

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA TOÁN - CƠ - TIN HỌC



BÁO CÁO CUỐI KÌ
PHÁT HIỆN VIÊM PHỔI TỪ ẢNH X-QUANG

Nguyễn Hữu An
Nguyễn Văn Dũng
Nguyễn Tiến Đạt
Vũ Đức Quý

Mã học phần: MAT3508
Học kỳ 1, Năm học 2025-2026

Hà Nội, 2025

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA TOÁN - CƠ - TIN HỌC

BÁO CÁO CUỐI KÌ
PHÁT HIỆN VIÊM PHỔI TỪ ẢNH X-QUANG

Nguyễn Hữu An - 23001493
Nguyễn Văn Dũng - 23001515
Nguyễn Tiến Đạt - 23001508
Vũ Đức Quý - 23001553

Giảng viên bộ môn:
TS. HOÀNG ANH ĐỨC

Hà Nội, 2025

THÔNG TIN

- **Môn học:** MAT3508 - Nhập môn Trí tuệ Nhân tạo
- **Học kỳ:** Học kì 1 - 2025-2026
- **Trường:** VNU-HUS (Đại học Quốc gia Hà Nội - Trường Đại học Khoa học Tự nhiên)
- **Tiêu đề:** Phát hiện Viêm phổi từ ảnh X-quang
- **Ngày nộp:** 30/11/2025
- **Báo cáo PDF:** Liên kết báo cáo
- **Slide thuyết trình:** Liên kết Slide
- **Kho lưu trữ:** Source Code

Họ và tên	Mã SV	GitHub	Đóng góp
Nguyễn Hữu An	23001493	ankkun13	Xây dựng Demo, Xây dựng mô hình
Nguyễn Tiến Đạt	23001515	tiendat228	Thiết kế Slide, Làm báo cáo
Nguyễn Văn Dũng	23001508	NguyenDung2301	Xây dựng mô hình, Xây dựng Demo
Vũ Đức Quý	23001553	VUDUCQUY	Xây dựng mô hình, Thiết kế Slide

MỤC LỤC

THÔNG TIN	i
MỤC LỤC.....	ii
DANH SÁCH HÌNH ẢNH	iv
DANH SÁCH BẢNG	v
DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT	vi
MỞ ĐẦU	viii
CHƯƠNG 1. CƠ SỞ LÝ THUYẾT	1
1.1. Cơ sở Y Sinh và Bối cảnh thực tế	1
1.1.1. Tổng quan về bệnh Viêm phổi.....	1
1.1.2. Thực trạng và Thách thức	2
1.2. Tổng quan về Học sâu và Mạng Nơ-ron Tích chập	2
1.2.1. Học sâu.....	2
1.2.2. Mạng Nơ-ron Tích chập	3
1.3. Các kiến trúc CNN phổ biến	5
1.3.1. VGGNet	5
1.3.2. ResNet	6
1.4. Kiến trúc DenseNet	7
1.4.1. Giới thiệu	7
1.4.2. Thành phần kiến trúc chính	8
1.4.3. Ưu điểm	9
1.4.4. Hạn chế.....	9
1.5. Kỹ thuật huấn luyện mô hình	9
1.5.1. Học chuyển tiếp.....	9
1.5.2. Hàm mất mát	10
1.5.3. Thuật toán tối ưu hóa	10
CHƯƠNG 2. XÂY DỰNG MÔ HÌNH VÀ TRIỂN KHAI	11
2.1. Mô tả tập dữ liệu	11
2.1.1. Đặc điểm và phân phối.....	11
2.2. Tiền xử lý dữ liệu.....	12
2.2.1. Kỹ thuật biến đổi và Chuẩn hóa ảnh	12
2.2.2. Xử lý mất cân bằng dữ liệu	13
2.2.3. Tổ chức và Tải dữ liệu	14

2.3. Xây dựng mô hình	14
2.3.1. Tăng cường dữ liệu	17
2.3.2. Cấu hình phần cứng và Tối ưu hóa.....	18
CHƯƠNG 3. THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ	20
3.1. Thiết lập môi trường và cấu hình thực nghiệm	20
3.1.1. Phần cứng và phần mềm	20
3.1.2. Tối ưu hóa và Thiết lập siêu tham số.....	20
3.1.3. Quá trình huấn luyện và Chống hiện tượng quá khớp	21
3.2. Tiêu chí đánh giá	22
3.3. Kết quả huấn luyện	24
3.4. Ưu điểm và hạn chế	28
3.4.1. Ưu điểm	28
3.4.2. Hạn chế và Thách thức.....	28
3.5. Hướng phát triển	29
3.5.1. Nâng cao chất lượng đầu vào và Kiến trúc mô hình.....	29
3.5.2. Mở rộng phạm vi chẩn đoán	29
3.5.3. Tăng cường khả năng tổng quát hóa	30
3.5.4. Triển khai ứng dụng thực tế	30
KẾT LUẬN	31
TÀI LIỆU THAM KHẢO	33

DANH SÁCH HÌNH ẢNH

1.1	<i>Ảnh viêm phổi do COVID-19.</i>	1
1.2	<i>Cấu trúc mạng nơ-ron nhân tạo.</i>	3
1.3	<i>Cấu trúc mạng nơ-ron tích chập (CNN).</i>	4
1.4	<i>Cấu trúc VGGNet.</i>	5
1.5	<i>Kiến trúc mạng ResNet.</i>	6
1.6	<i>Kiến trúc mạng DenseNet.</i>	7
2.1	<i>Biểu đồ phân bố sự mất cân bằng dữ liệu.</i>	11
2.2	<i>Một số hình ảnh từ tập dữ liệu.</i>	12
2.3	<i>Kiến trúc mạng DenseNet.</i>	15
2.4	<i>Cấu trúc DenseBlock.</i>	16
2.5	<i>Cấu trúc Transition Layer.</i>	16
2.6	<i>Minh họa kỹ thuật tăng cường dữ liệu: Ảnh gốc và các biến thể qua phép xoay, lật và chỉnh tương phản.</i>	17
3.1	<i>Quá trình huấn luyện.</i>	24
3.2	<i>Ma trận nhầm lẫn.</i>	25
3.3	<i>Biểu đồ đường cong ROC.</i>	26

DANH SÁCH BẢNG

3.1	Cấu hình siêu tham số cho quá trình huấn luyện	21
3.2	So sánh hiệu năng giữa mô hình DenseNet và các mô hình khác trên tập kiểm thử.....	27

DANH MỤC THUẬT NGỮ VÀ TỪ VIẾT TẮT

Từ viết tắt	Tiếng Anh (Gốc)	Giải thích (Tiếng Việt)
AI	Artificial Intelligence	Trí tuệ nhân tạo
ANN	Artificial Neural Network	Mạng nơ-ron nhân tạo
AUC	Area Under the Curve	Diện tích dưới đường cong ROC
CAD	Computer-Aided Diagnosis	Hệ thống hỗ trợ chẩn đoán
CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
CPU	Central Processing Unit	Bộ xử lý trung tâm
CXR	Chest X-Ray	Ảnh X-quang lồng ngực
DICOM	Digital Imaging and Communications in Medicine	Tiêu chuẩn hình ảnh và truyền thông kỹ thuật số trong y tế
FN	False Negative	Âm tính giả (Bỏ sót bệnh)
FP	False Positive	Đương tính giả (Báo động giả)
GDPR	General Data Protection Regulation	Quy định bảo vệ dữ liệu chung
GPU	Graphics Processing Unit	Bộ xử lý đồ họa
HIPAA	Health Insurance Portability and Accountability Act	Đạo luật về Trách nhiệm giải trình và Cung cấp bảo hiểm y tế
PACS	Picture Archiving and Communication System	Hệ thống lưu trữ và truyền hình ảnh y tế
RAM	Random Access Memory	Bộ nhớ truy cập ngẫu nhiên
ReLU	Rectified Linear Unit	Đơn vị chỉnh lưu tuyến tính (Hàm kích hoạt)
ResNet	Residual Networks	Mạng nơ-ron thăng dư
ROC	Receiver Operating Characteristic	Đường cong đặc trưng hoạt động của bộ thu
SGD	Stochastic Gradient Descent	Tối ưu hóa Gradient Descent ngẫu nhiên
SOTA	State-of-the-Art	Các phương pháp hiện đại nhất
SVM	Support Vector Machine	Máy vector hỗ trợ
TN	True Negative	Âm tính thật (Dự đoán đúng ca bình thường)
TP	True Positive	Đương tính thật (Dự đoán đúng ca bệnh)
ViT	Vision Transformer	Mô hình Transformer cho thị giác máy tính
VRAM	Video Random Access Memory	Bộ nhớ truy cập ngẫu nhiên video (Bộ nhớ GPU)
WHO	World Health Organization	Tổ chức Y tế Thế giới

MỞ ĐẦU

Trong kỷ nguyên chuyển đổi số và Cách mạng Công nghiệp 4.0, Trí tuệ Nhân tạo (AI) không còn dừng lại ở mức độ thử nghiệm hay xu hướng công nghệ đơn thuần, mà đã trở thành một lực lượng sản xuất cốt lõi, tác động sâu sắc và toàn diện đến mọi mặt của đời sống kinh tế - xã hội. Những tiến bộ vượt bậc của AI, đặc biệt là sự bùng nổ của nhánh Học sâu (Deep Learning) với khả năng mô phỏng cơ chế nhận thức của não bộ con người, đã tạo ra bước ngoặt lịch sử trong các tác vụ xử lý dữ liệu phức tạp, nhận dạng giọng nói và thị giác máy tính. Trong đó, lĩnh vực Y tế – nơi luôn đòi hỏi sự chính xác tuyệt đối, tốc độ xử lý nhanh và khả năng phân tích chuyên sâu – đang chứng kiến một cuộc cách mạng thực sự nhờ sự hỗ trợ của các mô hình AI tiên tiến.

Một trong những mũi nhọn ứng dụng quan trọng nhất của AI trong y tế chính là hỗ trợ chẩn đoán hình ảnh. Các kỹ thuật cận lâm sàng như chụp X-quang, CT-Scan, MRI đang tạo ra lượng dữ liệu hình ảnh khổng lồ mỗi ngày. Tuy nhiên, năng lực xử lý nguồn dữ liệu này lại đang gặp phải nút thắt cổ chai lớn: sự phụ thuộc hoàn toàn vào con người. Việc đọc và phân tích hình ảnh y khoa đòi hỏi bác sĩ phải có trình độ chuyên môn cao, kinh nghiệm dày dặn và sự tập trung cao độ. Thực tế cho thấy, trong các tình huống cấp cứu khẩn cấp, hoặc tại các cơ sở y tế tuyến dưới, vùng sâu vùng xa đang thiếu hụt trầm trọng bác sĩ chẩn đoán hình ảnh, quy trình này thường bị chậm trễ hoặc xảy ra sai sót do áp lực công việc và sự mệt mỏi (fatigue-related errors). Những sai sót này, dù nhỏ, cũng có thể dẫn đến việc bỏ sót bệnh hoặc chẩn đoán sai, ảnh hưởng trực tiếp đến phác đồ điều trị và tính mạng người bệnh.

Trong số các bệnh lý đường hô hấp, viêm phổi (Pneumonia) là căn bệnh phổ biến và nguy hiểm nhất, đặc biệt đối với trẻ nhỏ và người cao tuổi. Theo thống kê của Tổ chức Y tế Thế giới (WHO), viêm phổi chiếm hàng triệu ca tử vong mỗi năm và là một trong những nguyên nhân hàng đầu dẫn đến nhập viện tại nhiều quốc gia. Phương pháp sàng lọc tiêu chuẩn hiện nay là chụp X-quang lồng ngực do chi phí thấp và tốc độ nhanh. Tuy nhiên, việc đọc phim X-quang phổi không hề dễ dàng. Các dấu hiệu tổn thương như nốt mờ, đám thâm nhiễm thường có độ tương phản thấp, ranh giới không rõ ràng và dễ bị che khuất bởi xương sườn, bóng tim hoặc các cấu trúc giải phẫu khác. Sự chồng lấn hình ảnh này khiến ngay cả các chuyên gia cũng có thể gặp khó khăn trong việc phân biệt viêm phổi với các bệnh lý khác như tràn dịch màng phổi hay lao phổi.

Trong bối cảnh đó, khả năng tự động phân tích ảnh X-quang bằng AI trở thành một hướng tiếp cận đầy tiềm năng. Các mô hình mạng nơ-ron tích chập đã chứng

minh hiệu quả vượt trội trong các bài toán nhận dạng hình ảnh và gần như trở thành tiêu chuẩn vàng trong lĩnh vực thị giác máy tính. Tuy nhiên, dựa trên điều kiện thực tế của các bệnh viện và phòng khám không cho phép chúng ta áp dụng mô hình mạng nơ-ron lớn với hàng trăm triệu tham số, độ phức tạp tính toán cao và thời gian trả kết quả lâu. Do vậy, việc lựa chọn mô hình mạng nơ-ron với bộ tham số nhỏ, kết quả trả ra tức thời, độ chính xác cao và có thể chạy được ở trên những cấu hình máy tính thấp đến trung bình là vấn đề quan trọng và cấp thiết để áp dụng rộng rãi ở lâm sàng và cận lâm sàng.

Nhận thấy tầm quan trọng của bài toán và lợi ích thực tiễn của công nghệ AI đối với y tế cộng đồng, chúng tôi quyết định thực hiện đề tài “*Phát hiện viêm phổi từ ảnh X-quang bằng mô hình Học sâu*” như một công trình nghiên cứu mang lại lợi ích to lớn cho cộng đồng và xã hội, hướng tới việc hỗ trợ các bác sĩ và các cơ sở y tế trong công tác chuẩn đoán hình ảnh, cải thiện tốc độ công việc và giảm tải khối lượng công việc không đáng có.

Trong bài báo cáo này, chúng tôi sẽ trình bày một quy trình nghiên cứu toàn diện: từ việc hệ thống hóa cơ sở lý thuyết, phân tích dữ liệu, xử lý vấn đề mất cân bằng dữ liệu y tế, đến việc xây dựng, huấn luyện và tối ưu hóa mô hình mạng nơ-ron (tập trung vào kiến trúc DenseNet). Cuối cùng là việc đánh giá khắt khe hiệu năng mô hình và triển khai thử nghiệm ứng dụng Web. Cấu trúc báo cáo được chia thành 3 phần chính như sau:

- **Chương 1: Cơ sở lý thuyết:** Trình bày tổng quan về bệnh học viêm phổi, các thách thức trong chẩn đoán hình ảnh và nền tảng lý thuyết về Deep Learning, mạng CNN cùng các kỹ thuật tối ưu hóa hiện đại.
- **Chương 2: Xây dựng mô hình và Triển khai:** Mô tả chi tiết quy trình xử lý dữ liệu (tiền xử lý, tăng cường dữ liệu), thiết kế kiến trúc mạng nơ-ron tối ưu cho thiết bị cấu hình thấp và quy trình huấn luyện thực tế.
- **Chương 3: Thực nghiệm và Đánh giá:** Phân tích các chỉ số đánh giá (Accuracy, Recall, F1-Score, AUC), so sánh hiệu năng với các mô hình khác, những ưu điểm cũng như hạn chế còn tồn đọng của mô hình và vạch ra hướng phát triển lâu dài trong tương lai.

Chúng tôi hy vọng rằng đề tài không chỉ đáp ứng yêu cầu học thuật mà còn là nền tảng để tiếp tục nghiên cứu chuyên sâu hơn về AI trong lĩnh vực y tế, phát triển thêm nhiều tính năng, góp phần vào các hệ thống hỗ trợ chẩn đoán thông minh trong tương lai.

Chương 1

CƠ SỞ LÝ THUYẾT

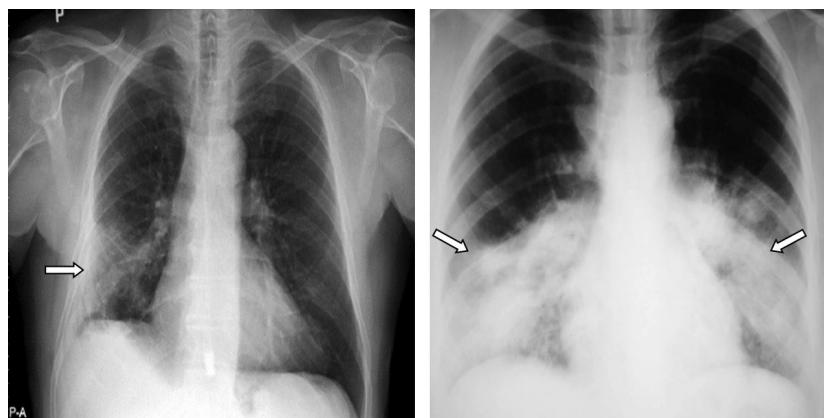
1.1. Cơ sở Y Sinh và Bối cảnh thực tế

1.1.1. Tổng quan về bệnh Viêm phổi

Viêm phổi là một dạng nhiễm trùng đường hô hấp cấp tính gây tổn thương nhu mô phổi, đặc biệt là tại các phế nang (túi khí nhỏ). Khi mắc bệnh, các phế nang này chứa đầy dịch mủ hoặc chất nhầy thay vì không khí, gây khó khăn cho việc hấp thụ oxy và hô hấp của người bệnh.

Nguyên nhân gây bệnh có thể do nhiều tác nhân gây ra, phổ biến nhất là vi khuẩn (*Streptococcus pneumoniae*), virus (như virus cúm, virus hợp bào hô hấp) hoặc nấm. Các triệu chứng thường gặp như: Ho dai dẳng, sốt cao đột ngột, khó thở, thở gấp, đau ngực khi hít sâu, mệt mỏi nghiêm trọng, v.v..

Theo Tổ chức Y tế Thế giới (WHO), viêm phổi là nguyên nhân gây tử vong hàng đầu ở trẻ em dưới 5 tuổi trên toàn cầu, chỉ riêng năm 2019, viêm phổi đã gây hơn 740.000 ca tử vong ở trẻ em dưới 5 tuổi, chiếm khoảng 14% tổng số tử vong trẻ em trên toàn cầu [6]. Việc phát hiện sớm và điều trị kịp thời là yếu tố tiên quyết để giảm thiểu tỷ lệ tử vong.



Hình 1.1: *Ảnh viêm phổi do COVID-19.*

Trong chẩn đoán lâm sàng, chụp X-quang lồng ngực (CXR) là phương pháp cận lâm sàng phổ biến nhất, được xem là "bước sàng lọc đầu tiên" do chi phí thấp và tốc độ nhanh. Trên phim X-quang, sự khác biệt giữa phổi bình thường và phổi bị viêm được thể hiện qua mật độ điểm ảnh:

- **Phổi bình thường:** Chứa đầy không khí nên tia X đi qua dễ dàng, tạo ra các vùng màu đen hoặc xám tối trên phim.
- **Phổi bị viêm:** Do sự tích tụ của dịch lỏng, tế bào viêm và xác vi khuẩn (hiện tượng đông đặc phổi), tia X bị cản lại nhiều hơn, tạo ra các vùng mờ đục màu trắng hoặc các nốt thâm nhiễm.

1.1.2. Thực trạng và Thách thức

Mặc dù X-quang là công cụ đắc lực, việc chẩn đoán thủ công bởi con người đang đối mặt với những thách thức lớn, tạo tiền đề cho sự can thiệp của Trí tuệ nhân tạo. Thách thức đầu tiên đó là: Các vùng mờ này thường không có ranh giới rõ ràng, độ tương phản thấp và dễ bị nhầm lẫn với các cấu trúc khác như xương sườn, bóng tim hoặc mạch máu phổi. Đây không chỉ là vấn đề riêng của các y bác sĩ mà còn là bài toán trích xuất đặc trưng của các mô hình Deep Learning cần giải quyết. Bên cạnh đó, có nhiều yếu tố khách quan không đáng có:

- **Sự quá tải tại các cơ sở y tế:** Trong các đợt bùng phát dịch bệnh hô hấp, số lượng phim chụp X-quang tăng đột biến. Các bác sĩ chẩn đoán hình ảnh thường phải làm việc với cường độ cao, dẫn đến tình trạng quá tải và mệt mỏi.
- **Yếu tố chủ quan và Sai sót của con người:** Việc đọc phim X-quang phụ thuộc nhiều vào kinh nghiệm và trạng thái tinh thần của bác sĩ. Chẩn đoán nhầm người khỏe mạnh thành có bệnh, gây lo lắng và lãng phí chi phí điều trị. Bỏ sót các dấu hiệu mờ nhạt ở giai đoạn đầu của bệnh, dẫn đến việc chậm trễ trong điều trị, gây nguy hiểm đến tính mạng bệnh nhân.
- **Hạn chế về nguồn nhân lực:** Tại các cơ sở y tế tuyến dưới hoặc vùng sâu vùng xa, việc tiếp cận với các chuyên gia chẩn đoán hình ảnh giàu kinh nghiệm là rất khó khăn.

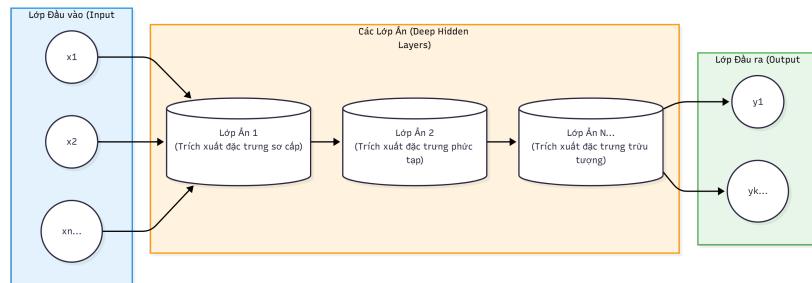
Từ những thực trạng trên, nhu cầu cấp thiết đặt ra là xây dựng một hệ thống Hỗ trợ chẩn đoán (CAD) dựa trên Học sâu. Hệ thống này không nhằm thay thế bác sĩ, mà đóng vai trò như một "trợ lý ảo" giúp sàng lọc nhanh, khoanh vùng nghi ngờ và đưa ra cảnh báo với độ chính xác và tính nhất quán cao, giúp giảm tải áp lực cho hệ thống y tế.

1.2. Tổng quan về Học sâu và Mạng Nơ-ron Tích chập

1.2.1. Học sâu

Học sâu (Deep Learning) là một tập con của Học máy (Machine Learning), lấy cảm hứng từ cấu trúc và chức năng của bộ não con người, được gọi là mạng nơ-ron

nhân tạo (ANN). Phương pháp Học sâu là sự cải tiến vượt bậc so với Học máy truyền thống, giúp máy tính có thể hiểu sâu hơn các đặc trưng của dữ liệu qua đó đưa ra những quyết định chính xác hơn.



Hình 1.2: Cấu trúc mạng nơ-ron nhân tạo.

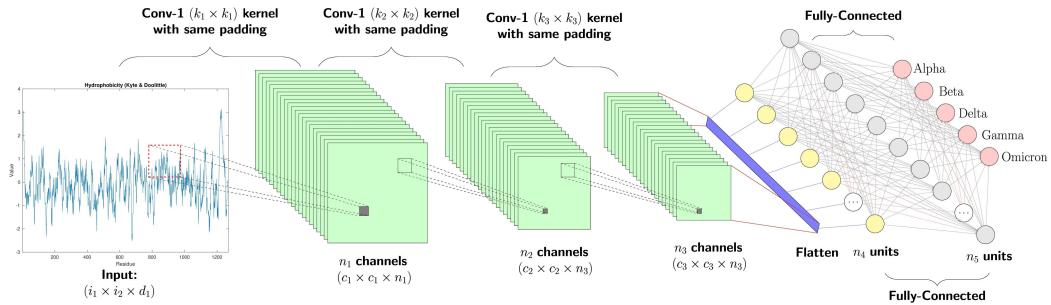
Sự khác biệt cốt lõi của Học máy và Học sâu

Trong các phương pháp Học máy truyền thống (như SVM, Random Forest), việc trích xuất đặc trưng thường phải được thực hiện thủ công bởi các chuyên gia. Ví dụ, với ảnh X-quang, con người phải tự định nghĩa các đặc điểm như độ sáng, cạnh, kết cấu để đưa vào máy tính. Ngược lại, Học sâu có khả năng tự động trích xuất đặc trưng từ dữ liệu thô. Thông qua việc sử dụng nhiều lớp ẩn (hidden layers) xếp chồng lên nhau, mô hình có thể học các đặc trưng từ cấp độ thấp (như cạnh, góc) đến cấp độ cao (như hình dạng phổi, vùng tổn thương) mà không cần sự can thiệp của con người.

Bài toán nhận diện hình ảnh nói chung và ảnh y tế nói riêng là ứng dụng cơ bản nhất của Học sâu trong vấn đề đời sống. Ảnh X-quang y tế chứa đựng các cấu trúc phức tạp và các biến thể sinh học rất tinh vi mà các thuật toán truyền thống khó có thể mô tả hết. Học sâu với khả năng học các hàm phi tuyến tính phức tạp từ lượng dữ liệu lớn, đã chứng minh hiệu quả vượt trội trong việc phát hiện các bệnh lý hình ảnh, trong đó có viêm phổi.

1.2.2. Mạng Nơ-ron Tích chập

Mạng Nơ-ron Tích chập (CNN) là một kiến trúc chuyên biệt của Học sâu, được thiết kế đặc biệt để xử lý dữ liệu dạng lưới, tiêu biểu là hình ảnh. CNN hiện là "tiêu chuẩn vàng" trong lĩnh vực Thị giác máy tính (Computer Vision). Khác với mạng nơ-ron truyền thống (Fully Connected Networks) nơi mỗi nơ-ron kết nối với tất cả nơ-ron lớp trước gây bùng nổ tham số khi xử lý ảnh, CNN sử dụng cơ chế chia sẻ trọng số và kết nối cục bộ. Một mạng CNN điển hình bao gồm các thành phần chính sau:



Hình 1.3: *Cấu trúc mạng nơ-ron tích chập (CNN).*

Lớp Tích chập (Convolutional Layer)

Đây là thành phần nòng cốt của CNN. Lớp này sử dụng các bộ lọc trượt qua toàn bộ bức ảnh đầu vào. Tại mỗi vị trí, nó thực hiện phép nhân chập giữa bộ lọc và vùng ảnh cục bộ. Mục tiêu của lớp Tích chập là tạo ra các bản đồ đặc trưng. Các lớp đầu tiên thường học các đường nét cơ bản, trong khi các lớp sâu hơn học các cấu trúc phức tạp hơn.

Công thức của một phép tích chập tại vị trí (i, j) :

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n)$$

(Trong đó I là ảnh đầu vào, K là kernel).

Hàm kích hoạt (Activation Function)

Sau mỗi lớp tích chập, một hàm kích hoạt phi tuyến tính được áp dụng. Phổ biến nhất là hàm **ReLU** (Rectified Linear Unit).

$$f(x) = \max(0, x)$$

Công thức này giúp mô hình học được các mối quan hệ phức tạp, phi tuyến tính của dữ liệu, đồng thời giảm chi phí tính toán và hỗ trợ hội tụ nhanh hơn.

Lớp gộp (Pooling Layer)

Thường được đặt sau các lớp tích chập để giảm kích thước không gian của bản đồ đặc trưng nhưng vẫn giữ lại các thông tin quan trọng nhất. Phương pháp phổ biến nhất là **Max Pooling** - Lấy giá trị lớn nhất trong vùng quét.

Fương pháp này giúp giảm số lượng tham số tính toán, kiểm soát hiện tượng quá khớp và giúp mô hình đạt được tính bất biến dịch chuyển (translation invariance) - tức là dù vật thể nằm ở đâu trong ảnh, mô hình vẫn nhận diện được.

Lớp Kết nối đầy đủ (Fully Connected Layer)

Nằm ở cuối mạng CNN. Sau khi các đặc trưng đã được trích xuất và giảm chiều, chúng được "dát phẳng" thành một vector và đưa vào mạng nơ-ron truyền thống.

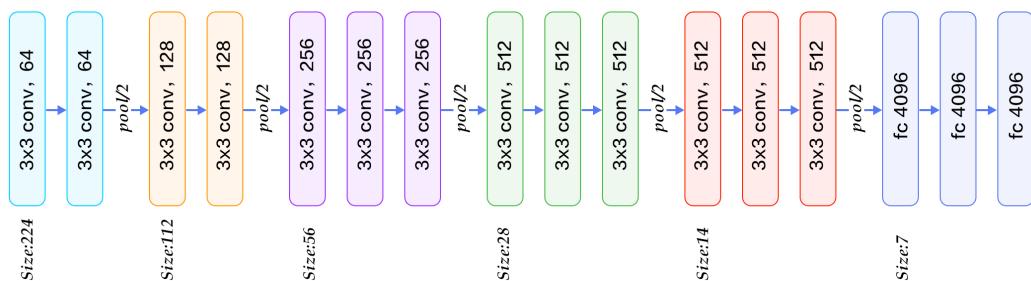
Lớp này giúp mô hình tổng hợp các đặc trưng lại để đưa ra quyết định phân loại cuối cùng. Đầu ra cuối cùng thường đi qua hàm Softmax (cho đa lớp) hoặc Sigmoid (cho nhị phân) để trả về xác suất dự đoán.

1.3. Các kiến trúc CNN phổ biến

Trong quá trình phát triển của Trí tuệ nhân tạo, nhiều kiến trúc CNN đã được đề xuất để giải quyết bài toán phân loại ảnh với độ chính xác ngày càng cao. Trong phạm vi đề tài, chúng tôi nghiên cứu và thực nghiệm trên hai kiến trúc nền tảng quan trọng là VGGNet và ResNet để làm cơ sở so sánh.

1.3.1. VGGNet

VGGNet được phát triển bởi nhóm Visual Geometry Group thuộc Đại học Oxford và đạt giải nhì trong cuộc thi ImageNet năm 2014 [5]. Đây là một trong những mô hình đầu tiên chứng minh rằng "độ sâu" của mạng là yếu tố then chốt để đạt hiệu suất cao.



Hình 1.4: Cấu trúc VGGNet.

VGGNet không sử dụng các bộ lọc kích thước lớn (như 7x7 hay 11x11 trong AlexNet), nó thiết lập một tiêu chuẩn mới bằng việc sử dụng đồng bộ các bộ lọc kích thước rất nhỏ (3x3) xuyên suốt toàn bộ mạng. Việc chồng nhiều lớp tích chập 3x3 liên tiếp cho phép mô hình có vùng cảm nhận (receptive field) tương đương với một lớp tích chập lớn, nhưng lại giảm được số lượng tham số và tăng tính phi tuyến tính nhờ nhiều hàm kích hoạt hơn.

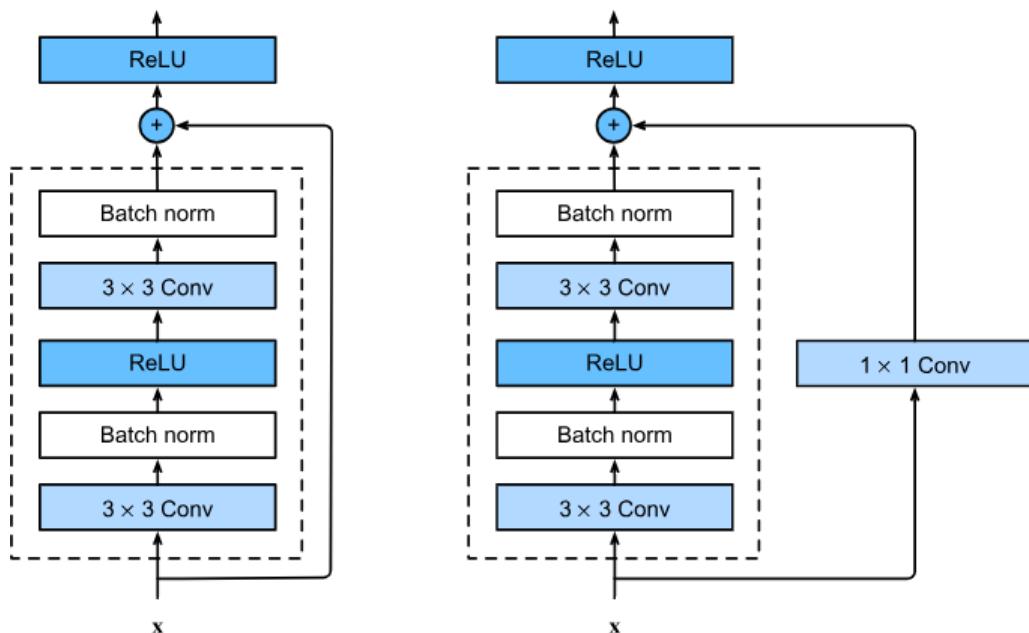
Ưu điểm và hạn chế

- Ưu điểm:** Kiến trúc đơn giản, đồng nhất, dễ cài đặt và hiểu quả trích xuất đặc trưng rất tốt.

- **Hạn chế:** Nhược điểm lớn nhất của VGGNet là số lượng tham số khổng lồ, dẫn đến kích thước mô hình nặng và tốn nhiều tài nguyên tính toán cũng như thời gian huấn luyện.

1.3.2. ResNet

Khi các mạng nơ-ron trở nên quá sâu, chúng gặp phải vấn đề "*Biến mất đạo hàm*" (Vanishing Gradient): thông tin đạo hàm dùng để cập nhật trong số bị giảm dần về 0 khi truyền ngược từ lớp cuối về lớp đầu, khiến các lớp đầu không thể học được gì. ResNet (Residual Networks) được Microsoft giới thiệu vào năm 2015 đã giải quyết triệt để vấn đề này [1].



Hình 1.5: *Kiến trúc mạng ResNet*.

ResNet giới thiệu khái niệm "*Kết nối tắt*" (Skip Connection / Shortcut Connection). Thay vì chỉ truyền dữ liệu qua các lớp phi tuyến tính, ResNet cộng trực tiếp đầu vào x vào đầu ra của lớp đó. Cơ chế này cho phép luồng thông tin đi "tắt" qua các lớp mà không bị suy giảm, giúp huấn luyện được các mạng cực sâu (lên tới 152 lớp hoặc hơn).

$$y = F(x, \{W_i\}) + x$$

Trong đó $F(x)$ là hàm ánh xạ dư (residual mapping) cần học.

Ưu điểm và hạn chế

- **Ưu điểm:** Giải quyết tốt vấn đề "*Biến mất đạo hàm*", dễ dàng huấn luyện mạng

rất sâu, độ chính xác cao hơn VGG nhưng lại nhẹ hơn đáng kể về số lượng tham số.

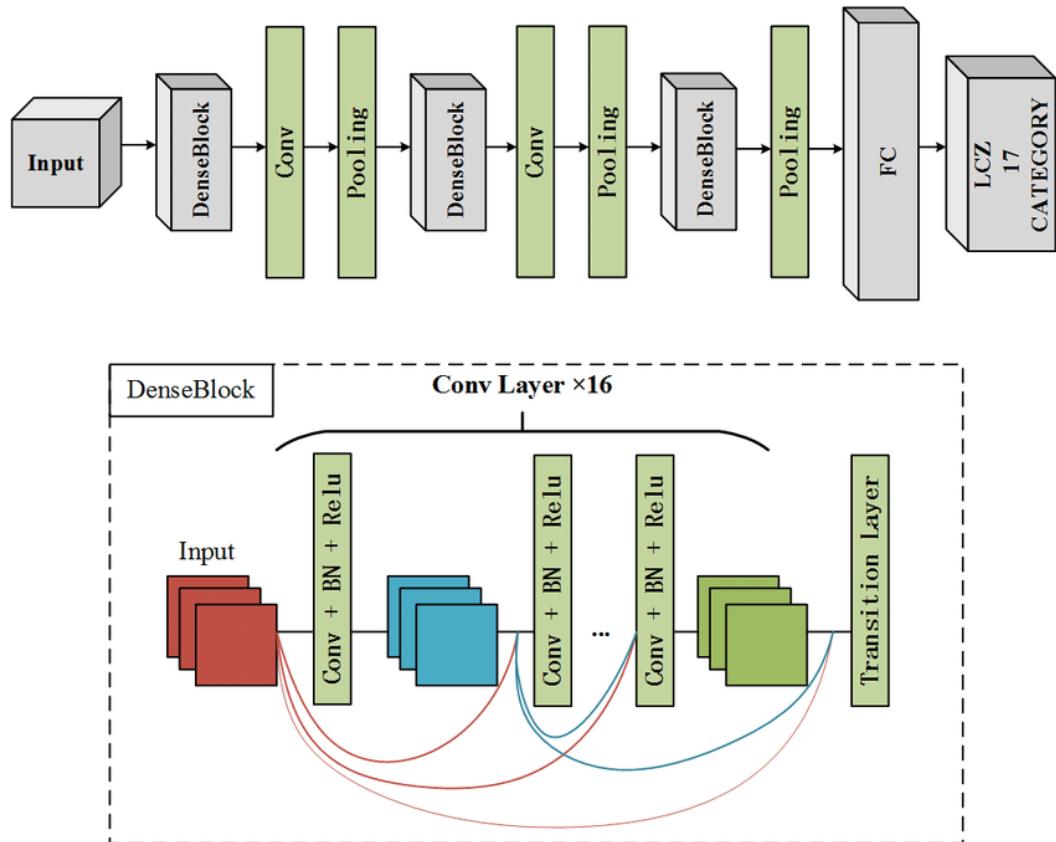
- **Hạn chế:** Mặc dù hiệu quả, nhưng các kết nối cộng gộp trong ResNet có thể làm mất đi một số đặc trưng ban đầu khi mạng quá sâu, dẫn đến việc thông tin không được khai thác triệt để.

1.4. Kiến trúc DenseNet

1.4.1. Giới thiệu

DenseNet (Densely Connected Convolutional Networks) là một bước đột phá trong kiến trúc mạng nơ-ron sâu, được giới thiệu bởi Gao Huang và cộng sự tại hội nghị CVPR 2017 (Best Paper Award) [2].

Khác với các mạng CNN truyền thống (như VGG) nơi lớp thứ L chỉ nhận đầu vào từ lớp $L - 1$, hay ResNet sử dụng phép cộng (summation) để gộp đặc trưng, DenseNet đề xuất một cơ chế kết nối hoàn toàn mới.



Hình 1.6: *Kiến trúc mạng DenseNet.*

Trong một khối Dense Block, mỗi lớp nhận đầu vào từ tất cả các lớp trước đó

và truyền bản đồ đặc trưng (feature maps) của chính nó cho tất cả các lớp phía sau. Thay vì cộng gộp, DenseNet sử dụng phép Ghép nối (Concatenation) theo chiều sâu (channel axis). Nếu H_l là hàm biến đổi phi tuyến tính (bao gồm Batch Norm, ReLU, Conv) của lớp thứ l , và x_l là đầu ra của lớp đó:

- Ở ResNet: $x_l = H_l(x_{l-1}) + x_{l-1}$ (Phép cộng element-wise).
- Ở DenseNet: $x_l = H_l([x_0, x_1, \dots, x_{l-1}])$ (Trong đó [...] thể hiện phép ghép nối các feature map từ lớp 0 đến lớp $l - 1$).

Điều này có nghĩa là lớp cuối cùng của mạng có thể "nhìn thấy" trực tiếp các đặc trưng nguyên bản từ những lớp đầu tiên cũng như các đặc trưng trùu tượng ở các lớp giữa.

1.4.2. Thành phần kiến trúc chính

Mô hình DenseNet được cấu thành từ các thành phần sau:

Khối dày đặc (Dense Block)

Đây là module chính của mạng. Trong khối này, kích thước của các bản đồ đặc trưng (chiều rộng x chiều cao) được giữ nguyên để có thể thực hiện phép ghép nối. Mỗi lớp trong một Dense Block thường bao gồm một lớp chuẩn hóa theo batch, một lớp kích hoạt ReLU, và một lớp tích chập. Các Dense Block giúp giảm độ phức tạp của mô hình mà vẫn giữ được độ chính xác cao nhờ vào việc tái sử dụng các đặc trưng đã học.

Lớp chuyển tiếp (Transition Layer)

Do phép tích chập trong CNN thường làm giảm kích thước ảnh, nên không thể ghép nối liên tục từ đầu đến cuối mạng. DenseNet sử dụng các lớp Transition nằm giữa các Dense Block để thực hiện việc kiểm soát kích thước và độ phức tạp của mô hình, ngăn chặn sự phình to của số lượng đặc trưng khi mạng trở nên quá sâu. Cấu trúc lớp bao gồm một lớp tích chập 1x1 để giảm số lượng kênh/channel và một lớp 2x2 Average Pooling để giảm kích thước không gian.

Growth Rate

Đây là siêu tham số quan trọng kiểm soát độ rộng của mạng. Nếu mỗi lớp H_l tạo ra k bản đồ đặc trưng, thì sau lớp thứ l , đầu vào của lớp tiếp theo sẽ tăng thêm k kênh. DenseNet thường sử dụng k rất nhỏ (ví dụ $k = 32$). Nhờ cơ chế tái sử dụng đặc trưng, mạng không cần phải quá rộng nhưng vẫn chứa đựng lượng thông tin lớn ("Collective Knowledge").

1.4.3. Ưu điểm

Trong bối cảnh bài toán phân loại viêm phổi, DenseNet thể hiện những ưu điểm vượt trội so với ResNet hay VGG:

- **Giảm thiểu hiện tượng Biến mất đạo hàm:** Nhờ kết nối trực tiếp, đạo hàm từ hàm mất mát có thể truyền thẳng đến các lớp đầu tiên mà không bị suy giảm qua các phép nhân ma trận phức tạp. Điều này giúp mạng dễ dàng hội tụ dù có cấu trúc rất sâu.
- **Tăng cường lan truyền đặc trưng:** Trong ảnh X-quang, các dấu hiệu viêm phổi đôi khi rất mờ nhạt như đám mờ kính đục hoặc rất nhỏ. Việc DenseNet cho phép các lớp cuối cùng truy cập lại các đặc trưng cấp thấp từ các lớp đầu giúp mô hình không bị mất đi các chi tiết nhỏ quan trọng này trong quá trình trừu tượng hóa.
- **Hiệu quả tham số và Giảm hiện tượng Quá khớp:** DenseNet yêu cầu ít tham số hơn đáng kể so với ResNet để đạt được cùng mức độ chính xác do các kenh hẹp và tái sử dụng đặc trưng. Với tập dữ liệu y tế thường có kích thước khiêm tốn, việc giảm số lượng tham số là cực kỳ quan trọng để tránh hiện tượng Quá khớp.

1.4.4. Hạn chế

Mặc dù hiệu quả, DenseNet cũng tồn tại nhược điểm về tài nguyên bộ nhớ. Do phải lưu trữ toàn bộ các feature map của các lớp trước đó để thực hiện ghép nối, DenseNet tiêu tốn nhiều RAM/VRAM GPU trong quá trình huấn luyện hơn so với ResNet.

1.5. Kỹ thuật huấn luyện mô hình

Để đảm bảo mô hình hoạt động hiệu quả trên tập dữ liệu và đạt độ chính xác cao nhất cần áp dụng các kỹ thuật huấn luyện tiên tiến.

1.5.1. Học chuyển tiếp

Học chuyển tiếp (Transfer Learning) là kỹ thuật sử dụng lại một mô hình đã được huấn luyện trước trên một tập dữ liệu lớn (như ImageNet với 1.2 triệu ảnh) để giải quyết một bài toán khác có liên quan nhưng ít dữ liệu hơn. Trong kỹ thuật này, chúng tôi giữ lại phần Lớp trích xuất đặc trưng của mô hình DenseNet đã học từ ImageNet và chỉ thay thế, huấn luyện lại phần Lớp phân loại cuối cùng để phù hợp với bài toán.

Việc áp dụng kỹ thuật này vào mô hình mang lại kết quả khả quan và tiện ích đáng ghi nhận:

- **Khắc phục thiếu hụt dữ liệu:** Tập dữ liệu X-quang hiện có chỉ khoảng 5,000 ảnh, quá nhỏ để huấn luyện một mạng sâu từ đầu mà không bị Quá khớp.
- **Tăng tốc độ hội tụ:** Do các trọng số đã được khởi tạo tốt từ trước, mô hình không cần tốn thời gian học lại cách "nhìn" các đường nét cơ bản, giúp quá trình huấn luyện nhanh hơn đáng kể.

1.5.2. Hàm mất mát

Hàm mất mát là thước đo đánh giá độ sai lệch giữa dự đoán của mô hình (\hat{y}) và nhãn thực tế (y). Mục tiêu của quá trình huấn luyện là tối thiểu hóa giá trị này. Trong bài toán phân loại nhị phân này, hàm mất mát sử dụng là hàm Cross-Entropy Loss:

$$L = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)]$$

Trong đó: y_i là nhãn thực tế (0 hoặc 1), \hat{y}_i là xác suất dự đoán lớp 1.

Cross-Entropy phạt rất nặng các dự đoán sai mà có độ tin cậy cao. Điều này thúc đẩy mô hình điều chỉnh trọng số mạnh mẽ để tránh những sai lầm nghiêm trọng, rất phù hợp cho các bài toán y tế đòi hỏi độ chính xác cao.

1.5.3. Thuật toán tối ưu hóa

Để cập nhật trọng số của mạng nhằm giảm thiểu hàm mất mát, thuật toán Adam (Adaptive Moment Estimation) là một lựa chọn phù hợp nhằm tối ưu thời gian chạy [4]. Đây là một trong những thuật toán tối ưu hóa phổ biến và hiệu quả nhất hiện nay cho Học sâu.

Adam là sự kết hợp ưu điểm của hai thuật toán mở rộng của Gradient Descent:

- **Momentum:** Giúp thuật toán vượt qua các điểm cực tiểu địa phương bằng cách giữ đà chuyển động của gradient.
- **RMSProp:** Tự động điều chỉnh tốc độ học cho từng tham số riêng biệt dựa trên độ lớn của gradient trong quá khứ.

Chương 2

XÂY DỰNG MÔ HÌNH VÀ TRIỂN KHAI

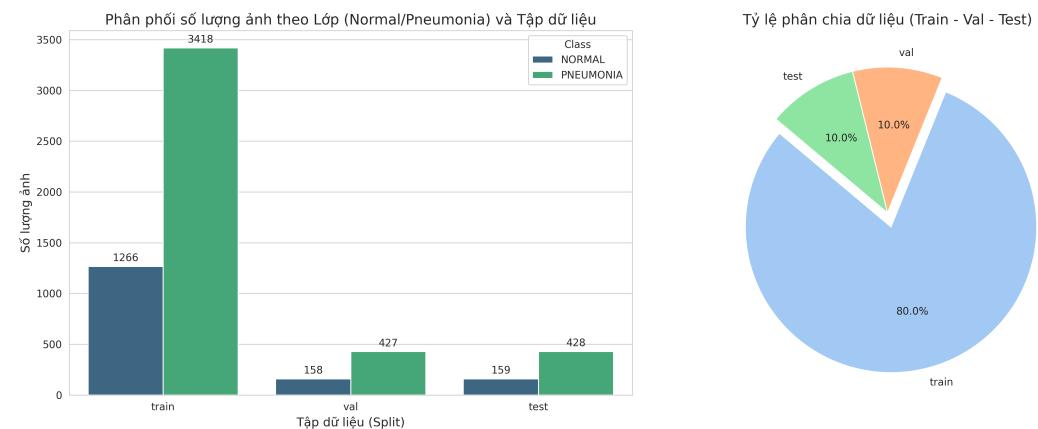
2.1. Mô tả tập dữ liệu

Dự án sử dụng bộ dữ liệu chuẩn hóa "*Chest X-Ray Images (Pneumonia)*", được cung cấp bởi Trung tâm Y tế Phụ nữ và Trẻ em Quảng Châu (Guangzhou Women and Children's Medical Center) [3]. Đây là tập dữ liệu mở phổ biến trên nền tảng Kaggle, thường được sử dụng làm chuẩn cho các nghiên cứu về ứng dụng AI trong chẩn đoán hình ảnh y tế.

Hình ảnh X-quang ngực được chọn từ các nhóm bệnh nhi từ một đến năm tuổi. Để phân tích hình ảnh X-quang ngực, tất cả các phim chụp X-quang ban đầu được sàng lọc để kiểm soát chất lượng bằng cách loại bỏ tất cả các ảnh chụp chất lượng thấp hoặc không thể đọc được. Chẩn đoán hình ảnh sau đó được hai bác sĩ chuyên khoa phân loại trước khi được phê duyệt để huấn luyện hệ thống AI.

2.1.1. Đặc điểm và phân phối

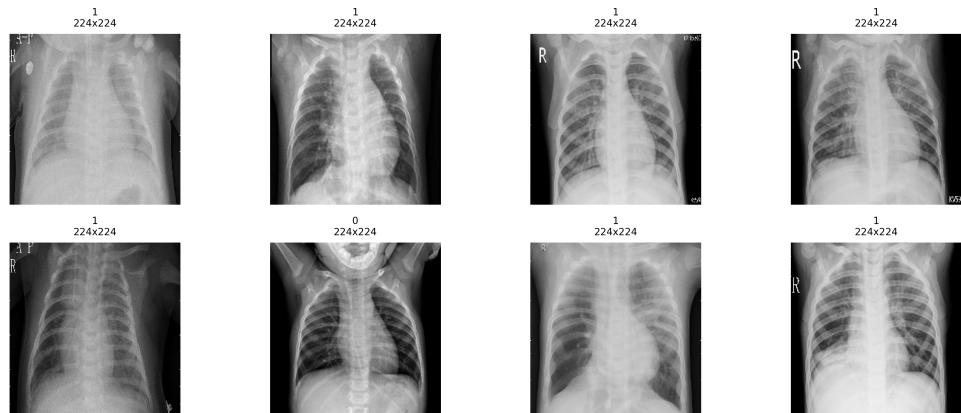
Các ảnh được đặt trong hai thư mục: NORMAL và PNEUMONIA tương tự hai nhãn trong bài toán phân loại nhị phân. Thư mục NORMAL gồm các ảnh X-quang phổi bình thường trong đó thư mục PNEUMONIA gồm các ảnh X-quang phổi bị viêm (bao gồm cả các ca viêm phổi do vi khuẩn và virus).



Hình 2.1: Biểu đồ phân bố sự mất cân bằng dữ liệu.

Bộ dữ liệu này có tổng cộng 5,856 ảnh X-quang. Dữ liệu thực tế cho thấy sự mất cân bằng lớp đáng kể, với số lượng ca bệnh Viêm phổi chiếm đa số. Qua quá trình tổng

hợp ảnh từ toàn bộ thư mục dữ liệu, ta thấy được tổng số ảnh Normal: 1,583 ảnh, tổng số ảnh Pneumonia: 4,273 ảnh. Có thể thấy rằng, số ảnh mang nhãn Pneumonia lớn hơn rất nhiều số ảnh mang nhãn Normal, gấp gần 2.7 lần số ảnh bình thường.



Hình 2.2: Một số hình ảnh từ tập dữ liệu.

Quá trình khám phá toàn bộ tập dữ liệu, chúng ta có thể thấy các tập huấn luyện, tập xác thực, tập kiểm thử không đồng đều về số lượng ảnh. Nếu giữ nguyên tỷ lệ tập dữ liệu để huấn luyện mô hình và đánh giá, kết quả trả ra có thể bị sai lệch. Để đảm bảo quá trình huấn luyện và đánh giá mô hình khách quan, dữ liệu không sử dụng cấu trúc thư mục mặc định mà được gộp lại và chia ngẫu nhiên theo tỷ lệ 80% Train - 10% Validation - 10% Test:

- **Tập Huấn luyện:** 4,684 ảnh (3,418 Pneumonia / 1,266 Normal).
- **Tập Xác thực:** 585 ảnh (427 Pneumonia / 158 Normal).
- **Tập Kiểm thử:** 587 ảnh (428 Pneumonia / 159 Normal).

2.2. Tiền xử lý dữ liệu

Dữ liệu ảnh y tế thường có kích thước không đồng nhất và chứa nhiều nhiễu. Để đảm bảo mô hình DenseNet hoạt động hiệu quả, nhóm nghiên cứu đã xây dựng một quy trình tiền xử lý gồm hai giai đoạn chính: Biến đổi hình ảnh và Xử lý mất cân bằng dữ liệu.

2.2.1. Kỹ thuật biến đổi và Chuẩn hóa ảnh

Trước khi đưa vào mô hình, các ảnh X-quang vốn có kích thước không đồng nhất đều được xử lý đưa về kích thước chuẩn là 224x224 pixels và chuyển sang định dạng

ảnh xám (Grayscale) thay vì ảnh màu RGB, giúp giảm độ phức tạp tính toán mà vẫn giữ được các đặc trưng quan trọng của ảnh X-quang.

Báo cáo sử dụng thư viện `torchvision.transforms` để thiết lập các chuỗi xử lý tự động cho từng ảnh trước khi đưa vào mô hình. Quy trình cụ thể bao gồm:

- **Đồng bộ kích thước:** Các ảnh X-quang gốc có độ phân giải rất đa dạng. Để phù hợp với đầu vào tiêu chuẩn của kiến trúc DenseNet, tất cả ảnh được thay đổi kích thước về cố định 224 x 224 pixels. Việc này giúp giảm khối lượng tính toán nhưng vẫn giữ được các cấu trúc giải phẫu chính của phổi.
- **Chuyển đổi Tensor:** Dữ liệu ảnh được chuyển đổi sang dạng Tensor của PyTorch. Bước này đồng thời thay đổi phạm vi giá trị pixel từ [0, 255] về [0.0, 1.0], tạo điều kiện thuận lợi cho các phép tính đại số tuyến tính trên GPU.
- **Chuẩn hóa dữ liệu:** Để tăng tốc độ hội tụ của mô hình, các ảnh đầu vào được chuẩn hóa theo phân phối chuẩn dựa trên bộ dữ liệu ImageNet - tập dữ liệu mà DenseNet đã được huấn luyện trước đó.

2.2.2. Xử lý mất cân bằng dữ liệu

Như đã phân tích ở phần mô tả dữ liệu, số lượng ảnh Viêm phổi (Pneumonia) gấp gần 3 lần ảnh Bình thường (Normal). Nếu huấn luyện thông thường, mô hình sẽ có xu hướng "học vẹt" bằng cách luôn dự đoán là Viêm phổi để đạt độ chính xác cao, dẫn đến bỏ sót các ca bình thường. Để giải quyết vấn đề này, nhóm không sử dụng phương pháp nhân bản dữ liệu (Oversampling) gây trùng lặp, mà áp dụng kỹ thuật Lấy mẫu có trọng số (Weighted Random Sampler) trong quá trình tạo Batch huấn luyện.

Quy trình giải quyết vấn đề mất cân bằng dữ liệu được thực hiện như sau:

- **Tính toán trọng số lớp:** Trọng số của mỗi lớp được tính nghịch đảo với số lượng mẫu của lớp đó. Lớp nào càng ít dữ liệu thì trọng số càng cao.

$$W_{class} = \frac{1}{\text{Số lượng mẫu của class}}$$

- **Gán trọng số cho từng mẫu:** Mỗi bức ảnh trong tập huấn luyện được gán một giá trị trọng số tương ứng với nhãn của nó.
- **Cơ chế lấy mẫu:** Trong mỗi epoch, thay vì lấy dữ liệu ngẫu nhiên đều nhau, cơ chế sẽ bốc mẫu dựa trên xác suất tỉ lệ thuận với trọng số. Dù trong tập dữ liệu gốc ảnh Normal rất ít, nhưng trong mỗi Batch được đưa vào mô hình, tỉ lệ

giữa Normal và Pneumonia sẽ xấp xỉ 50:50. Điều này buộc mô hình phải học đặc trưng của cả hai lớp một cách công bằng.

2.2.3. Tổ chức và Tải dữ liệu

Sau khi dữ liệu đã được chuyển đổi sang dạng Tensor, bước tiếp theo là thiết lập cơ chế tải dữ liệu tự động để đưa vào quy trình huấn luyện. Trong PyTorch, `DataLoader` đóng vai trò là "cầu nối" hiệu quả giữa dữ liệu thô và mô hình. Nó giải quyết các nhiệm vụ:

- **Đóng gói dữ liệu:** Quản lý việc truy cập vào danh sách ảnh và nhãn.
- **Chia nhỏ:** Thay vì nạp toàn bộ hàng nghìn bức ảnh vào RAM cùng lúc (gây tràn bộ nhớ), `DataLoader` chia dữ liệu thành các gói nhỏ (Batch) để xử lý tuần tự.
- **Xáo trộn và Lấy mẫu:** Quyết định chiến lược chọn ảnh nào để đưa vào mô hình, giúp giải quyết vấn đề mất cân bằng dữ liệu.

Do số lượng ảnh PNEUMONIA chiếm đa số so với NORMAL, nếu sử dụng phương pháp lấy mẫu ngẫu nhiên thông thường, mô hình sẽ bị thiên lệch. Báo cáo này đã áp dụng kỹ thuật Lấy mẫu ngẫu nhiên có trọng số (Weighted Random Sampler) đã được nhắc phía trên. Bộ ba Loader được khởi tạo với cấu hình khác nhau cho từng giai đoạn:

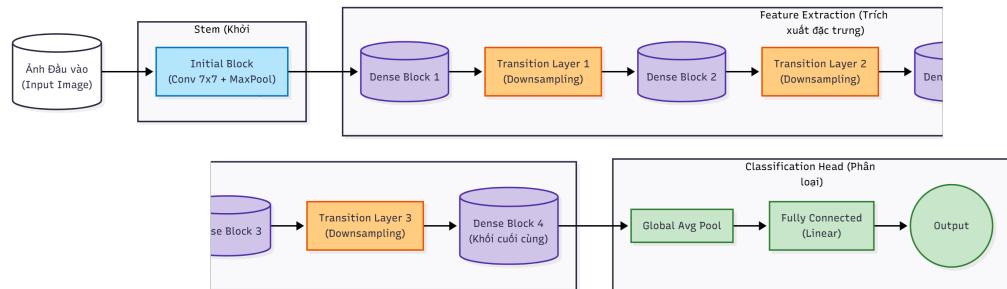
- **Train Loader:** `batch_size = 32, sampler = weighted_sampler`. Dảm bảo mô hình luôn học được số lượng ảnh Normal và Pneumonia cân bằng nhau trong mỗi bước cập nhật trọng số.
- **Validation & Test Loaders:** `batch_size = 32, sampler = False`. Đánh giá mô hình trên phân phối dữ liệu thực tế để có cái nhìn khách quan nhất về hiệu năng, và giữ cố định thứ tự để dễ dàng so sánh kết quả giữa các lần chạy.

2.3. Xây dựng mô hình

Bài toán đặt ra là làm sao có thể tối ưu lượng tài nguyên tính toán để có thể chạy được trên nhiều thiết bị từ thấp đến cao, song, mô hình phải có khả năng tính toán với thông lượng lớn và độ chính xác cao. Qua đó, báo cáo này quyết định chọn mô hình mạng DenseNet là mô hình nền tảng để giải quyết bài toán đặt ra.

Thay vì sử dụng các mô hình đã huấn luyện sẵn trên ImageNet vốn thường yêu cầu đầu vào ảnh màu 3 kênh (RGB) và có kích thước mô hình lớn, nhóm đã quyết

định tự thiết kế và cài đặt lại kiến trúc DenseNet (Densely Connected Convolutional Networks) từ nền tảng PyTorch.



Hình 2.3: *Kiến trúc mạng DenseNet.*

Việc này giúp tối ưu hóa mô hình cho dữ liệu ảnh X-quang đen trắng (Grayscale - 1 kênh), giảm thiểu dư thừa tham số và cho phép kiểm soát hoàn toàn luồng xử lý dữ liệu. Mô hình DenseNet được xây dựng với cấu trúc gồm 4 khối Dense Block, tuân theo nguyên lý "Feature Reuse" của DenseNet gốc.

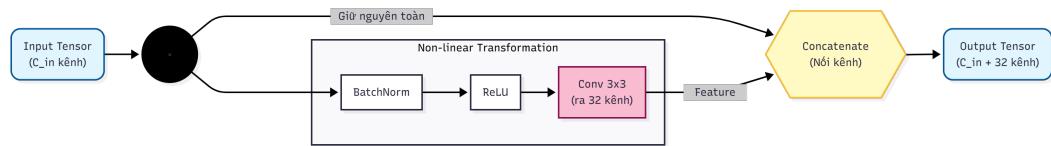
Khối xử lý đầu vào

Đây là cổng vào của mạng, chịu trách nhiệm giảm kích thước không gian của ảnh gốc và trích xuất các đặc trưng sơ cấp đầu tiên.

- **Đầu vào:** Ảnh Tensor kích thước (*Batch*, 1, 224, 224).
- **Quy trình xử lý:**
 - **Convolution 7×7:** Sử dụng bộ lọc kích thước lớn với stride=2 và padding=3. Bước này giảm kích thước ảnh xuống còn một nửa (112×112) và tăng độ sâu lên 64 kênh.
 - **Max Pooling 3 × 3:** Tiếp tục giảm kích thước ảnh xuống còn một nửa nữa (56×56) thông qua lớp gộp cực đại với stride=2.
- **Kết quả:** Đầu ra của khối này là một Feature Map có kích thước (64, 56, 56), đóng vai trò là đầu vào cho chuỗi Dense Blocks phía sau.

Khối dày đặc (Dense Layer)

Đây là trái tim của kiến trúc, nơi diễn ra quá trình "Tái sử dụng đặc trưng". Mô hình bao gồm 4 Dense Block với số lượng lớp lần lượt là [4, 4, 8, 8]. Cấu trúc chi tiết bên trong mỗi lớp con (DenseLayer) của một Block:

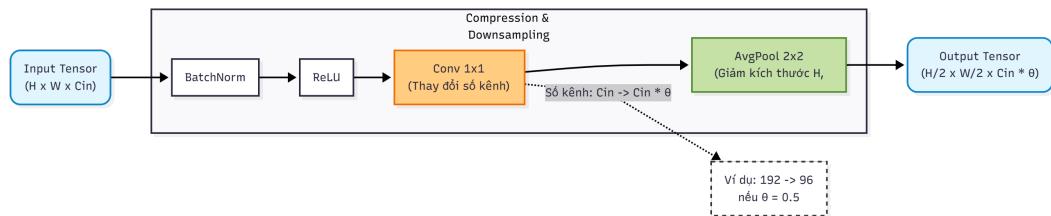


Hình 2.4: *Cấu trúc DenseBlock.*

- **Pre-activation:** Thay vì Conv → ReLU như CNN truyền thống, nhóm sử dụng thứ tự: Batch Normalization → ReLU → Convolution. Thứ tự này giúp luồng gradient truyền ngược ổn định hơn.
- **Convolution 3×3 :** Thực hiện tích chập để tạo ra k đặc trưng mới (với $k = 32$ là Growth Rate).
- **Cơ chế Concatenation (Ghép nối):** Đầu ra của lớp này không đè lên đầu vào mà được ghép nối (stack) vào phía sau theo chiều sâu (channel).

Lớp chuyển tiếp (Transition Layers)

Do cơ chế ghép nối liên tục làm số lượng kênh tăng lên rất nhanh gây bùng nổ tham số, các lớp Transition được đặt giữa các Dense Block để thực hiện nhiệm vụ "nén" mô hình. Vị trí của các lớp chuyển tiếp nằm sau các Dense Block. Các thành phần của lớp chuyển tiếp gồm:



Hình 2.5: *Cấu trúc Transition Layer.*

- **Conv 1×1 :** Đóng vai trò giảm số lượng kênh đi một nửa (`out_channels = in_channels // 2`). Ví dụ: Nếu Dense Block trước đó xuất ra 256 kênh, lớp này sẽ nén xuống còn 128 kênh.
- **Average Pooling 2×2 :** Giảm kích thước không gian của ảnh đi một nửa (ví dụ từ 56×56 xuống 28×28 , rồi 14×14 , 7×7).

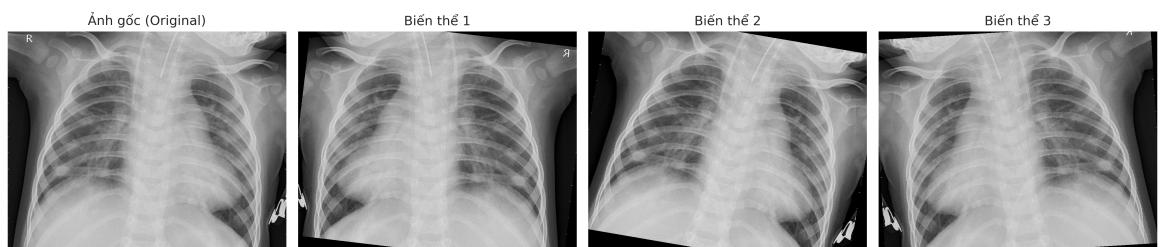
Khối phân loại

Sau khi đi qua 4 Dense Block, kích thước ảnh chỉ còn (7×7) nhưng chứa đựng độ sâu thông tin rất lớn. Khối cuối cùng thực hiện việc đưa ra quyết định:

- **Batch Norm & ReLU cuối:** Chuẩn hóa đặc trưng lần cuối.
- **Adaptive Average Pooling:** Biến đổi Feature Map từ kích thước ($C, 7, 7$) thành vector ($C, 1, 1$). Bước này giúp mô hình chấp nhận được các kích thước đầu vào khác nhau (nếu có) và loại bỏ hoàn toàn thông tin không gian, chỉ giữ lại sự hiện diện của đặc trưng.
- **Flatten:** Dát phẳng thành vector 1 chiều.
- **Linear Layer (Fully Connected):** Lớp kết nối đầy đủ duy nhất trong mạng, ánh xạ số lượng đặc trưng về 2 giá trị Logits tương ứng với lớp Normal và Pneumonia.

2.3.1. Tăng cường dữ liệu

Để chống lại hiện tượng Quá khớp khi huấn luyện trên tập dữ liệu y tế, một pipeline tăng cường dữ liệu mạnh mẽ được áp dụng chỉ riêng cho tập huấn luyện:



Hình 2.6: Minh họa kỹ thuật tăng cường dữ liệu: Ảnh gốc và các biến thể qua phép xoay, lật và chỉnh tương phản.

- **Ngẫu nhiên xoay:** Xoay ảnh trong khoảng ± 10 độ để mô phỏng sự sai lệch tư thế chụp.
- **Lật ngang:** Lật ảnh trái/phải với xác suất 50%.
- **Biến đổi màu sắc:** Thay đổi nhẹ độ sáng (brightness) và độ tương phản (contrast) $\pm 10\%$ để mô hình thích nghi với các điều kiện phơi sáng khác nhau của máy chụp X-quang.
- **Chuẩn hóa:** Dựa pixel về phân phối chuẩn với mean=0.5, std=0.5.

Tập Validation và Test chỉ được chuẩn hóa, không áp dụng các phép biến đổi ngẫu nhiên để đảm bảo tính nhất quán khi đánh giá.

2.3.2. Cấu hình phần cứng và Tối ưu hóa

Để đảm bảo mô hình hội tụ hiệu quả và đạt độ chính xác cao nhất trên tập dữ liệu kiểm thử, nhóm nghiên cứu đã thiết lập một chiến lược huấn luyện nghiêm ngặt với các thành phần cụ thể như sau:

Môi trường tính toán

Trước khi bắt đầu vòng lặp huấn luyện, toàn bộ kiến trúc mô hình và dữ liệu được chuyển sang hoạt động trên GPU (Graphics Processing Unit) thông qua cơ chế của PyTorch. Sử dụng hàm `.to(device)` để đẩy các Tensor từ RAM (CPU) sang VRAM (GPU). Việc này tận dụng khả năng tính toán song song của hàng nghìn lõi CUDA, giúp tăng tốc độ huấn luyện lên gấp 10-20 lần so với sử dụng CPU thông thường.

Hàm mất mát

Mô hình sử dụng hàm CrossEntropyLoss (`nn.CrossEntropyLoss`). Hàm này trừng phạt rất nặng các dự đoán sai mà mô hình lại "tự tin" (ví dụ: ảnh là Viêm phổi nhưng mô hình dự đoán 90% là Bình thường). Điều này thúc đẩy gradient thay đổi mạnh mẽ để sửa lỗi nhanh chóng trong giai đoạn đầu. Đây là hàm mục tiêu tiêu chuẩn cho bài toán phân loại. Trong PyTorch, hàm này đã tích hợp sẵn hai bước: LogSoftmax (để chuyển đổi logit thành xác suất) và NLLLoss (Negative Log Likelihood).

Thuật toán tối ưu hóa Nhóm lựa chọn thuật toán Adam (Adaptive Moment Estimation) thay vì SGD truyền thống. Tốc độ học ở mức $\alpha = 0.001$. Đây là mức khởi đầu an toàn, đủ lớn để mô hình học nhanh nhưng đủ nhỏ để không bị dao động quá mạnh quanh điểm cực tiểu. Một điểm nhấn quan trọng trong cấu hình này là việc sử dụng hàm hiệu chỉnh L2 (L2 Regularization). DenseNet là một mạng sâu với nhiều tham số, rất dễ bị quá khớp trên tập dữ liệu nhỏ. Giải pháp là cộng thêm một lượng nhỏ giá trị trọng số `weight_decay=1e-4` vào hàm loss ($\lambda \sum w^2$). Điều này ép các trọng số của mô hình phải nhỏ và phân tán, giúp mô hình đơn giản hóa và tổng quát hóa tốt hơn trên dữ liệu lạ.

Điều chỉnh tốc độ học

Thay vì giữ nguyên tốc độ học suốt quá trình, nhóm áp dụng chiến lược "giảm tốc khi gặp khó" bằng **ReduceLROnPlateau**. Bộ điều phối sẽ giám sát độ chính xác trên tập Validation (`val_acc`). Thiết lập `patience=3`. Nghĩa là nếu sau 3 epochs liên tiếp mà độ chính xác không tăng, hệ thống coi như mô hình đã bị kẹt hoặc dao động quanh điểm tối ưu. Khi điều kiện trên xảy ra, Learning Rate sẽ được nhân với 0.5 (giảm đi một nửa). Việc giảm nhỏ bước nhảy giúp mô hình "lách" sâu hơn vào điểm

cực tiểu toàn cục để đạt độ chính xác cao nhất có thể.

Chương 3

THỰC NGHIỆM VÀ ĐÁNH GIÁ KẾT QUẢ

3.1. Thiết lập môi trường và cấu hình thực nghiệm

Để đảm bảo tính khách quan, độ tin cậy và khả năng tái lập kết quả nghiên cứu, quá trình thực nghiệm huấn luyện mô hình DenseNet được thiết lập trên một quy trình chuẩn hóa nghiêm ngặt về cả phần cứng lẫn chiến lược tối ưu hóa. Các thiết lập này được duy trì nhất quán trong suốt quá trình thử nghiệm để đánh giá chính xác hiệu năng của kiến trúc mô hình đề xuất.

3.1.1. Phần cứng và phần mềm

Do đặc thù của kiến trúc DenseNet là một mạng nơ-ron tích chập sâu (Deep CNN) với hàng triệu phép tính ma trận phức tạp, việc sử dụng tài nguyên tính toán hiệu năng cao là điều kiện tiên quyết. Nhóm nghiên cứu đã triển khai thực nghiệm trên nền tảng đám mây **Google Colab**. Cấu hình phần cứng như sau:

- **GPU:** NVIDIA Tesla T4 (VRAM 16GB)
- **RAM:** 12GB

Về mặt phần mềm, mô hình được xây dựng hoàn toàn trên framework PyTorch [0], kết hợp với thư viện `torchvision` để xử lý dữ liệu ảnh và `scikit-learn` để tính toán các chỉ số đánh giá. Song, thư viện `matplotlib` dùng để trực quan hóa kết quả huấn luyện mô hình. Môi trường thực nghiệm sử dụng ngôn ngữ Python 3.10+, đảm bảo tương thích tốt nhất với các thư viện Deep Learning hiện hành.

3.1.2. Tối ưu hóa và Thiết lập siêu tham số

Việc lựa chọn các siêu tham số đóng vai trò quyết định đến khả năng hội tụ của mô hình. Trong nghiên cứu này, thay vì sử dụng phương pháp tìm kiếm ngẫu nhiên, nhóm đã thiết lập các thông số dựa trên các thực nghiệm sơ bộ và các khuyến nghị chuẩn trong cộng đồng nghiên cứu thị giác máy tính:

Dầu tiên, hàm mục tiêu được lựa chọn là Cross-Entropy Loss. Đối với bài toán phân loại nhị phân (Bình thường vs. Viêm phổi), hàm mất mát này cực kỳ hiệu quả trong việc trừng phạt các dự đoán sai lệch có độ tin cậy cao, đồng thời thúc đẩy mô hình tối đa hóa xác suất log-likelihood của nhãn đúng.

Để cực tiểu hóa hàm mất mát, nhóm sử dụng thuật toán tối ưu hóa Adam (Adaptive Moment Estimation). So với SGD truyền thống, Adam có ưu điểm vượt trội nhờ cơ

Tham số (Parameter)	Giá trị thiết lập
Thuật toán tối ưu (Optimizer)	Adam
Tốc độ học (Learning Rate)	0.001
Kích thước Batch (Batch Size)	32
Số vòng lặp (Epochs)	100
Kích thước ảnh đầu vào (Input Size)	224×224
Hàm mất mát (Loss Function)	CrossEntropy
Suy giảm trọng số (Weight Decay)	$1e^{-4}$

Bảng 3.1: Cấu hình siêu tham số cho quá trình huấn luyện.

chế tự động điều chỉnh tốc độ học cho từng tham số dựa trên moment bậc nhất và bậc hai của gradient. Tốc độ học khởi tạo được thiết lập ở mức $\alpha = 0.001 (1e^{-3})$, một giá trị cân bằng giúp mô hình học nhanh trong giai đoạn đầu nhưng vẫn đảm bảo sự ổn định.

Đặc biệt, để giải quyết vấn đề mô hình bị dao động hoặc mắc kẹt tại các điểm tối ưu cục bộ, chiến lược điều phối tốc độ học **ReduceLROnPlateau** đã được áp dụng. Cơ chế này hoạt động như một hệ thống giám sát: nếu độ chính xác trên tập xác thực (**val_acc**) không cải thiện trong 3 chu kỳ liên tiếp (**patience=3**), tốc độ học sẽ tự động giảm đi một nửa (**factor=0.5**). Điều này cho phép mô hình thực hiện các bước cập nhật trọng số nhỏ hơn và tinh vi hơn khi tiến gần đến điểm hội tụ toàn cục.

3.1.3. Quá trình huấn luyện và Chống hiện tượng quá khớp

Một trong những thách thức lớn nhất khi huấn luyện mạng nơ-ron sâu trên tập dữ liệu y tế quy mô nhỏ (khoảng 5.000 ảnh) là hiện tượng quá khớp. Để khắc phục điều này, quy trình thực nghiệm đã tích hợp đồng thời ba kỹ thuật điều chuẩn (Regularization) mạnh mẽ:

Thứ nhất, kỹ thuật Tăng cường dữ liệu được áp dụng trực tiếp trong quy trình tải dữ liệu. Các ảnh huấn luyện được biến đổi ngẫu nhiên thông qua các phép xoay (**RandomRotation 10 độ**), lật ngang (**RandomHorizontalFlip**) và thay đổi độ tương phản (**ColorJitter**). Điều này ép buộc mô hình phải học các đặc trưng bất biến của bệnh viêm phổi thay vì học thuộc lòng các chi tiết nhiễu của ảnh gốc.

Thứ hai, tham số **weight_decay** (Suy giảm trọng số) được thiết lập ở mức $1e^{-4}$ trong bộ tối ưu hóa Adam. Đây thực chất là kỹ thuật L2 Regularization, giúp hạn chế giá trị của các trọng số trở nên quá lớn, từ đó làm giảm độ phức tạp của mô hình và tăng khả năng tổng quát hóa.

Cuối cùng, quy trình huấn luyện được thực hiện trong vòng 100 epochs với kích

thước Batch Size là 32. Kích thước batch này được lựa chọn để vừa vặn với bộ nhớ GPU, đồng thời đủ nhỏ để tạo ra nhiều ngẫu nhiên trong quá trình ước lượng gradient, giúp mô hình thoát khỏi các điểm yên ngựa. Cơ chế Model Checkpointing cũng được sử dụng để chỉ lưu lại bộ trọng số có độ chính xác cao nhất trên tập xác thực, đảm bảo mô hình cuối cùng được chọn là phiên bản tối ưu nhất chứ không phải là phiên bản bị overfitting ở các epoch cuối cùng.

3.2. Tiêu chí đánh giá

Trong các bài toán chẩn đoán y tế, việc chỉ dựa vào độ chính xác đơn thuần thường không phản ánh đúng hiệu năng của mô hình, đặc biệt khi dữ liệu có sự mất cân bằng giữa các lớp. Do đó, nhóm nghiên cứu sử dụng hệ thống các tiêu chí đánh giá đa chiều dựa trên Ma trận nhầm lẫn để có cái nhìn khách quan nhất.

Ma trận nhầm lẫn (Confusion Matrix)

Đây là bảng thống kê đối chiếu giữa kết quả dự đoán của mô hình và nhãn thực tế của dữ liệu kiểm thử. Với bài toán phân loại nhị phân (Normal vs. Pneumonia), ma trận này bao gồm 4 thành phần cơ bản:

- **True Positive (TP):** Số ca bệnh Viêm phổi được mô hình dự đoán đúng là Viêm phổi.
- **True Negative (TN):** Số ca Bình thường được mô hình dự đoán đúng là Bình thường.
- **False Positive (FP):** Số ca Bình thường nhưng mô hình dự đoán nhầm là Viêm phổi (Báo động giả).
- **False Negative (FN):** Số ca Viêm phổi nhưng mô hình dự đoán nhầm là Bình thường (Bỏ sót bệnh).

Nhận xét: Trong y tế, FN (Bỏ sót bệnh) là lỗi nguy hiểm nhất vì có thể dẫn đến việc bệnh nhân không được điều trị kịp thời. Mục tiêu tối thượng của mô hình là giảm thiểu chỉ số này về mức thấp nhất có thể.

Các chỉ số định lượng

Dựa trên các giá trị của Ma trận nhầm lẫn, các chỉ số hiệu năng được tính toán như sau:

a. Độ chính xác toàn cục (Accuracy): Là tỷ lệ tổng số lần dự đoán đúng trên tổng số mẫu thực nghiệm.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)$$

Ý nghĩa: Cho biết khả năng phân loại chung của mô hình. Tuy nhiên, nếu tập dữ liệu mất cân bằng (ví dụ 90% mẫu là bệnh), Accuracy có thể rất cao dù mô hình không phát hiện được ca bình thường nào.

b. Độ chính xác dương tính (Precision): Là tỷ lệ số ca thực sự mắc bệnh trong số các ca mà mô hình dự đoán là có bệnh.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.2)$$

Ý nghĩa: Chỉ số này cho biết độ tin cậy của mô hình khi đưa ra cảnh báo. Precision cao đồng nghĩa với việc ít có "báo động giả" (False Alarm).

c. Độ nhạy (Recall / Sensitivity): Là tỷ lệ số ca bệnh được phát hiện đúng trên tổng số ca bệnh thực tế.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.3)$$

Ý nghĩa: Đây là chỉ số quan trọng nhất trong bài toán sàng lọc y tế. Recall cao chứng tỏ mô hình có khả năng "quét" sạch các ca bệnh, hạn chế tối đa việc bỏ sót bệnh nhân. Nhóm chấp nhận đánh đổi một phần Precision để đạt được Recall cao.

d. Điểm F1 (F1-Score): Là trung bình điều hòa giữa Precision và Recall.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.4)$$

Ý nghĩa: F1-Score cung cấp một đánh giá cân bằng hơn khi có sự đánh đổi giữa độ chính xác và độ nhạy. Chỉ số này đặc biệt hữu ích để so sánh xếp hạng giữa các mô hình khác nhau (ví dụ so sánh DenseNet với VGG16).

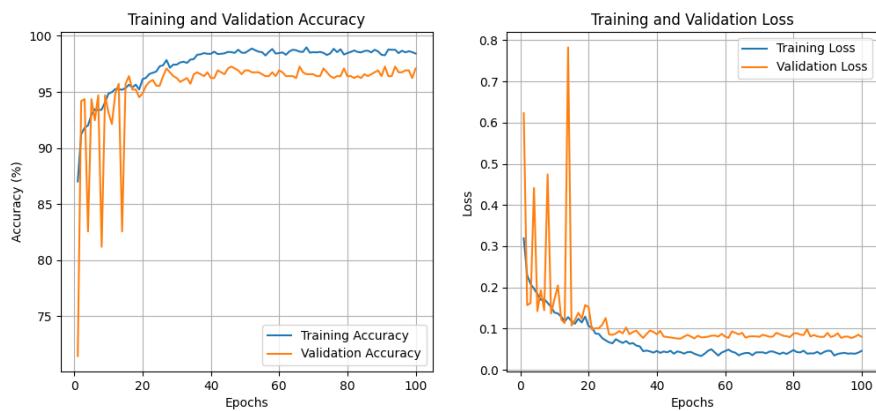
Biểu đồ trực quan hóa

Bên cạnh các con số, nhóm sử dụng các biểu đồ để đánh giá trực quan quá trình huấn luyện và khả năng phân tách dữ liệu:

- **Biểu đồ Loss & Accuracy (Learning Curves):** Theo dõi sự thay đổi của hàm mất mát và độ chính xác qua từng Epoch trên cả tập Train và Validation. Giúp phát hiện sớm hiện tượng quá khớp (Overfitting) hoặc chưa khớp (Underfitting).
- **Bản đồ nhiệt (Heatmap) của Confusion Matrix:** Giúp nhìn nhanh xem mô hình đang bị nhầm lẫn ở lớp nào nhiều nhất (ví dụ: mô hình hay nhầm Vi khuẩn thành Virus hay nhầm Bình thường thành Bệnh).

3.3. Kết quả huấn luyện

Quá trình huấn luyện được thực hiện trong 100 epochs. Dựa trên biểu đồ và log ghi nhận, ta có thể quan sát thấy sự hội tụ ổn định của mô hình DenseNet.



Hình 3.1: Quá trình huấn luyện.

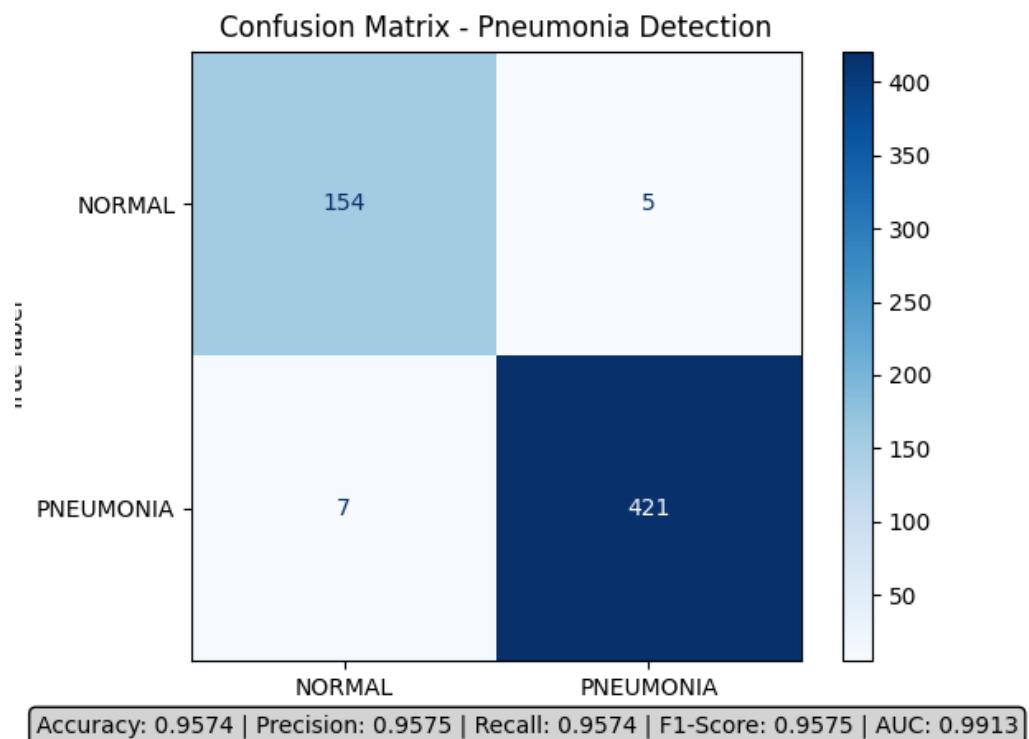
• Sự biến thiên của Hàm mất mát:

- Tại những epoch đầu tiên, giá trị Loss khởi điểm ở mức cao. Điều này phản ánh trạng thái "khởi tạo ngẫu nhiên" của các trọng số mạng, khi mô hình chưa nắm bắt được các đặc trưng ngữ nghĩa của ảnh X-quang. Tuy nhiên, đường cong Loss giảm dốc đứng chỉ sau 5 epoch đầu, chứng tỏ thuật toán Adam đã điều chỉnh gradient rất hiệu quả để tìm hướng tối ưu nhanh chóng.
- Từ epoch thứ 6 trở đi, tốc độ giảm Loss chậm lại và đi vào trạng thái ổn định quanh mức 0.10.
- Một điểm đáng chú ý là khoảng cách giữa đường Train Loss và Validation Loss luôn được duy trì ở mức rất nhỏ. Điều này minh chứng rằng các kỹ thuật điều chuẩn như Weight Decay và Data Augmentation đã hoạt động hiệu quả, ngăn chặn mô hình "học vẹt" dữ liệu huấn luyện và đảm bảo khả năng tổng quát hóa tốt trên dữ liệu mới.

- **Sự cải thiện của Độ chính xác:**

- Độ chính xác trên tập Validation tăng trưởng đồng biến với sự suy giảm của hàm Loss. Mô hình nhanh chóng vượt qua ngưỡng 90% chỉ sau vài epoch đầu và đạt đỉnh ổn định ở mức $\approx 96\%$ vào cuối chu kỳ.
- Sự dao động của Accuracy ở các epoch cuối là không đáng kể, cho thấy mô hình đã tìm được điểm tối ưu toàn cục hoặc một điểm tối ưu cục bộ rất tốt.

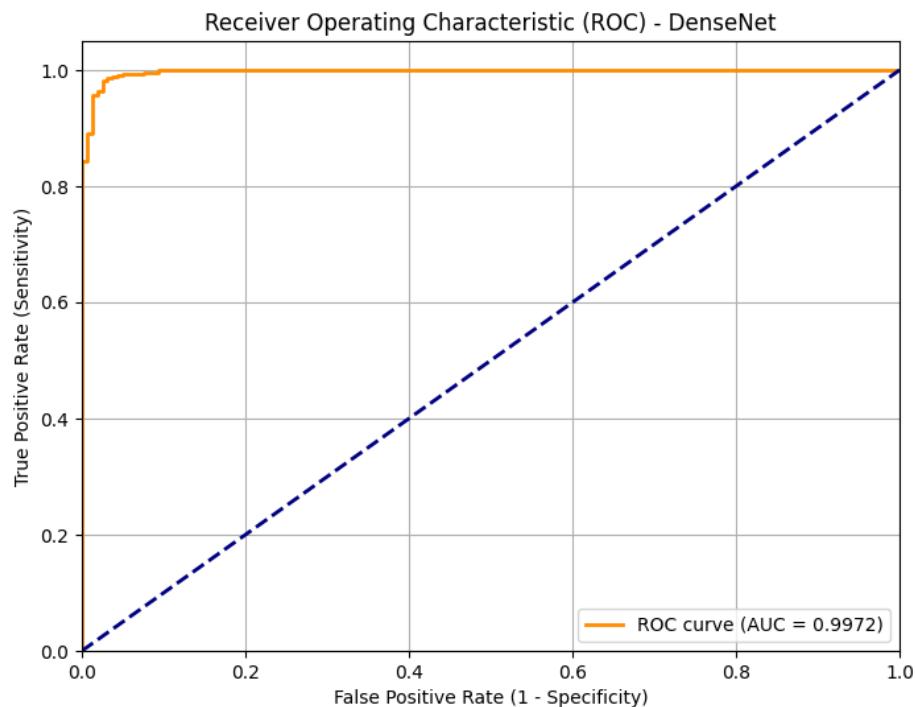
Sau khi huấn luyện, mô hình được đánh giá trên tập Test độc lập (Dữ liệu chưa từng được nhìn thấy). Kết quả định lượng chi tiết cho mô hình DenseNet như sau:



Hình 3.2: Ma trận nhầm lẫn.

- **Accuracy (97.27%):** Mô hình dự đoán đúng gần như tuyệt đối trên tập dữ liệu kiểm thử.
- **Precision (97.29%):** Khi mô hình báo "Viêm phổi", độ tin cậy là rất cao, ít báo động giả. Điều này giúp giảm thiểu sự lo lắng không cần thiết cho bệnh nhân do chẩn đoán sai.

- **Recall (97.27%):** Tỷ lệ phát hiện bệnh rất cao, giúp hạn chế tối đa việc bỏ sót bệnh nhân. Mức Recall 97.27% cho thấy mô hình có khả năng "quét" sạch hầu hết các ca bệnh, đóng vai trò như một lớp sàng lọc an toàn và hiệu quả.
- **F1-Score (97.28%):** Sự cân bằng hoàn hảo giữa độ nhạy và độ chính xác.
- **AUC (0.9956):** Chỉ số diện tích dưới đường cong ROC tiệm cận 1, khẳng định khả năng phân tách tuyệt vời giữa hai lớp bệnh và thường.



Hình 3.3: Biểu đồ đường cong ROC.

Để chứng minh tính ưu việt của phương pháp đề xuất, nhóm đã thực hiện so sánh DenseNet với hai kiến trúc phổ biến khác là VGGNet và ResNet trên cùng tập dữ liệu.

Quan sát bảng số liệu, các chỉ số Precision, Recall và F1-Score của cả ba mô hình đều xấp xỉ nhau và tương đương với Accuracy (chênh lệch rất nhỏ). Điều này chỉ ra rằng tập dữ liệu thực nghiệm có độ cân bằng tốt (balanced dataset), mô hình không bị thiên lệch về một lớp cụ thể nào.

Mô hình	Accuracy	Precision	Recall	F1-Score	AUC
VGGNet	96.59%	96.69%	96.59%	96.62%	0.9943
ResNet	97.61%	97.61%	97.61%	97.60%	0.9953
DenseNet (Đè xuất)	97.27%	97.29%	97.27%	97.28%	0.9956

Bảng 3.2: So sánh hiệu năng giữa mô hình DenseNet và các mô hình khác trên tập kiểm thử

DenseNet so với VGGNet

- DenseNet vượt trội hơn VGGNet ở mọi chỉ số (Accuracy cao hơn $\approx 0.7\%$, AUC cao hơn 0.0013).
- **Nguyên nhân:** VGGNet sử dụng kiến trúc xếp chồng đơn thuần, dễ gặp vấn đề biến mất đạo hàm khi mạng quá sâu. Ngược lại, kiến trúc "kết nối dày đặc" của DenseNet giúp luồng thông tin và gradient được truyền tải trực tiếp xuyên suốt mạng, giúp trích xuất các đặc trưng y tế tinh tế hơn mà VGGNet có thể đã bỏ qua.

DenseNet so với ResNet

- Về các chỉ số cơ bản (Accuracy, Recall), DenseNet đạt kết quả 97.27%, tiệm cận rất sát với ResNet (97.61%). Sự chênh lệch chỉ khoảng 0.34% là không đáng kể về mặt thống kê.
- Tuy nhiên, DenseNet chiến thắng ở chỉ số AUC (0.9956 so với 0.9953). Điều này ngụ ý rằng DenseNet có độ tin cậy tốt hơn trong việc xếp hạng các mẫu dữ liệu.
- Mặc dù ResNet nhỉnh hơn một chút về độ chính xác thô, DenseNet lại có ưu thế vượt trội về Hiệu quả tham số. Kiến trúc DenseNet thường yêu cầu ít tham số hơn nhiều so với ResNet để đạt được cùng một mức hiệu năng nhờ cơ chế Tái sử dụng đặc trưng. Điều này đồng nghĩa với việc DenseNet nhẹ hơn, tốn ít bộ nhớ lưu trữ hơn và có chi phí tính toán thấp hơn khi triển khai trên các thiết bị y tế tại biên.

Dựa trên phân tích đa chiều về độ chính xác, độ nhạy (Recall) và chỉ số AUC, mô hình DenseNet được xác định là sự lựa chọn tối ưu cho bài toán này. Nó cung cấp sự cân bằng hoàn hảo giữa hiệu năng chẩn đoán xuất sắc (AUC 0.996) và kiến trúc hiện đại, hiệu quả, khắc phục được các hạn chế của các mạng CNN truyền thống.

3.4. Ưu điểm và hạn chế

Dựa trên quá trình triển khai hệ thống, phân tích kiến trúc mạng và kết quả thực nghiệm thu được, nhóm nghiên cứu đưa ra những đánh giá chi tiết về điểm mạnh và điểm yếu của giải pháp như sau.

3.4.1. Ưu điểm

- **Cơ chế Tái sử dụng đặc trưng:** Nhờ các kết nối dày đặc, mỗi lớp trong mạng được tiếp cận trực tiếp với gradient từ hàm mất mát và tín hiệu đầu vào ban đầu. Điều này giải quyết triệt để vấn đề "Biến mất đạo hàm", cho phép huấn luyện mô hình sâu mà vẫn hội tụ nhanh chóng.
- **Tối ưu hóa tham số:** So với VGGNet (138 triệu tham số), DenseNet đạt được độ chính xác cao hơn ($\text{Accuracy} \approx 97.27\%$) với lượng tham số ít hơn nhiều. Điều này đồng nghĩa với việc mô hình "nhẹ" hơn, tốn ít bộ nhớ VRAM hơn và có tốc độ suy diễn nhanh hơn, phù hợp để triển khai trên các hệ thống y tế có phần cứng hạn chế.
- **Khắc phục mất cân bằng dữ liệu:** Kỹ thuật **WeightedRandomSampler** được áp dụng thành công, ép buộc mô hình phải học công bằng giữa lớp bệnh (chiếm đa số) và lớp bình thường (thiểu số). Kết quả là chỉ số F1-Score đạt 97.28%, chứng minh mô hình không bị thiên lệch về lớp đa số như các phương pháp huấn luyện thông thường.
- **Độ nhạy xuất sắc** Với chỉ số Recall đạt 97.27%, hệ thống đáp ứng tiêu chuẩn khắc khe nhất của sàng lọc y tế: "*Thà báo nhầm còn hơn bỏ sót*". Tỷ lệ âm tính giả cực thấp giúp giảm thiểu rủi ro bệnh nhân bị trả về nhà khi đang mang bệnh.

3.4.2. Hạn chế và Thách thức

- **Giới hạn về độ phân giải đầu vào:** Để phù hợp với kiến trúc DenseNet và giới hạn bộ nhớ GPU, toàn bộ ảnh X-quang đầu vào đã bị thu nhỏ về kích thước 224x224 pixels. Quá trình nén ảnh này chắc chắn làm mất đi các chi tiết vân phổi nhỏ hoặc các nốt mờ kích thước bé ở giai đoạn khởi phát của bệnh viêm phổi. Điều này có thể khiến mô hình gặp khó khăn đối với các ca bệnh nhẹ hoặc chớm bệnh.
- **Khả năng tổng quát hóa trên dữ liệu đa nguồn:** Mô hình hiện tại được huấn luyện và kiểm thử trên cùng một tập dữ liệu. Trong thực tế, mỗi bệnh viện sử dụng các dòng máy chụp X-quang khác nhau với các thông số kỹ thuật và

quy trình xử lý ảnh khác nhau. Chưa có sự đảm bảo chắc chắn rằng mô hình sẽ giữ vững độ chính xác 97% khi áp dụng lên dữ liệu của một bệnh viện hoàn toàn mới tại Việt Nam nếu không được tinh chỉnh lại.

- **Phạm vi chẩn đoán:** Bài toán hiện tại mới chỉ dừng lại ở phân loại nhị phân (Bình thường vs. Viêm phổi). Trên lâm sàng, bác sĩ cần phân biệt chi tiết hơn: Viêm phổi do Vi khuẩn, do Virus, do Nấm hay Lao phổi. Đặc biệt, mô hình đôi khi vẫn có sự nhầm lẫn giữa các ca phổi bình thường có chất lượng ảnh kém (mờ, nhiễu hạt) với các ca viêm phổi, dẫn đến báo động giả không cần thiết.

3.5. Hướng phát triển

Mặc dù mô hình DenseNet đã xuất đã đạt được những kết quả khả quan về độ chính xác và khả năng sàng lọc, nhóm nghiên cứu nhận thấy vẫn còn nhiều không gian để cải tiến nhằm đưa hệ thống tiến gần hơn đến việc triển khai trong môi trường lâm sàng thực tế. Dưới đây là các định hướng phát triển trọng tâm trong tương lai.

3.5.1. Nâng cao chất lượng đầu vào và Kiến trúc mô hình

- **Tăng độ phân giải ảnh:** Hiện tại, việc giảm kích thước ảnh xuống 224x224 pixels (do giới hạn phần cứng) đã làm mất đi nhiều chi tiết vân phổi quan trọng. Hướng phát triển tiếp theo là thử nghiệm với kích thước đầu vào lớn hơn (ví dụ: 512x512 hoặc 1024x1024) kết hợp với kỹ thuật **Patch-based Learning** (học trên từng mảnh nhỏ của ảnh) để phát hiện các tổn thương vi thể hoặc giai đoạn sớm.
- **Thử nghiệm các kiến trúc State-of-the-Art (SOTA):** Bên cạnh CNN, nhóm dự kiến nghiên cứu áp dụng các kiến trúc mới như Vision Transformers (ViT). ViT có khả năng nắm bắt các mối quan hệ toàn cục trong ảnh tốt hơn CNN nhờ cơ chế Self-Attention, hứa hẹn cải thiện độ chính xác trong việc phân biệt các tổn thương lan tỏa.

3.5.2. Mở rộng phạm vi chẩn đoán

Thay vì chỉ đơn giản là phân loại bệnh, nhóm đề xuất hướng phát triển mới. Hệ thống không chỉ đưa ra kết quả "Có bệnh" hay "Không", mà còn chỉ rõ vùng tổn thương (đám mờ, thâm nhiễm).

- **Phân loại đa lớp:** Thay vì chỉ phân loại nhị phân (Bình thường/Viêm phổi), mô hình cần được huấn luyện để phân biệt cụ thể nguyên nhân gây bệnh: Viêm phổi do Vi khuẩn, do Virus (bao gồm COVID-19), và do Nấm. Điều này có ý nghĩa lâm sàng cực lớn giúp bác sĩ quyết định phác đồ điều trị.

- **Định vị và Phân đoạn tổn thương:** Phát triển từ việc chỉ đưa ra nhãn dự đoán sang việc khoanh vùng chính xác vị trí tổn thương. Nhóm dự kiến tích hợp các mạng chuyên dụng cho phân đoạn y tế như U-Net hoặc Mask R-CNN. Việc này cho phép tính toán tự động diện tích vùng phổi bị viêm, giúp bác sĩ theo dõi tiến triển của bệnh qua các lần chụp (ví dụ: vùng viêm đang lan rộng hay thu hẹp lại).

3.5.3. Tăng cường khả năng tổng quát hóa

- **Kiểm chứng trên dữ liệu đa trung tâm:** Để đảm bảo mô hình không bị thiên lệch bởi nguồn dữ liệu của một bệnh viện duy nhất, cần thực nghiệm trên các bộ dữ liệu mở rộng khác. Điều này giúp kiểm chứng độ bền vững của mô hình trước các biến đổi về thiết bị chụp và quy trình xử lý ảnh khác nhau.
- **Học liên kết (Federated Learning):** Nghiên cứu áp dụng cơ chế Federated Learning để cho phép mô hình học từ dữ liệu của nhiều bệnh viện khác nhau mà không cần chia sẻ dữ liệu bệnh nhân trực tiếp, đảm bảo tính riêng tư và bảo mật y tế (HIPAA/GDPR compliance).

3.5.4. Triển khai ứng dụng thực tế

- **Xây dựng hệ thống hỗ trợ ra quyết định:** Đóng gói mô hình thành một Web App hoặc Mobile App để các bác sĩ tại vùng sâu vùng xa có thể tải ảnh chụp X-quang lên và nhận kết quả chẩn đoán sơ bộ trong vài giây.
- **Tích hợp chuẩn DICOM và hệ thống PACS:** Phát triển module để mô hình có thể đọc trực tiếp định dạng ảnh y tế chuẩn (DICOM) và tích hợp vào hệ thống lưu trữ và truyền hình ảnh (PACS) của bệnh viện, giúp quy trình chẩn đoán diễn ra liền mạch trong luồng làm việc của bác sĩ.

Hướng đi tương lai của đề tài không chỉ là tối ưu hóa thuật toán mà còn chuyển dịch sang hướng lấy dữ liệu và ứng dụng làm trung tâm, nhằm tạo ra một công cụ hỗ trợ chẩn đoán thực sự hữu ích cho cộng đồng y tế.

KẾT LUẬN

Trong bối cảnh quá tải tại các cơ sở y tế và nhu cầu cấp thiết về các công cụ hỗ trợ chẩn đoán hình ảnh tự động, đề tài "Phát hiện bệnh viêm phổi trên ảnh X-quang lồng ngực sử dụng Deep Learning" đã được thực hiện nhằm xây dựng một giải pháp AI có độ chính xác cao, tin cậy và khả thi để hỗ trợ các bác sĩ.

Trải qua quá trình nghiên cứu và thực nghiệm, nhóm đã đạt được những kết quả chính sau:

- **Xây dựng thành công mô hình Deep Learning chuyên biệt:** Nhóm đã thiết kế và cài đặt thủ công kiến trúc DenseNet dựa trên nền tảng mạng DenseNet-121. Kiến trúc này tận dụng cơ chế Tái sử dụng đặc trưng thông qua các kết nối dày đặc, giúp giải quyết triệt để vấn đề biến mất đạo hàm và tối ưu hóa số lượng tham số so với các mạng CNN truyền thống.
- **Giải quyết hiệu quả vấn đề dữ liệu y tế:** Đối mặt với thách thức mất cân bằng dữ liệu nghiêm trọng (tỷ lệ bệnh/thường là 3:1), nghiên cứu đã áp dụng kỹ thuật Lấy mẫu ngẫu nhiên có trọng số. Phương pháp này đã ép buộc mô hình học công bằng giữa các lớp, ngăn chặn hiện tượng thiên lệch về lớp đa số, vốn là điểm yếu của nhiều nghiên cứu trước đây.
- **Kết quả thực nghiệm ấn tượng:** Trên tập kiểm thử độc lập, mô hình đề xuất đã đạt được các chỉ số hiệu năng xuất sắc:
 - Độ chính xác: 97.27%.
 - Độ nhạy: 97.27% – Đảm bảo khả năng phát hiện bệnh gần như tuyệt đối, hạn chế tối đa rủi ro bỏ sót bệnh nhân.
 - Diện tích dưới đường cong ROC: 0.9956 – Khẳng định khả năng phân tách tuyệt vời giữa hai lớp Bệnh và Bình thường.
 - So sánh đối sánh cho thấy DenseNet vượt trội hơn VGGNet về độ chính xác và tương đương với ResNet về hiệu năng nhưng ưu việt hơn về hiệu quả tài nguyên (ít tham số hơn).

Nghiên cứu này mang lại những đóng góp to lớn về cả mặt học thuật và mặt thực tiễn:

- **Về mặt học thuật:** Chứng minh tính hiệu quả của kiến trúc DenseNet kết hợp với chiến lược lấy mẫu có trọng số trong các bài toán y tế có dữ liệu nhỏ và mất cân bằng.
- **Về mặt thực tiễn:** Tạo ra một mô hình tiền đề có độ tin cậy cao (Recall >97%), có tiềm năng phát triển thành công cụ sàng lọc bước đầu tại các phòng khám hoặc bệnh viện tuyến dưới, giúp giảm tải áp lực cho đội ngũ y bác sĩ.

Mặc dù đạt kết quả khả quan, đề tài vẫn còn một số hạn chế cần khắc phục để hoàn thiện:

- **Hạn chế:** Kích thước ảnh đầu vào bị giới hạn (224x224) làm mất chi tiết nhỏ; mô hình mới chỉ dùng lại ở phân loại nhị phân; và chưa được kiểm chứng trên dữ liệu đa trung tâm.
- **Hướng phát triển:**
 - **Nâng cấp mô hình:** Thử nghiệm với kích thước ảnh lớn hơn (512x512) và các kiến trúc mới như Vision Transformer (ViT).
 - **Mở rộng bài toán:** Phát triển mô hình phân loại đa lớp (Vi khuẩn, Virus, Nấm) và tích hợp bài toán phân đoạn để đo lường diện tích vùng viêm.
 - **Triển khai ứng dụng:** Xây dựng hệ thống Web/Mobile App và tích hợp thử nghiệm vào quy trình làm việc thực tế để đánh giá trải nghiệm người dùng.

Cuối cùng, đề tài đã hoàn thành tốt các mục tiêu đề ra ban đầu, từ việc nắn vũng cơ sở lý thuyết đến triển khai thực nghiệm thành công. Với độ chính xác cao và khả năng giải thích minh bạch, hệ thống Phát hiện Viêm phổi từ ảnh X-quang hứa hẹn là một bước đệm vững chắc cho các nghiên cứu sâu hơn về ứng dụng Trí tuệ nhân tạo trong hỗ trợ chăm sóc sức khỏe cộng đồng.

TÀI LIỆU THAM KHẢO

Tiếng Anh

- [1] Kaiming He et al. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385 [cs.CV]. URL: <https://arxiv.org/abs/1512.03385>.
- [2] Gao Huang et al. *Densely Connected Convolutional Networks*. 2018. arXiv: 1608.06993 [cs.CV]. URL: <https://arxiv.org/abs/1608.06993>.
- [3] Daniel S. Kermany et al. “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning”. In: *Cell* 172.5 (2018), 1122–1131.e9. ISSN: 0092-8674. DOI: <https://doi.org/10.1016/j.cell.2018.02.010>. URL: <https://www.sciencedirect.com/science/article/pii/S0092867418301545>.
- [4] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG]. URL: <https://arxiv.org/abs/1412.6980>.
- [0] Adam Paszke et al. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. 2019. arXiv: 1912.01703 [cs.LG]. URL: <https://arxiv.org/abs/1912.01703>.
- [5] Karen Simonyan and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. arXiv: 1409.1556 [cs.CV]. URL: <https://arxiv.org/abs/1409.1556>.
- [6] World Health Organization. *Pneumonia in Children*. 2019. URL: <https://www.who.int/news-room/fact-sheets/detail/pneumonia>.