

INVESTIGATING THE EFFECTIVENESS OF ADVANCED DEEP LEARNING MODELS IN PULMONARY DISEASE CLASSIFICATION USING CHEST X-RAY IMAGES

Nguyen Huu Khanh

Thai Nguyen University

ARTICLE INFO		ABSTRACT
Received:	25/7/2023	According to the WHO's Global Tuberculosis Report 2022, an estimated 10.6 million people are infected with Tuberculosis and 1.6 million people die from Tuberculosis worldwide. Meanwhile, pneumonia is estimated to cause more than 700,000 deaths each year globally according to the UNICEF. In Vietnam, these two diseases are still serious health problems with a large number of cases and deaths. To support effective lung disease diagnosis, the study applied deep learning techniques, especially typical CNN models such as Resnet, DenseNet, Xception, MobileNet and InceptionV3 to classify chest X-ray images. These models are applied fine-tune transfer learning and then trained and compared to find the model with the highest accuracy. The results show that the MobileNet model is the best, with an accuracy of up to 98.31%, the highest compared to other models. The research results can be deployed and developed into a medical support system that helps doctors and medical staff quickly and accurately identify lung diseases from X-ray images, thereby helping to improve the diagnosis and treatment process for patients.
Revised:	30/8/2023	
Published:	31/8/2023	
KEYWORDS		
Deep learning		
Convolutional Neural Network		
Pulmonary Disease		
X-Ray Images		
Transfer Learning		

NGHIÊN CỨU HIỆU QUẢ CỦA CÁC MÔ HÌNH HỌC SÂU TIỀN TIẾN TRONG PHÂN LOẠI BỆNH PHỔI BẰNG HÌNH ẢNH X-QUANG NGỰC

Nguyễn Hữu Khánh

Đại học Thái Nguyên

THÔNG TIN BÀI BÁO		TÓM TẮT
Ngày nhận bài:	25/7/2023	Theo báo cáo bệnh lao toàn cầu năm 2022 của WHO, ước tính có khoảng 10,6 triệu người mắc lao và 1,6 triệu người tử vong do lao trên toàn thế giới. Trong khi đó, viêm phổi được ước tính gây ra hơn 700.000 ca tử vong mỗi năm trên toàn cầu theo UNICEF. Tại Việt Nam, hai bệnh này vẫn là vấn đề y tế nghiêm trọng với số ca mắc và tử vong không nhỏ. Để hỗ trợ chẩn đoán bệnh phổi hiệu quả, nghiên cứu đã áp dụng các kỹ thuật học sâu, đặc biệt là các mô hình CNN tiêu biểu như Resnet, DenseNet, Xception, MobileNet và InceptionV3, để phân loại ảnh X-quang phổi. Các mô hình này được áp dụng kỹ thuật học chuyển tiếp tinh chỉnh mô hình sau đó được huấn luyện và so sánh để tìm ra mô hình có độ chính xác cao nhất. Kết quả cho thấy mô hình MobileNet là xuất sắc nhất, với độ chính xác lên đến 98,31%, cao nhất so với các mô hình khác. Kết quả nghiên cứu có thể được triển khai, phát triển thành hệ thống hỗ trợ y tế để giúp bác sĩ và nhân viên y tế xác định chính xác và nhanh chóng các loại bệnh phổi từ ảnh X-quang, từ đó giúp cải thiện quá trình chẩn đoán và điều trị cho bệnh nhân.
Ngày hoàn thiện:	30/8/2023	
Ngày đăng:	31/8/2023	
TỪ KHÓA		
Học sâu		
Mạng nơron tích chập		
Bệnh về phổi		
Ảnh X-quang		
Học chuyển tiếp		

DOI: <https://doi.org/10.34238/tnu-jst.8408>

Email: khanhnh@tnu.edu.vn

<http://jst.tnu.edu.vn>

45

Email: jst@tnu.edu.vn

1. Giới thiệu

1.1. Bối cảnh nghiên cứu

Viêm phổi và lao phổi là hai bệnh truyền nhiễm nguy hiểm ảnh hưởng đến phổi của con người. Theo báo cáo toàn cầu về lao của Tổ chức Y tế Thế giới (WHO) năm 2022, năm 2021 có khoảng 10,6 triệu người mắc lao và 1,6 triệu người tử vong do lao trên toàn thế giới [1]. Còn viêm phổi được ước tính gây ra hơn 700.000 ca tử vong mỗi năm trên toàn cầu theo Quỹ Nhi đồng Liên Hợp Quốc – Unicef [2]. Theo thống kê của WHO, mỗi năm thế giới ghi nhận khoảng 150 triệu trẻ bị viêm phổi, trong đó có tới 11 triệu trẻ phải nhập viện điều trị [3]. Tại Việt Nam, bệnh lao phổi vẫn là một trong những vấn đề y tế quan trọng, với khoảng 169.000 ca mắc mới và 12.000 ca tử vong trong năm 2021 [1]. Bệnh viêm phổi cũng không kém phần nguy hiểm tại Việt Nam, khi hàng năm có khoảng 2,9 triệu lượt trẻ mắc và hơn 4.000 trẻ chết vì viêm phổi [3].

Các bệnh phổi thường gặp như viêm phổi, lao phổi hay COVID-19 đều có thể gây ra các biến chứng nguy hiểm cho sức khỏe và tính mạng của người bệnh nếu không được phát hiện và điều trị kịp thời. Một trong những phương pháp chẩn đoán bệnh phổi hiệu quả và phổ biến là chụp X-quang phổi. Tuy nhiên, việc đọc và phân tích ảnh X-quang phổi đòi hỏi sự am hiểu và kinh nghiệm của các bác sĩ chuyên khoa, đồng thời cũng có thể bị ảnh hưởng bởi các yếu tố khách quan như điều kiện ánh sáng hay sự khác biệt về hình ảnh giữa các loại bệnh phổi. Do đó, việc áp dụng các kỹ thuật học sâu, cụ thể là các mô hình CNN (Convolutional Neural Network), để hỗ trợ chẩn đoán bệnh phổi từ ảnh X-quang là một hướng nghiên cứu có ý nghĩa và tiềm năng.

1.2. Những nghiên cứu liên quan

Trong lĩnh vực chẩn đoán hình ảnh y học, việc sử dụng công nghệ học sâu và mô hình CNN đã thúc đẩy nhiều nghiên cứu về phân loại ảnh X-quang để hỗ trợ các bác sĩ và chuyên gia y tế trong việc đánh giá và chẩn đoán các bệnh phổi, một vài nghiên cứu có liên quan được tổng hợp trong Bảng 1.

Bảng 1. Các nghiên cứu liên quan

Tác giả	Mô hình / Phương pháp	Nhãn dữ liệu	Hiệu quả (độ chính xác)
Muhammad Abubakar và cộng sự [4]	- CNN - VGG16 - Inception Net	- Bình thường - Viêm phổi - Lao phổi	92,97%
Rachita Mogaveera và cộng sự [5]	ConvNets	- Bình thường - Viêm phổi - Lao phổi	95% (cao nhất)
Sunny Kudum và cộng sự [6]	- InceptionResNetV2 - DenseNet121 - ResNet50 - AlexNet - ResNet152 - InceptionV3	- Bình thường - Lao phổi - Viêm phổi	94,42% (DenseNet121 cao nhất)
A.Pranaya và cộng sự [7]	CNN (ConvNet)	- Viêm phổi - Bình thường	91%
Aditya Nair và cộng sự [8]	CNN	- Lao phổi - Bình thường	98,14%

1.3. Mục tiêu của nghiên cứu

Nghiên cứu này sử dụng dataset ảnh X-quang phổi chia làm 3 nhóm: viêm phổi, lao phổi, và bình thường để huấn luyện và so sánh hiệu suất của các mô hình CNN nổi tiếng như Resnet [9], DenseNet [10], Xception [11], MobileNet [12], InceptionV3 [13]. Các mô hình này được chọn vì chúng đã được chứng minh là có độ chính xác cao và tốc độ nhanh trong các cuộc thi và nghiên cứu.

cứu trước đây. Nghiên cứu mong muốn tìm ra mô hình CNN tốt nhất để phân loại ảnh X-quang phổi và góp phần cải thiện chất lượng chẩn đoán và điều trị bệnh phổi.

2. Phương pháp nghiên cứu

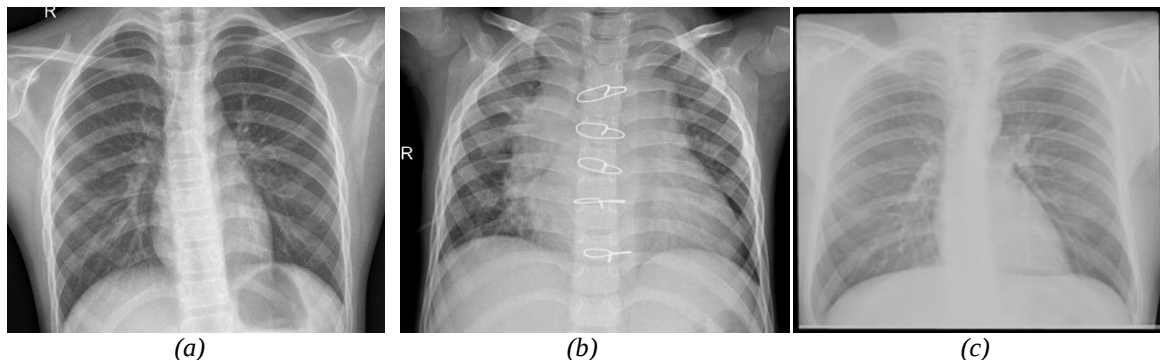
2.1. Dataset

Do ảnh X-quang phổi của các bệnh nhân là thông tin cá nhân, việc công bố công khai rất hạn chế nên nguồn dữ liệu công khai về ảnh X-quang phổi cũng bị giới hạn. Toàn bộ ảnh dữ liệu được dùng cho nghiên cứu này vì thế cũng phải được xây dựng từ hai nguồn dữ liệu khác nhau được công bố công khai trên website Kaggle, hai nguồn dữ liệu này được gán nhãn chính xác bởi các bác sĩ chuyên gia trong lĩnh vực y học [14], [15]. Thông tin chi tiết của từng nguồn dữ liệu hình thành nên dataset được ghi rõ trong Bảng 2.

Bảng 2. Thông tin chi tiết nguồn dữ liệu

Dataset	Lao phổi	Viêm phổi	Bình thường
Tuberculosis (TB) Chest X-ray Database [14]	3500	0	3500
Chest X-Ray Images (Pneumonia) [15]	0	4273	1583
Tổng hợp	3500	4273	5083

Từ hai nguồn dữ liệu trên, một tập dữ liệu mới được hình thành với 3500 ảnh lao phổi, 4273 ảnh viêm phổi và 5083 ảnh bình thường (người không bị nhiễm bệnh). Toàn bộ số ảnh trên được lấy ra ngẫu nhiên 30% dành cho việc kiểm thử (testing). Số lượng ảnh còn lại được chia tiếp ra 70% cho huấn luyện (training) và 30% cho việc xác thực (validation).



Hình 1. Ảnh X-Quang về phổi được công khai trên Kaggle (a) Bình thường (b) Viêm phổi (c) Lao phổi

Dữ liệu ảnh X-quang công khai trên đây được thu thập từ nhiều nguồn khác nhau nên kích thước của ảnh không đồng đều, có rất nhiều loại kích thước ảnh khác nhau. Tác giả thống kê được 4399 loại kích thước ảnh trong tổng số 12856 ảnh dữ liệu. Điều này gây khó khăn nhất định cho việc huấn luyện và xác định độ chính xác của các mô hình. Để giải quyết vấn đề này, tiền xử lý dữ liệu là yêu cầu bắt buộc.

2.2. Tiền xử lý dữ liệu – Data preprocessing

Tiền xử lý dữ liệu là một bước quan trọng trong việc xây dựng các mô hình học sâu cho phân loại ảnh, nghiên cứu này sử dụng hai kỹ thuật tiền xử lý ảnh đó là thay đổi độ phân giải (resize) và rescaling.

2.2.1. Thay đổi độ phân giải – Resize

Các mô hình CNN được sử dụng trong nghiên cứu này có hai độ phân giải của ảnh đầu vào là 224x224 (MobileNet, ResNet, DenseNet) và 299x299 (InceptionV3, Xception). Thêm vào đó, như đã trình bày ở trên, độ phân giải của ảnh trong dataset rất khác nhau. Để đảm bảo sự nhất quán và khả năng so sánh giữa các mô hình, cần phải thay đổi toàn bộ ảnh về một độ phân giải

duy nhất. Vì đa số mô hình CNN được nghiên cứu sử dụng độ phân giải 224×224 làm độ phân giải đầu vào nên độ phân giải này được chọn làm độ phân giải tiêu chuẩn của ảnh trong dataset. Việc chọn một kích thước cố định cho dataset như vậy giúp dữ liệu được đồng nhất và đảm bảo rằng các mô hình CNN tiếp nhận ảnh đầu vào một cách nhất quán.

2.2.2. Rescaling

Nghiên cứu này thực hiện rescaling của ảnh bằng cách chia giá trị mỗi điểm ảnh cho 255. Quá trình này giúp đưa các giá trị điểm ảnh về khoảng từ 0 đến 1, làm giảm độ phức tạp tính toán và tăng hiệu suất của mô hình. Điều này làm giảm phạm vi giá trị điểm ảnh ban đầu và chuẩn hóa dữ liệu thành một phạm vi phù hợp cho quá trình huấn luyện, giúp cải thiện hiệu suất và sự ổn định của quá trình huấn luyện và kiểm thử các mô hình CNN.

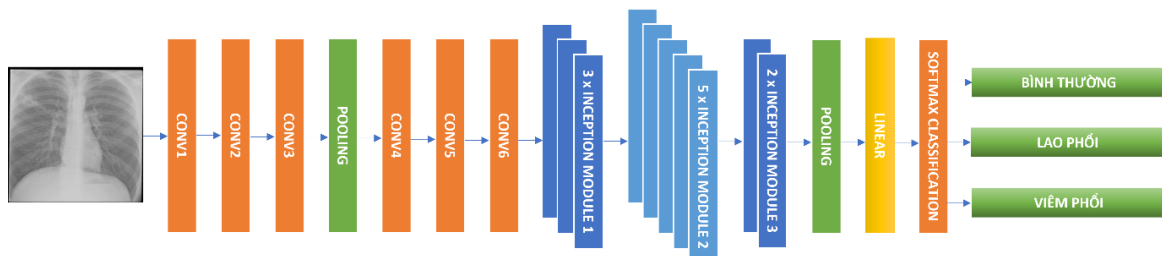
2.3. Học chuyển tiếp – Transfer learning

Học chuyển tiếp là một kỹ thuật hữu ích để tận dụng các mô hình CNN đã được huấn luyện trước trên các tập dữ liệu lớn và phong phú, như ImageNet, để giải quyết các bài toán phân loại ảnh có kích thước và chất lượng dữ liệu giới hạn. Có hai hướng học chuyển tiếp là: trích xuất đặc trưng (feature extraction) và tinh chỉnh mô hình (fine-tuning).

Nghiên cứu này áp dụng hướng học chuyển tiếp thông qua việc tinh chỉnh mô hình từ một mô hình đã được huấn luyện trên bộ dữ liệu ImageNet. Cụ thể, tác giả chọn mở băng (unfreeze) 20 lớp cuối của các mô hình, bởi vì các lớp đầu tiên thường học được những đặc trưng cơ bản như cạnh, góc, và các đặc điểm hình học đơn giản. Việc mở băng các lớp cuối cùng cho phép cập nhật trọng số của các lớp này dựa trên dữ liệu mới của nghiên cứu. Qua việc tinh chỉnh các lớp này, mô hình có khả năng học các đặc trưng phức tạp và có thể thích ứng với tác vụ mới mà không cần phải huấn luyện lại toàn bộ mô hình từ đầu giúp tiết kiệm thời gian và công sức đáng kể trong quá trình huấn luyện các mô hình.

2.4. Các mô hình CNN được sử dụng

2.4.1. InceptionV3



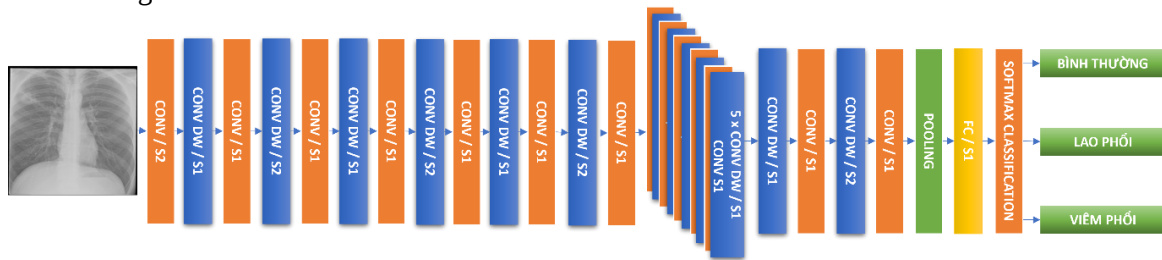
Hình 2. Kiến trúc mô hình InceptionV3 áp dụng trong nghiên cứu

InceptionV3 là mô hình mạng nơ-ron tích chập được tạo bởi tác giả Christian Szegedy và cộng sự vào năm 2015 [13]. Mô hình này có khoảng 48 lớp và là phiên bản cải tiến của InceptionV1. InceptionV3 nổi bật với khả năng mạnh mẽ trong việc xử lý hình ảnh và nhận dạng đối tượng, nó đã đạt được hiệu suất cao và trở thành một trong những mô hình phổ biến trong lĩnh vực học sâu. Hình 2 giải thích mô hình InceptionV3 được tinh chỉnh của nghiên cứu này và áp dụng vào phân loại lao phổi, viêm phổi và bình thường.

2.4.2. MobileNet

MobileNet là một mô hình mạng nơ-ron tích chập được giới thiệu bởi tác giả Andrew G. Howard và cộng sự vào năm 2017 [12]. Mô hình này ra đời nhằm giải quyết bài toán nhận dạng hình ảnh trên các thiết bị di động với tốc độ nhanh và tài nguyên thấp. Với số lượng lớp thấp, mô hình này thích hợp cho các ứng dụng có tài nguyên hạn chế. Hình 3 giải thích mô hình

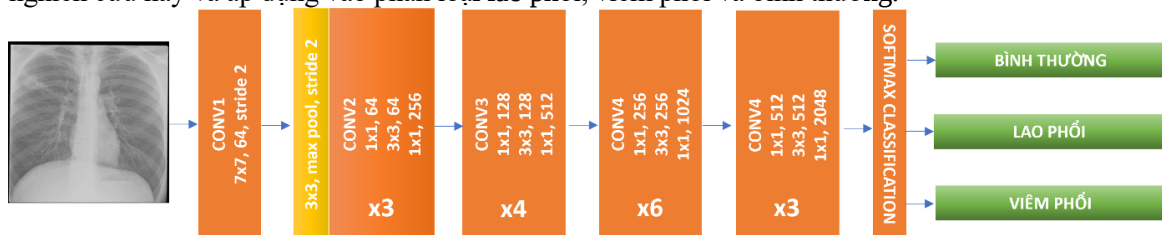
MobileNet được tinh chỉnh của nghiên cứu này và áp dụng vào phân loại lao phổi, viêm phổi và bình thường.



Hình 3. Kiến trúc mô hình MobileNet áp dụng trong nghiên cứu

2.4.3. ResNet

ResNet (Residual Neural Network) là mô hình nổi tiếng trong lĩnh vực học sâu, được tạo ra bởi Kaiming He và cộng sự vào năm 2015 [9]. Mô hình này đã đem đến một đột phá quan trọng trong giải quyết vấn đề vanishing gradient (vấn đề khi huấn luyện mạng nơ ron nhiều lớp, giá trị đạo hàm của hàm mất mát trở nên vô cùng nhỏ tại các lớp đầu tiên khiến việc cập nhật trọng số mạng không thể xảy ra dẫn đến huấn luyện không hiệu quả) khi xây dựng mạng nơ-ron sâu. Với ý tưởng của "residual blocks", ResNet cho phép xây dựng các mạng nơ-ron có số lượng lớp lớn, lên tới hàng trăm hay hàng nghìn lớp, mà vẫn duy trì hiệu suất tốt và dễ dàng huấn luyện. Nghiên cứu này sử dụng ba mô hình được đề xuất trong bài báo [9] là ResNet50, ResNet101, ResNet152 với số lớp tương ứng lần lượt là 50, 101 và 152. Hình 4 giải thích mô hình ResNet50 được tinh chỉnh của nghiên cứu này và áp dụng vào phân loại lao phổi, viêm phổi và bình thường.



Hình 4. Kiến trúc mô hình Resnet50 áp dụng trong nghiên cứu

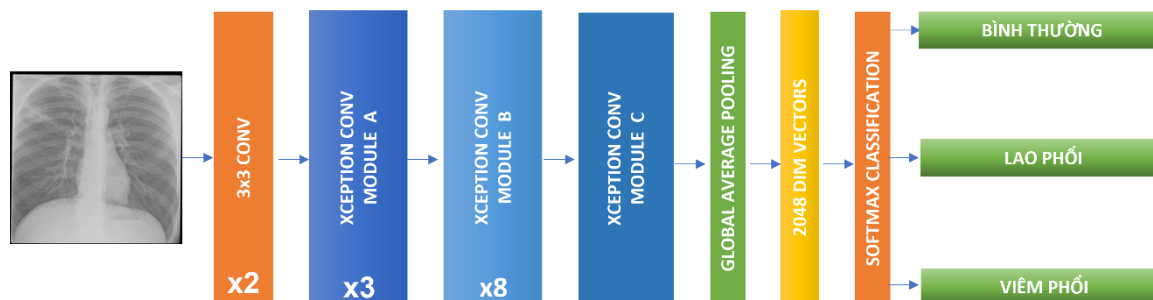
2.4.4. DenseNet

DenseNet là mô hình mạng nơ-ron tích chập đáng chú ý, được giới thiệu bởi Gao Huang và cộng sự vào năm 2016 [10]. Mô hình này sử dụng kiến trúc đặc biệt có tên là "Dense Blocks", trong đó các lớp trong mỗi block kết nối chặt chẽ với nhau thông qua việc nối đầu vào với đầu ra của các lớp trước đó. Điều này giúp giảm số lượng tham số và khắc phục vấn đề vanishing gradient, đồng thời cải thiện hiệu suất huấn luyện. DenseNet đã chứng tỏ hiệu quả và ứng dụng rộng rãi trong việc giải quyết các bài toán học sâu. Nghiên cứu này sử dụng ba mô hình được đề xuất trong bài báo [10] là DenseNet121, DenseNet169, DenseNet201 với số lớp tương ứng lần lượt là 121, 169 và 201. Hình 5 giải thích mô hình DenseNet201 được tinh chỉnh của nghiên cứu này và áp dụng vào phân loại lao phổi, viêm phổi và bình thường.



Hình 5. Kiến trúc mô hình DenseNet121 áp dụng trong nghiên cứu

2.4.5. Xception



Hình 6. Kiến trúc mô hình Xception áp dụng trong nghiên cứu

Xception là mô hình mạng nơ-ron tích chập tiên tiến, được giới thiệu bởi tác giả François Chollet vào năm 2016. Tên gọi "Xception" xuất phát từ việc kết hợp hai từ "Extreme Inception", mô hình này có 36 lớp tích chập được cấu trúc thành 14 mô-đun. Điều này giúp giảm số lượng tham số và tăng hiệu suất mạng, làm cho Xception trở thành một trong những mô hình tiên tiến trong lĩnh vực thị giác máy tính. Hình 6 giải thích mô hình Xception được tinh chỉnh của nghiên cứu này và áp dụng vào phân loại lao phổi, viêm phổi và bình thường.

2.5. Huấn luyện các mô hình

Tất cả các mô hình CNN được áp dụng trong nghiên cứu (MobileNet, ResNet, DenseNet, InceptionV3, Xception) đều được huấn luyện trên cùng một cấu hình phần cứng, thư viện hỗ trợ cũng như các cài đặt thông số mô hình.

Về phần cứng và thư viện hỗ trợ, quá trình huấn luyện của các mô hình đã được thực hiện trên một máy tính sử dụng CPU Core i7, RAM 16GB và GPU RTX 3060 với dung lượng 12GB. Để thực hiện triển khai nhanh chóng thuật toán các mô hình CNN, nghiên cứu đã sử dụng thư viện Tensorflow 2.11 cùng với Keras API. Thư viện này đặc biệt được thiết kế và phát triển cho mục đích học sâu, cung cấp các công cụ và chức năng quan trọng để xây dựng và huấn luyện các mô hình mạng nơ-ron. Tổ hợp phần cứng và thư viện hỗ trợ này đã cho phép nghiên cứu tiến hành huấn luyện các mô hình CNN trong việc phân loại ảnh X-quang phổi một cách hiệu quả và nhanh chóng.

Về cài đặt thông số, các mô hình CNN được thiết lập với lớp kết nối đầy đủ cuối cùng bao gồm 3 lớp đầu ra để thực hiện phân loại. Chúng được sử dụng hàm kích hoạt Softmax. Quá trình huấn luyện của các mô hình đã được thực hiện trong 50 epochs (vòng huấn luyện), với một tốc độ học (learning rate) ban đầu là 0,001 trong 20 epochs đầu tiên, sau đó giảm đi 10 lần sau mỗi 20 epochs nhằm tránh tình trạng overfitting.

Tác giả cũng đã triển khai kỹ thuật Checkpoint, dựa trên độ chính xác trên bộ xác thực (val_accuracy: độ chính xác trên tập xác thực), nhằm lưu lại bộ trọng số tốt nhất cho mỗi mô hình. Điều này rất hữu ích trong việc đánh giá và so sánh hiệu suất của các mô hình.

Bằng việc lưu lại bộ trọng số tốt nhất, nghiên cứu có thể sử dụng chúng sau này để đánh giá và so sánh các mô hình một cách nhất quán. Quá trình này giúp nghiên cứu có được một cơ sở xác đáng cho việc so sánh các mô hình với nhau và đánh giá hiệu suất của chúng trong nghiên cứu này.

2.6. Đánh giá các mô hình

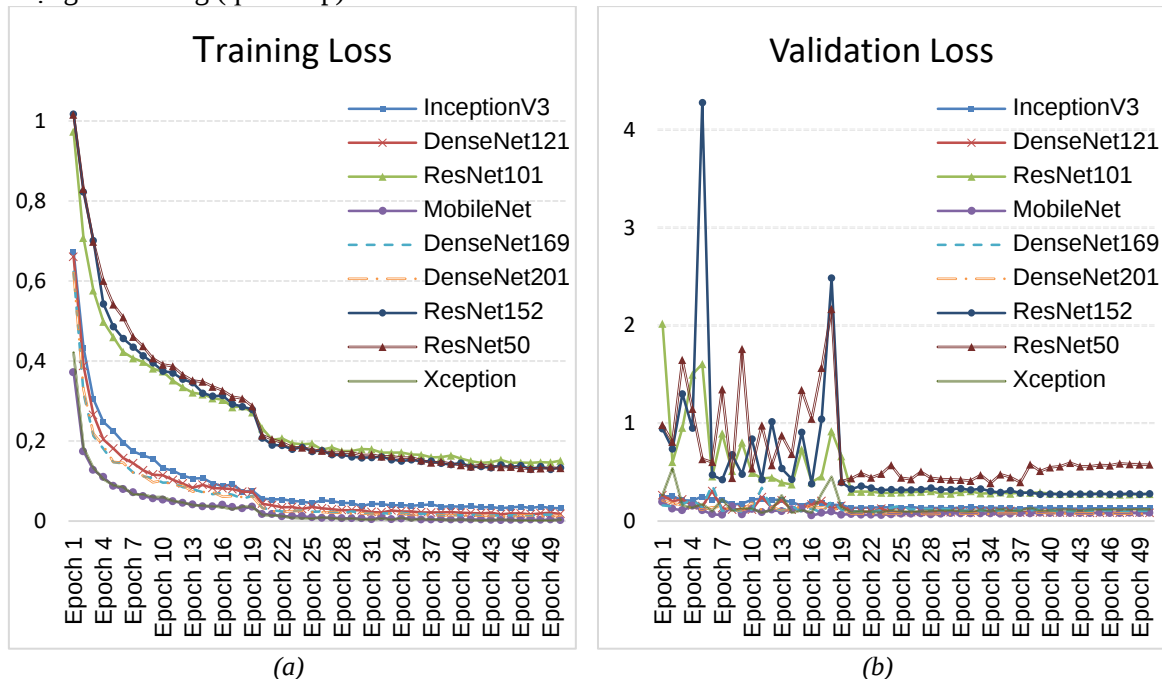
Sau khi hoàn tất quá trình huấn luyện và thu thập được bộ trọng số tối ưu dựa trên độ chính xác trên tập dữ liệu xác thực (validation), các mô hình CNN đã được đánh giá trên bộ dữ liệu kiểm tra bằng chức năng Keras Evaluation. Để đảm bảo tính công bằng trong việc so sánh, nghiên cứu tiếp tục sử dụng cùng một hệ thống phần cứng và thư viện hỗ trợ đã được sử dụng trong quá trình huấn luyện. Điều này đảm bảo rằng các mô hình được đánh giá dưới cùng một

điều kiện và đồng nhất. Các mô hình được đánh giá dựa trên hai độ đo phổ biến là độ chính xác (accuracy), độ lỗi (loss). Điều này giúp chúng ta có cái nhìn toàn diện về khả năng phân loại của từng mô hình và đánh giá được hiệu suất của chúng trong bối cảnh ứng dụng thực tế.

3. Kết quả và bàn luận

Sau quá trình huấn luyện các mô hình CNN với tập dữ liệu huấn luyện là 12856 ảnh X-quang phổi được chia thành ba nhãn dữ liệu khác nhau như trong Bảng 2, nghiên cứu đã thu được một số kết quả rất khả quan thông qua hai chỉ số cơ bản là độ lỗi và độ chính xác của quá trình huấn luyện và xác thực.

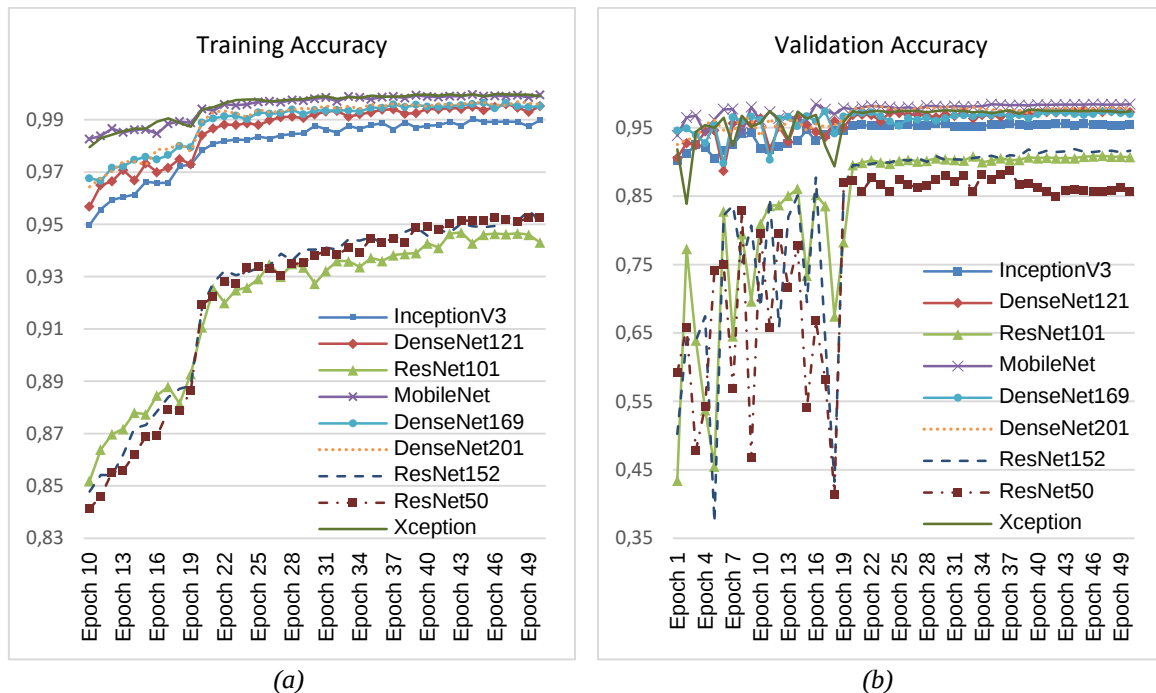
Hình 7 cho thấy độ lỗi của các mô hình tiến dần về 0, điều này thể hiện rằng mô hình đang học cách dự đoán đúng kết quả và hiệu suất của nó đang được cải thiện. Đây là một dấu hiệu tích cực và phản ánh sự tiến bộ của mô hình trong việc học từ dữ liệu huấn luyện và đánh giá độ chính xác của nó trên dữ liệu xác thực. Tuy nhiên, có thể thấy độ lỗi các mô hình ResNet50, ResNet152, ResNet101 không hội tụ trên dữ liệu xác thực, đặc biệt mô hình ResNet50 có hiện tượng overfitting (quá khớp).



Hình 7. Độ lỗi trong quá trình huấn luyện (a) và xác thực (b)

Hình 8 cho thấy độ chính xác của các mô hình tiến dần về 1, điều này thể hiện rằng mô hình đang có hiệu suất rất cao trong việc dự đoán đúng kết quả trên dữ liệu huấn luyện và dữ liệu xác thực. Tuy nhiên, cũng giống như độ lỗi, các mô hình ResNet50, ResNet101, ResNet152 cho thấy độ chính xác không hội tụ trên tập dữ liệu xác thực và có hiện tượng overfitting.

Sau quá trình đánh giá mô hình, Bảng 3 tổng hợp kết quả và cho thấy rõ được độ chính xác và độ lỗi của các mô hình đối với những dữ liệu mới. Kết quả cho thấy các mô hình đều đạt hiệu quả tốt khi ứng dụng để phân loại trong thực tế và mô hình tối ưu nhất cho dataset trong nghiên cứu là mô hình MobileNet với độ chính xác cao nhất lên tới 98,31% trong khi độ lỗi là thấp nhất chỉ với xấp xỉ 0,06376. Đây cũng là kết quả tốt nhất thu được khi so sánh với các nghiên cứu trước đó được tổng hợp tại Bảng 1.



Hình 8. Độ chính xác trong quá trình huấn luyện (a) và xác thực (b)

Bảng 3. Kết quả đánh giá các mô hình CNN

STT	Mô hình	Độ lỗi	Độ chính xác
1	DenseNet121	0,095534056	97,08%
2	DenseNet169	0,091507986	97,00%
3	DenseNet201	0,088077836	97,03%
4	InceptionV3	0,135143042	95,50%
5	MobileNet	0,063760318	98,31%
6	ResNet101	0,329211026	89,44%
7	ResNet152	0,294413775	90,53%
8	ResNet50	0,384447634	88,37%
9	Xception	0,124627501	97,27%

4. Kết luận

Nghiên cứu này đã thực hiện việc so sánh và đánh giá hiệu suất của các mô hình CNN tiêu biểu như Resnet, DenseNet, Xception, MobileNet và InceptionV3 trong việc phân loại ảnh X-quang thành ba nhóm: viêm phổi, lao phổi và bình thường. Kết quả cho thấy mô hình MobileNet đã đạt độ chính xác cao nhất, đến 98,31%, vượt trội so với các mô hình khác trong việc phân loại các loại bệnh phổi. Sự so sánh giữa các mô hình đã giúp rút ra các kết luận chính xác và đáng tin cậy về hiệu quả của từng mô hình trong ứng dụng y tế.

Dựa trên những kết quả tích cực này, tác giả đề xuất mở rộng nghiên cứu để tăng cường độ chính xác và tính ổn định của mô hình MobileNet trong việc phân loại các bệnh phổi. Điều này có thể được thực hiện thông qua việc tăng cường dữ liệu, thử nghiệm các siêu tham số khác nhau và điều chỉnh kiến trúc mô hình, đặc biệt là kết hợp bài toán phân vùng ảnh (Image Segmentation) để có thể phân tách phổi từ ảnh X-quang và giảm thiểu các yếu tố nhiễu.

Bên cạnh đó, nghiên cứu cũng khuyến nghị các nghiên cứu tương lai nghiên cứu sâu hơn về khía cạnh diễn giải của mô hình, nhằm giải thích quyết định của mô hình và cung cấp thông tin hữu ích cho bác sĩ và chuyên gia y tế. Điều này sẽ giúp cải thiện tính minh bạch và đáng tin cậy của mô hình trong việc hỗ trợ quyết định chẩn đoán.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] T. Kasaeva, K. Floyd, A. Baddeley, S. d. Boon, A. Dean, H. M. Dias, D. Falzon, K. Floyd, I. G. Baena, N. Gebreselassie, P. Glaziou, M. Lalli, I. Law, and P. Nguhiu, *Global tuberculosis report 2022*, World Health Organization, 2022.
- [2] UNICEF, "Childhood pneumonia: Everything you need to know," UNICEF, November 08, 2022. [Online]. Available: <https://www.unicef.org/stories/childhood-pneumonia-explained>. [Accessed July 15, 2023].
- [3] P. Thuong, "Experts show how to detect children with pneumonia early at home (In Vietnamese)," *Health and Life*, November 03, 2022. [Online]. Available: <https://suckhoedoisong.vn/chuyen-gia-chi-cach-phat-hien-tre-mac-viem-phoi-som-nhat-tai-nha-169221102231244035.htm>. [Accessed July 15, 2023].
- [4] M. Abubakar, I. Shah, W. Ali, and F. Bashir, "Classification of Pneumonia and Tuberculosis from Chest X-rays," March 25, 2021. [Online]. Available: <https://arxiv.org/abs/2103.14562>. [Accessed July 15, 2023].
- [5] R. Mogaveera, R. Maur, Z. Qureshi, and Y. Mane, "Multi-class Chest X-ray classification of Pneumonia, Tuberculosis and Normal X-ray images using ConvNets," *ITM Web of Conferences*, vol. 44, 2022, Art. no. 03007, doi: 10.1051/itmconf/20224403007.
- [6] S. Kudum, D. Shah, I. Vaish, and P. Tirumala, "Determining the best convolutional neural network for identifying tuberculosis and pneumonia in chest x-rays," *The Journal of Emerging Investigators*, vol. 5, 2022. [Online]. Available: <https://emerginginvestigators.org/articles/21-047>. [Accessed July 15, 2023].
- [7] A. Pranaya, D. V. Sowmya, L. Poojitha, P. Grace, K. Bhavya and Dr. N. V. Ganapathi Raju, "Pneumonia Detection Using Deep Learning," *E3S Web of Conferences*, vol. 391, 2023, doi: 10.1051/e3sconf/202339101067.
- [8] A. Nair, A. Mohan, J. J. Fernandez, S. B. Seema, and V. G. Sreena, "Chest X-ray Image Classification for Tuberculosis using Deep Convolutional Neural Network," *National Conference on Smart Systems and Technologies*, vol. 8, no. 7, pp. 96-103, 2021.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770-778, doi: 10.48550/arXiv.1512.03385.
- [10] G. Huang, Z. Liu, L. v. d. Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700-4708. [Online]. Available: <https://arxiv.org/abs/1608.06993>. [Accessed July 15, 2023].
- [11] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1251-1258. [Online]. Available: <https://arxiv.org/abs/1610.02357>. [Accessed July 15, 2023].
- [12] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv*, 2017. [Online]. Available: <https://arxiv.org/abs/1704.04861>. [Accessed July 15, 2023].
- [13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818-2826. [Online]. Available: <https://arxiv.org/abs/1512.00567>. [Accessed July 15, 2023].
- [14] T. Rahman, A. Khandakar, M. A. Kadir, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Ayari, and M. E. H. Chowdhury, "Reliable Tuberculosis Detection using Chest X-ray with Deep Learning, Segmentation and Visualization," *IEEE Access*, vol. 8, pp. 191586 - 191601, 2020.
- [15] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification," *Mendeley Data*, 2018. [Online]. Available: <https://dx.doi.org/10.17632/rscbjbr9sj.2>. [Accessed July 15, 2023].