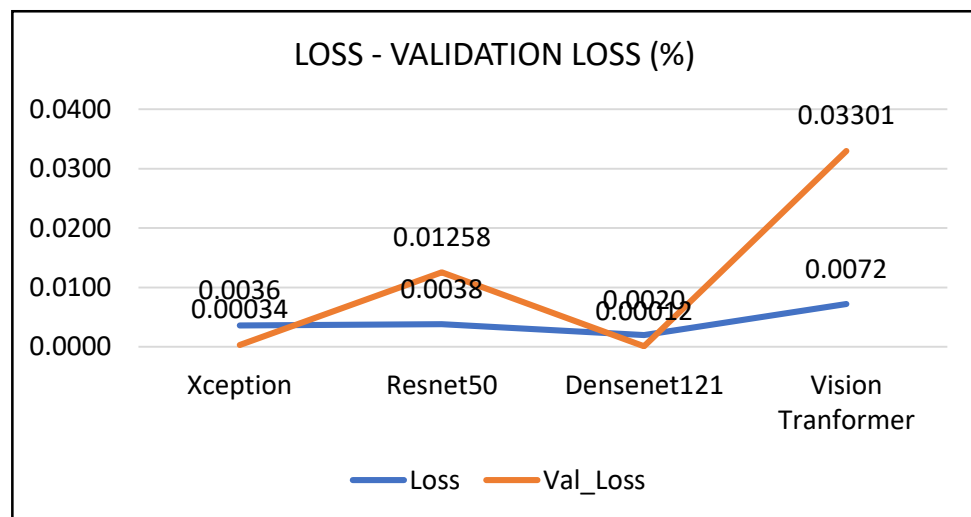
4.4.1.2 Độ mất mát

Bảng 4.4. Độ mất mát các mô hình pha phân loại

Kịch bản	Mô hình mạng	Biểu đồ độ mất mát
1	Densenet121	
2	Resnet50	



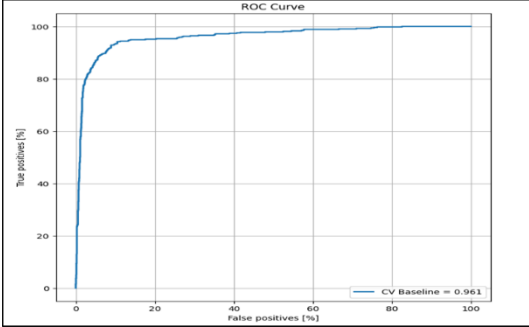
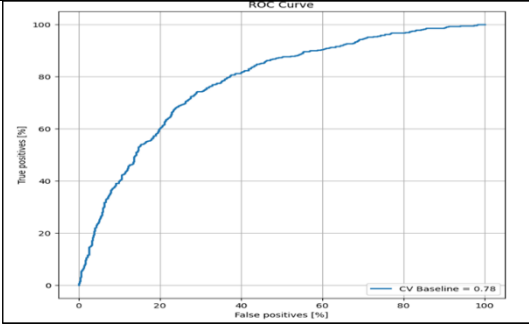
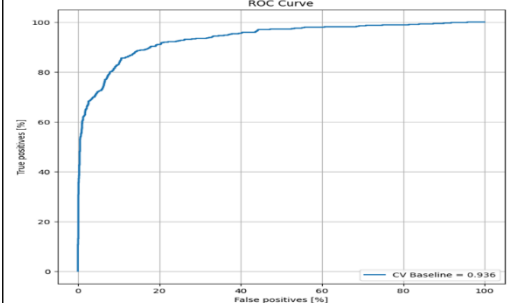
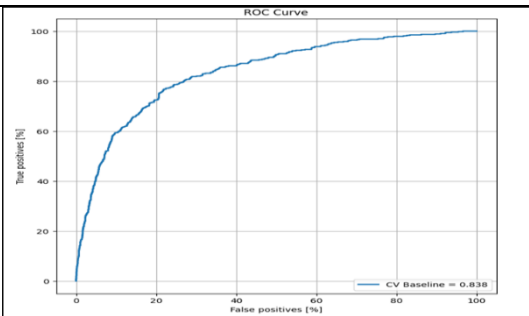
Sau khi huấn luyện qua 150 epochs thì ta thu được kết quả sau: Độ chính xác của các mô hình Densenet121, ResNet50, Xception, Vision Trasformer ở tập thăm định lần lượt đạt 0.00034, 0.01258, 0.00029, 0.03301 . Mô hình Xception đạt được độ mất mát thấp nhất với 0.00029 ở tập thăm định.



Hình 4.2. Biểu đồ so sánh độ mất mát các mô hình phân loại

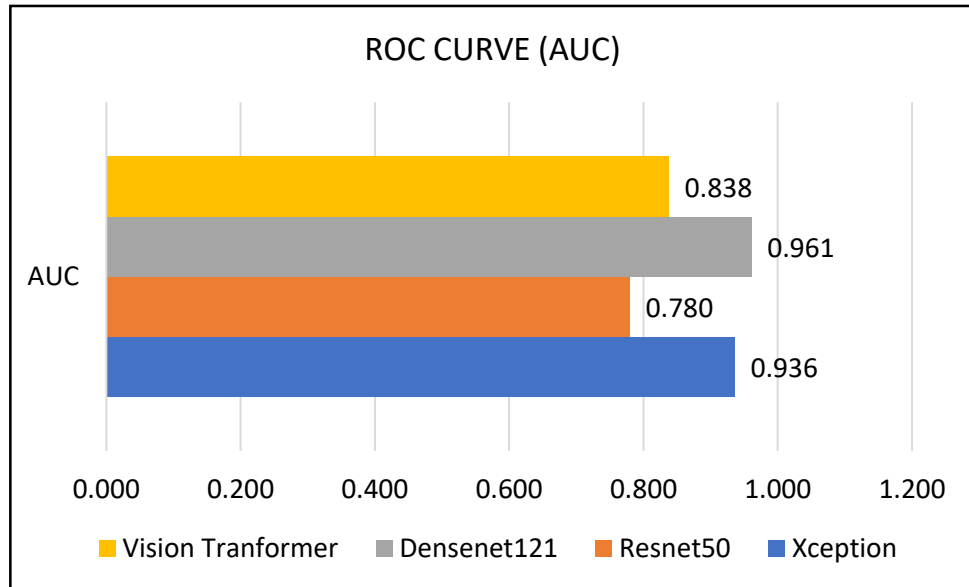
4.4.1.3 ROC-AUC

Bảng 4.5. ROC-AUC các mô hình pha phân loại

Kịch bản	Mô hình mạng	Biểu đồ ROC-AUC
1	Densenet121	 <p>The ROC Curve for Densenet121 shows a blue line representing the model's performance. The y-axis is 'True positives [%]' and the x-axis is 'False positives [%]'. The curve rises sharply from (0,0) to approximately (5, 80) and then levels off towards (100, 100). A legend indicates 'CV Baseline = 0.961'.</p>
2	Resnet50	 <p>The ROC Curve for Resnet50 shows a blue line representing the model's performance. The y-axis is 'True positives [%]' and the x-axis is 'False positives [%]'. The curve rises from (0,0) to approximately (100, 100) in a more gradual, S-shaped manner. A legend indicates 'CV Baseline = 0.78'.</p>
3	Xception	 <p>The ROC Curve for Xception shows a blue line representing the model's performance. The y-axis is 'True positives [%]' and the x-axis is 'False positives [%]'. The curve rises sharply from (0,0) to approximately (5, 60) and then levels off towards (100, 100). A legend indicates 'CV Baseline = 0.936'.</p>
4	Vision Transformer	 <p>The ROC Curve for Vision Transformer shows a blue line representing the model's performance. The y-axis is 'True positives [%]' and the x-axis is 'False positives [%]'. The curve rises from (0,0) to approximately (100, 100) in a more gradual, S-shaped manner. A legend indicates 'CV Baseline = 0.838'.</p>

Sử dụng phương pháp đánh giá mô hình bằng ROC-AUC sau khi huấn luyện qua 150 epochs thì ta thu được kết quả sau: AUC của các mô hình Densenet121, ResNet50, Xception, Vision Trasformer ở tập thử định lần lượt đạt 0.961, 0.78,

0.963, 0.838. Mô hình Densenet121 đạt được AUC cao nhất 0.961 cho thấy Densenet121 đạt được hiệu suất cao trong phân loại.



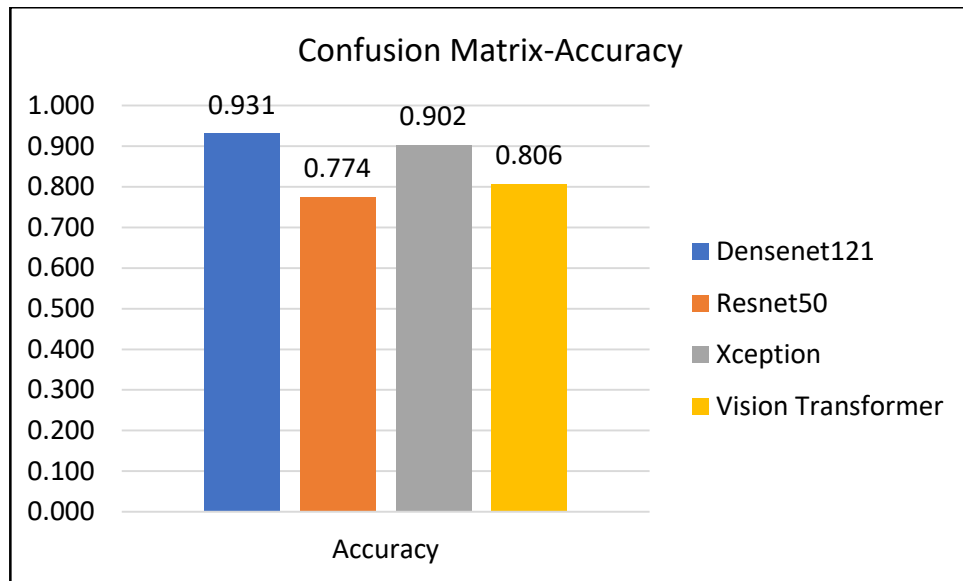
Hình 4.3. Biểu đồ so sánh AUC các mô hình phân loại

4.4.1.4 Confusion matrix

Bảng 4.6. Kết quả Confusion matrix các mô hình phân loại

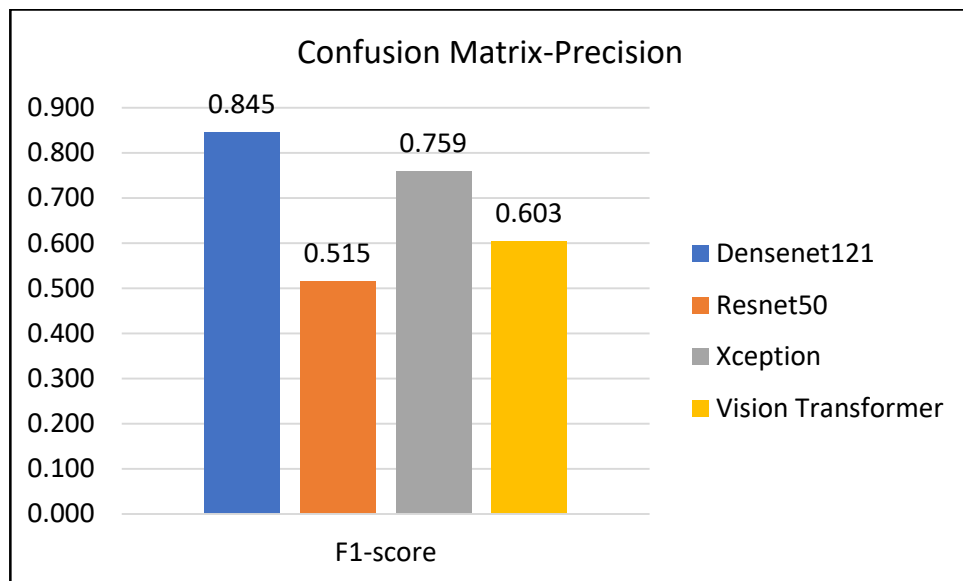
Tên mô hình	Confusion matrix							
	TN	FN	TP	FP	Accuracy	Precision	Recall	F1-score
Densenet121	1788	79	455	88	0.931	0.838	0.852	0.845
Resnet50	1575	243	291	301	0.774	0.491	0.545	0.517
Xception	1801	161	373	75	0.902	0.832	0.698	0.759
Vision Transformer	1587	178	356	289	0.806	0.551	0.667	0.603

Số lượng ảnh sử dụng để thẩm định bằng phương pháp Confusion matrix là 2410 ảnh với 534 ảnh bệnh và 1876 ảnh không bị bệnh ta được các kết quả *Bảng 4.6*. Với Densenet121 đạt được các thông số như: TN(True Negatives) đạt 1788 ảnh, FN(False Negatives) đạt 79 ảnh, TP(True Positives) đạt 455 ảnh và FP(False Positives) đạt 88 ảnh. Dựa vào các thông số trên ta có thể tính ra được các chỉ số quan trọng như: Accuracy ,Precision ,Recall ,F1-score.



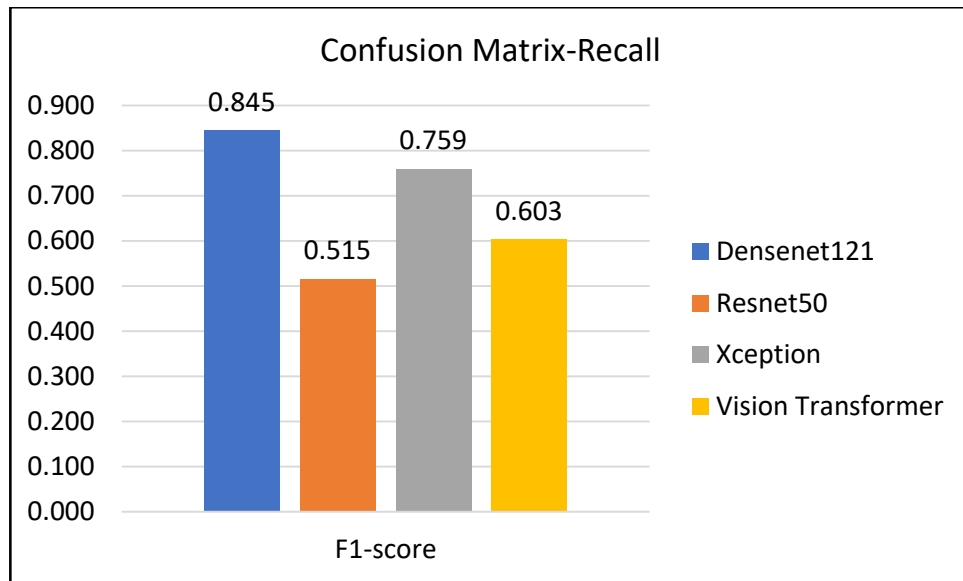
Hình 4.4. Biểu đồ so sánh Confusion matrix-Accuracy phân loại

Với Accuracy đạt được là 0.931 mô hình Densenet121 có chỉ số tốt hơn so với 3 mô hình Resnet50, Xception, Vision Transformer với chỉ số đạt lần lượt là: 0.774, 0.902, 0.806.



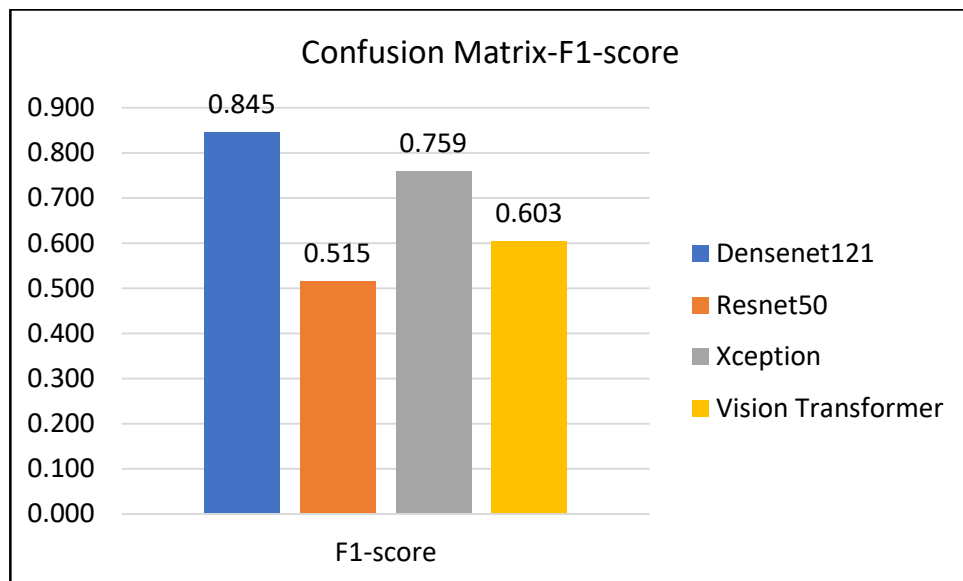
Hình 4.5. Biểu đồ so sánh Confusion matrix-Precision phân loại

Với Precision đạt được là 0.838 mô hình Densenet121 có chỉ số tốt hơn so với 3 mô hình Resnet50, Xception, Vision Transformer với chỉ số đạt lần lượt là: 0.491, 0.832, 0.551.



Hình 4.6. Biểu đồ so sánh Confusion matrix-Recall phân loại

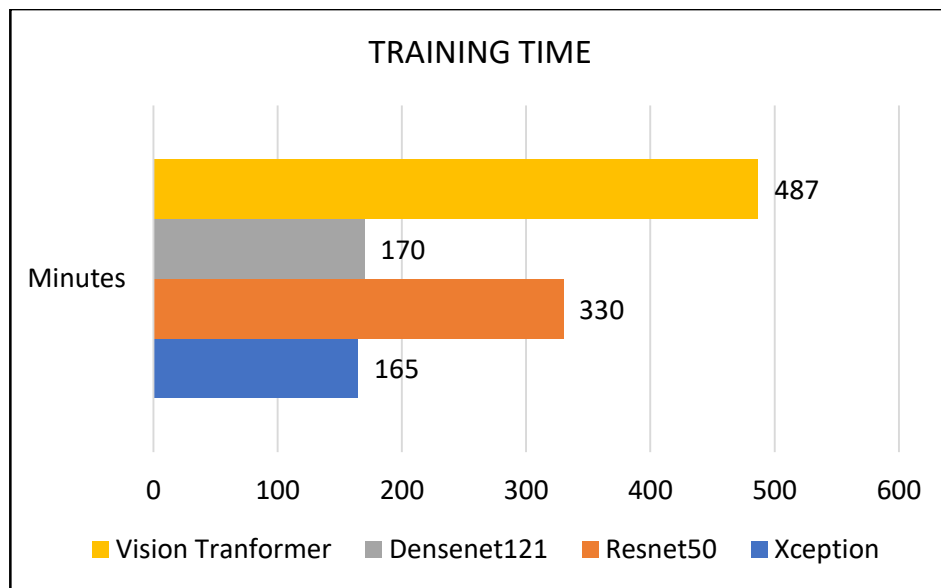
Với Recall đạt được là 0.852 mô hình Densenet121 có chỉ số tốt hơn so với 3 mô hình Resnet50, Xception, Vision Transformer với chỉ số đạt lần lượt là: 0.545, 0.698, 0.667.



Hình 4.7. Biểu đồ so sánh Confusion matrix-F1-score phân loại

Với F1-score đạt được là 0.845 mô hình Densenet121 có chỉ số tốt hơn so với 3 mô hình Resnet50, Xception, Vision Transformer với chỉ số đạt lần lượt là: 0.515, 0.759, 0.603.

4.4.1.5 Thời gian huấn luyện



Hình 4.8. Biểu đồ so sánh thời gian huấn luyện các mô hình phân loại

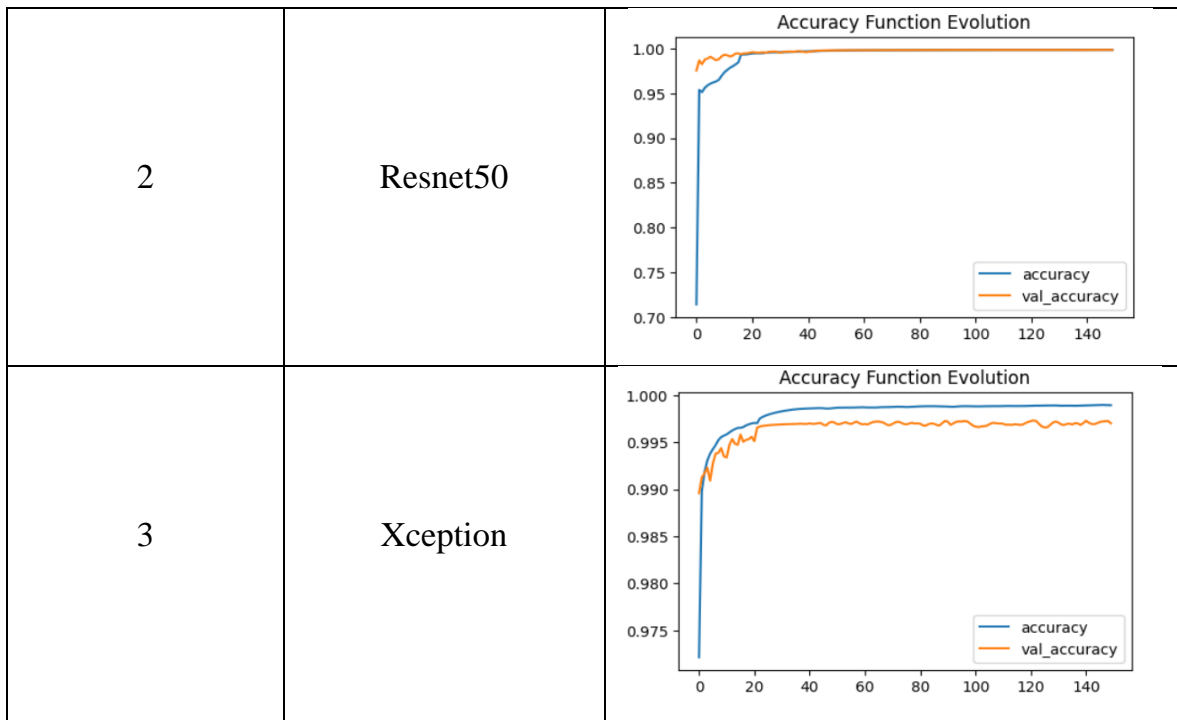
Thời gian huấn luyện lần lượt ở các mô hình Densenet121, Resnet50, Xception, Visio Transformer lần lượt là 170 phút, 330 phút, 165 phút, 487 phút. Xception là mô hình có thời gian huấn luyện nhanh nhất.

4.4.2 Kết quả huấn luyện pha phân đoạn

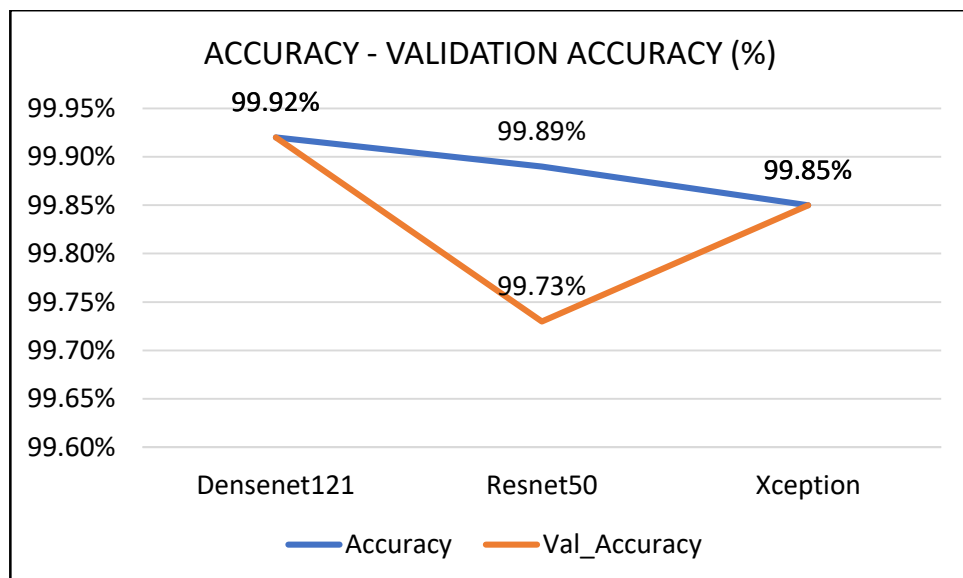
4.4.2.1 Độ chính xác

Bảng 4.7. Độ chính xác các mô hình pha phân đoạn

Kịch bản	Mô hình mạng	Biểu đồ độ chính xác
1	Densenet121	<p>The graph shows the evolution of accuracy and validation accuracy over 140 epochs. The y-axis represents accuracy values from 0.975 to 1.000. The x-axis represents epochs from 0 to 140. The blue line (accuracy) and orange line (val_accuracy) both start at approximately 0.975 at epoch 0. They rise sharply, with some initial fluctuations, reaching a plateau of approximately 0.998 by epoch 20. From epoch 20 to 140, both lines remain very close to each other and near the 1.000 mark, indicating high and stable performance.</p>



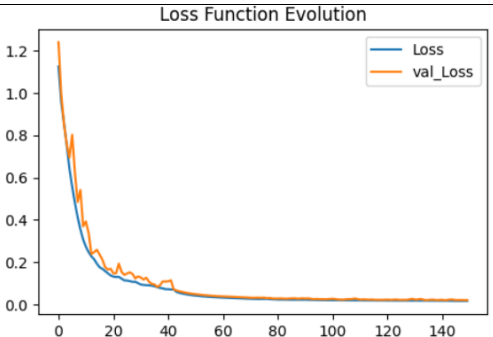
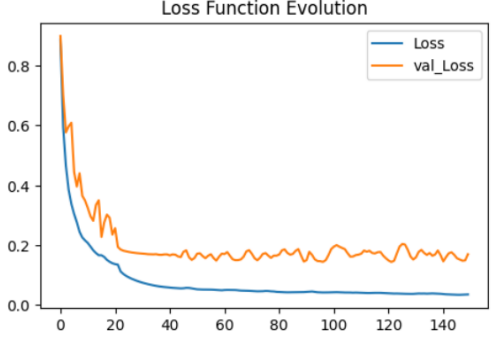
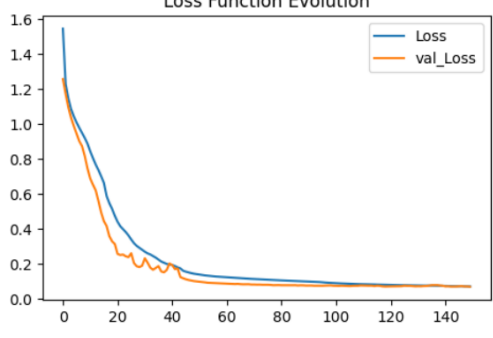
Sau khi huấn luyện qua 150 epochs thì ta thu được kết quả sau: Độ chính xác của các mô hình Densenet121 với mạng phân vùng Unet, ResNet50 với mạng phân vùng DeeplabV3+, Xception với mạng phân vùng Unet, khi phân đoạn ở tập thử định lần lượt đạt 99.92%, 99.73%, 99.85%. Mô hình Densenet121 với mạng phân vùng Unet đạt được kết quả tốt nhất với độ chính xác ở tập thử định đạt 99.92% so với 2 mô hình còn lại.



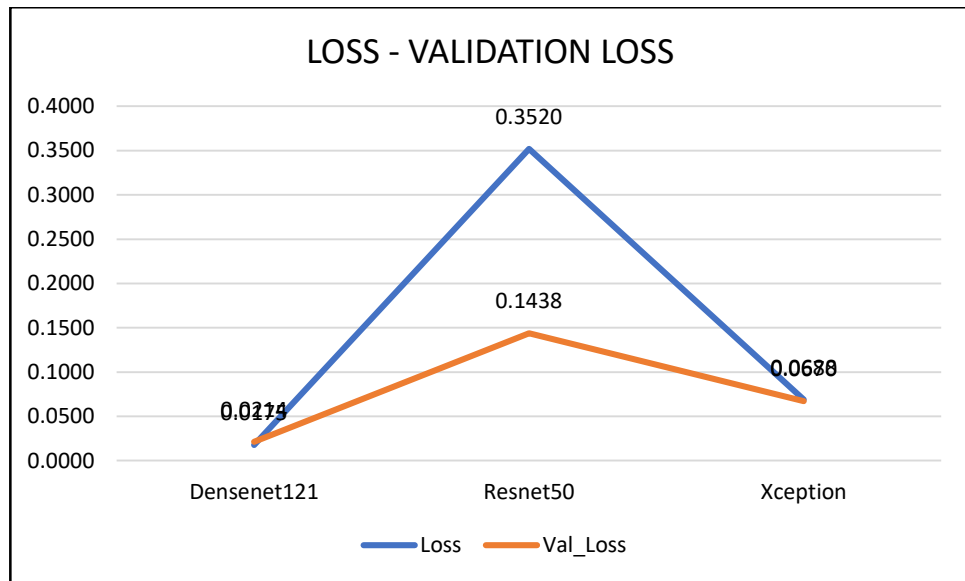
Hình 4.9. Biểu đồ so sánh độ chính xác các mô hình phân đoạn

4.4.2.2 Độ mất mát

Bảng 4.8. Độ mất mát các mô hình phân đoạn

Kịch bản	Mô hình mạng	Biểu đồ độ mất mát
1	Densenet121	
2	Resnet50	
3	Xception	

Sau khi huấn luyện qua 150 epochs thì ta thu được kết quả: độ mất mát của các mô hình Densenet121 với mạng phân vùng Unet, Resnet50 với mạng phân vùng DeeplabV3+, Xception với mạng phân vùng Unet, khi phân đoạn ở tập thử định lần lượt đạt 0.0214, 0.1438, 0.0670. Mô hình Densenet121 với mạng phân vùng Unet đạt được kết quả tốt nhất với độ mất mát ở tập thử định đạt 0.0214 so với 2 mô hình còn lại.

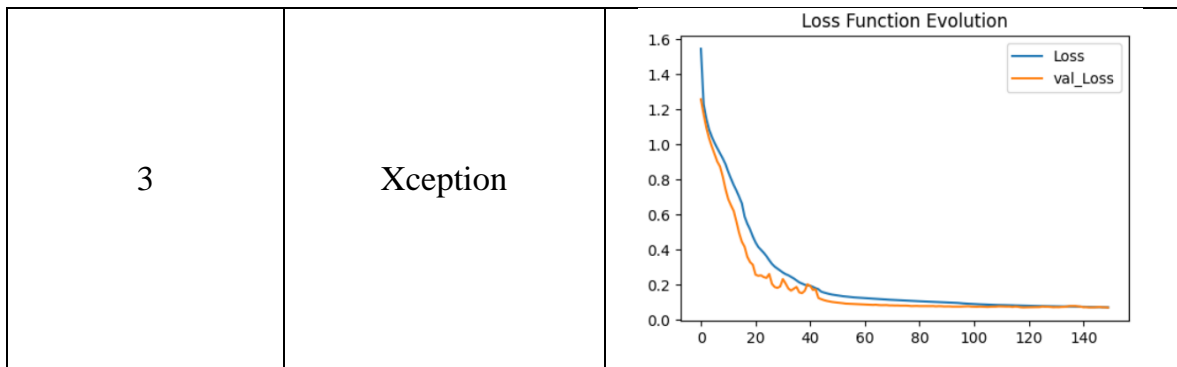


Hình 4.10. Biểu đồ so sánh độ mất mát các mô hình phân đoạn

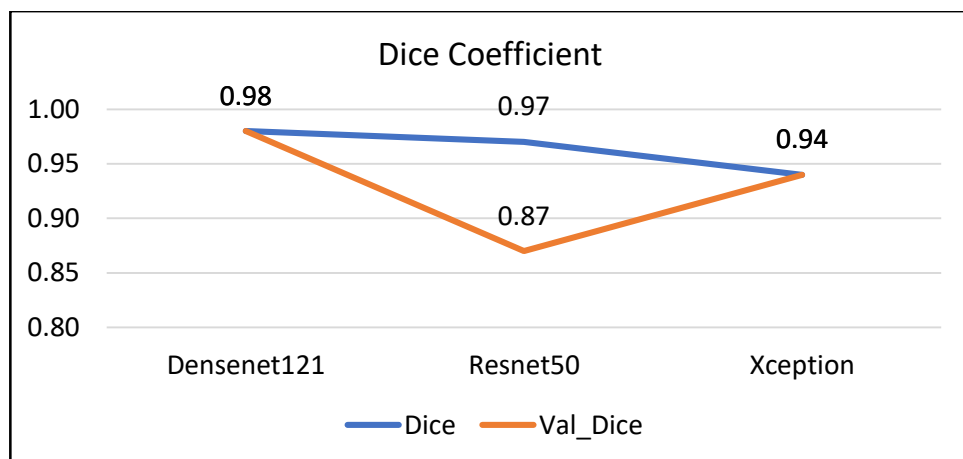
4.4.2.3 Dice Coefficient

Bảng 4.9. Dice coefficient các mô hình phân đoạn

Kịch bản	Mô hình mạng	Biểu đồ dice coefficient
1	Densenet121	
2	Resnet50	

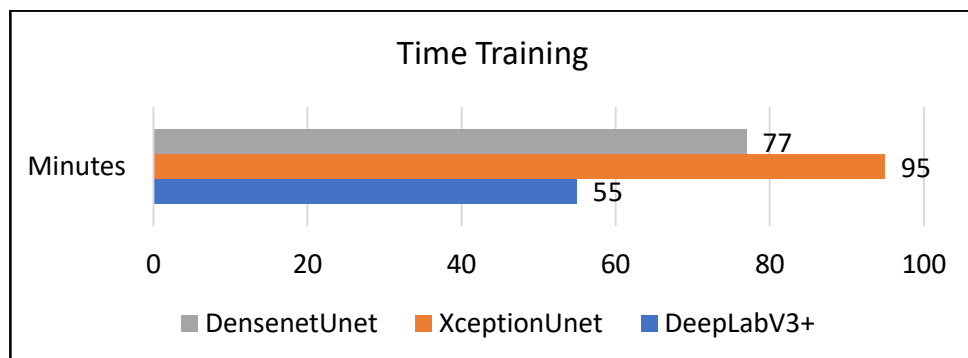


Sau khi huấn luyện qua 150 epochs thì ta thu được kết quả: Dice của các mô hình Densenet121 với mạng phân vùng Unet, Resnet50 với mạng phân vùng DeeplabV3+, Xception với mạng phân vùng Unet, khi phân đoạn ở tập thử định lần lượt đạt 0.98, 0.87, 0.94. Mô hình Densenet121 với mạng phân vùng Unet đạt được kết quả tốt nhất với Dice ở tập thử định đạt 0.98 so với 2 mô hình còn lại.



Hình 4.11. Biểu đồ so sánh Dice coefficient các mô hình phân đoạn

4.4.2.4 Thời gian huấn luyện



Hình 4.12. Biểu đồ so sánh thời gian huấn luyện các mô hình phân đoạn

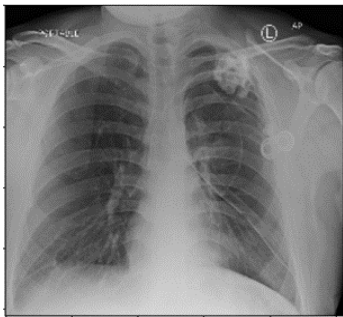
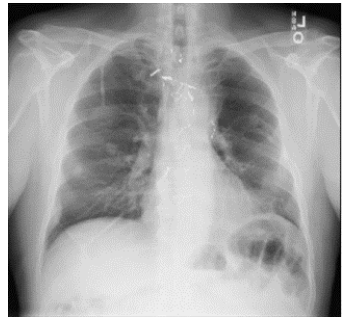
Thời gian huấn luyện lần lượt ở các mô hình Densenet121 với mạng phân vùng Unet, Resnet50 với mạng phân vùng DeeplabV3+, Xception với mạng phân vùng

Unet lần lượt là 77 phút, 95 phút, 55 phút. Resnet50 với mạng phân vùng DeeplabV3+ là mô hình có thời gian huấn luyện nhanh nhất.

4.2 Kết quả kiểm thử

4.2.1 Phân loại tràn khí màng phổi

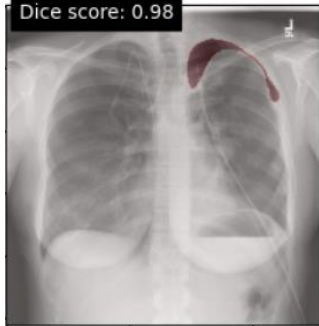
Bảng 4.10. Kết quả kiểm thử các mô hình phân loại

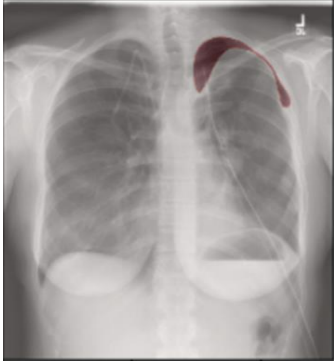

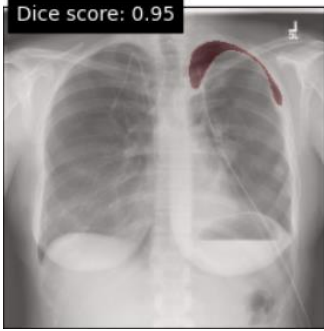
<div> <div>Ảnh kiểm thử</div> <div>Mô hình</div> </div>	Pneumothorax	No Pneumothorax
		
Densenet121	Pneumothorax Found	No Pneumothorax
Resnet50	No Pneumothorax	No Pneumothorax
Xception	Pneumothorax Found	No Pneumothorax
Vision Transformer	No Pneumothorax	No Pneumothorax

Với mô hình Densenet121 và Xception cho ra kết quả chính xác ở cả hai ảnh kiểm thử. Resnet50 và Vision Transformer cho kết quả sai ở ảnh Pneumothorax

4.2.2 Phân vùng tràn khí màng phổi

Bảng 4.11. Kết quả kiểm thử các mô hình phân đoạn

Mô hình	Mặt nạ thật	Mặt nạ dự đoán	Dice coefficient
Densenet121			98%

Resnet50			88%
Xception			95%

Sau quá trình kiểm thử, ta thu được kết quả sau: Tất cả 3 mô hình đều cho kết quả kiểm thử khá tốt. Trong đó, mô hình Densenet121 với mạng phân vùng Unet đưa ra mặt nạ dự đoán có độ tương đồng so với mặt nạ thật cao nhất với Dice coefficient đạt 98%. Mô hình Resnet50 với mạng phân vùng Deeplabv+ đưa ra mặt nạ dự đoán có độ tương đồng so với mặt nạ thật thấp nhất với Dice coefficient đạt 88%.

CHƯƠNG 5: KẾT LUẬN

5.1 Kết quả đạt được

Trong lúc thực hiện, tôi cũng đã tiếp thu thêm được nhiều kiến thức bổ ích thông qua những lúc thầy góp ý hàng tuần. Qua thực nghiệm đã thực hiện trong đề tài có một số đóng góp như sau:

- Tập hợp dữ liệu về tràn khí màng phổi từ nguồn dữ liệu Kaggle
- Tiến hành tiền xử lý ảnh qua việc lật ảnh và xoay ảnh để có thể tạo ra bộ dữ liệu mới nhằm tăng sự đa dạng cho dữ liệu
- Cung cấp các kết quả thực nghiệm, đánh giá các mô hình phân loại Densenet, Xception, Resnet50, Vision Transformer dựa trên phương pháp đánh giá ROC và Confusion Matrix
- Xây dựng thành công các mô hình phân đoạn Resnet50-DeepLabV3+, Densenet121-Unet, Xception-Unet với khả năng phân vùng bệnh với độ chính xác cao
- So sánh đánh giá các mô hình phân loại và phân đoạn
- Thực hiện cải tiến mô hình phân đoạn với phương pháp Transfer Learning

5.2 Hạn chế của đề tài

Bên cạnh những mặt đã đạt được thì có một số hạn chế còn gặp phải. Vì lượng kiến thức còn hạn hẹp nên thực hiện đề tài cũng còn thiếu một số ý kiến, không thể hoàn thành tối đa những mục tiêu đã đặt ra. Một số hạn chế có thể kể đến như:

- Tập dữ liệu còn hạn chế, chưa kết hợp được với nhiều tập dữ liệu khác
- Chưa xây dựng được mô hình phân đoạn với mạng trích xuất là Vision Transformer
- Mô hình phân loại chưa đạt được độ chính xác cần thiết

5.2 Hướng phát triển

- Tìm kiếm thu thập thêm nhiều dữ liệu mới
- Tìm hiểu và sử dụng thêm nhiều mô hình mạng mới
- Áp dụng thêm một số phương pháp cải tiến khác nhằm tối ưu mô hình.

TÀI LIỆU KHAM KHẢO

- [1] Trần Thị Kim Ngọc (2021), “*Tràn khí màng phổi tự phát: nguyên nhân và phương pháp điều trị*”, Bệnh viện đa khoa MEDLATEC, Truy cập vào ngày 25/5/2023 từ <https://medlatec.vn/tin-tuc/tran-khi-mang-phoi-tu-phat-nguyen-nhan-va-phuong-phap-dieu-tri-s64-n22274>. (bài báo)
- [2] Trần Thị Kim Ngọc (2021), “*Tràn khí màng phổi tự phát: dấu hiệu nhận biết, biến chứng và phương pháp điều trị*”, Bệnh viện đa khoa MEDLATEC, Truy cập vào ngày 25/5/2023 từ <https://medlatec.vn/tin-tuc/tran-khi-mang-phoi-dau-hieu-nhan-biet-bien-chung-va-cach-dieu-tri-s64-n21724>. (Bài báo)
- [3] Richard W. Light(2021), “*Tràn khí màng phổi*” ,MSD MANUAL, Truy cập vào 25/5/2023 từ <https://www.msmanuals.com/vi-vn/chuyen-gia/roi-loan-chuc-nang-ho-hap/viem-mang-phoi/tran-kho-mang-phoi>. (Bài báo)
- [4] Trần Thị Diễm Trang (2020), “*Nguyên nhân và chẩn đoán tràn khí màng phổi*”, VINMEC, Truy cập vào 25/5/2023 từ <https://www.vinmec.com/vi/tin-tuc/thong-tin-suc-khoe/suc-khoe-tong-quat/nguyen-nhan-va-chan-doan-tran-khi-mang-phoi>. (Bài báo)
- [5] VBD (2022), “*Tổng quan về Vision Transformer (ViT)*”, VinBigData, Truy cập vào 8/8/2023 từ [https://vinbigdata.com/camera-ai/tong-quan-ve-vision-transformer vit.ht](https://vinbigdata.com/camera-ai/tong-quan-ve-vision-transformer-vit.ht) . (Bài báo)
- [6] “*Thuật toán CNN là gì? Cấu trúc mạng Convolutional Neural Network*” ,TopDeV, Truy cập vào 25/5/2023 từ <https://topdev.vn/blog/thuat-toan-cnn-convolutional-neural-network/>.(Bài báo)
- [7] “*Top 10 thuật toán machine learning dành cho newbie*”, TopDEV. Truy cập vào 25/5/2023 từ <https://topdev.vn/blog/top-10-thuat-toan-machine-learning-danh-cho-newbie/>.(Bài báo)
- [8] Phan Anh Cang (2021), *Giáo trình Trí tuệ nhân tạo*, Trường Đại học Sư phạm Kỹ thuật Vĩnh Long.(Sách)
- [9] *SIIM-ACR Pneumothorax Segmentation Identify Pneumothorax disease in chest x-rays* (2019), Society for Imaging Informatics in Medicine(SIIM), Truy cập vào 26/4/2023 từ <https://www.kaggle.com/c/siim-acr-pneumothorax-segmentation>

- [10] “*Convolutional Neural Networks cheatsheet*”,Stanford. Truy cập vào 26/5/2023 từ <https://stanford.edu/~shervine/teaching/cheatsheet-convolutional-neural-networks>. (Bài báo)
- [11]Jame Le (2021),”*How to do Semantic Segmentation using Deep learning,Nanonets*”, Truy cập vào 27/5/2023 từ <https://nanonets.com/blog/how-to-do-semantic-segmentation-using-deep-learning>. (Bài báo)
- [12]Alexey Tolkachev, Ilyas Sirazitdinov, Maksym Kholiavchenko, Tamerlan Mustafaev, Bulat Ibragimov (2020), *Deep Learning for Diagnosis and Segmentation of Pneumothorax: The Results on the Kaggle Competition and Validation Against Radiologists*.(Luận văn)
- [13]André Gooßen, Hrishikesh Deshpande, Tim Harder, Evan Schwab, Ivo Baltruschat, Thusitha Mabotuwana, Nathan Cross, Axel Saalbach (2019), *Deep Learning for Pneumothorax Detection and Localization in Chest Radiographs*.(Luận văn)
- [14]Xiao Han,Jing Zhang,Minghui Wang,Jianhui He,Yuqi Fang,Jun Lan,Sen Yang,Xiyue Wang (2020), *Automatic Segmentation of Pneumothorax in Chest Radiographs Based on a Two-Stage Deep Learning Method* .(Luận văn)
- [15]Hongyu Wang,Hong Gu,Pan Qin,Jia Wang (2020), *CheXLocNet: Automatic localization of pneumothorax in chest radiographs using deep convolutional neural networks*.(Luận văn)
- [16]Crosby, Jennie,Rhines,Thomas , Li, Feng , MacMahon, Heber , Giger, Maryellen (2020),*Deep learning for pneumothorax detection and localization using networks fine-tuned with multiple institutional datasets*.(Luận văn)
- [17]Yongil Cho, Jong Soo Kim, Tae Ho Lim, Inhye Lee, Jongbong Choi (2021), *Detection of the location of pneumothorax in chest X-rays using small artificial neural networks and a simple training process*.(Luận văn)
- [18]Pulkit Sharm (2020), “*Computer Vision Tutorial: A Step-by-Step Introduction to Image Segmentation Techniques*” ,Analytics Vidhya,Truy cập vào 27/5/2023 từ <https://www.analyticsvidhya.com/blog/2019/04/introduction-image-segmentation-techniques-python/>.(Bài báo)

- [19] “*Semantic Segmentation Using Deep Learning*”, MathWorks, Truy cập vào 28/5/2023 từ <https://www.mathworks.com/help/vision/ug/semantic-segmentation-using-deep-learning.html>. (Bài báo)
- [20] Michael A. Nielsen (2015) ,*Neural Network and Deep Learning*. (Sách)
- [21] Adrian Rosebrock (31/2018) ,*Keras Conv2D and Convolutional Layers*.(Sách)
- [22] Waseem Rawat and Zenghui Wang (9/6/2017) *Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review* .(Luận văn)
- [23] Olaf Ronneberger, Philipp Fischer và Thomas Brox (2015) ,*U-Net: Convolutional Networks for Biomedical Image Segmentation*. (Luận văn)
- [24] Kaiming He, Xiangyu Zhang, Shaoqing Ren và Jian Sun (2016),*Deep Residual Learning for Image Recognition*. (Luận văn)
- [25] Aakash Kaushik (2020), “*Understanding ResNet50 architecture*”, OpenGenus. Truy cập vào 27/6/2023 từ <https://iq.opengenus.org/resnet50-architecture>.(Bài báo)
- [26] Gao Huang, Zhuang Liu, Laurens van der Maaten và Kilian Q. Weinberger (2017), *Densely Connected Convolutional Networks*. (Luận văn)
- [27] “*Architecture of DenseNet-121*”, IQ, truy cập vào 28/6/2023 từ <https://iq.opengenus.org/architecture-of-densenet121/>.(Bài báo)
- [28] “*Xception Model and Depthwise Separable Convolutions*”, MAEL, truy cập vào 28/6/2023 từ <https://maelfabien.github.io/deeplearning/xception>. (Bài báo)
- [29] “*The Annotated ResNet-50*”, Toward data Science, truy cập vào 28/6/2023 từ <https://medium.com/m/global-identity-towardsdatascience.com%2Fthe-annotated-resnet-50>. (Bài báo)
- [30] “*UNet — Line by Line Explanation*”, Toward Data Science, Truy cập vào 26/6/2023 từ <https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5>. (Bài báo)
- [31] Phạm Đình Khánh (2020), “*Bài 41 - DeepLab Sentiment Segmentation*”, Truy cập vào 26/6/2023 từ <https://phamdinhhkhanh.github.io/2020/06/18/DeepLab.html>. (Bài báo)

- [32]Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, Neil Houlsby (2020), “*An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*”, Truy cập vào 26/6/2023 từ <https://paperswithcode.com/method/vision-transformer>. (Bài báo)
- [33]Nguyễn Hoàng Nam (2021). “*Tìm hiểu chi tiết về AUC - ROC trong Machine Learning?*”, Truy cập vào 26/6/2023 từ <https://www.noron.vn/post/tim-hieu-chi-tiet-ve-auc---roc-trong-machine-learning-1fz9nhqo5ut>. (Bài báo)
- [34]Minh Vũ (2020),”*Tìm hiểu về Confusion matrix trong Machine Learning?*”, Truy cập vào 26/6/2023 từ <https://viblo.asia/p/tim-hieu-ve-confusion-matrix-trong-machine-learning-Az45bRpo5xY>. (Bài báo)
- [35]Pratik Bhavsar(2019), “*An Ultimate Guide To Transfer Learning In NLP*”, Truy cập vào 26/6/2023 từ <https://www.topbots.com/transfer-learning-in-nlp/>. (Bài báo)