**TRƯỜNG ĐẠI HỌC CÔNG NGHIỆP TPHCM**

**KHOA CÔNG NGHỆ THÔNG TIN**



**ĐỒ ÁN MÔN HỌC**

**TÊN ĐỀ TÀI**

**GIẢNG VIÊN HƯỚNG DẪN:**

**SINH VIÊN THỰC HIỆN 1:**

**MÃ SỐ SINH VIÊN:**

**SINH VIÊN THỰC HIỆN 2:**

**MÃ SỐ SINH VIÊN:**

*TP HỒ CHÍ MINH, Tháng 04 năm 2025*

**INDUSTRIAL UNIVERSITY OF HO CHI MINH CITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**GRADUATION THESIS**

**PNEUMONIA DETECTION USING DEEP LEARNING MODELS**

**SUPERVISOR: MSc. Vo Quang Hoang Khang**

**STUDENT NAME:**

**STUDENT CODE:**

**STUDENT NAME:**

**STUDENT CODE: 20079801**

*HO CHI MINH CITY, Month 04 year 2025*

**TÓM TẮT**

Nghiên cứu này trình bày một mô hình học sâu cải tiến nhằm hỗ trợ chẩn đoán viêm phổi từ ảnh X-quang lồng ngực, một vấn đề cấp thiết, đặc biệt sau đại dịch COVID-19. Mô hình được đánh giá trên tập dữ liệu X-quang từ Kaggle, sử dụng các chỉ số như Accuracy, F1 score, Precision, Recall, Specificity và AUC. Kết quả cho thấy mô hình đạt độ chính xác cao, hứa hẹn khả năng ứng dụng trong lâm sàng để hỗ trợ bác sĩ chẩn đoán chính xác hơn, giảm thiểu sai sót.

**LỜI CẢM ƠN**

Trong suốt thời gian làm đồ án tốt nghiệp, chúng em đã luôn nhận được nhiều sự quan tâm, hướng dẫn và giúp đỡ tận tình của các thầy cô giáo trong khoa Công nghệ thông tin cùng với sự động viên giúp đỡ từ bạn bè và gia đình.

Lời đầu tiên em xin chân thành cảm ơn Ban giám hiệu Trường Đại học Công nghiệp thành phố Hồ Chí Minh, Ban chủ nhiệm khoa Công nghệ thông tin đã luôn tận tình quan tâm giúp đỡ chúng em trong suốt thời gian học tập tại trường.

Đặc biệt chúng em xin gửi lời cảm ơn chân thành và sâu sắc tới thầy hướng dẫn ThS. Võ Quang Hoàng Khang đã trực tiếp hướng dẫn, giúp đỡ chúng em hoàn thành khoá luận này.

Chúng em cũng xin gửi lời cảm ơn đến gia đình, người thân và bạn bè đã luôn ở bên quan tâm, giúp đỡ và động viên chúng em hoàn thành khoá luận tốt nghiệp này.

Chúng em xin trân trọng cảm ơn!

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN**

**TP. Hồ Chí Minh, ngày…. tháng …. năm….**

CHỮ KÝ CỦA GIẢNG VIÊN

**NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN 1**

**TP. Hồ Chí Minh, ngày…. Tháng …. năm….**

CHỮ KÝ CỦA GIẢNG VIÊN

**NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN 2**

**TP. Hồ Chí Minh, ngày…. tháng …. năm….**

CHỮ KÝ CỦA GIẢNG VIÊN

**LỜI MỞ ĐẦU**

Trong thời đại kỷ nguyên số hiện nay, các mô hình máy học và học sâu đóng vai trò rất quan trọng trong mọi lĩnh vực đời sống. Để đáp ứng nhu cầu ngày càng tăng cao về các mô hình công nghệ, ngành Khoa học máy tính luôn cần những chuyên gia có kiến thức chuyên sâu và kỹ năng thực hành để phát triển các mô hình máy học và học sâu mới, từ đó đáp ứng được các thách thức, nhu cầu của thế giới ngày nay.

Trong bối cảnh đó, chúng tôi chọn chủ đề “Phát hiện viêm phổi bằng các mô hình học sâu” cho khoá luận tốt nghiệp của chúng tôi. Đây là một chủ đề đang rất được quan tâm và yêu thích trong dịch COVID-19 vừa qua. Nó giúp thời gian nhận diện, chuẩn đoán và chữa bệnh của các bác sĩ nhanh hơn giảm thiểu số ca tử vong do phát hiện quá muộn của bệnh COVID-19.

Bài khoá luận tốt nghiệp của chúng tôi tập trung vào việc nghiên cứu và cải tiến mô hình phát hiện viêm phổi sử dụng các cơ chế chú ý, tập trung và nhóm tích chập. Mô hình cải tiến được đánh giá trên tập dữ liệu về ảnh chụp X-quang lồng ngực có sẵn trên kaggle sử dụng các chỉ số đánh giá phổ biến như: Độ chính xác (accuracy), F1 score, precision, recall, specsificity, AUC. Kết quả thực nghiệm cho thấy mô hình đề xuất có khả năng phát hiện viêm phổi trên ảnh chụp X-quang phổi với độ chính xác cao, từ đó làm nổi bật khả năng đầy hứa hẹn của nó trong việc triển khai mô hình lâm sàng trong thực tế.

Cuối cùng, chúng em xin bày tỏ lòng biết ơn sâu sắc đến nhà trường, các giảng viên khoa Công nghệ thông tin, gia đình và bạn bè đã động viên, hỗ trợ chúng tôi trong suốt quá trình hoàn thiện bài khoá luận này. Chúng tôi hy vọng rằng những kiến thức và kinh nghiệm từ luận văn này sẽ giúp chúng tôi và các sinh viên khác trong ngành Khoa học máy tính tăng cường chuyên môn và đóng góp tích cực vào sự phát triển của mô hình máy học, học sâu trong lĩnh vực chuẩn đoán, phát viêm phổi trong y tế.

**MỤC LỤC**

[**CHƯƠNG 1.**  **TỔNG QUAN** 1](#_Toc185685365)

[**1.1 Lý do chọn đề tài** 1](#_Toc185685366)

[**1.2 Mục tiêu nghiên cứu** 2](#_Toc185685367)

[**1.3 Phạm vi nghiên cứu** 2](#_Toc185685368)

[**1.4 Phương pháp nghiên cứu** 3](#_Toc185685369)

[**1.5 Kết cấu đồ án.** 4](#_Toc185685370)

[**CHƯƠNG 2. CƠ SỞ LÝ THUYẾT** 5](#_Toc185685371)

[**2.1. Học sâu (Deep learning)** 5](#_Toc185685372)

[**2.2. Tăng cường dữ liệu (Data augmentation)** 5](#_Toc185685373)

[**2.3. Cơ chế chú ý (Attention Mechanisms)** 5](#_Toc185685374)

[**2.4. Kỹ thuật nhóm tích chập (Grouped Convolution)** 6](#_Toc185685375)

[**1. Squeeze:** 6](#_Toc185685376)

[**2. Excitation:** 7](#_Toc185685377)

[**3. Scale:** 7](#_Toc185685378)

[**2.6. Các phương pháp đánh giá mô hình** 7](#_Toc185685379)

[**1. Độ chính xác (Accuracy)** 7](#_Toc185685380)

[**2**. **Precision (Độ chính xác dự đoán dương)** 8](#_Toc185685381)

[**3. Recall (Độ nhạy)** 8](#_Toc185685382)

[**4.F1-Score** 9](#_Toc185685383)

[**5. AUC-ROC (Area Under Curve - Receiver Operating Characteristic)** 9](#_Toc185685384)

[**6. Specificity (Độ đặc hiệu)** 9](#_Toc185685385)

[**CHƯƠNG 3. MÔ HÌNH ĐỀ XUẤT** 10](#_Toc185685386)

[**3.1 Mô hình tổng quát** 10](#_Toc185685387)

[**3.2 Đặc trưng của mô hình đề xuất** 12](#_Toc185685388)

[**3.2.2. Hợp nhất và tăng cường đặc trưng dữ liệu** 15](#_Toc185685389)

[**3.2.3. Khối phân loại** 21](#_Toc185685390)

[**CHƯƠNG 4. THỰC NGHIỆM** 22](#_Toc185685391)

[**4.1 Môi trường thực nghiệm** 22](#_Toc185685392)

[**4.1.1. Cấu hình phần cứng** 22](#_Toc185685393)

[**4.1.2. Cấu hình phần mềm** 22](#_Toc185685394)

[**4.1.3. Thiết lập môi trường** 22](#_Toc185685395)

[**4.1.4. Lý do chọn Kaggle** 23](#_Toc185685396)

[**4.2 Tập dữ liệu** 23](#_Toc185685397)

[**4.2.1. Nguồn dữ liệu** 23](#_Toc185685398)

[**4.2.2. Số lượng dữ liệu** 23](#_Toc185685399)

[**4.2.3. Tiền xử lý dữ liệu** 23](#_Toc185685400)

[**4.2.4. Tăng cường dữ liệu** 24](#_Toc185685401)

[**4.2.5. Chia dữ liệu** 24](#_Toc185685402)

[**4.2.6. Lý do chọn tập dữ liệu** 25](#_Toc185685403)

[**4.3 Ứng dụng thực nghiệm** 25](#_Toc185685404)

[**4.3.1. Quy trình huấn luyện mô hình** 25](#_Toc185685405)

[**4.3.2. Cấu hình huấn luyện** 25](#_Toc185685406)

[**4.3.3. Kết quả thực nghiệm** 26](#_Toc185685407)

[**4.4 Đánh giá kết quả** 28](#_Toc185685408)

[**4.4.1. So sánh với các mô hình pretrain** 28](#_Toc185685409)

[**4.4.2 So sánh với các mô hình khi thay đổi các khối** 30](#_Toc185685410)

[**4.4.3 So sánh với mô hình gốc** 33](#_Toc185685411)

[**CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN** 35](#_Toc185685412)

[**5.1 Kết luận** 35](#_Toc185685413)

[**5.2 Hướng phát triển** 35](#_Toc185685414)

**MỤC LỤC HÌNH ẢNH**

[Hình 1. Kiến trúc mô hình 11](#_Toc184148871)

[Hình 2. Độ chính xác trong quá trình huấn luyện 27](#_Toc184148872)

[Hình 3. Loss trong quá trình huấn luyện 27](#_Toc184148873)

[Hình 4. Ma trận nhầm lẫn của mô hình đề xuất 28](#_Toc184148874)

[Hình 5. Biểu đồ so sánh số liệu các mô hình phổ biến với mô hình đề xuất 29](#_Toc184148875)

[Hình 6. Biểu đồ so sánh số liệu các mô hình thay đổi khối khác với mô hình đề xuất 32](#_Toc184148876)

**MỤC LỤC BẢNG BIỂU**

[Bảng 1. Các phương pháp và tham số tăng cường dữ liệu 24](#_Toc184207958)

[Bảng 2. Kết quả thử nghiệm và xác thực chéo của mô hình đề xuất 26](#_Toc184207959)

[Bảng 3. So sánh các tiêu chí đánh giá của mô hình đề xuất với các biến thể của nó 31](#_Toc184207960)

[Bảng 4: So sánh các tiêu chí đánh giá của mô hình gốc với mô hình đề xuất 33](#_Toc184207961)

**DANH MỤC CÁC THUẬT NGỮ VIẾT TẮT**

|  |  |  |
| --- | --- | --- |
| Từ viết tắt | Từ đầy đủ | Nghĩa |
| **GAP** | Global Average Pooling | Kỹ thuật gôp trung bình toàn cầu dùng để giảm kích thước của các đặc trưng dữ liệu và chuẩn bị dữ liệu cho các lớp tích chập sau bằng cách tính trung bình các giá trị pixel trong feature map tạo ra một vector có kích thước bằng số lượng feature maps, mỗi phần tử trong vector đại diện cho giá trị trung bình của một feature map. |
| **BN** | Batch Normalization | Chuẩn hoá hàng loạt là một kỹ thuật trong học sâu. Nó tính toán trung bình và phương sai của dữ liệu trong một batch (một nhóm dữ liệu nhỏ), sau đó chuẩn hóa dữ liệu dựa trên các giá trị này. Từ đó làm cho dữ liệu phân phối ổn định giúp các mạng nơ-ron học nhanh, hiệu quả hơn. |
| **CNN** | Convolutional Neural Network | Mạng nơ-ron tích chập, một mô hình học sâu được sử dụng chủ yếu trong xử lý ảnh và thị giác máy tính. |
| **LSTM** | Long Short-Term Memory | Mạng nơ-ron hồi tiếp đặc biệt dùng cho dữ liệu tuần tự hoặc chuỗi thời gian. |

# **CHƯƠNG 1.** **TỔNG QUAN**

## **1.1 Lý do chọn đề tài**

Viêm phổi là tình trạng nhu mô phổi bị nhiễm trùng (sưng) bao gồm viêm phế nang (túi khí nhỏ), túi phế nang, ống phế nang, tổ chức liên kết khe kẽ và viêm tiểu phế quản tận cùng do vi khuẩn, virus, nấm gây nên [1]. Viêm phổi là nguyên nhân gây tử vong hàng đầu ở trẻ em trên toàn thế giới, chiếm 14% tổng số trẻ em dưới 5 tuổi tử vong [2]. Theo Tổ chức Y tế Thế giới (WHO), ước tính hàng năm có khoảng 150 triệu đợt viêm phổi xảy ra ở trẻ em các nước đang phát triển, với khoảng 11 triệu trẻ nhập viện [3]. Vì thế việc chuẩn đoán nhanh chóng và chính xác bệnh là rất quan trọng để có thể phòng ngừa, can thiệp nhanh và hiệu quả nhất, tuy nhiên các phương pháp chuẩn đoán phổ thông, truyền thống thường đòi hỏi rất nhiều thời gian, nguồn lực và kinh nghiệm của các bác sĩ, chuyên gia y tế trong lĩnh vực X-quang.

Ngày nay chúng ta đã áp dụng các kỹ thuật máy tính nhằm hỗ trợ tìm kiếm, định vị, phân loại và nhận dạng các vùng bị viêm phổi để giảm thiểu thời gian và các lỗi chủ quan từ con người. Các mô hình phân loại, định vị, nhận dạng, phân đoạn các hình ảnh y tế như CNN, VGG16, ResNet, InceptionV3 và DenseNet121[4] đang chứng minh hiệu quả vượt trội trong việc xử lý các ảnh y khoa và được ứng dụng rộng rãi trong các nghiên cứu và thực tiễn lâm sàng. Mô hình trong bài báo [5] sử dụng mạng nơ-ron tích chập sâu kết hợp hiệu quả các điểm mạnh của EfficientNetB0 và DenseNet121, đồng thời được cải tiến bởi các cơ chế tập trung (attention mechanisms) nhằm phân loại hình ảnh viêm phổi một cách tinh tế. Trong khi đó, mô hình được đề xuất trong nghiên cứu của Brauwers và Frasincar [6] sử dụng các mô-đun tự tập trung đa đầu (multi-head, self-attention modules) để khai thác đặc trưng chi tiết từ hình ảnh X-quang, đồng thời áp dụng chiến lược trộn và hợp nhất đặc trưng dựa trên sự tập trung kênh (channel-attention-based feature fusion) [7], Các cải tiến này giúp nâng cao hiệu quả mô hình bằng cách kết hợp với khối dư (residual block) và chiến lược tăng cường đặc trưng, đặc biệt là với phân bổ động và sử dụng attention, như đã chỉ ra trong nghiên cứu của Vaswani và cộng sự [8].

Bài nghiên cứu này đề xuất cải tiến mô hình học sâu để chẩn đoán bệnh viêm phổi[9]. Để đánh giá hiệu suất mô hình đề xuất, phương pháp Stratified Five-Fold Cross-Validation được sử dụng giúp xác định độ chính xác và khả năng tổng quát của mô hình. Cuối cùng, mô hình của bài báo được so sánh đánh giá với các mô hình tiền huấn luyện như VGG16, ResNet50, MobileNet, InceptionV3 và Dense121 [10] và từ đó sẽ lựa chọn mô hình phù hợp nhất cho bài toán chẩn đoán bệnh viêm phổi.

## **1.2 Mục tiêu nghiên cứu**

1. Xây dựng mô hình học sâu có khả năng phân loại hình ảnh X-quang lồng ngực nhằm phát hiện viêm phổi với độ chính xác cao.
2. Tích hợp các cơ chế tập trung (attention mechanisms), tối ưu hoá hiệu năng và chiến lược kết hợp đặc trưng để cải thiện hiệu quả của mô hình.
3. So sánh hiệu suất của mô hình đề xuất với các mô hình tiền huấn luyện khác như VGG16, ResNet50, MobileNet, InceptionV3 và DenseNet121.
4. Ứng dụng các kỹ thuật kiểm tra như Stratified Five-Fold Cross-Validation để đảm bảo tính ổn định và chính xác của mô hình.
5. Tìm hiểu các vấn đề thiếu sót trong mô hình từ đó đưa ra hướng giải quyết trong tương lai và khả năng áp dụng của mô hình.

## **1.3 Phạm vi nghiên cứu**

* **Đối tượng nghiên cứu**: Ảnh X-quang lồng ngực, trong đó tập trung vào các hình ảnh có dấu hiệu viêm phổi.
* **Phạm vi kỹ thuật**: Sử dụng các mô hình học sâu, cụ thể là mạng nơ-ron tích chập (CNN) kết hợp với EfficientNetB0 và DenseNet121, cùng với việc sử dụng nhóm tích chập (Grouped Convolution) để tối ưu hoá hiệu năng kết hợp thêm các cơ chế chú ý như multi-head self-attention, channel-attention và spatial-attention cho mô hình.
* **Phạm vi kiểm tra**: Thực hiện kiểm tra mô hình trên các tập dữ liệu ảnh X-quang lồng ngực với phương pháp Stratified Five-Fold Cross-Validation.

## **1.4 Phương pháp nghiên cứu**

1. **Nghiên cứu tài liệu**:

* Tìm hiểu các nguyên lý về viêm phổi và chẩn đoán từ ảnh X-quang lồng ngực.
* Tham khảo các mô hình học sâu hiện đại trong xử lý ảnh y khoa.

1. **Xây dựng mô hình**:

* Kết hợp EfficientNetB0 và DenseNet121 để tận dụng các ưu điểm của hai mô hình.
* Tích hợp các cơ chế attention để tăng cường khả năng trích xuất đặc trưng từ ảnh.
* Sử dụng kỹ thuật nhóm tích chập (Grouped Convolution) để tăng tính đa dạng của đặc trưng của ảnh và tối ưu hoá hiệu năng của mô hình.

1. **Kiểm thử và đánh giá**:

* Ứng dụng Stratified Five-Fold Cross-Validation để kiểm tra độ ổn định và hiệu suất mô hình.
* So sánh kết quả với các mô hình khác như VGG16, ResNet50, MobileNet, InceptionV3 và DenseNet121.

1. **Phân tích và cải tiến**:

* Sử dụng chiến lược trộn và hợp nhất đặc trưng (feature fusion) dựa trên attention để tối ưu hiệu quả mô hình.
* Áp dụng các chiến lược tăng cường dữ liệu và tối ưu hóa thông qua các khối dư (residual block) và phân bổ động (dynamic allocation).

## **1.5 Kết cấu đồ án.**

Chương 1: Tổng quan

Chương 2: Cơ sở lý thuyết

Chương 3: Phân tích yêu cầu và thiết kế mô hình

Chương 4: Thực nghiệm

Chương 5: Kết luận và hướng phát triển

Tài liệu tham khảo.

# **CHƯƠNG 2. CƠ SỞ LÝ THUYẾT**

## **2.1. Học sâu (Deep learning)**

Học sâu (Deep Learning) là một nhánh con của học máy (Machine Learning) và là một lĩnh vực nghiên cứu quan trọng trong trí tuệ nhân tạo. Học sâu sử dụng các mô hình học máy với các lớp mạng nơ-ron nhiều tầng để tự động học các đặc trưng phức tạp từ dữ liệu mà không cần sự can thiệp của con người. Các mô hình học sâu có thể học được các đặc trưng chi tiết hơn từ dữ liệu thô, giúp mô hình cải thiện độ chính xác trong việc nhận diện và phân loại. [9]. Học sâu đã có những ứng dụng quan trọng trong nhiều lĩnh vực, đặc biệt là trong xử lý hình ảnh, nhận diện giọng nói, và xử lý ngôn ngữ tự nhiên. Trong y học, học sâu được ứng dụng mạnh mẽ trong việc phân tích ảnh y khoa, chẳng hạn như phân loại bệnh viêm phổi từ ảnh chụp X-quang, giúp tăng độ chính xác và giảm thời gian chẩn đoán của bác sĩ. Một trong những thành tựu nổi bật của học sâu trong xử lý ảnh là khả năng nhận diện các đặc trưng ẩn trong hình ảnh, chẳng hạn như các dấu hiệu viêm phổi, mà mắt thường khó phát hiện.

## **2.2. Tăng cường dữ liệu (Data augmentation)**

Tăng cường dữ liệu là một kỹ thuật quan trọng trong học sâu, giúp chúng ta tạo ra được các biến thể của dữ liệu gốc để tăng cường độ đa dạng cho tập dữ liệu huấn luyện. Kỹ thuật này giúp giảm hiện tượng quá khớp (overfitting) và cải thiện khả năng tổng quát của mô hình. Tăng cường dữ liệu đặc biệt hữu ích khi bạn có tập dữ liệu hạn chế, giúp mô hình học được nhiều đặc trưng hơn.

## **2.3. Cơ chế chú ý (Attention Mechanisms)**

Cơ chế chú ý (Attention) là một kỹ thuật trong học sâu cho phép mô hình tập trung vào các phần quan trọng trong dữ liệu. Điều này đặc biệt quan trọng khi xử lý các dữ liệu phức tạp như ảnh, khi các vùng quan trọng có thể không nằm ở vị trí cố định. Cơ chế chú ý giúp mô hình học được các mối quan hệ phức tạp giữa các phần của dữ liệu, tăng cường khả năng của mô hình trong việc phân tích và ra quyết định.

* **Cơ chế chú ý nhiều đầu (Multi-Head Attention)**

Multi-Head Attention là một phương pháp chú ý tiên tiến, cho phép mô hình học được nhiều mối quan hệ phức tạp trong dữ liệu đầu vào. Cơ chế này giúp mô hình đồng thời học và tập trung vào nhiều vùng quan trọng trong ảnh, giúp cải thiện khả năng phân loại khi xử lý các dữ liệu phức tạp.

## **2.4. Kỹ thuật nhóm tích chập (Grouped Convolution)**

Kỹ thuật nhóm tich chập là một kỹ thuật trong mạng thần kinh tích chập (CNN) trong đó các kênh của feature map được chia thành các nhóm. Mỗi nhóm sẽ được xử lý bởi một bộ lọc riêng biệt, sau đó nhân tích chập với từng nhóm nhỏ và cuối cùng sử dụng một lớp concatenate để nối chúng lại với nhau. Kỹ thuật này giúp giảm số lượng tham số của mạng, giảm độ phức tạp tính toán và cải thiện hiệu suất của mô hình, đặc biệt là trên các thiết bị có tài nguyên hạn chế [11].

**2.5. Phương pháp Squeeze and Excitation (SE)**

Squeeze and Excitation là một phương pháp cải thiện hiệu suất của mạng nơ-ron tích chập[7]. Nó gồm một vài lớp nhằm học cách điều chỉnh tác động và tăng cường đặc trưng chứa thông tin giữa các kênh. SE sử dụng toàn bộ thông tin sau đó nhấn mạnh có chọn lọc vào từng kênh có đặc trưng quan trọng và ít chú ý vào những kênh ít quan trọng hơn. Các bước thực hiện của khối SE thường như sau:

#### **1. Squeeze:**

* Tóm tắt thông tin không gian của mỗi kênh bằng **Global Average Pooling (GAP)**:

Zc =

Trong đó:

* Zc: Giá trị tóm tắt (scalar) của kênh c.
* : Giá trị tại vị trí (i, j) trong kênh c.
* H và W: Kích thước chiều cao và chiều rộng của feature map.

Kết quả là một vector Z có c chiều (số lượng kênh).

#### **2. Excitation:**

* Lấy vector Z qua hai lớp Fully Connected (FC) để học trọng số cho các kênh:

s =

Trong đó:

* W1và W2: Ma trận trọng số của hai lớp FC.
* : Hàm kích hoạt ReLU.
* : Hàm kích hoạt sigmoid để chuẩn hóa trọng số trong khoảng [0,1].
* s: Vector trọng số học được cho từng kênh.

#### **3. Scale:**

* Nhân trọng số s với từng kênh của feature map ban đầu:

= sc ⋅Xc

Trong đó:

* : Kênh được điều chỉnh.
* sc : Trọng số cho kênh c.

## **2.6. Các phương pháp đánh giá mô hình**

Để đánh giá hiệu quả của mô hình học sâu nhiều chỉ số đánh giá khác nhau được sử dụng, trong đó có các chỉ số như **Accuracy**, **Precision**, **Recall**, **F1-Score**, và **AUC-ROC**. Các chỉ số này giúp xác định độ chính xác của mô hình trong việc phân loại các đối tượng từ dữ liệu đầu vào.

#### **1. Độ chính xác (Accuracy)**

Độ chính xác đo lường tỷ lệ dự đoán đúng trên tổng số các dự đoán, bao gồm cả các dự đoán âm và dương đúng. Đây là chỉ số đơn giản nhưng không phải lúc nào cũng phản ánh chính xác hiệu quả của mô hình, đặc biệt khi tập dữ liệu mất cân bằng.

Accuracy=

Trong đó:

* TP: Số lượng dự đoán dương đúng (True Positives).
* TN: Số lượng dự đoán âm đúng (True Negatives).
* FP: Số lượng dự đoán dương sai (False Positives).
* FN: Số lượng dự đoán âm sai (False Negatives).

#### **2**. **Precision (Độ chính xác dự đoán dương)**

Precision đo lường khả năng của mô hình trong việc dự đoán đúng các trường hợp dương. Đặc biệt trong bài toán nhận diện viêm phổi, Precision giúp đánh giá khả năng mô hình không nhầm lẫn các trường hợp viêm phổi thành bình thường.

Precision=

Trong đó:

* TP: Dự đoán dương đúng (True Positives).
* FP: Dự đoán dương sai (False Positives).

#### **3. Recall (Độ nhạy)**

Recall đo lường khả năng của mô hình trong việc nhận diện tất cả các trường hợp dương, tức là độ nhạy trong việc phát hiện viêm phổi. Đây là chỉ số quan trọng trong việc đánh giá khả năng của mô hình trong việc phát hiện mọi trường hợp viêm phổi, đặc biệt là trong các tình huống mà các tổn thương có thể rất nhỏ hoặc không dễ nhận diện.

Recall=

Trong đó:

* TP: Dự đoán dương đúng (True Positives).
* FN: Số lượng trường hợp dương bị bỏ sót (False Negatives).

#### **4.F1-Score**

F1-Score là trung bình điều hòa giữa Precision và Recall. F1-Score là chỉ số quan trọng trong các bài toán phân loại không cân bằng, khi mà cả Precision và Recall đều quan trọng. F1-Score giúp cân bằng giữa việc không bỏ sót các trường hợp viêm phổi (Recall) và không nhầm lẫn quá nhiều các trường hợp bình thường thành viêm phổi (Precision).

F1 =

#### **5. AUC-ROC (Area Under Curve - Receiver Operating Characteristic)**

AUC-ROC đo lường khả năng phân biệt của mô hình giữa các lớp, với AUC có giá trị từ 0 đến 1. Một mô hình có AUC gần 1 cho thấy khả năng phân biệt giữa các lớp là rất tốt, trong khi AUC gần 0.5 có nghĩa là mô hình hoạt động gần như ngẫu nhiên.

* **Công thức AUC-ROC**:

AUC là diện tích dưới đường cong ROC, được tính thông qua việc vẽ đồ thị với tỷ lệ **True Positive Rate (TPR)** và **False Positive Rate (FPR)**:

TPR= (True Positive Rate, hay còn gọi Recall)

FPR= (False Positive Rate)

#### **6. Specificity (Độ đặc hiệu)**

Độ đặc hiệu đo lường khả năng của một mô hình phân loại trong việc đoán chính xác các mẫu âm (negative instances). Công thức của nó được viết như sau:

Specificity =

Trong đó:

* TN: Dự đoán âm đúng (True Negatives).
* FP: Số lượng trường hợp âm bị bỏ sót (False Positives).

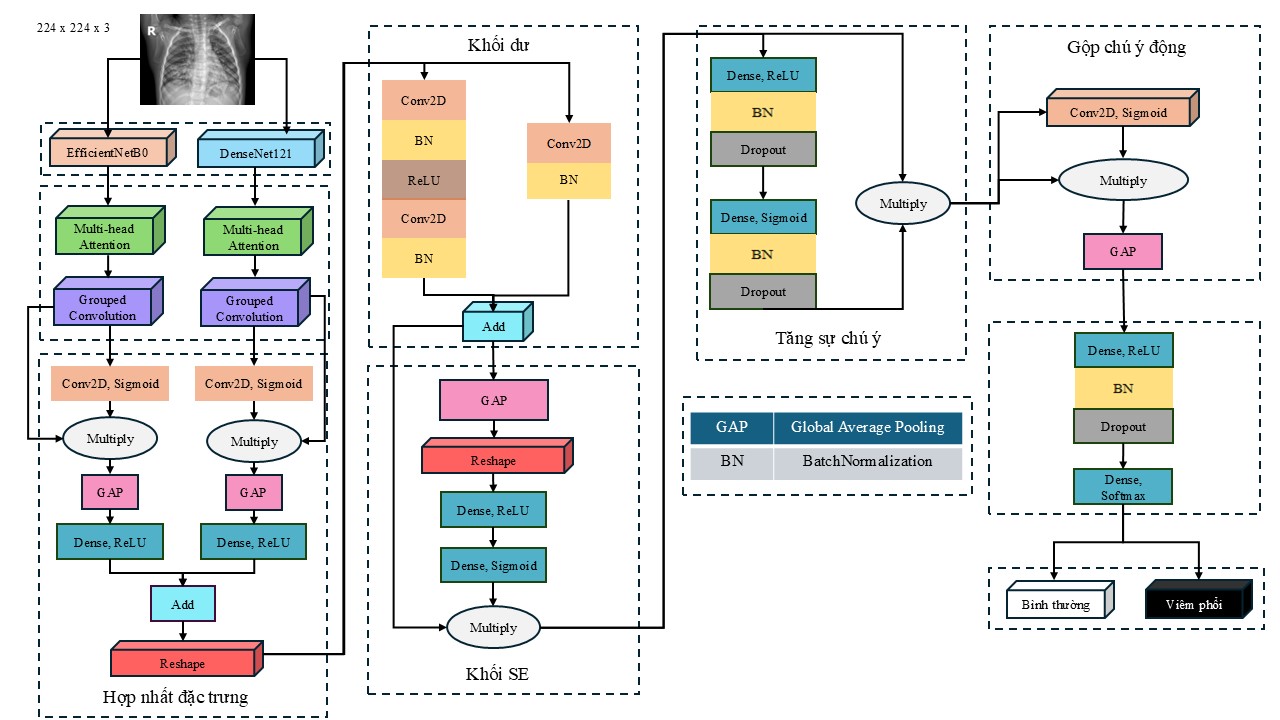
# **CHƯƠNG 3. MÔ HÌNH ĐỀ XUẤT**

## **3.1 Mô hình tổng quát**

Trong mô hình nghiên cứu mô hình học sâu hiện đại, sự kết hợp thông tin và các tính năng từ nhiều nguồn khác nhau đã được chứng minh là có thể tăng cường hiệu suất của mô hình, đặc biệt là trong lĩnh vực phân loại hình ảnh. Một trong số đó có mô hình sử dụng mạng nơ-ron tích chập sâu kết hợp với các phương pháp chú ý đặc trưng đã được đề xuất trong bài báo “A Deep Convolutional Neural Network for Pneumonia Detection in X-ray Images with Attention Ensemble” [9]. Trên cơ sở đó, chúng tôi đã thực hiện các cải tiến để nâng cao hiệu suất của mô hình trong việc phát hiện viêm phổi trên hình ảnh X-quang bao gồm:

1. Cơ chế nhóm tích chập (Grouped Convolution): Chúng tôi thêm tám lớp tích chập có áp dụng cơ chế này nhằm giảm số lượng tham số, tăng tốc độ tính toán và cải thiện khả năng tổng quát của mô hình.
2. Khối SE (Squeeze and Excitation Block): Chúng tôi tích hợp thêm một khối SE trước khối tăng sự chú ý nhằm giúp mô hình tập trung vào các kênh quan trọng và chuẩn bị các đặc trưng tốt hơn cho khối tăng sự chú ý.

Kiến trúc mô hình cải tiến được thể hiện trong **Hình 1**.



Hình 1. Kiến trúc mô hình

* Đầu vào của mô hình là ảnh chụp X-quang đã qua bước tiền xử lý và tăng cường dữ liệu để có kích thước là 224x224x3.
* Đầu ra của mô hình là phân phối xác suất trên hai lớp là: bình thường (NORMAL) và viêm phổi (PNEUMONIA). Dạng đầu ra này cung cấp cái nhìn sâu hơn cho chẩn đoán, vượt xa việc chỉ đưa ra nhãn phân loại đơn giản.
* Mô hình được chia thành 3 giai đoạn chính gồm:
* Trích xuất đặc trưng: Tại bước này, các đặc trưng của ảnh đầu vào được trích xuất và là yếu tố then chốt trong phân tích ảnh y khoa. Mô hình sử dụng EfficientNet-B0 và DenseNet-121, hai kiến trúc tiên tiến, để thực hiện việc trích xuất đặc trưng toàn diện. Tiếp đó, đầu ra của mỗi mạng trích xuất đặc trưng được xử lý qua các kỹ thuật tập trung đa đầu (multi-head attention) nhằm cải thiện khả năng biểu diễn của các vùng quan trọng. Cuối cùng đi qua các lớp tích chập được áp dụng cơ chế nhóm tích chập (Grouped Convolution) để tăng tốc độ tính toán cũng như tăng khả năng học được đa dạng đặc trưng từ các đặc trưng đầu vào.
* Hợp nhất và tăng cường đặc trưng: Sau khi qua giai đoạn trích xuất đặc trưng, mô hình bắt đầu hợp nhất các đặc trưng, hình thành một đại diện đặc trưng thống nhất. Ngoài ra, các khối dư (residual block), khối SE (Squeeze and Excitation), khối tăng cường dựa trên attention và phân bổ động (dynamic pooling) được tích hợp để gia tăng thêm sức mạnh biểu diễn của các đặc trưng, các kênh đã hợp nhất.
* Quyết định phân loại: Sau khi hoàn tất toàn bộ quy trình trích xuất, hợp nhất và tăng cường đặc trưng, mô hình xử lý thông tin qua mạng nơ-ron và cuối cùng đưa ra một phân phối xác suất trên các lớp bình thường và viêm phổi.
* Cấu trúc của mô hình này được thiết kế nhằm đảm bảo rằng tất cả thông tin có giá trị trích xuất từ ảnh được khai thác hiệu quả, với mục tiêu đạt được hiệu suất tối ưu trong nhiệm vụ phát hiện viêm phổi.

## **3.2 Đặc trưng của mô hình đề xuất**

**3.2.1. Trích xuất đặc trưng (Feature Extraction)**

1. Mạng cơ bản

Các hình ảnh X-quang lồng ngực chứa rất nhiều thông tin bên trong ảnh như: hình dạng, kết cấu và hoa văn, những thông tin này ảnh hưởng rất nhiều đến mô hình phân loại hình ảnh y tế. Để lấy được những thông tin đặc trưng phức tạp đó một cách hiệu quả ta sử dụng hai kiến trúc mạnh mẽ là EfficientNet và DenseNet

+ EfficientNetB0: EfficientNet là các dòng mô hình được thiết kế để mở rộng kiến trúc mạng nơ-ron và nổi tiếng với độ chính xác cao và tối ưu khi sử dụng ít tham số hơn. Kiến trúc này sử dụng phương pháp mở rộng tổng hợp (compound scaling), trong đó độ sâu, độ rộng và độ phân giải của mạng được mở rộng theo cách phối hợp.

+ DenseNet121: Mạng tích chập dày đặc (DenseNet) tận dụng các mẫu kết nối giữa các lớp thúc đẩy việc tái sử dụng các đặc trưng và giảm nguy cơ quá khớp (overfitting). Mỗi lớp trong DenseNet lấy thêm các đầu vào từ tất cả các lớp trước đó và truyền bản đồ đặc trưng của nó cho tất cả các lớp tiếp theo. Kết nối phức tạp này đảm bảo rằng mạng nắm bắt được nhiều đặc trưng không gian, do đó làm cho nó đặc biệt phù hợp với tập dữ liệu của chúng tôi. Trong nghiên cứu này, chúng tôi đã chọn DenseNet121 làm một trong những mô hình cơ sở.

1. Cơ chế chú ý nhiều đầu (Multi-Head Attention Mechanism)

Cơ chế chú ý nhiều đầu giúp mô hình học được mối quan hệ phức tạp giữa các thành phần khác nhau của dữ liệu. Điểm mạnh của nó nằm ở khả năng ưu tiên các phần quan trọng, cụ thể của dữ liệu đầu vào. Khả năng này được xem là điểm then chốt trong việc làm nổi bật lên các vùng có khả năng chỉ ra bệnh viêm phổi cao hơn trên hình ảnh X-quang.

Đối với mỗi đầu vào mô hình sẽ tính toán một tập hợp các điểm chú ý. Các điểm này xác định mức độ quan trọng cần được dành cho từng phần của đầu vào khi tạo ra đầu ra. Công thức toán học của nó:

Attention(𝑄,𝐾,𝑉)=softmax𝑉

Trong đó:

* Q là ma trận truy vấn (queries).
* K là ma trận khoá (keys).
* V là ma trận giá trị (values).
* dk là kích thước của các vector khóa/truy vấn.

Hàm softmax được sử dụng để chuẩn hoá các giá trị tương quan thành xác suất sau đó nhân với ma trận giá trị V từ đó tạo ra đầu ra có trọng số hoá.

Khi đã tính được các điểm chú ý trong các tập chú ý. Chúng ta kết hợp các điểm chú ý đó lại để có được một ma trận đầu ra với trọng số chú ý. Công thức toán học được miêu tả lại như sau:

MultiHead(Q, K,V) = Concat(head1, head2, ..., headh)Wo

Trong đó:

* Q, K, V là các ma trận Query, Key, và Value đầu vào.
* h là số lượng các “head” (lớp con Attention).
* Wo là ma trận trọng số đầu ra Multi-Head Attention.
* headi là kết quả của từng Attention với các trọng số .

Chúng tôi sử dụng lớp MultiHeadAttention của thư viện keras để tăng cường sự chú ý vào các đặc trưng cụ thể của đầu ra cho EfficientNetB0 và DenseNet121. Nhằm đảm bảo mô hình làm nổi bật các vùng trong ảnh X-quang có chứa các chỉ số quan trọng của bệnh viêm phổi.

1. Kỹ thuật nhóm tích chập (Grouped Convolution)

Kỹ thuật nhóm tích chập giúp tăng khả năng học được đa dạng các đặc trưng trong ảnh X-quang từ các nhóm kênh khác nhau. Nhóm các kênh cũng giúp mô hình tập trung vào các khía cạnh khác nhau của đầu vào, tăng tính đa dạng của đặc trưng.

- Mã giả của Grouped Convolution được cho như sau:

|  |
| --- |
| **Thuật toán 1:** Hàm Grouped\_Convolution |
| **1:** **function** Grouped\_Convolution (Finput, num\_groups, filters\_per\_group, kernel\_size, strides, padding)  **2:** group\_channels ← Finput.channel // num\_groups  **3:**  Gconv ← []  **4:** for each group in Groups do  **5:**  group ← SLICE(Finput, start\_index=i \* group\_channels, end\_index=(i + 1) \* group\_channels, axis=channel)  **6:** G ← Conv(filters=filters\_per\_group, kernel\_size=kernel\_size, strides=strides, padding=padding)(group)  **8:** Gconv.append(G)  **9:** end for  **10:** Gconcat ← Concatenate(Gconv, axis=channel)  **11:** return Gconcat  **12:** **end function** |

### **3.2.2. Hợp nhất và tăng cường đặc trưng dữ liệu**

Giai đoạn này hỗ trợ nâng cao khả năng phân biệt của các đặc trưng được học. Sự kết hợp này rất quan trọng nhằm tăng cường khả năng của mô hình trong việc phát hiện và phân tích các dấu hiệu viêm phổi từ ảnh X-quang ngực một cách hiệu quả.

1. Hợp nhất đặc trưng (Feature Fusion)

Trong giai đoạn này các bản đồ đặc trưng được kết hợp nhịp nhàng để tận dụng hết các điểm mạnh của từng mạng cơ bản. Nó sử dụng các cơ chế chú ý để tinh chỉnh quy trình tích hợp, do đó đảm bảo rằng các đặc trưng hợp nhất được căn chỉnh tối ưu cho nhiệm vụ đang thực hiện.

Các nhiệm vụ chính trong giai đoạn hợp nhất đặc trưng:

+ Các cơ chế chú ý theo kênh (Channel-attention mechanisms) được sử dụng để tính toán trọng số chú ý cho từng bản đồ đặc trưng.

+ Các bản đồ đặc trưng đã được gán trọng số chú ý từ EfficientNetB0 và DenseNet121 được nhân từng phần tử để tăng cường chọn lọc đặc trưng.

+ Các đặc trưng sau đó được chiếu lên một không gian đặc trưng chung để chuẩn hóa các kích thước kênh. Số lượng neuron trong lớp dense được xác định bởi quá trình định hình lại sau đó. Để đảm bảo rằng các đặc trưng được định hình lại bảo tồn các đặc điểm vật lý của đầu vào ở mức tối đa, chúng tôi chọn cấu hình kích thước 7 × 7.

+ Các bản đồ đặc trưng kết quả được cộng lại để tạo thành một bản đồ đặc trưng hợp nhất.

+ Các đặc trưng đã được tăng cường được định hình lại để phù hợp với các lớp tiếp theo của kiến trúc mạng.

* Phác thảo thuật toán:

|  |
| --- |
| **Thuật toán 2:** Hàm Feature Fusion |
| **1:**  **function** FeatureFusion(𝐹𝑒𝑓𝑓, 𝐹𝑑𝑒𝑛𝑠𝑒)  **2:**         𝐹𝑒𝑓𝑓1 ← Conv(𝐹chiều eff, (1,1), ′𝑠𝑖𝑔𝑚𝑜𝑖𝑑′)(𝐹𝑒𝑓𝑓)  **3:**         𝐹𝑑𝑒𝑛𝑠𝑒1 ← Conv(𝐹chiều dense, (1,1),′𝑠𝑖𝑔𝑚𝑜𝑖𝑑′)(𝐹𝑑𝑒𝑛𝑠𝑒)  **4:**         𝐹𝑒𝑓𝑓 ←𝐹𝑒𝑓𝑓·𝐹𝑒𝑓𝑓1  **5:**         𝐹𝑑𝑒𝑛𝑠𝑒 ←𝐹𝑑𝑒𝑛𝑠𝑒·𝐹𝑑𝑒𝑛𝑠𝑒1  **6:**        𝐹𝑒𝑓𝑓 ← GlobalAveragePooling(𝐹𝑒𝑓𝑓)  **7:**         𝐹𝑑𝑒𝑛𝑠𝑒 ← GlobalAveragePooling(𝐹𝑑𝑒𝑛𝑠𝑒)  **8:**         𝐹𝑒𝑓𝑓 ánh xạ ← Dense(784,′𝑟𝑒𝑙𝑢′)(𝐹𝑒𝑓𝑓)  **9:**        𝐹𝑑𝑒𝑛𝑠𝑒 ánh xạ ← Dense(784,′𝑟𝑒𝑙𝑢′)(𝐹𝑑𝑒𝑛𝑠𝑒)  **10:**       𝐹𝑒𝑓𝑓 ánh xạ ← Dropout(0.23)(𝐹𝑒𝑓𝑓 ánh xạ)  **11:**       𝐹𝑑𝑒𝑛𝑠𝑒 ánh xạ ← Dropout(0.23)(𝐹𝑑𝑒𝑛𝑠𝑒 ánh xạ)  **12:**      𝐹hợp nhất←𝐹𝑒𝑓𝑓 ánh xạ + 𝐹𝑑𝑒𝑛𝑠𝑒 ánh xạ  **13:**       𝐹𝑟𝑒𝑠ℎ𝑎𝑝𝑒𝑑 ← Reshape((7,7,16))(𝐹hợp nhất)  **14:**       **return** 𝐹𝑟𝑒𝑠ℎ𝑎𝑝𝑒𝑑  **15:**  **end function** |

* Cách tiếp cận này không chỉ giữ được các điểm mạnh độc đáo của EfficientNet và DenseNet mà chúng còn được kết hợp lại một cách nhịp nhàng. Điều này tạo ra một tập hợp tính năng không gian phong phú và toàn diện giúp tăng cường đáng kể khả năng của mô hình trong việc phát hiện và phân tích bệnh viêm phổi từ hình ảnh X-quang.

1. Khối dư (Residual Block)

Các đặc trưng sau khi được hợp nhất sẽ đi qua một khối dư nhằm tăng cường sự lan truyền tính năng và giảm bớt vấn đề độ dốc biến mất (vanishing gradient).

Khối dư này gồm các lớp tích chập, các lớp chuẩn hoá hàng loạt và hàm kích hoạt ReLU. Nó cũng có thêm một kết nối ngắn (shortcut connection) để ánh xạ danh tính (identity mapping). Tất cả các lớp tích chập có trong khối này đều được chúng tôi thiết lập 256 bộ lọc.

1. Khối SE (SE Block)

Sau khi tích hợp khối dư, các đặc trưng sẽ được đưa vào khối SE nhằm tăng khả năng tập trung vào các kênh đóng vai trò quan trọng, giảm ảnh hưởng từ các kênh không liên quan giúp tăng trọng số kênh quan trọng. Nó cũng giúp chuẩn bị đặc trưng tốt hơn cho bước Attention, tối ưu hóa khả năng học tập trung.

SE Block nhận đầu ra của Residual Block, thực hiện ba giai đoạn:

* **Squeeze:** Tóm tắt thông tin toàn bộ không gian (spatial) bằng Global Average Pooling.
* **Excitation:** Điều chỉnh tầm quan trọng của từng kênh thông qua hai lớp Dense với tỷ lệ nén (reduction).
* **Scale:** Nhân từng kênh đầu vào với trọng số quan trọng đã tính.

- Mã giả của khối SE như sau:

|  |
| --- |
| **Thuật toán 3:** Hàm SE\_Block |
| **1:** **function** SE\_Block (*Fdư, reduction*)  **2:** SEGAP ← GlobalAveragePooling (*Fdư*)  **3:** SEreshape ← Reshape((1,1, *Fdư*.channel))( SEGAP)  **4:** SE ← Dense(*Fdư*.channel // reduction, ′relu′) ( SEreshape)  **5:** SE ← Dense(*Fdư*.channel, ′𝑠𝑖𝑔𝑚𝑜𝑖𝑑′)(SE)  **6:** SEmultiply ← *Fdư* · SE  **7:** **end function** |

1. Tăng cường sự chú ý (Attention Augmentation)

Sau khi tăng khả năng tập trung vào các kênh đóng vai trò quan trọng bằng khối SE, các đặc trưng sẽ được đưa vào khối tăng cường sự chú ý (Attention Augmentation) để tinh chỉnh, cải thiện hơn nữa việc biểu diễn các đặc trưng bằng cách áp dụng một cơ chế chú ý bổ sung, cơ chế này sẽ nhấn mạnh một cách có chọn lọc các đặc trưng nổi bật và loại bỏ các đặc trưng không liên quan.

Nó gồm các bước sau:

+ Một lớp dense theo sau là một lớp chuẩn hoá (batch normalization) và một lớp Dropout nhằm tạo ra một bản đồ chú ý (attention map) từ các đặc trưng đã được nối lại.

+ Bản đồ chú ý trải qua quá trình kích hoạt sigmoid, batch normalization, và thêm một dropout khác để ổn định quá trình huấn luyện.

+ Cuối cùng, bản đồ chú ý sẽ được áp dụng lên các đặc trưng đã được nối lại trong đầu ra của khối SE để tạo ra các đặc trưng có trọng số chú ý (attention-weighted features).

- Mã giả của khối tăng cường sự chú ý được trình bày trong Algorithm 3 như sau:

|  |
| --- |
| **Thuật toán 4:** Hàm Attention Augmentation |
| **1:** **function** AttentionAugmentation(*F*)  **2:**      𝐴 ← Dense(𝐹kênh,′𝑟𝑒𝑙𝑢′)(𝐹)  **3:**      𝐴 ← BatchNormalization(𝐴)  **4:**      𝐴 ← Dropout(0.23)(𝐴)  **5:**      𝐴 ← Dense(𝐹kênh,′𝑠𝑖𝑔𝑚𝑜𝑖𝑑′)(𝐴)  **6:**      𝐴 ← BatchNormalization(𝐴)  **7:**      𝐴 ← Dropout(0.23)(𝐴)  **8:**     𝐹trọng số chú ý ← 𝐹·𝐴  **9:**      **return** 𝐹trọng số chú ý  **10:** **end function** |

- Sự kết hợp chú ý (attention fusion) có thể được tóm tắt như sau:

𝐹chú ý=𝐴⊙𝐹hợp nhất

Trong đó:

* A đại diện cho attention map
* ⊙ biểu thị cho tích Hadamard

Do đó đảm bảo rằng mỗi đặc trưng có trong Fhợp nhất đều được đánh giá theo mức độ quan trọng của nó.

1. Khối pooling chú ý động (Dynamic Attention Pooling)

Các lớp pooling truyền thống trong mạng nơ-ron tích chập có chức năng giảm kích thước không gian của các bản đồ đặc trưng bằng cách tổng hợp thông tin. Tuy nhiên, các phương pháp này thường sử dụng các phép tính tĩnh nên thường hay dẫn đến việc mất mát các đặc trưng, chi tiết quan trọng. Để giải quyết vấn đề đó khối pooling chú ý động ra đời cho phép mô hình quyết định một cách thích ứng các vùng quan trọng trong bản đồ đặc trưng trước khi thực hiện pooling, từ đó bảo toàn các thông tin quan trọng.

Khối pooling chú ý động kết thúc quy trình tăng cường đặc trưng bằng cách áp dụng một cơ chế chú ý không gian, cho phép mô hình tập trung thích ứng vào các phần thông tin quan trọng nhất của bản đồ đặc trưng. Quá trình này bao gồm các bước sau:

+ Tạo ra một bản đồ chú ý không gian thông qua một lớp tích chập với hàm kích hoạt sigmoid.

+ Nhân bản đồ đặc trưng ban đầu với bản đồ chú ý để thu được bản đồ đặc trưng có trọng số (weighted feature map).

+ Thực hiện global average pooling trên bản đồ đặc trưng có trọng số để trích xuất mô tả đặc trưng được pooling.

- Quá trình pooling chú ý động có thể được mô tả như sau:

𝑃động=𝐴gộp⊙𝑃𝑜𝑜𝑙(𝐹chú ý)

Trong đó:

* Agộp là bản đồ chú ý để pooling. Nó có thể điều chỉnh một cách linh hoạt sự ảnh hưởng của từng vị trí không gian trong quá trình pooling.
* Pool là các phép tính pooling truyền thống (ví dụ như global average pooling)

- Quá trình của khối pooling chú ý động (Dynamic Attention Pooling)

|  |
| --- |
| **Thuật toán 5:** Hàm Dynamic Attention Pooling |
| **1:** **function** DynamicAttentionPooling(*F*)  **2:**      𝐴𝑚𝑎𝑝 ← Conv2D(1,(1,1),′𝑠𝑖𝑔𝑚𝑜𝑖𝑑′)(𝐹)  **3:**      𝐹trọng số ← 𝐹·𝐴𝑚𝑎𝑝  **4:**      𝐹gộp ← GAP(𝐹trọng số)  **5:**      **return** 𝐹gộp  **6:** **end function** |

### **3.2.3. Khối phân loại**

* Trong khối cuối cùng này, đầu vào sẽ được truyền qua một lớp dense có kết nối đầy đủ (fully connected dense layer) với 128 nơ-ron sử dụng hàm kích hoạt 'relu’ để đưa về dưới dạng phi tuyến tính, điều này rất quan trọng để học các mẫu phức tạp. Hơn nữa việc điều chuẩn L2 (L2 regularization) được áp dụng để giảm thiểu tình trạng quá khớp (overfitting) thông qua việc phạt các trọng số lớn.
* Tiếp đó, một lớp chuẩn hoá hàng loạt (batch normalization) được áp dụng. Lớp này chuẩn hoá các hoạt động từ lớp trước thông qua việc giải quyết các thay đổi biến phụ thuộc nội bộ bằng cách ổn định quá trình học và đẩy nhanh quá trình hội tụ.
* Sau khi chuẩn hoá hàng loạt (batch normalization) ta đưa các đặc trưng đã được chuẩn hoá vào trong lớp dropout để ngăn tình trạng overfitting trong quá trình đào tạo giúp mô hình mạnh mẽ hơn.
* Quá trình phân loại kết thúc với việc thực hiện một lớp dense gồm 2 nơ-ron tương ứng với số lượng lớp, cùng hàm kích hoạt ‘softmax’.
* Quá trình chi tiết của lớp phân loại như sau:

|  |
| --- |
| **Thuật toán 6** Lớp phân loại mạng Nơ-ron |
| **1:** 𝑥←Dense(128,‘relu’,kernel\_regularizer=𝑙2(0.006))  **2:** 𝑥←BatchNormalization  **3:** 𝑥←Dropout(0.23)  **4:** 𝑜𝑢𝑡𝑝𝑢𝑡←Dense(2,‘softmax’) |

# **CHƯƠNG 4. THỰC NGHIỆM**

## **4.1 Môi trường thực nghiệm**

Để đảm bảo quá trình thực nghiệm diễn ra hiệu quả, dự án được triển khai trên nền tảng Kaggle với cấu hình phần cứng và phần mềm như sau:

#### **4.1.1. Cấu hình phần cứng**

* **Nền tảng**: Kaggle
* **Bộ xử lý đồ họa (GPU)**: NVIDIA Tesla P100

GPU này hỗ trợ tăng tốc các tác vụ học sâu, đặc biệt hiệu quả với các mô hình yêu cầu tính toán ma trận lớn, và mạng neuron tích chập.

#### **4.1.2. Cấu hình phần mềm**

* **Hệ điều hành**: Môi trường mặc định của Kaggle (Linux)
* **Ngôn ngữ lập trình**: Python 3.10
* **Thư viện chính sử dụng**:
  + TensorFlow/Keras: Huấn luyện mô hình học sâu.
  + NumPy và Pandas: Xử lý dữ liệu và phân tích.
  + Matplotlib/Seaborn: Trực quan hóa dữ liệu.
  + scikit-learn: Cung cấp các công cụ bổ trợ như chia tách dữ liệu, tính toán các chỉ số đánh giá mô hình.

#### **4.1.3. Thiết lập môi trường**

Toàn bộ các thí nghiệm được thực hiện trên nền tảng Kaggle, tận dụng các môi trường cài đặt sẵn, giúp tiết kiệm thời gian cài đặt và tối ưu hóa tài nguyên.

Các thiết lập khác bao gồm:

* Bộ nhớ RAM: 29 GB.
* Dung lượng lưu trữ tạm thời: 2.1T.
* Thời gian thực nghiệm: Khoảng 15 phút cho mỗi lần huấn luyện mô hình với khoảng 8.000 mẫu dữ liệu.

#### **4.1.4. Lý do chọn Kaggle**

Kaggle được chọn làm môi trường thực nghiệm vì nhiều lý do vượt trội:

* **Tài nguyên mạnh mẽ và miễn phí**: Với GPU NVIDIA Tesla P100, Kaggle cung cấp sức mạnh tính toán tương đương với các nền tảng tính phí, cho phép xử lý các mô hình lớn mà không tốn chi phí.
* **Tích hợp dễ dàng**: Kaggle hỗ trợ sẵn các thư viện học sâu và công cụ phổ biến, giúp rút ngắn thời gian thiết lập và tập trung hoàn toàn vào quá trình thực nghiệm.
* **Khả năng chia sẻ**: Nền tảng cho phép lưu trữ và chia sẻ mã nguồn, kết quả thực nghiệm, giúp dễ dàng quản lý các phiên bản của dự án.
* **Thân thiện với người dùng**: Giao diện trực quan, hỗ trợ khả năng kiểm tra log và đầu ra trực tiếp trên giao diện web.

## **4.2 Tập dữ liệu**

### **4.2.1. Nguồn dữ liệu**

Dữ liệu sử dụng trong dự án được lấy từ kaggle, gồm các ảnh chụp X-quang lồng ngực để có thể giúp chuẩn đoán tình trạng sức khỏe của phổi, tập dữ liệu gồm 2 nhóm chính là:

* Ảnh chuẩn đoán bình thường (NORMAL).
* Ảnh chuẩn đoán bị viêm phổi (PNEUMONIA).

### **4.2.2. Số lượng dữ liệu**

Tập dữ liệu ban đầu có tổng cộng 5856 ảnh được phân bổ như sau:

* 1583 ảnh phổi bình thường.
* 4273 ảnh bị viêm phổi.

### **4.2.3. Tiền xử lý dữ liệu**

Để chuẩn bị cho các bước tiếp theo thì ta sẽ tiến hành tiền sử lý dữ liệu trước, ở đây ta sẽ áp dụng kỹ thuật cân bằng histogram (Histogram Equalization) cho toàn bộ dữ liệu . Kỹ thuật này sẽ giúp chúng ta cải thiện được độ tương phản của hình ảnh và giúp mô hình dễ dàng nhận diện được các đặc trưng quan trọng trong các vùng tối hoặc là vùng sáng hơn.

### **4.2.4. Tăng cường dữ liệu**

Tập dữ liệu có tình trạng mất cân bằng giữa 2 nhóm phổi bình thường và bị viêm phổi nên ta áp dụng các kỹ thuật tăng cường ảnh để giải quyết vấn đề mất cân bằng dữ liệu đó Các kỹ thuật tăng cường ảnh trong tập train được sử dụng để giúp cho mô hình học được các đặt trưng khác và tránh overfit được liệt kê trong **Bảng 1**:

Bảng 1. Các phương pháp và tham số tăng cường dữ liệu

|  |  |
| --- | --- |
| **Phương pháp** | **Tham số** |
| Rescale | 1./255 |
| Rotation | ±15 |
| Shift | 0.1 |
| Shear | 0.2 |
| Zoom | 0.2 |
| Brightness | 0.8-1.2 |
| Fill mode | nearest |
| Horizontal flip | True |

Sau khi đã được tăng cường và chuẩn hóa lại ta được một tập dữ liệu mới với 8437 ảnh gồm: 4273 ảnh có nhãn viêm phổi và 4161 ảnh có nhãn bình thường.

### **4.2.5. Chia dữ liệu**

Tập dữ liệu sau khi được tăng cường sẽ được chia thành 3 phần chính:

* Tập huấn luyện( training set): 64% tập dữ liệu.
* Tập kiểm định(validation set): 16% tập dữ liệu.
* Tập kiểm tra(test set): 20% tập dữ liệu.

Quá trình chia dữ liệu được thực hiện ngẫu nhiên nhưng đảm bảo duy trì tỷ lệ cân đối giữa hai nhóm ảnh trong mỗi tập.

### **4.2.6. Lý do chọn tập dữ liệu**

Tập dữ liệu từ Kaggle được chọn vì:

* Độ tin cậy cao: Ảnh được thu thập từ các nguồn y tế uy tín, đảm bảo tính chính xác và chất lượng.
* Độ đa dạng và quy mô phù hợp: Bao gồm nhiều mẫu ảnh từ cả hai nhóm chẩn đoán, giúp mô hình học được nhiều đặc trưng hơn.
* Tính khả dụng: Tập dữ liệu có thể dễ dàng truy cập.

## **4.3 Ứng dụng thực nghiệm**

### **4.3.1. Quy trình huấn luyện mô hình**

Mô hình được triển khai trên nền tảng Kaggle, sử dụng GPU NVIDIA Tesla P100 để tăng tốc quá trình tính toán. Quy trình huấn luyện bao gồm các bước:

* Tiền xử lý và tăng cường dữ liệu từ tập ban đầu.
* Xây dựng mô hình.
* Huấn luyện mô hình trên tập dữ liệu đã được chuẩn bị.

### **4.3.2. Cấu hình huấn luyện**

Mô hình được huấn luyện với các tham số sau:

* Số epoch: 5 epoch.
* Batch size: 8.
* Tối ưu hóa (Optimizer): Adam với learning rate là 5.10-5.
* Hàm mất mát (Loss function): sử dụng Categorical Cross-Entropy.
* Metrics đánh giá: accuracy, precision, Recall, F1-score, AUC.
* Các tham số chính được sử dụng cho cơ chế chú ý nhiều đầu vào (Multi-Head Attention Mechanism) gồm:

**+ Số lượng Heads (num\_heads):** Chúng tôi đã chọn 4 head vì điều này biểu thị rằng cơ chế chú ý có thể đồng thời tính đến 4 phân phối trọng số khác nhau. Điều này hỗ trợ mô hình trong việc nắm bắt sự phụ thuộc trong các tính năng ở nhiều cấp độ khác nhau.

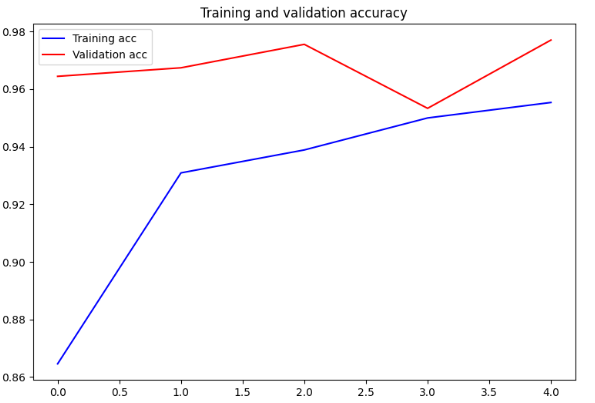
**+ Kích thước của Khóa (key\_dim):** Kích thước của các khóa được đặt ở mức 128. Tham số này chỉ định cách cơ chế chú ý cân nhắc thông tin đầu vào. Kích thước khóa lớn hơn giúp mô hình xem xét một tập thông tin phong phú hơn khi tính toán trọng số chú ý.

### **4.3.3. Kết quả thực nghiệm**

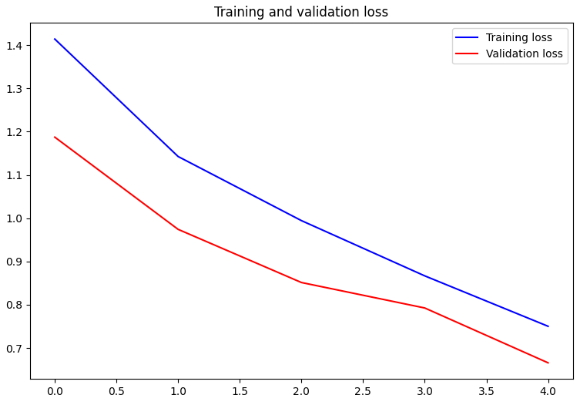
Sau khi training với 5 epoch, mỗi epoch mất khoảng 169-261 giây với tổng thời gian đào tạo là 942 giây ta và chạy Stratified Five-Fold Cross-Validation ta có được kết quả sau:

Bảng 2. Kết quả thử nghiệm và xác thực chéo của mô hình đề xuất

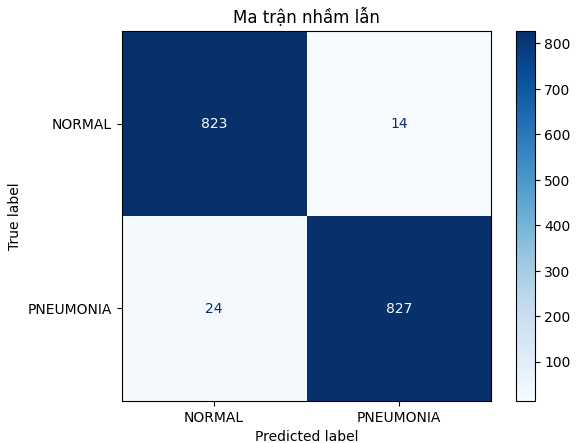
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Phương pháp** | **Accuracy(%)** | **Precision(%)** | **Recall** **(%)** | **F1-score (%)** | **Specificity(%)** | **AUC** |
| Thử nghiệm | 97.75 | 98.33 | 97.18 | 97.75 | 98.33 | 0.9956 |
| Xác thực chéo | 95.19 ± 2.34 | 95.34 ± 2.19 | 95.19 ± 2.34 | 95.17 ± 2.35 | 92.34 ± 4.62 | 0.9902 ± 0.0074 |



Hình 2. Độ chính xác trong quá trình huấn luyện



Hình 3. Loss trong quá trình huấn luyện



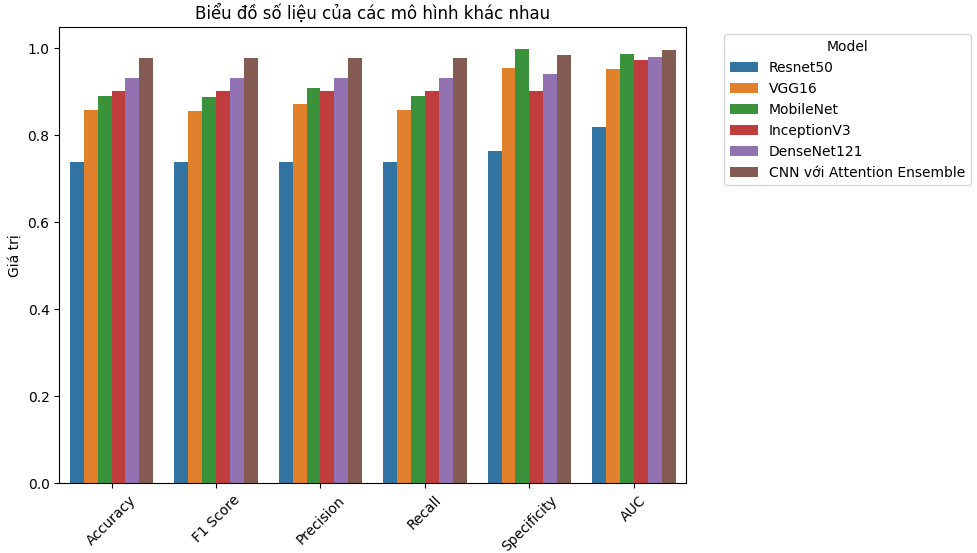
Hình 4. Ma trận nhầm lẫn của mô hình đề xuất

## **4.4 Đánh giá kết quả**

### **4.4.1. So sánh với các mô hình pretrain**

Sau khi huấn luyện mô hình, kết quả được đánh giá và so sánh với các mô hình tiền huấn luyện (pretrained models) như **ResNet50**, **VGG16, MobileNet, InceptionV3**,và **DenseNet**. Các mô hình này đã được sử dụng làm điểm chuẩn để kiểm tra xem các cải tiến trong mô hình có cải thiện hiệu suất so với các mô hình pretrain hay không.

Biểu đồ so sánh giữa các chỉ số **Accuracy**, **F1-Score**, **Precision**, **Recall**, **Specificity**, và **AUC** của mô hình đề xuất và các mô hình pretrained



Hình 5. Biểu đồ so sánh số liệu các mô hình phổ biến với mô hình đề xuất

**ResNet50**:

* + - **Accuracy** thấp nhất là **73.76%**, cho thấy khả năng phân loại hạn chế so với các mô hình khác.
    - **AUC** chỉ đạt **0.8198**, cho thấy mô hình khó phân biệt rõ ràng giữa các lớp.

**VGG16**:

* + - Hiệu suất cải thiện với **Accuracy** đạt **85.78%** và **AUC** là **0.9520**, tuy nhiên vẫn kém so với các mô hình hiện đại.
    - **Specificity** cao cho thấy khả năng tốt trong việc phát hiện ảnh **bình thường**, nhưng Precision và Recall chưa cân đối.

**MobileNet**:

* + - Mô hình nhẹ nhưng hiệu quả với **Accuracy** đạt **88.98%** và **AUC** là **0.9873**.
    - Đặc biệt, **Specificity** cao nhất trong các mô hình pretrained (**99.76%**), chứng tỏ khả năng mạnh mẽ trong việc phân loại đúng các trường hợp bình thường.

**InceptionV3**:

* + - **Accuracy** và **F1-Score** đều đạt **90.23%**, cùng với **AUC** là **0.9724**, cho thấy khả năng phân loại tốt và ổn định.
    - Các chỉ số Precision và Recall cân đối, chứng tỏ đây là một mô hình mạnh trong các kiến trúc pretrained.

**DenseNet121**:

* + - Đạt hiệu suất cao nhất trong các mô hình pretrained với **Accuracy** đạt **93.13%** và **AUC** là **0.9797**.
    - DenseNet có sự cân đối tốt giữa các chỉ số Precision và Recall, cho thấy khả năng phân loại toàn diện.

**Mô hình đề xuất:**

* **Accuracy**: Đạt **97.75%**, cao hơn so với DenseNet121 – mô hình tốt nhất trong nhóm pretrained.
* **F1-Score**: Đạt **97.75%**, cho thấy sự cân đối hoàn hảo giữa Precision và Recall.
* **Precision**: Đạt **97.76%**, cao hơn tất cả các mô hình pretrained, chứng minh khả năng phân loại chính xác các trường hợp viêm phổi.
* **Recall**: Đạt **97.75%**, đảm bảo không bỏ sót các trường hợp bệnh thực sự.
* **Specificity**: Đạt **98.33%**, chỉ kém **MobileNet** vì cơ chế pooling động trong mô hình khiến trọng số bị ưu tiên hơn vào các đặc trưng mạnh, thường liên quan đến lớp PNEUMONIA, do lớp này có các đặc điểm nổi bật hơn so với NORMAL. Tuy vậy mô hình vẫn chứng tỏ được khả năng phân loại chính xác các trường hợp bình thường.
* **AUC**: Đạt **0.9956**, cho thấy mô hình có khả năng phân biệt tốt giữa hai lớp.

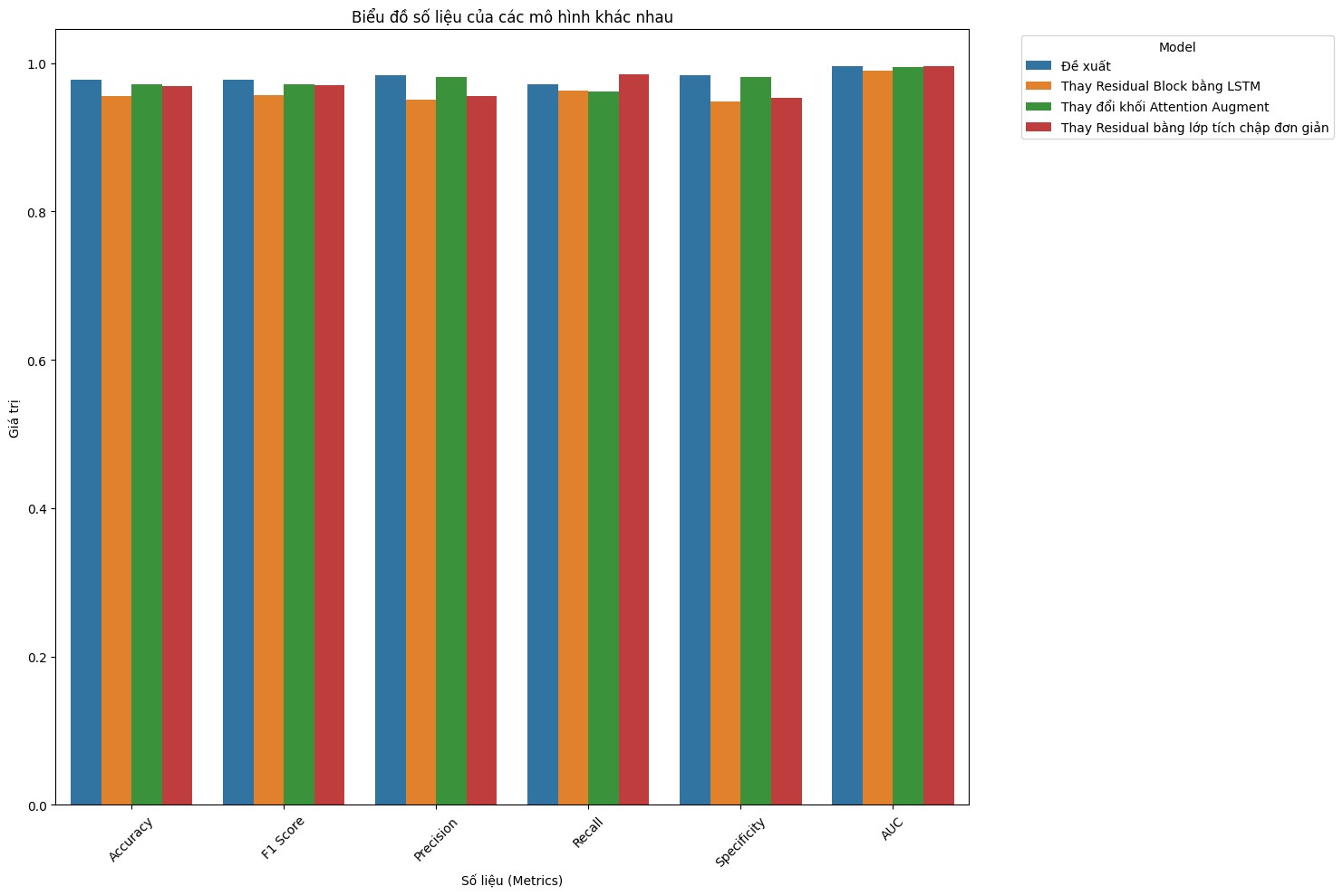
### **4.4.2 So sánh với các mô hình khi thay đổi các khối**

Các so sánh sẽ được sử dụng với mô hình đề xuất gồm **Thay Residual Block bằng LSTM, Thay khối Attention Augmentation và Thay Residual Block bằng Conv đơn giản:**

Bảng 3. So sánh các tiêu chí đánh giá của mô hình đề xuất với các biến thể của nó

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Biến thể** | **Accuracy** | **F1-Score** | **Precision** | **Recall** | **Specificity** | **AUC** |
| **Mô hình đề xuất** | 97.75% | 97.75% | 97.76% | 97.75% | 98.33% | 0.9956 |
| **Thay Residual Block bằng LSTM** | 95.56% | 95.56% | 95.56% | 95.56% | 94.86% | 0.9896 |
| **Thay khối Attention Augmentation** | 97.10% | 97.10% | 97.12% | 97.10% | 98.09% | 0.9941 |
| **Thay Residual Block bằng Conv đơn** | 96.92% | 96.92% | 96.96% | 96.92% | 95.34% | 0.9956 |

Biểu đồ so sánh giữa các chỉ số **Accuracy**, **F1-Score**, **Precision**, **Recall**, **Specificity**, và **AUC** của mô hình đề xuất và các thay đổi khối:



Hình 6. Biểu đồ so sánh số liệu các mô hình thay đổi khối khác với mô hình đề xuất

**Thay đổi Residual Block**:

* Khi thay **Residual Block** bằng **LSTM**, hiệu suất giảm do LSTM không phù hợp với dữ liệu không gian như ảnh.

**Thay đổi Attention Augmentation**:

* Thay đổi khối chú ý (Attention Augmentation) làm giảm nhẹ hiệu suất , nhưng vẫn giữ được mức hiệu quả cao, chứng tỏ vai trò của Attention trong việc nhấn mạnh các đặc trưng quan trọng.

**Thay Residual Block bằng lớp tích chập đơn giản:**

* Hiệu suất giảm vừa phải: Accuracy giảm còn 96.92% ,AUC vẫn giữ nguyên ở mức 0.9956, cho thấy khả năng phân biệt của mô hình không bị ảnh hưởng nhiều.

### **4.4.3 So sánh với mô hình gốc**

Để đánh giá hiệu quả của các cải tiến được thực hiện trong mô hình đề xuất, phần này tiến hành so sánh trực tiếp giữa mô hình gốc (chưa được cải tiến) và mô hình đề xuất. Các tiêu chí so sánh bao gồm: Accuracy, Precision, Recall, F1-Score, Specificity, và AUC.

Bảng 4: So sánh các tiêu chí đánh giá của mô hình gốc với mô hình đề xuất

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Mô hình** | **Accuracy** | **F1-Score** | **Precision** | **Recall** | **Specificity** | **AUC** |
| **Mô hình gốc** | 95.19% | 98.38% | 93.84% | 96.06% | 97.43% | 0.9556 |
| **Mô hình đề xuất** | 97.75% | 97.75% | 97.76% | 97.75% | 98.33% | 0.9956 |

* Độ chính xác (Accuracy) của mô hình đề xuất đạt **97.75%**, cao hơn đáng kể so với **95.19%** của mô hình gốc cho thấy các cải tiến đã giúp mô hình nhận diện hình ảnh viêm phổi chính xác hơn.
* Precision của mô hình gốc đạt **98.38%**, nhỉnh hơn một chút so với **97.76%** của mô hình đề xuất có thể do mô hình gốc có xu hướng ưu tiên độ chính xác cao hơn trong các dự đoán dương tính.
* Recall của mô hình đề xuất là **97.75%**, vượt xa **93.84%** của mô hình gốc. Đây là một cải tiến quan trọng vì nó cho thấy mô hình đề xuất có khả năng phát hiện hầu hết các trường hợp viêm phổi, giảm thiểu các trường hợp bỏ sót.
* F1-Score của mô hình đề xuất đạt **97.75%**, cao hơn **96.06%** của mô hình gốc cho thấy sự cân bằng tốt hơn giữa Precision và Recall.
* Specificity của mô hình đề xuất đạt **98.33%**, nhỉnh hơn **97.43%** của mô hình gốc chứng tỏ khả năng nhận diện chính xác các trường hợp bình thường cũng được cải thiện.
* AUC của mô hình đề xuất đạt **0.9956**, vượt trội so với **0.9564** của mô hình gốc cho thấy mô hình đề xuất phân biệt tốt hơn giữa hai lớp.

# **CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

## **5.1 Kết luận**

Trong nghiên cứu này, chúng tôi đã trình bày một mô hình học sâu cải tiến nhằm phát hiện viêm phổi từ ảnh chụp X-quang lồng ngực dựa trên cơ sở là mô hình mạng nơ-ron tích chập sâu kết hợp với Attention Ensemble. Bằng cách thêm vào các nhóm tích chập (Grouped Convolution) và khối SE để làm tăng khả năng học đa dạng các đặc trưng, các kênh của dữ liệu giúp cải thiện thêm việc chuẩn bị các thuộc tính đầu vào cho các lớp chú ý, mà mô hình đề xuất đã đạt được hiệu suất vượt trội với độ chính xác 97,75%, độ đặc hiệu 98,33%, độ nhạy 97,75%, điểm F1 97,75% và AUC đạt 0,9956 trên tập dữ liệu kiểm tra. Việc kết hợp thêm các nhóm tích chập và khối SE vào mô hình phát hiện viêm phổi là một phương pháp hiệu quả và tiềm năng. Những kết quả này không chỉ xác nhận khả năng phân biệt chính xác các tình trạng bệnh lý nguy kịch của mô hình mà còn làm nổi bật tiềm năng của mô hình như một công cụ đáng tin cậy trong việc nâng cao chẩn đoán bệnh viêm phổi, giảm thiểu thời gian chẩn đoán và hỗ trợ các y bác sĩ trong việc chăm sóc bệnh nhân đặc biệt trong bối cảnh dịch COVID-19 và các bệnh liên quan đến hệ hô hấp đang diễn biến phức tạp hiện nay.

## **5.2 Hướng phát triển**

Dựa trên các kết quả đạt được, chúng tôi đề xuất một số hướng nghiên cứu và cải tiến trong tương lai như sau:

1. **Tăng cường dữ liệu**:

* Thu thập thêm các tập dữ liệu từ nhiều nguồn khác nhau để mô hình học được đa dạng đặc trưng từ các trường hợp bệnh nhân thuộc nhiều độ tuổi, giới tính và điều kiện y tế khác nhau.
* Sử dụng các kỹ thuật tăng cường dữ liệu tiên tiến hơn như GAN (Generative Adversarial Networks) để tạo ra các hình ảnh giả lập có độ chân thực cao.

1. **Áp dụng trên các loại bệnh khác**:

* Mở rộng mô hình để phát hiện và phân loại nhiều loại bệnh phổi khác nhau như ung thư phổi, lao phổi hoặc viêm phổi do các nguyên nhân khác nhau (vi khuẩn, virus, nấm).

1. **Tích hợp mô hình vào hệ thống lâm sàng**:

* Xây dựng một hệ thống phần mềm tích hợp mô hình vào quy trình làm việc của bệnh viện, hỗ trợ bác sĩ trong việc chẩn đoán.
* Tối ưu hóa tốc độ xử lý của mô hình để phù hợp với các hệ thống thời gian thực.

1. **Cải tiến hiệu suất mô hình**:

* Áp dụng các kỹ thuật tinh chỉnh (fine-tuning) và tối ưu hóa siêu tham số (hyperparameter optimization) để cải thiện hiệu suất mô hình.
* Khám phá các kiến trúc mạng mới như Vision Transformers (ViTs) hoặc các mô hình tự giám sát (self-supervised learning) để tăng khả năng học đặc trưng.

1. **Giảm chi phí tính toán**:

* Tối ưu hóa mô hình để hoạt động hiệu quả trên các thiết bị phần cứng có tài nguyên hạn chế, ví dụ như thiết bị di động hoặc các máy tính tại các khu vực y tế thiếu thốn.

**TÀI LIỆU THAM KHẢO**

[1] Mai Mộc Thảo. Viêm phổi: nguyên nhân, triệu chứng, chẩn đoán và cách điều trị. VNVC 2022. <https://vnvc.vn/viem-phoi/>.

[2] *Tháng 11 nói về NGÀY VIÊM PHỔI THẾ GIỚI 12/11/2015*. (n.d). Hội Hô Hấp TP.HCM. <http://www.hoihohaptphcm.org/benh-nhan/221-thang11-noi-ve-ngay-viem-phoi-the-gioi>

[3] *Bệnh viêm phổi ở trẻ em*. (n.d). Hội Hô Hấp TP.HCM. <http://www.hoihohaptphcm.org/benh-nhan/146-benh-viem-phoi-o-tre-em>

[4] Mujahid, M.; Rustam, F.; Álvarez, R.; Luis Vidal Mazón, J.; Díez, I.d.l.T.; Ashraf, I. Pneumonia classification from X-ray images with inceptionV3-V3 and convolutional neural network. Diagnostics 2022, 12, 1280. <https://www.mdpi.com/2075-4418/12/5/1280>

[5] Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., & Geem, Z. W. (2020). Efficient Pneumonia Detection in Chest X-ray Images Using Deep Transfer Learning. Diagnostics, 10(6), 417. <https://www.mdpi.com/2075-4418/10/6/417>

[6] Brauwers, G.; Frasincar, F. A general survey on attention mechanisms in deep learning. IEEE Trans. Knowl. Data Eng. 2021, 35, 3279–3298. <https://ieeexplore.ieee.org/document/9609539>

[7] Jie Hu, Li Shen, Gang Sun, Samuel Albanie & Enhua Wu (2018). Squeeze-and-Excitation Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 7132–7141. <https://arxiv.org/pdf/1709.01507>

[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, L., & Polosukhin, I. (2017). Attention is All You Need. Advances in Neural Information Processing Systems (NeurIPS), 30. <https://arxiv.org/abs/1706.03762>

[9] An, Q., Chen, W., & Shao, W. (2024). A Deep Convolutional Neural Network for Pneumonia Detection in X-ray Images with Attention Ensemble. Diagnostics, 14(4), 390. <https://www.mdpi.com/2075-4418/14/4/390>

[10] Ben Atitallah, S.; Driss, M.; Boulila, W.; Koubaa, A.; Ben Ghezala, H. Fusion of convolutional neural networks based on Dempster– Shafer theory for automatic pneumonia detection from chest X-ray images. Int. J. Imaging Syst. Technol. 2022, 32, 658–672. <https://onlinelibrary.wiley.com/doi/10.1002/ima.22653>

[11] Tan, A., Guo, T., Zhao, Y. *et al.* Object detection based on polarization image fusion and grouped convolutional attention network. *Vis Comput* **40**, 3199–3215 (2024). <https://doi.org/10.1007/s00371-023-03022-6.>