TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH

KHOA ĐÀO TẠO QUỐC TẾ

NGÀNH KỸ THUẬT MÁY TÍNH

ADVANCED TOPICS IN COMPUTER ENGINEERING

MẠNG NƠ-RON TÍCH CHẬP CHÚ Ý TỌA ĐỘ TỰ ĐỘNG PHÁT HIỆN HÌNH ẢNH LAO PHỔI TRÊN X-QUANG NGỰC

Sinh viên: **MAI THANH LÂM**

MSSV: 20119137

MAI HỒNG PHONG

MSSV: 20119192

GVHD: **TS. HUỲNH THẾ THIỆN**

THÀNH PHỐ HỒ CHÍ MINH, 11/2023

TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT THÀNH PHỐ HỒ CHÍ MINH

KHOA ĐÀO TẠO QUỐC TẾ

NGÀNH KỸ THUẬT MÁY TÍNH

ADVANCED TOPICS IN COMPUTER ENGINEERING

MẠNG NƠ-RON TÍCH CHẬP CHÚ Ý TỌA ĐỘ TỰ ĐỘNG PHÁT HIỆN HÌNH ẢNH LAO PHỔI TRÊN X-QUANG NGỰC

Sinh viên: **MAI THANH LÂM**

MSSV: 20119137

MAI HỒNG PHONG

MSSV: 20119192

GVHD: **TS. HUỲNH THẾ THIỆN**

THÀNH PHỐ HỒ CHÍ MINH, 11/2023

LỜI CẢM ƠN

Em muốn bày tỏ lòng biết ơn chân thành đến người hướng dẫn của mình, Tiến sĩ Huỳnh Thế Thiện, với sự hỗ trợ và chỉ dẫn quý báu trong suốt môn học "Advanced Topics In Computer Engineering". Sự kiên nhẫn, hiểu biết, và động viên từ thầy đã mang lại giá trị to lớn đối với em. Em biết ơn những góp ý và lời khuyên mà thầy đã chia sẻ, giúp em cải thiện đáng kể kỹ năng nghiên cứu và kiến thức của mình. Em cũng biết ơn sự sẵn lòng trả lời câu hỏi và đưa ra phản hồi về công việc của em từ phía thầy.

Em cũng biết ơn bản thân mình vì sự cố gắng và tận tâm đặt vào dự án này. Em đã học được nhiều và tự hào với công việc đã làm. Em tin rằng dự án này sẽ là một bổ sung quý giá cho kiến thức của em và sẽ giúp em thành công trong sự nghiệp tương lai.

Nhóm thực hiện báo cáo
(Ký và ghi rõ họ tên)

Mai Hồng Phong Mai Thanh Lâm

LỜI CAM ĐOAN

Tôi, Mai Thanh Lâm và Mai Hồng Phong, là hai tác giả chịu trách nhiệm cho dự án này. Chúng tôi tuyên bố rằng tất cả các khía cạnh của công việc được trình bày là kết quả của sự nỗ lực độc lập của chúng tôi. Chúng tôi không sao chép hoặc sử dụng bất kỳ nội dung hoặc kết quả nào từ các dự án khác mà không có sự đồng thuận của tác giả. Mọi nguồn thông tin bên ngoài mà chúng tôi đề cập đến đều đã được ghi nhận chính xác thông qua các trích dẫn.

Nhóm thực hiện báo cáo
(Ký và ghi rõ họ tên)

Mai Hồng Phong Mai Thanh Lâm

TÓM TẮT NỘI DUNG

Lao phổi là một bệnh truyền nhiễm nguy hiểm đe dọa sức khỏe con người. Theo tác giả, việc chẩn đoán lao phổi hiện nay vẫn gặp phải nhiều hạn chế, chẳng hạn như tỷ lệ sai sót cao và thời gian chuẩn đoán lâu. Do đó, mục tiêu của nghiên cứu là đề xuất một giải pháp tự động, chi phí thấp để phát hiện lao phổi sớm thông qua chụp X-quang phổi.

Các tác giả đề xuất một thuật toán phân loại hình ảnh dựa trên mạng nơ-ron tích chập và học sâu để tính đến các tình huống cụ thể. Thuật toán nâng cao khả năng nhận dạng bằng cách chú ý đến thông tin liên kênh, vị trí và hướng của các điểm ảnh trong ảnh X-quang. Trong quá trình huấn luyện, để mô hình được huấn luyện nhanh hơn tác giả đã sử dụng kỹ thuật học transfer learning và freezing network.

Hiệu suất của phương pháp được đánh giá trên tập dữ liệu phân loại lao của Bệnh viện Thâm Quyến 3, Trung Quốc. So với ConvNet, FPN + Faster RCNN và các phương pháp khác , phương pháp của tác giả vượt trội hơn, có độ chính xác tốt hơn và có thể hỗ trợ bác sĩ chụp X-quang chẩn đoán sớm bệnh lao phổi. Tóm lại, đây là một nghiên cứu trong lĩnh vực y tế có ý nghĩa thực tiễn cao.

Mục lục

Danh sách bảng

Danh sách hình vẽ

Danh mục các từ viết tắt

Dưới đây là danh sách các từ viết tắt được sử dụng trong bài báo cáo.

Từ viết tắt	Nghĩa tiếng Anh	Định nghĩa
CNN	Convolutional Neural Network	Mạng nơ-ron tích chập
NLP	Natural Language Processing	Xử lý ngôn ngữ tự nhiện
CA	Coordinate Atention	Chú ý toạ độ
CXR	Chest X Ray	X-quang ngực
AP	Anteroposterior	Trước và sau

Chương 1

Giới thiệu

Bệnh lao phổi là một trong mười nguyên nhân hàng đầu gây tử vong trên toàn thế giới. Tại Trung Quốc, tỷ lệ mắc và tử vong do bệnh lao phổi đứng thứ hai trong các nguyên nhân tử vong. Nếu phát hiện không kịp thời, bệnh lao sẽ lan rộng trong cơ thể gây ra hoại tử các mô ở phổi, hình thành các vết loét và có xu hướng xuất huyết trầm trọng. Do đó, việc phát hiện và chẩn đoán nhanh chóng bệnh lao đã trở nên cực kỳ quan trọng. Chẩn đoán bệnh lao phụ thuộc vào việc phân tích X-quang ngực [1, 2] bởi các bác sĩ chuyên khoa. X-quang ngực không chỉ có thể phát hiện bệnh lao sớm, mà còn có thể tìm thấy các tổn thương nhỏ hơn hoặc các tổn thương tiềm ẩn. Tuy nhiên, có một tỷ lệ chẩn đoán sai lầm nhất định, thời gian chuẩn đoán kéo dài khi làm bằng phương pháp truyền thống và nó cũng đòi hỏi các chuyên gia phải có kiến thức chuẩn đoán dựa trên hình ảnh y tế thật phong phú.

Trong những năm gần đây, việc học sâu trong phát hiện và phân loại hình ảnh cũng cực kì phát triển đặc biệt là trong lĩnh vực y tế đã và đang khá tích cực trong việc triển khai ứng học học sâu vào thực tiễn. Một trong những mô hình phổ biến nhất là mô hình mạng nơ-ron tích chập (CNN), sử dụng phương pháp trích xuất đặc trưng để xác định và phân loại hình ảnh. Tuy nhiên, việc phát hiện và phân loại bệnh lao phổi đang phải đối mặt với những thách thức. Các hình ảnh bệnh lao phổi thường không rõ ràng xen lẫn với các khối mô tế bào khác. Thêm vào đó, các tập dữ liệu dành riêng cho lĩnh vực này chưa được phát triển cũng là một thách thức không hề nhỏ. Trong lĩnh vực hình ảnh y tế, một số nhà nghiên cứu đề xuất giải quyết các vấn đề tương tự bằng cách

thêm attention mechanism. Lấy cảm hứng từ điều này, bài báo này giới thiệu Coordinate Atention (CA) [3] dựa trên CNN truyền thống VGG16 [4]. Mặc dù Attention Mechanism lần đầu tiên được sử dụng trong xử lý ngôn ngữ tự nhiên (NLP), tuy nhiên nó cũng rất hiệu quả trong việc xử lý ảnh. Attention Mechanism cho phép CNN chọn những vị trí để tập trung trích xuất các đặc điểm, để tìm ra các đặc điểm phân biệt chung, các đặc tính đáng chú ý. Là một mô-đun plug-and-play, cơ chế chú ý cũng rất thích hợp để sửa đổi khối CNN điều này cũng giúp các đặc trưng thích ứng với sự sâu hơn của mạng [5]. So sánh với Channel Attention (ECANET) [6] và Spatial Atention (CBAM) [7], CA không chỉ thu thập thông tin qua các kênh, mà còn nhận thức về hướng và vị trí, cho phép mô hình xác định chính xác hơn khu vực mục tiêu. Việc này giúp đỡ tốt hơn cho việc phát hiện và phân loại bệnh lao phổi bằng X-quang ngực. Đồng thời, Các nhà nghiên cứu đào tạo mô hình bằng cách thêm học tập chuyển giao để đóng băng mạng để tăng tốc độ đào tạo và cải thiện độ chính xác khi phân loại. Cuối cùng, Họ đã vượt qua 5 lần xác nhận chéo để xác minh tính mạnh mẽ của phương pháp.

Trong bài báo này, một thuật toán dựa trên CNN với CA cho việc tự động phát hiện hình ảnh bệnh lao phổi trên tia X-quang ngực đã được đề xuất. Chúng tôi sử dụng các mô hình mạng lưới được đào tạo trước khác nhau để phát hiện kết quả phân loại bệnh laphổi trong hình ảnh X-quang ngực (CXR), so sánh với kết quả của hệ thống phân loại hệ thống chuyển học đóng băng để đánh giá hiệu quả của chương trình đào tạo. So với các chương trình end-to-end hiện có, phương pháp của tác giả đạt được kết quả tốt hơn.

Phần còn lại của bài báo được sắp xếp như sau:

- Phần 2: Nghiên cứu liên quan.
- Phần 3: Tập dữ liệu - Datasheet.
- Phần 4: Các Phương pháp huấn luyện.
- Phần 5: Kết quả thực nghiệm.
- Phần 6: Kết luận và thảo luận.

Chương 2

Các nghiên cứu liên quan

Trong phần này, tác giả đã khám phá một số chương trình chẩn đoán bệnh lao dựa trên học máy hoặc học sâu. Chẩn đoán bệnh lao chủ yếu dựa trên hình ảnh thu được bằng cách kiểm tra X-quang ngực. Tuy nhiên, việc xác định chẩn đoán hình ảnh y tế đòi hỏi các Bác sỹ chuyên nghành có kinh nghiệm và chuyên môn, đó là một trở ngại lớn đối với việc chẩn báo hiệu quả bệnh lao phổi. So với chẩn đoán của con người, các thuật toán máy tính có thể cung cấp kết quả quan trọng hơn với ít sai sót chuẩn đoán, tốn ít thời gian hơn, và sử dụng ít nguồn lực hơn để đạt được chuẩn đoán hình ảnh hiệu quả của bệnh lao với quy mô lớn.

Trong các nghiên cứu trước đây, chẳng hạn như CNN đã đạt được kết quả tương đối tốt trong việc phân loại và nhận dạng hình ảnh y tế về bệnh lao. Eman Showkatian và các cộng sự [8] đã đề xuất một kiến trúc CNN (ConvNet), đạt độ chính xác 88,0%, độ nhạy 87,0%, điểm F1 87.0% và AUC 87.0% trên các bộ dữ liệu phân loại bệnh lao ở Thâm Quyến và Montgomery, Trung Quốc. Qingchen, Zhang và các cộng sự [9] đạt được độ chính xác 87.71% trên tập dữ liệu phân loại bệnh lao Montgomery bằng cách thay đổi lớp global average pooling của mô hình thành một lớp adaptive dropout trên mạng ResNet50. Xie cùng các cộng sự [10] đề xuất một phương pháp phát hiện tổn thương lao đa lớp. Thuật toán này giới thiệu một cấu trúc kim tự tháp có thể mở rộng học tập trong Fast Region Convolutional Neural Network (Faster RCNN) và đạt được hiệu suất tốt trên hai bộ dữ liệu công cộng, tập dữ liệu Montgomery: AUC: 97.7% và Accuracy: 92.6%; tập hợp dữ liệu Thâm Quyến: AUC: 94.1% và Accuracy: 90.2%.

Abideen, Zain và các cộng sự [11] đề xuất một giải pháp cho việc xác định bệnh lao thông qua mạng lưới thần kinh chuyển động Bayesian (B-CNN). So với các phương pháp học máy hiện đại và CNN, B-CNN đạt được đọ 96,42% (Montgomery dataset) và 86,46% (Thâm Quyến Dataset) trên hai tập dữ liệu, tương ứng. Ngoài ra, còn có các báo cáo nghiên cứu về việc xử lý trước hình ảnh y tế thông qua việc tăng cường dữ liệu. Munadi và các cộng sự [12] đạt được độ chính xác phân loại 89.92% và AUC 94.8% trên các mô hình ResNet và EfficientNet thông qua các phương pháp tăng cường hình ảnh như Unsharp Masking (UM), High-Frequency Emphasis Filtering (HEF) và Contrast Limited Adaptive Histogram Equalization (CLAHE). Tất cả các kết quả được thu thập bằng cách sử dụng bộ dữ liệu công cộng của Thâm Quyến. Trong số các phương pháp truyền thống, đã có những nghiên cứu thông qua phương pháp như máy học. Jaeger cùng cộng sự [13] tính toán kết cấu và đặc điểm hình dạng ở vùng phổi được cắt bằng biểu đồ, và có được bộ dữ liệu được thu thập bởi dự án kiểm soát bệnh lao của bộ phận y tế quận địa phương ở Hoa Kỳ và dữ liệu của Thâm Quyến được thiết lập bằng phương pháp học máy hồi quy tuyến tính logistic (LLR). Khu vực dưới ROC curve (AUC) là 87% (78,3% chính xác), và AUC của tập dữ liệu Thâm Quyến là 90% (84% chính xác).

Trong các nghiên cứu hiện có, chúng tôi phát hiện ra rằng các phương pháp end-to-end hiện tại vẫn có những hạn chế trong việc phát hiện hình ảnh bệnh lao phổi. Do sự can thiệp của nhiều mô khác nhau trong hình ảnh bệnh lao phổi, một mạng lưới thần kinh và phương pháp tăng cường dữ liệu đơn giản không thể thu thập thông tin quan trọng trong các hình ảnh y tế, làm cho việc phát hiện hình ảnh lao vẫn còn khó khăn. Để giải quyết vấn đề này, bài báo này sẽ kết hợp một cơ chế chú ý để trích xuất các đặc điểm hình ảnh tốt hơn và để cải thiện độ chính xác phân loại của hình ảnh bệnh lao phổi.

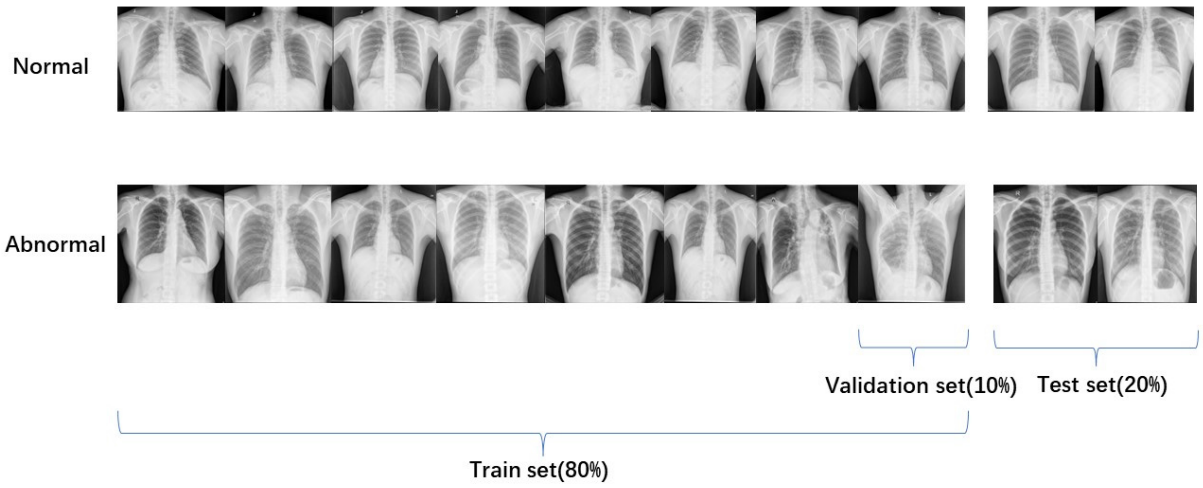
Chương 3

Tập dữ liệu

Tất cả các thí nghiệm được thực hiện trong bài báo này sử dụng Thâm Quyến tuberculosis CXR [14]. Bộ dữ liệu Thâm Quyến được thu thập với sự hợp tác của Bệnh viện Nhân dân Thượng Hải số 3, Đại học Y Quảng Đông, Thâm Quyến, Trung Quốc. X-quang ngực là từ các phòng khám ngoại trú và được chụp như là một phần của thói quen bệnh viện hàng ngày trong khoảng thời gian 1 tháng, chủ yếu là vào tháng 9 năm 2012, sử dụng hệ thống Philips DR Digital Diagnost. Bộ dữ liệu không có tiêu chí loại trừ, chẳng hạn như giới tính, tuổi tác hoặc chủng tộc. Bộ chứa 662 X-quang ngực phía trước, trong đó 326 là trường hợp bình thường và 336 là những trường hợp có biểu hiện của bệnh lao, bao gồm cả chụp X-quang ở trẻ em (AP). Các bức xạ được cung cấp ở định dạng PNG. Kích thước của chúng có thể khác nhau nhưng khoảng 3000×3000 pixel.

Mỗi mô hình mạng có một kích cỡ hình ảnh đầu vào cụ thể. Để đáp ứng các yêu cầu đầu vào của mô hình này, chúng tôi chuẩn hóa hình ảnh của bộ dữ liệu. Chúng tôi đồng đều chuyển đổi các hình ảnh tập dữ liệu thành độ phân giải 224×224 để đưa vào mạng đào tạo. Độ phân giải này đáp ứng các yêu cầu của mô hình, có thể giữ lại đầy đủ các chi tiết cấu trúc của hình ảnh, và giảm đáng kể chi phí tính toán của mẫu. Về phân chia tập dữ liệu, tác giả đã chọn ngẫu nhiên 20% kích cỡ mẫu từ các hình ảnh bình thường và bệnh lao làm tập kiểm tra để đảm bảo sự cân bằng giữa các mẫu dương và âm trong bộ dữ liệu. 10% khác được chọn ngẫu nhiên từ bộ đào tạo làm bộ xác thực cho việc điều chỉnh mô hình. Đồng thời, xét thấy đây là tập dữ liệu nhỏ, tác giả đã random tập dữ liệu, lật hình, thu phóng tỉ lệ, đảo chiều dài và rộng cho tập hình ảnh và thay đổi màu sắc, để

tránh overfitting và nâng cao khả năng tổng quát của mô hình. Sự phân chia của tập dữ liệu được hiển thị trong Hình 3.1.



Hình 3.1: Tập dữ liệu bệnh lao phổi Thâm Quyến.

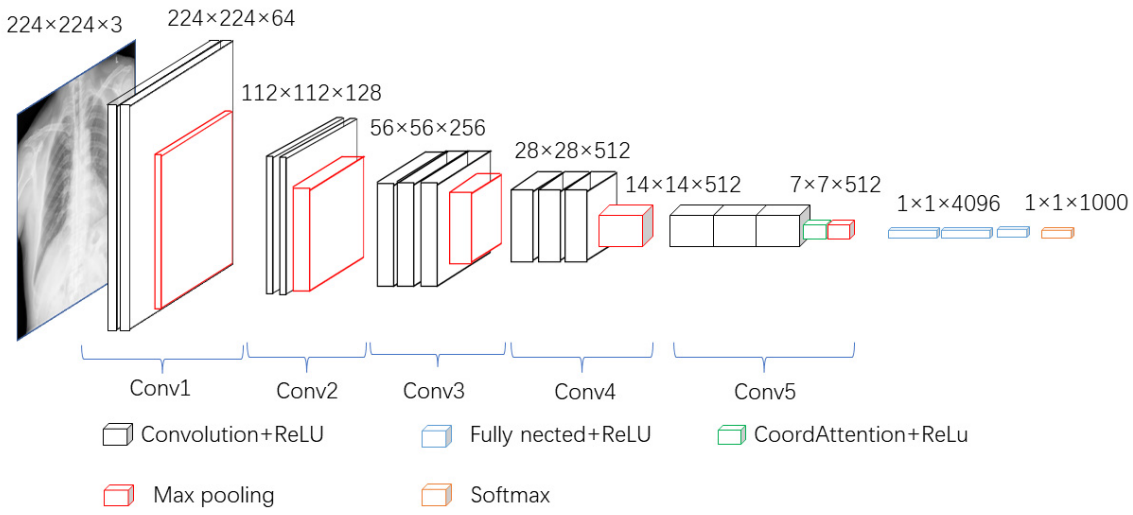
Chương 4

Phương pháp

Phương pháp của tác giả bao gồm một loạt các bước, bao gồm transfer learning, freezing network, khai thác tính năng và các nhiệm vụ phân loại bằng cách sử dụng học tập có giám sát. tác giả đã tiến hành ba thí nghiệm để xác minh tính khả thi của phương pháp của họ. Trong nghiên cứu đầu tiên, kết hợp cơ chế chú ý tọa độ với mạng VGG16, tác giả đề xuất một mô hình mạng của VGG16-CoordAttention, được đào tạo bằng cách đóng băng mạng để đánh giá hiệu ứng phân loại của nó trên tập dữ liệu Thâm Quyến. Trong nghiên cứu thứ hai, tác giả sử dụng bốn mô hình đại diện và mô hình của tác giả như là backbone của mạng để đánh giá hiệu quả của đào tạo mạng đống băng. Trong nghiên cứu thứ ba, tác giả tiến hành 5 lần kiểm tra chéo của phương pháp đề xuất để xác minh tính ổn định của mô hình.

4.1 Mô hình VGG16-CoordAttention

Mạng VGG16 là một CNN truyền thống. Mạng sử dụng các khối chuyển động nhỏ hơn. Bằng cách tăng độ sâu mạng, hiệu suất phân loại có thể được cải thiện hiệu quả. Để cải thiện độ chính xác phân loại của hình ảnh bệnh lao càng nhiều càng tốt, tác giả kết hợp mô hình VGG16 với CoordAttention để thiết lập mô hình mạng lưới deep learning về bệnh lao mới VGG16-CoordAttention. Cấu trúc mạng lưới được thể hiện trong Hình 4.1.



Hình 4.1: Kiến trúc mạng VGG16-CoordAttention.

Mạng VGG16 chủ yếu bao gồm convolution layer, lớp pooling và lớp fully connection. Mạng lưới yêu cầu hình ảnh đầu vào là một hình ảnh 3 kênh kích thước 224×224. Mạng lưới bao gồm năm khối tích chập (Conv). Trong Conv1, thực hiện xử lý 3×3 convolution và hàm kích hoạt ReLu trên hình ảnh đầu vào hai lần, lớp tính năng đầu ra là 64, và sau đó thông qua lớp 2×2 maxpooling, lớp maxpool nén chiều cao và chiều rộng của hình ảnh mà không thay đổi số lượng kênh để có được hình ảnh của $112 \times 112 \times 64$. Conv2 giống như Conv1, và mạng đầu ra là $56 \times 56 \times 128$. Conv3, Conv4 và Conv5 đều thực hiện ba lần xử lý 3×3 convolution và hàm kích hoạt ReLu trên hình ảnh đầu vào, và sau đó thực hiện khai thác tính năng toàn cầu thông qua lớp maxpool 2×2. Kết quả cuối cùng của mạng là $7 \times 7 \times 512$. Để đạt được mục tiêu phân loại cuối cùng, lớp đặc trưng tạo ra được phân loại sau khi đưa qua một khối fully connection.

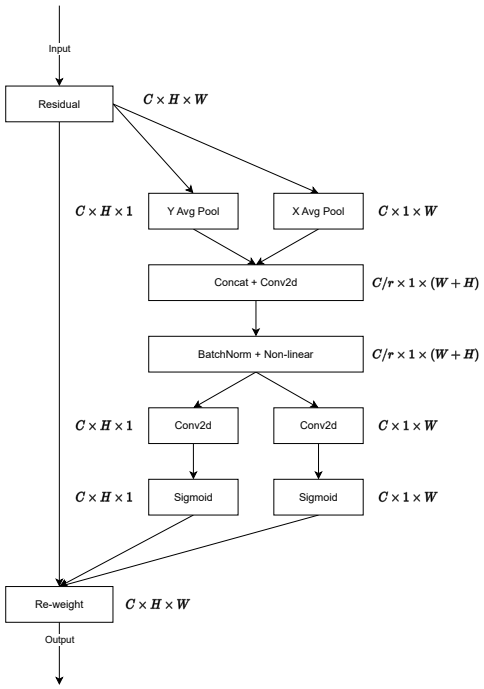
Ở cuối mạng backbone VGG16, trước lớp maxpool của Conv5, tác giả đã thêm cơ chế chú ý tọa độ và hàm kích hoạt ReLu, được gọi dưới đây là CA. CA là một cơ chế chú ý đơn giản và hiệu quả. Cấu trúc mạng lưới của nó được minh họa trong hình 4.2. Bằng cách nhúng thông tin vị trí vào sự chú ý kênh, mạng lưới có thể có được thông tin của một khu vực lớn hơn và xác định mục tiêu quan tâm chính xác hơn. Đồng thời, mỗi bản đồ chú ý (attention map) ghi lại sự phụ thuộc xa dọc theo một hướng không gian của bản

đồ đặc trưng (feature map) đầu vào. Cuối cùng, các bản đồ chú ý theo hai hướng không gian được gắn vào bản đồ đặc trưng đầu vào thông qua phép toán tích chập (convolution operation) để khôi phục số lượng kênh đầu vào, nhằm tăng cường khả năng trích xuất đặc trưng.

Ngoài ra, CA linh hoạt và nhẹ, có thể đóng vai trò cắm và phát huy tác dụng. Nó chỉ cần xác định số lượng kênh đầu vào và đầu ra. Nó có thể kết hợp tốt với mạng nơ-ron tích chập truyền thống và mạng nhẹ, nhằm tăng cường đặc trưng bằng cách tăng cường đại diện thông tin.

So với mô-đun chú ý kênh ECANET và mô-đun chú ý không gian CBAM trước đây, ECANET chỉ xem xét thông tin kênh nội bộ và bỏ qua tầm quan trọng của thông tin vị trí, thiếu sự tương tác thông tin giữa các kênh; Mặc dù mô-đun chú ý CBAM cố gắng giới thiệu thông tin vị trí bằng cách lấy trung bình toàn cục trên kênh, phương pháp này chỉ có thể thu được thông tin cục bộ, nhưng không thể thu được thông tin phụ thuộc xa.

Do sự can thiệp của các mô khác trong hình ảnh lao phổi, CA có thể trích xuất đặc trưng của hình ảnh theo hai hướng để xác định tốt hơn vị trí của tổn thương, nhằm nâng cao độ chính xác phân loại.

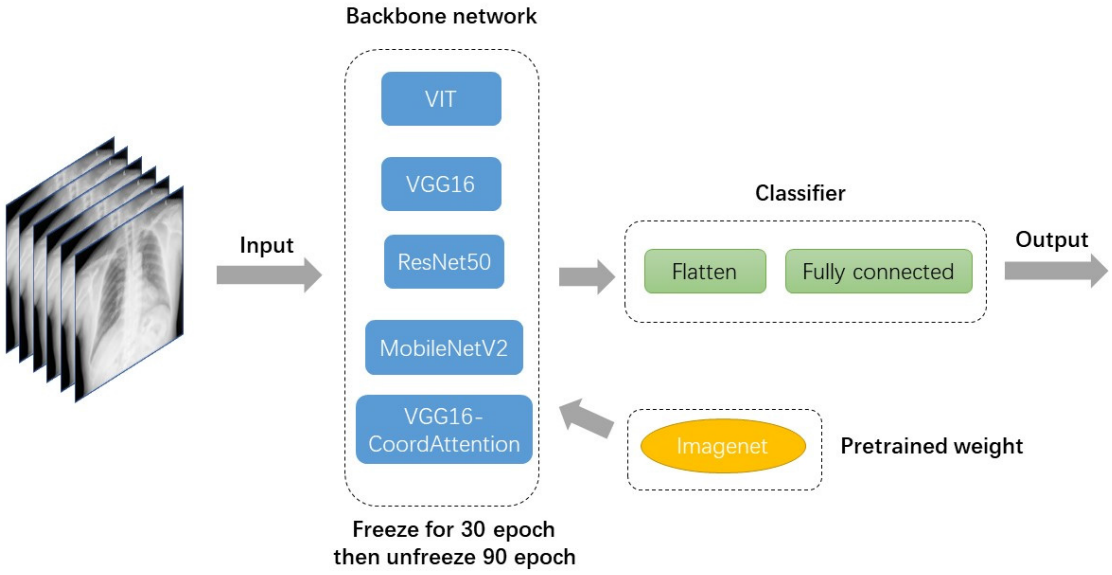


Hình 4.2: Kiến trúc CoordAttention.

4.2 Phương pháp huấn luyện dựa trên Transfer learning

Trong phần này, tác giả hy vọng đánh giá hiệu quả của phương pháp huấn luyện mạng transfer learning freezing thông qua các thực nghiệm. Phương pháp freezing network thuộc về một loại transfer learning. Khi tác giả cung cấp trọng số tiền huấn luyện cho mô hình, tác giả đóng băng mô hình mạng nền ở giai đoạn ban đầu của quá trình huấn luyện, để mô hình không thay đổi trọng số mạng nền ở giai đoạn ban đầu huấn luyện và tập trung vào huấn luyện bộ phân loại, để đặt nhiều tài nguyên hơn vào các tham số mạng ở phần sau của quá trình huấn luyện, điều này có thể cải thiện rất lớn việc sử dụng thời gian và tài nguyên.

Đồng thời, điều này có thể ngăn chặn việc trọng số bị phá hủy ở giai đoạn đầu huấn luyện, để mô hình có thể hội tụ nhanh chóng. Các thực nghiệm của tác giả cho thấy trên các mô hình của tác giả, phương pháp huấn luyện mạng đóng băng cũng có thể cải thiện độ chính xác phân loại.



Hình 4.3: Quá trình đào tạo Freeze network.

Trong nghiên cứu này, tác giả đã chọn bốn mô hình mạng đại diện và mô hình của tác giả để đánh giá hiệu quả của phương pháp đào tạo mạng đóng băng, Vision-Transformer (VIT) [15] đại diện cho mạng dựa trên cơ chế chú ý, VGG16 đại diện cho mạng nơ-ron

tích chập truyền thống, ResNet50 [16] đại diện cho mạng residual, và MobileNetV2 [17] là mạng nhẹ. Quá trình đào tạo được hiển thị trong hình 4.3. Quá trình huấn luyện được thể hiện trong Hình 4. Năm mạng trên sử dụng trọng số Imagenet đã huấn luyện sẵn để khởi tạo mô hình, và sử dụng mạng đóng băng và không đóng băng để huấn luyện lần lượt. Năm chỉ số đánh giá, Độ chính xác Top1, diện tích dưới đường cong ROC (AUC), độ nhạy (recall), độ chính xác (precision) và điểm F1, được sử dụng để đánh giá hiệu suất của các mạng nền khác nhau.

Trong đó, độ chính xác, độ nhạy (recall), độ chính xác (precision) và điểm F1 được tính như công thức (4.1),(4.2), (4.3), (4.4) :

$$\text{ACC} = \frac{TP+TN}{TP+TN+FP+FN}, \tag{4.1}$$

$$\text{Recall (Sensitivity)} = \frac{TP}{TP+FN}, \tag{4.2}$$

$$\text{Precision} = \frac{TP}{TP+FP}, \tag{4.3}$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{4.4}$$

Trong đó, TP (True Positive) có nghĩa là dự đoán các lớp tích cực là các số tích cực, TN (True Negative) có nghĩa là dự đoán các lớp tiêu cực là các số tiêu cực, FP (False Positive) có nghĩa là dự đoán các lớp tiêu cực là các số tích cực, và FN (False Negative) có nghĩa là dự đoán các lớp tích cực là các số tiêu cực.

Trong các thí nghiệm so sánh của hai nhóm, tác giả chạy epoch cho tổng cộng 120 lần để so sánh. Họ đóng băng các thông số mạng backbone 30 lần và mở khóa đào tạo mạng 90 lần, để đánh giá tác động của phương pháp đào tạo này. Learning rate được đặt là $1e-3$, batch-size là 8, bộ tối ưu hóa Adam và giảm độ dốc thuật toán được sử dụng, và overfitting được ngăn chặn bởi weight decay.

4.3 Xác thực chéo

Để xác minh độ bền của mô hình, Tác giả đánh giá mô hình bằng cách 5-fold cross validation. Họ chia tập dữ liệu thành 5 phần trung bình, mỗi lần lấy 4 phần làm tập huấn và 1 phần làm bộ xác minh. Cuối cùng, Tác giả có được trung bình kết quả của năm lần để xác minh khả năng tổng quát của mô hình.

Chương 5

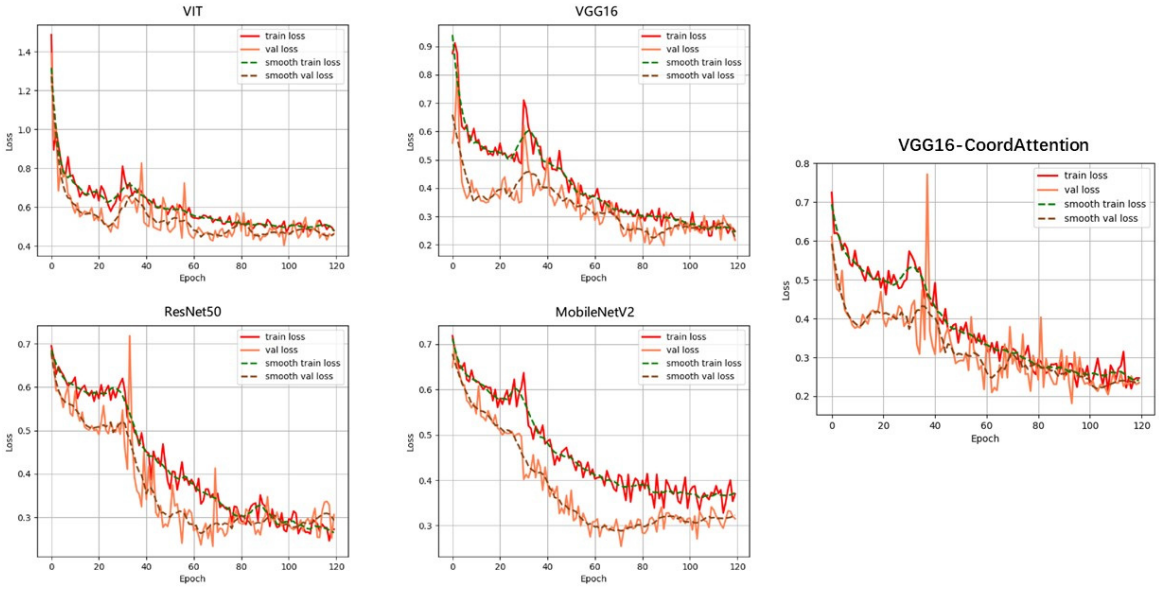
Kết quả thực nghiệm

5.1 Kết quả của phương pháp Transfer Learning

Dưới 5 mạng backbone khác nhau, tác giả sử dụng hàm cross entropy loss và phương pháp huấn luyện đóng băng mạng để đầu ra các đường cong hàm mất trên tập huấn luyện và tập xác thực, như được thể hiện trong Hình 5.1. Chúng ta có thể thấy rằng không phụ thuộc vào mô hình nào, trong giai đoạn sớm của quá trình huấn luyện sau 30 epoch, tỷ lệ mất trên tập huấn luyện và tập xác thực sẽ tăng đáng kể. Điều này xảy ra vì việc mở đóng đóng băng làm cho trọng số của thân mô hình được cập nhật và lặp lại, và những biến động xảy ra trong một khoảng thời gian ngắn, sau đó mô hình có thể nhanh chóng hội tụ và dần dần ổn định.

5.2 Lựa chọn Backbone Network

Tác giả tiến hành một thí nghiệm so sánh bằng cách sử dụng freeze network để đánh giá hiệu quả của việc đào tạo bằng phương pháp này. Kết quả được trình bày trong Bảng 5.1 và 5.2. Từ hai bảng dưới, chúng ta có thể tìm thấy rằng khi sử dụng VIT, VGG16, ResNet50 và VGG16-CoordAttention như là mạng backbone, phương pháp đào tạo của freeze network có thể cải thiện các chỉ số đánh giá của dự đoán các vấn đề phân loại đến một số mức độ, trong đó sự cải tiến của mạng CNN truyền thống VGG16 là rõ ràng nhất. Dưới phương pháp đào tạo này, hiệu ứng phân loại của mạng MobileNetV2 giảm, mà có



Hình 5.1: Đường cong hàm loss của tập huấn luyện và tập xác thực của năm mạng đường backbone bị đóng băng trên tập dữ liệu Thâm Quyến

thể là do thực tế rằng mạng này nhẹ có ít thông số và freeze network làm cho mô hình backbone không thể trích xuất các đặc điểm hình ảnh chính.

Như có thể thấy từ bảng 2, trong tập dữ liệu Thâm Quyến, phương pháp có thể đạt được độ chính xác cao nhất bằng cách đóng băng đào tạo mạng. Mặc dù độ chính xác của mạng lưới còn lại ResNet50 cũng cao, từ đường cong hàm mất mát trong 5.1, ảnh hưởng của mô hình đối với tập dữ liệu không ổn định và dao động rất nhiều. Trong vài lần lặp lại, tỷ lệ mất mát của bộ xác minh bắt đầu tăng dần, và có hiện tượng over fitting. Vì vậy cuối cùng tác giả đã chọn VGG16 mô hình như là mạng backbone gốc.

	Top1-ACC(%)	AUC(%)	Recall(%)	Precision(%)	F1(%)
VIT	78.79	84.32	78.87	79.20	77.78
VGG16	81.82	88.77	62.02	64.29	79.31
ResNet50	89.39	95.66	89.35	89.56	89.86
MobileNetV2	88.64	96.49	88.76	89.66	87.80
VGG16 - CoorAttention	**85.61**	**93.16**	**85.61**	**85.61**	**85.71**

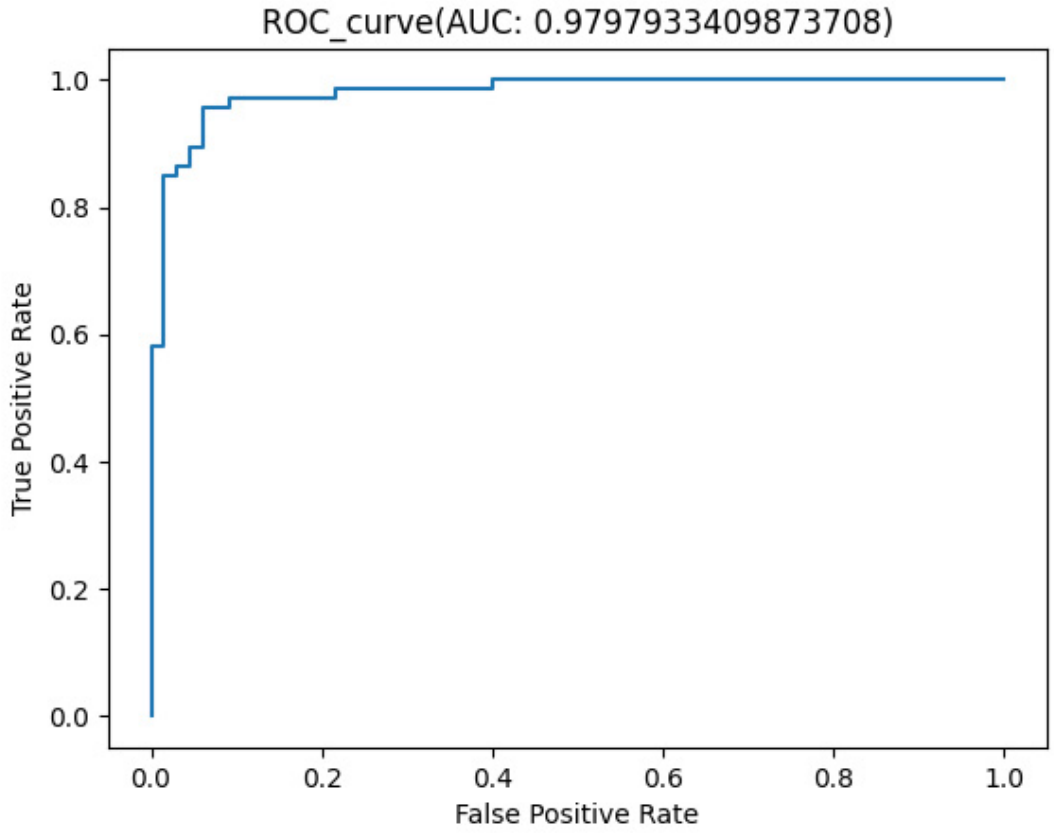
Bảng 5.1: Hiệu suất của năm mạng backbone trong quá trình đào tạo không bị đóng băng (In đậm là phương pháp của tác giả).

	Top1-ACC(%)	AUC(%)	Recall(%)	Precision(%)	F1(%)
VIT	80.30	90.56	80.46	81.68	78.33
VGG16	90.91	96.33	90.91	90.91	91.04
ResNet50	90.15	95.98	90.91	90.91	91.04
MobileNetV2	87.12	95.11	87.20	87.50	86.61
VGG16 - CoorAttention	**93.18**	**97.98**	**93.19**	**93.18**	**93.23**

Bảng 5.2: Hiệu suất của năm mạng backbone trong quá trình đào tạo bị đóng băng (In đậm là phương pháp của tác giả).

5.3 Thí nghiệm cắt bỏ

Trên tập dữ liệu Thâm Quyến, thí nghiệm cắt bỏ đã được thực hiện để xác minh tính hiệu quả của phương pháp của tác giả. Từ bảng 5.1, phương pháp này được xây đựng mà không cần đóng băng mạng, VGG16-CoordAttention so với VGG16, do thêm coordinate atteention, mô hình có thể tốt hơn về hướng và vị trí thông tin của hình ảnh, mà làm cho hiệu ứng phân loại của thuật toán được cải thiện.



Hình 5.2: Đường cong ROC trên bộ thử nghiệm.

Từ bảng 5.2, có thể thấy rằng bằng cách đóng băng mạng, mạng backbone giữ lại thông tin ngữ nghĩa cấp cao của trọng số được đào tạo trước ở giai đoạn đầu của việc đào tạo, và dành nhiều nguồn lực hơn để đào tạo phân loại, làm cho hiệu suất phân loại của VGG16-CoordAttention được cải thiện hơn nữa. tác giả đã sử dụng mô hình VGG16-CoordAttention với mạng đóng băng để có được kết quả tốt nhất hiện nay, với Top1-ACC: 93.18%, AUC: 97.98%, thu hồi: 93.19%, độ chính xác: 93,18% và điểm F1: 93,23%. Chỉ số AUC gần 98%. Hình 5.2 cho thấy đường cong ROC trên thiết lập thử nghiệm. Đường ROC cho thấy sự cân bằng giữa Recall và độ đặc hiệu (specificity) [18]. AUC được coi là phương pháp hiệu quả để hiển thị độ chính xác của ROC do mỗi mô hình tạo ra, điều này chứng minh rằng mô hình của chúng tôi có khả năng phân loại xuất sắc trong nhiệm vụ phát hiện lao.

5.4 Thí nghiệm so sánh

	Top1-ACC(%)	AUC(%)	Recall(%)	Precision(%)	F1(%)
1	80.30	90.56	80.46	81.68	78.33
2	90.91	96.33	90.91	90.91	91.04
3	90.15	95.98	90.91	90.91	91.04
4	87.12	95.11	87.20	87.50	86.61
5	87.12	95.11	87.20	87.50	86.61
AVG	**93.18**	**97.98**	**93.19**	**93.18**	**93.23**

Bảng 5.3: Kết quả xác thực chéo 5 lần trên tập dữ liệu Thâm Quyến (Phần in đậm thể hiện kết quả trung bình).

	ACC (%)	AUC (%)
ConvNet [8]	87	87
ResNet50+ AdaptiveDropout [9]	81.80	-
FPN+Faster RCNN [10]	90.20	94.10
B-CNN [11]	86.46	-
EfficientNet+HEF [12]	89.92	94.80
LLR [13]	84.00	90
Author's method	**92.73**	**97.71**

Bảng 5.4: So với VGG16-CoordAttention và các phương pháp hiện có trên tập dữ liệu Thâm Quyến (Những phương pháp in đậm thể hiện kết quả tốt nhất)

Phương pháp của tác giả đã đạt được kết quả xuất sắc trong phân loại bệnh lao phổi trong tập dữ liệu Thâm Quyến. Theo thứ tự để xác minh tính mạnh mẽ và khả năng tổng quát của mô hình, họ đã đánh giá mô hình thông qua 5 phương pháp xác nhận chéo. Hiển thị kết quả của 5 lần xác nhận chéo trong bảng 5.3. Chỉ số đánh giá của họ là trên 91%, chứng minh rằng mô hình của chúng ta có khả năng tổng quát nhất định.

Hơn nữa tác giả đã so sánh kết quả trung bình của 5 lần xác nhận chéo với các phương pháp end-to-end hiện có khác, và đánh giá chúng với các chỉ số độ chính xác và AUC đại diện cho các vấn đề phân loại. Hiệu suất so sánh là được trình bày trong bảng 5.4. Kết quả cho thấy phương pháp của họ tốt hơn kết quả hơn so với các phương pháp từ end-to-end khác trên tập dữ liệu Thâm Quyến.

Chương 6

Kết luận và thảo luận

Trong nghiên cứu của tác giả, họ đề xuất một mô hình CoordinateAttention vào mạng nơ-ron tích chập VGG16. So với các mô hình phổ biến hiện tại, việc thêm cơ chế chú ý cho phép phương pháp của họ tập trung tốt hơn vào thông tin vị trí và hướng trong hình ảnh lao phổi, nhằm đạt được độ chính xác phân loại tốt hơn. Đồng thời, họ sử dụng phương pháp Freeze Network để tăng tốc quá trình huấn luyện mô hình và cải thiện hiệu suất của mạng. So với các phương pháp end-to-end hiện tại, phương pháp của họ có hiệu quả tốt hơn. Phương pháp của họ không cần sử dụng học tập tập hợp cho sự hợp nhất nhiều mô hình, cũng không cần tiêu thụ nguồn lực tính toán lớn. So với các phương pháp truyền thống sử dụng nhiều phương pháp tiền xử lý dữ liệu để cải thiện hiệu suất, phương pháp của họ cũng tránh các phương pháp tiền xử lý hoặc mở rộng dữ liệu của các nhiệm vụ cụ thể. Kết quả của việc kiểm định chéo năm lần cũng cho thấy phương pháp của họ có khả năng tổng quát nhất định. Ngoài ra, độ chính xác phân loại, AUC, độ chính xác, độ nhớ và điểm F1 của mô hình của họ trong việc phát hiện lao phổi đều trên 91%. Kết quả thu được từ phương pháp của họ có hiệu suất tốt hơn, có thể giúp các bác sĩ chẩn đoán hình ảnh lao phổi. Trong nghiên cứu tiếp theo, họ hy vọng đạt được hiệu suất tương tự thông qua việc sử dụng các mạng nhẹ như mobilenetv2, để việc tính toán cũng có thể được thực hiện trên các thiết bị di động, điều này tiện lợi hơn cho các bác sĩ hỗ trợ trong việc chẩn đoán.

Tài liệu tham khảo

[1] K. R. Steingart, M. Henry, V. Ng, P. C. Hopewell, A. Ramsay, J. Cunningham, R. Urbanczik, M. Perkins, M. A. Aziz, and M. Pai, "Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review," *The Lancet Infectious Diseases*, vol. 6, no. 9, pp. 570–581, 2006.

[2] W. H. Organization, *Chest Radiography in Tuberculosis Detection: Summary of Current WHO Recommendations and Guidance on Programmatic Approaches*. Geneva, Switzerland: World Health Organization, 2016.

[3] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," 2021.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[5] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," 2017.

[6] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[7] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

[8] E. Showkatian, M. Salehi, H. Ghaffari, R. Reiazi, and N. Sadighi, "Deep learning-based automatic detection of tuberculosis disease in chest x-ray images," *Pol. J. Radiol.*, vol. 87, no. 1, pp. e118–e124, Feb. 2022.

[9] Q. Zhang, C. Bai, Z. Liu, L. T. Yang, H. Yu, J. Zhao, and H. Yuan, "A gpu-based residual network for medical image classification in smart medicine," *Information Sciences*, vol. 536, pp. 91–100, 2020.

[10] Y. Xie, Z. Wu, X. Han, H. Wang, Y. Wu, L. Cui, J. Feng, Z. Zhu, and Z. Chen, "Computer-aided system for the detection of multicategory pulmonary tuberculosis in radiographs," *J. Healthc. Eng.*, vol. 2020, p. 9205082, Aug. 2020.

[11] Z. Ul Abideen, M. Ghafoor, K. Munir, M. Saqib, A. Ullah, T. Zia, S. A. Tariq, G. Ahmed, and A. Zahra, "Uncertainty assisted robust tuberculosis identification with bayesian convolutional neural networks," *IEEE Access*, vol. 8, pp. 22 812–22 825, 2020.

[12] K. Munadi, K. Muchtar, N. Maulina, and B. Pradhan, "Image enhancement for tuberculosis detection using deep learning," *IEEE Access*, vol. 8, pp. 217 897–217 907, 2020.

[13] S. Jaeger, A. Karargyris, S. Candemir, L. Folio, J. Siegelman, F. Callaghan, Z. Xue, K. Palaniappan, R. K. Singh, S. Antani, G. Thoma, Y.-X. Wang, P.-X. Lu, and C. J. McDonald, "Automatic tuberculosis screening using chest radiographs," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 233–245, 2014.

[14] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, "Two public chest x-ray datasets for computer-aided screening of pulmonary diseases," *Quant. Imaging Med. Surg.*, vol. 4, no. 6, pp. 475–477, Dec. 2014.

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[17] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.

[18] D. L. Streiner and J. Cairney, “What’s under the ROC? an introduction to receiver operating characteristics curves,” *Can. J. Psychiatry*, vol. 52, no. 2, pp. 121–128, Feb. 2007.