

ENHANCING THE EFFECTIVENESS OF TUBERCULOSIS DIAGNOSIS MODELS BASED ON VISUALIZATION TECHNIQUES

Nguyen Trong Vinh¹, Pham Thi Bay², Lam Thanh Hien¹, Do Nang Toan^{3*}

¹Lac Hong University, ²Lilama 2 International Technology College

³Institute of Information Technology - Vietnam Academy of Science and Technology

ARTICLE INFO		ABSTRACT
Received:	28/5/2024	The robust development of deep learning models has solved many practical problems, including supporting disease diagnosis through images. Although much progress has been achieved, explaining the decisions made by deep learning models remains a significant challenge. In some cases, these models use information outside the diagnostic area. This paper proposes a solution through visualization in diagnosing tuberculosis from chest X-ray images. It highlights the regions within the images that the deep learning model utilizes, aiming to detect discrepancies within the training dataset images. This enables data normalization and the application of techniques to improve the model's accuracy. The proposal has been implemented in trials and has shown effectiveness with deep learning models for chest X-ray images, aiding in the diagnosis of tuberculosis. It not only improves the reliability but also increases the accuracy of the model by adjusting and normalizing the training data based on real findings from the visualization process.
Revised:	26/6/2024	
Published:	26/6/2024	
KEYWORDS		
Imaging diagnosis		
Data visualization		
Deep learning		
Score-CAM		
Transfer learning		

NÂNG CAO HIỆU QUẢ MÔ HÌNH CHẨN ĐOÁN BỆNH LAO DỰA TRÊN KỸ THUẬT TRỰC QUAN HÓA

Nguyễn Trọng Vinh¹, Phạm Thị Bầy², Lâm Thành Hiện¹, Đỗ Năng Toàn^{3*}

¹Trường Đại học Lạc Hồng, ²Trường Cao đẳng Công nghệ quốc tế Lilama 2

³Viện Công nghệ thông tin - Viện Hàn lâm Khoa học và Công nghệ Việt Nam

THÔNG TIN BÀI BÁO		TÓM TẮT
Ngày nhận bài:	28/5/2024	Sự phát triển mạnh mẽ của các mô hình học sâu đã giúp giải quyết nhiều vấn đề trong thực tiễn trong đó có hỗ trợ chẩn đoán bệnh dựa vào hình ảnh. Mặc dù đã đạt được nhiều tiến bộ, nhưng việc giải thích quyết định của các mô hình học sâu vẫn còn là một thách thức lớn. Trong một số trường hợp, mô hình học sâu sử dụng thông tin nằm ngoài vùng cần chẩn đoán. Bài báo này đề xuất một giải pháp thông qua việc trực quan hóa trong chẩn đoán bệnh lao từ ảnh X-quang lồng ngực để làm nổi bật các vùng trong ảnh mà mô hình học sâu sử dụng, để phát hiện sự không ăn khớp của các ảnh trong tập dữ liệu huấn luyện. Nhờ đó có thể chuẩn hóa dữ liệu và áp dụng các kỹ thuật nâng cao độ chính xác của mô hình. Đề xuất đã được cài đặt thử nghiệm và tỏ ra hiệu quả với mô hình học sâu cho ảnh chụp X-quang lồng ngực, hỗ trợ chẩn đoán bệnh lao. Không chỉ cải thiện khả năng tin cậy, mà còn tăng độ chính xác của mô hình bằng cách điều chỉnh và chuẩn hóa dữ liệu huấn luyện theo những phát hiện thực tế từ quá trình trực quan hóa.
Ngày hoàn thiện:	26/6/2024	
Ngày đăng:	26/6/2024	
TỪ KHÓA		
Chẩn đoán hình ảnh		
Trực quan hóa dữ liệu		
Học sâu		
Score-CAM		
Học chuyển tiếp		

DOI: <https://doi.org/10.34238/tnu-jst.10483>

* Corresponding author. Email: donangtoan@gmail.com

1. Giới thiệu

Trong thập kỷ qua, sự phát triển của các mô hình học sâu đã thúc đẩy sự tiến bộ vượt bậc của trí tuệ nhân tạo (AI) và học máy đã mở ra những cơ hội mới mẻ để cải thiện đáng kể chất lượng và hiệu quả của chẩn đoán y tế. Trong số các bệnh nhiễm trùng gây tử vong hàng đầu trên toàn cầu, lao phổi tiếp tục là một trong những thách thức lớn nhất cho hệ thống y tế toàn cầu, với hàng triệu ca mới được chẩn đoán mỗi năm. Việc phát hiện sớm và chính xác lao phổi thông qua ảnh X-quang là một bước quan trọng trong việc kiểm soát và điều trị bệnh, giảm thiểu nguy cơ lây nhiễm và cải thiện kết quả điều trị cho bệnh nhân. Từ thực tế đó, việc ứng dụng các mô hình học sâu trong phân loại ảnh lao phổi trở thành một công cụ mạnh mẽ, hứa hẹn thay đổi cách thức chúng ta tiếp cận với việc chẩn đoán và điều trị lao.

Trong bối cảnh dữ liệu bệnh nhân Việt Nam, chúng tôi đã cung cấp một giải pháp để chẩn đoán bệnh lao phổi trên bộ dữ liệu hình ảnh X-Ray của Việt Nam, được thu thập từ một bệnh viện địa phương ở Việt Nam với sự trợ giúp của VRPACS [1]. Từ đó, chúng tôi đã thiết kế các chiến lược khác nhau để xây dựng hệ thống phân loại tự động với các mô hình khác nhau bằng cách sử dụng phương pháp học chuyển giao [2]. Hình 1 dưới đây thể hiện kết quả chuẩn đoán bệnh lao của hệ thống mà chúng tôi đã xây dựng.



Hình 1. Hệ thống hỗ trợ chẩn đoán lao phổi: Ảnh gốc (hình trái), kết quả chẩn đoán được trực quan hóa (hình phải)

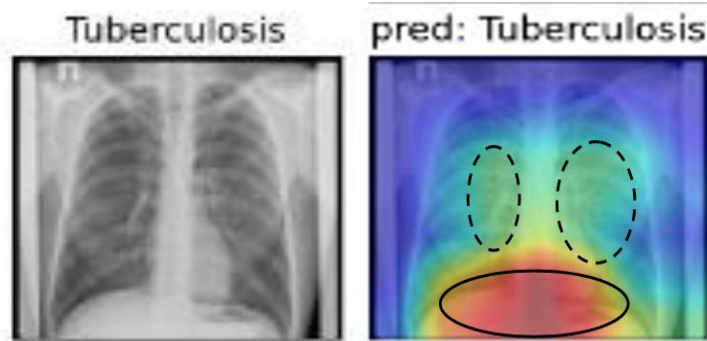
Tuy kết quả đạt được bước đầu khá tốt [2] nhưng việc khoanh vùng nghi nhiễm lao trong một số trường hợp lại nằm ngoài vùng tổn thương của phổi, điều này có thể do một số dữ liệu được đưa vào huấn luyện có phần sai lệch so với số chung của mô hình. Trong bài báo này, chúng tôi đề xuất một giải pháp thông qua việc trực quan hóa trong chẩn đoán bệnh lao từ ảnh X-quang lồng ngực để làm nổi bật các vùng trong ảnh mà mô hình học sâu sử dụng, để phát hiện sự không ăn khớp của các ảnh trong tập dữ liệu huấn luyện. Nhờ đó có thể chuẩn hóa dữ liệu và áp dụng các kỹ thuật nâng cao độ chính xác của mô hình.

Trong những năm gần đây, sự tiến bộ của học sâu [3], [4] trong chẩn đoán hình ảnh X-quang đã mang lại những cải tiến đáng kể. Bolei Zhou và các cộng sự đã đặt nền móng quan trọng với công trình "Learning Deep Features for Discriminative Localization" được trình bày tại CVPR năm 2016 [5]. Công trình này mở rộng khả năng của mô hình học sâu không chỉ trong việc phân loại mà còn trong việc xác định vị trí dấu hiệu bệnh lý trên ảnh X-quang. Điều này đã cải thiện độ chính xác của quyết định chẩn đoán bằng cách làm nổi bật các khu vực có tính chẩn đoán cao trên hình ảnh.

Tiếp nối thành công, Tawsifur Rahman và đồng nghiệp đã đưa ra nghiên cứu trong hội nghị CVPR 2020 với tên gọi "Reliable Tuberculosis Detection using Chest X-ray with Deep Learning Segmentation and Visualization" [6], trong đó họ đã thể hiện khả năng của mô hình học sâu trong việc phân loại và xác định vị trí các tổn thương do bệnh lao từ ảnh X-quang. Nghiên cứu này không chỉ chứng tỏ hiệu quả của các mô hình trong việc phát hiện bệnh lao mà còn cho thấy tiềm năng ứng dụng rộng rãi trong việc chẩn đoán các bệnh lý khác từ hình ảnh X-quang.

Mỗi bước tiến trong nghiên cứu không chỉ là một sự cải tiến kỹ thuật mà còn là một bước tiến quan trọng trong việc hỗ trợ các bác sĩ, giúp họ chẩn đoán sớm và chính xác hơn. Các nghiên cứu này cũng góp phần định hình lại cách thức mà chúng ta tiếp cận và xử lý dữ liệu y tế, từ đó nâng cao chất lượng điều trị cho bệnh nhân.

Bằng cách liên kết các nghiên cứu quan trọng trước đây, chúng ta có thể nhìn thấy làn sóng của sự thay đổi mà học sâu mang lại cho chẩn đoán hình ảnh y tế. Mỗi nghiên cứu không chỉ giải quyết một vấn đề cụ thể mà còn đưa ra cơ hội mới để phát triển các công nghệ hỗ trợ y tế hiệu quả hơn, thể hiện kết quả rõ ràng hơn và dễ tiếp cận hơn. Tuy nhiên, các công trình này thường tập trung vào việc phát triển các mô hình phức tạp với lượng dữ liệu lớn, nhằm cho ra độ chính xác cao trong việc phát hiện các dấu hiệu bệnh, nhưng chúng thường gặp khó khăn trong việc giải thích tại sao một khu vực nhất định lại được chọn để đưa ra kết luận. Các mô hình có thể chú ý đến những vùng không liên quan đến bệnh lý để đưa ra quyết định của mình, điều này có thể do một số dữ liệu trong tập huấn luyện bị nhiễu hoặc không đồng nhất. Hình 2, cho thấy trường hợp, mô hình học sâu tập trung vào các vùng của hình ảnh không chứa tổn thương phổi mà lại bị ảnh hưởng bởi các yếu tố như chất lượng ảnh, chế độ chụp, hoặc những chi tiết nền không liên quan.



Hình 2. Mô hình học sâu tập trung vào vùng của hình ảnh không chứa tổn thương phổi (nét liền), vùng tổn thương (nét đứt)

Trong khi các mô hình học sâu đã đạt được những bước tiến vượt bậc trong việc chẩn đoán bệnh từ ảnh X-quang, một thách thức lớn vẫn còn tồn tại là khả năng giải thích các quyết định của chúng. Việc các mô hình này đôi khi dựa vào thông tin nằm ngoài vùng lâm sàng cần thiết để đưa ra kết luận làm dấy lên mối quan ngại về độ tin cậy của quá trình chẩn đoán tự động. Một vấn đề cốt lõi là các mô hình học sâu thường được "học" một cách không giám sát hoặc bán giám sát từ dữ liệu huấn luyện lớn mà không có đủ kiểm soát về các đặc điểm lâm sàng có ý nghĩa. Điều này dẫn đến tình trạng mô hình có thể nhận dạng những mẫu trong dữ liệu mà không thể lý giải tại sao hoặc những mẫu đó có liên quan như thế nào đến tình trạng bệnh lý cụ thể. Trong y tế, điều này không chỉ làm giảm chất lượng chẩn đoán mà còn có thể dẫn đến những hậu quả nghiêm trọng cho sức khỏe bệnh nhân.

Để giải quyết vấn đề này, các nghiên cứu gần đây đang hướng tới việc phát triển các kỹ thuật trực quan hóa hình ảnh [7] – [9], không chỉ các mô hình học sâu đã đạt được những tiến bộ vượt bậc trong chẩn đoán hình ảnh y tế, mà còn có sự phát triển đáng kể trong lĩnh vực trực quan hóa để làm cho các quyết định của mô hình trở nên rõ ràng và minh bạch hơn. Selvaraju và các cộng sự đã đề xuất công trình “Visual Explanations from Deep Networks via Gradient-based Localization” [10] vào năm 2017 và công trình “Leveraging explanations to make vision and language models more grounded” [11] vào năm 2019, họ đề xuất một kỹ thuật gọi là Grad-Cam để tạo ra 'lời giải thích trực quan' cho các quyết định từ một lớp lớn các mô hình dựa trên Mạng thần kinh chuyển đổi (CNN), làm cho chúng trở nên minh bạch và dễ giải thích hơn. Cách tiếp cận của chúng tôi – Ảnh xạ kích hoạt lớp có trọng số theo độ dốc (Grad-CAM), sử dụng độ dốc của bất kỳ khái niệm mục tiêu nào (giả sử 'con chó' trong mạng phân loại hoặc một chuỗi các từ

trong mạng phụ đề) chảy vào lớp chấp cuối cùng để tạo ra một bản đồ thô bản đồ định vị làm nổi bật các vùng quan trọng trong ảnh để dự đoán khái niệm. Tiếp đó, Haofan Wang và các cộng sự đã đưa ra một bước tiến quan trọng với Score-CAM, một kỹ thuật trực quan hóa không dựa vào gradient, cho phép hiển thị chính xác các vùng mà mô hình sử dụng để đưa ra quyết định. Kỹ thuật này được giới thiệu trong công trình "Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks" [12] vào năm 2020, nâng cao khả năng giải thích và minh bạch của mô hình thông qua trực quan hóa trực tiếp đầu ra của mô hình.

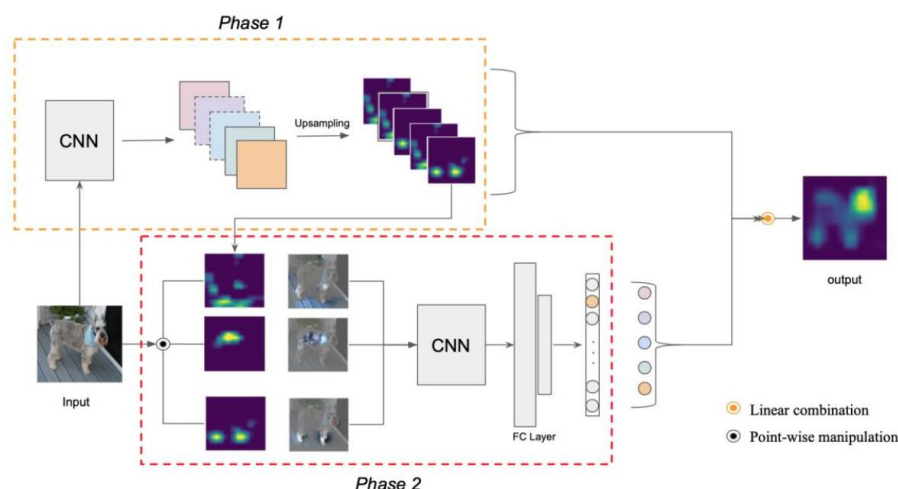
Tiếp nối nghiên cứu này, các phương pháp khác như LIFT-CAM và Ablation-CAM cũng đã được phát triển để cải thiện thêm khả năng trực quan hóa. LIFT-CAM, một sự cải tiến được giới thiệu bởi Hyungsik Jung và đồng nghiệp, áp dụng phương pháp DeepLIFT [13] để tính toán các trọng số của activation maps, qua đó cung cấp một phương pháp giải thích dựa trên đóng góp của từng đặc trưng cho quyết định cuối cùng của mô hình. Điều này không chỉ giúp các nhà nghiên cứu hiểu rõ hơn về cách thức mà mô hình làm việc, mà còn giúp các bác sĩ kiểm soát chất lượng quyết định một cách chính xác hơn.

Phần còn lại của bài báo được cấu trúc như sau: phần 2 là kỹ thuật đề xuất, tiếp đến là thử nghiệm và cuối cùng là kết luận về kỹ thuật đề xuất.

2. Kỹ thuật đề xuất

Trong lĩnh vực y tế, trực quan hóa hình ảnh là một công cụ quan trọng không chỉ giúp các bác sĩ dễ dàng nhận diện các bệnh lý từ ảnh chụp X-quang mà còn hỗ trợ việc hiểu và giải thích quyết định của các mô hình học sâu. Các mô hình thường được huấn luyện để tự động nhận dạng và tập trung vào những đặc trưng quan trọng trong ảnh mà chúng cho là có khả năng dự đoán nhãn lớp một cách chính xác. Các đặc trưng này có thể bao gồm cạnh, góc, hình dạng đặc thù, hoặc các khu vực có màu sắc và cường độ sáng khác biệt. Tất cả đều có ý nghĩa trong việc xác định dấu hiệu của bệnh lý trong ảnh X-quang. Tuy nhiên, quá trình này không phải lúc nào cũng hoàn hảo và có thể dễ dàng bị ảnh hưởng bởi các yếu tố nhiễu khác.

Kỹ thuật Score-CAM, bằng cách không dựa vào gradient mà sử dụng các đầu ra trực tiếp từ mô hình, cho phép chúng ta trực quan hóa các vùng quan trọng mà mô hình sử dụng để đưa ra quyết định, giúp các bác sĩ và bệnh nhân có cái nhìn rõ ràng và dễ hiểu hơn về lý do tại sao một hình ảnh được phân loại vào trạng thái bệnh hoặc bình thường. Bằng cách này, Score-CAM không chỉ giải quyết các thách thức về trực quan hóa mà còn tạo điều kiện cho việc chuẩn hóa và cải thiện tập dữ liệu huấn luyện, từ đó nâng cao độ chính xác và tin cậy của mô hình. Điều này là hết sức quan trọng trong lĩnh vực y tế, nơi mỗi quyết định chẩn đoán có thể ảnh hưởng trực tiếp đến sức khỏe và cuộc sống của bệnh nhân.



Hình 3. Mô hình trực quan hóa hình ảnh sử dụng Score-CAM

Trong đề xuất của mình chúng tôi sử dụng quy trình tiếp cận đa pha và sử dụng kỹ thuật trực quan hóa Score-CAM. Hình 3 cho thấy trong giai đoạn đầu tiên, mạng CNN được sử dụng để trích xuất các đặc trưng và tạo ra các bản đồ đặc trưng với quá trình upsampling để tăng cường độ phân giải. Giai đoạn thứ hai sử dụng các bản đồ này, áp dụng các phép toán như tổ hợp tuyến tính và thao tác điểm để tính toán đóng góp của từng vùng trong bức ảnh đối với quyết định cuối cùng của mô hình. Quá trình này nhằm khám phá khả năng cải thiện hiệu suất mô hình và tăng cường tính rõ ràng, dễ hiểu của kết quả chẩn đoán. Việc cung cấp cái nhìn trực quan về các đặc điểm mà mô hình học sâu sử dụng để nhận dạng và phân loại các tổn thương liên quan đến bệnh lao, giúp xác định các vùng trọng yếu không dựa vào gradient mà thông qua các đầu ra trực tiếp của mô hình. Điều này, giúp cho việc hiệu chỉnh tập huấn luyện sát với mô hình hơn, từ đó nâng cao chất lượng phân loại.

2.1. Kỹ thuật nâng cao hiệu quả phân loại lao phổi dựa trên trực quan hóa

Kỹ thuật được thực hiện qua bốn pha chính: Huấn luyện mô hình chẩn đoán hình ảnh, Áp dụng Score-CAM, Chuẩn hóa dữ liệu, Tái huấn luyện Mô hình. Cụ thể như sau:

Pha 1: Huấn luyện mô hình

Chúng tôi bắt đầu với việc huấn luyện một mô hình học sâu để nhận dạng và phân loại các dấu hiệu của lao phổi từ ảnh X-quang. Mô hình này được thiết kế để tự động phát hiện các tổn thương và các đặc điểm bệnh lý, dựa trên học sâu và các thuật toán phân tích ảnh tiên tiến. Trong giai đoạn này, mô hình học từ một tập dữ liệu lớn, bao gồm cả ảnh chụp của bệnh nhân bị nghi ngờ nhiễm lao và những người không nhiễm bệnh, để phát triển khả năng phân biệt chính xác giữa các trạng thái bệnh và không bệnh.

Pha 2: Áp dụng Score-CAM

Sau khi mô hình đã được huấn luyện, chúng tôi áp dụng kỹ thuật Score-CAM để trực quan hóa kết quả phân loại. Đây là bước đột phá, thuật toán Score-CAM không dựa vào gradient mà sử dụng các đầu ra trực tiếp từ mô hình để xác định tầm quan trọng của các feature maps. Phương pháp này cho phép chúng tôi hiển thị rõ ràng các vùng ảnh mà mô hình dựa vào để đưa ra quyết định, từ đó giúp các bác sĩ hiểu được cơ sở của các quyết định chẩn đoán, cũng như cung cấp một cái nhìn trực tiếp và chính xác về cách thức phân loại ảnh.

Pha 3: Chuẩn hóa bộ dữ liệu

Trong quá trình áp dụng Score-CAM, chúng tôi phát hiện ra sự không ăn khớp trong các ảnh của tập dữ liệu huấn luyện, điều này có thể dẫn đến việc mô hình học các đặc điểm sai lệch hoặc không liên quan. Để giải quyết vấn đề này, chúng tôi tiến hành tinh chỉnh và chuẩn hóa tập dữ liệu bằng cách loại bỏ nhiễu và điều chỉnh các mẫu không phù hợp, đảm bảo rằng mô hình học được từ những thông tin chính xác và có giá trị dự đoán cao.

Pha 4: Tái huấn luyện mô hình

Cuối cùng, sau khi đã tinh chỉnh và chuẩn hóa dữ liệu, chúng tôi tiến hành huấn luyện lại mô hình với tập dữ liệu đã được cải thiện. Quá trình này nhằm tăng cường khả năng và độ chính xác của mô hình trong việc phân loại bệnh. Bằng cách áp dụng Score-CAM, mô hình không chỉ cải thiện về mặt độ chính xác mà còn tăng cường khả năng giải thích, làm cho các quyết định chẩn đoán trở nên rõ ràng và dễ hiểu hơn, từ đó hỗ trợ hiệu quả hơn trong việc chẩn đoán và điều trị bệnh lao.

2.2. Thuật toán trực quan hóa sử dụng Score-CAM

Đầu vào: Mô hình mạng nơ-ron đã qua huấn luyện `model`, tensor hình ảnh đầu vào `x`, tên lớp `layer_name`, số lượng tối đa các bản đồ kích hoạt tùy chọn `max_N`

Đầu ra: Chỉ số lớp dự đoán `cls`, Bản đồ kích hoạt lớp `cam`

Bắt đầu

Bước 1: Tính chỉ số lớp dự đoán từ đầu ra của mô hình cho đầu vào `x`.

$$cls \leftarrow \text{argmax}(f(\text{model}(x)))$$

Bước 2: Trích xuất các bản đồ kích hoạt cho `layer\_name` từ `model` với đầu vào `x`.

$A \leftarrow model(input=model.input, output=model.get_layer(layer_name).output)(x)$

Bước 3: Nếu `max\_N` được chỉ định, chọn các bản đồ kích hoạt hàng đầu `max\_N` với độ lệch chuẩn cao nhất.

If $max_N \neq -1$

$S \leftarrow \{std(a_i) \mid i = 1, \dots, N, a_i \in A\}$

$A \leftarrow \{a_i \mid a_i \in S_{top_max_N}\}$

Bước 4: Mở rộng các bản đồ kích hoạt đã chọn đến kích thước không gian tương tự như hình ảnh đầu vào `x`.

$A' \leftarrow \{resize(a_i, d) \mid a_i \in A\}$

$A'' \leftarrow \{normalize(a_i') \mid a_i' \in A'\}$

Bước 5: Tạo các phiên bản đã che của hình ảnh đầu vào `x` bằng cách nhân từng phần tử với từng bản đồ kích hoạt đã chuẩn hóa.

$M \leftarrow \{x \otimes a_i'' \mid a_i'' \in A''\}$

Bước 6: Thực hiện dự đoán mô hình trên các đầu vào đã che và áp dụng hàm softmax cho các giá trị logits.

$P \leftarrow softmax(f(model(M)))$

$w \leftarrow \{p_cls \mid p \in P\}$

Bước 7: Tính Bản đồ Kích hoạt Lóp như là tổng có trọng số của các bản đồ kích hoạt đã chuẩn hóa.

$cam \leftarrow \sum(a_i'' * w_i \mid a_i'' \in A'', w_i \in w)$

$cam \leftarrow ReLU(cam)$

$cam \leftarrow cam / max(cam)$

Bước 8: Trả về chỉ số lớp dự đoán và Bản đồ Kích hoạt Lóp.

Return (cls, cam)

Kết thúc

2.3. Độ phức tạp của thuật toán

- Chi phí huấn luyện mô hình $T_1 = O(n)$
- Chi phí trực quan hóa tại Pha 2 $T_2 = O(n^2)$
- Chi phí tái huấn luyện mô hình $T_3 = O(n)$
- Do đó, tổng chi phí thực hiện là: $T = T_1 + T_2 + T_3 = O(n^2)$

3. Thử nghiệm và đánh giá kết quả

Chúng tôi đã cài đặt thử nghiệm của kỹ thuật đề xuất, với bộ dữ liệu VRTB được chuẩn bị với sự hỗ trợ của nhóm [2], gồm 2000 hình ảnh X-quang ngực được phân loại cẩn thận thành hai trạng thái: “bệnh” và “bình thường” được mô tả trong Bảng 1. Sự cân nhắc giữa số lượng hình ảnh “bệnh” và “bình thường” đảm bảo sự cân đối và đối sánh công bằng trong quá trình thực nghiệm.

Bảng 1. Số lượng mẫu trong tập dữ liệu VRTB

Tập dữ liệu	Bệnh lao	Bình thường
VRTB	500	500

KaggleTB, được công bố trên nền tảng Kaggle bởi Rahman và cộng sự [14], cung cấp một lượng dữ liệu hình ảnh phong phú từ các nguồn uy tín, tạo điều kiện thuận lợi cho việc áp dụng và so sánh các kết quả thu được từ Score-CAM. Đặc biệt, việc chọn lọc 4200 hình ảnh có sẵn từ nền tảng này cho phép chúng tôi tiếp cận với dữ liệu bệnh lao từ bệnh nhân quốc tế, mở rộng phạm vi và tính đa dạng của nghiên cứu. Chi tiết về bộ dữ liệu VRTBCombineTrain như trong Bảng 2.

Bảng 2. Mô tả dữ liệu trong VRTBCombineTrain

VRTBCombineTrain	Bệnh lao	Bình thường
Ảnh X-quang	1200	4000
Tập huấn luyện	768	2560
Tập xác thực	192	640
Tập kiểm tra (20%)	240	800

Sau trực quan hóa và đánh giá bộ dữ liệu, chúng tôi đã chọn lọc và chuẩn hóa các mẫu dữ liệu để đảm bảo chúng phản ánh một cách chính xác các tính chất cần thiết cho mô hình học sâu thể hiện ở Bảng 3.

Bảng 3. Bộ dữ liệu VRTBCombineTrain sau chuẩn hóa

VRTBCombineTrain	Bệnh lao	Bình thường
Ảnh X-quang	1185	4015
Tập huấn luyện	758	2570
Tập xác thực	190	642
Tập kiểm tra (20%)	237	803

Sau khi chuẩn hóa dữ liệu, chúng tôi đã tiến hành huấn luyện lại mô hình và tiến hành chạy so sánh thuật toán trước sử dụng mô hình huấn luyện bộ dữ liệu chưa chuẩn hóa và mô hình huấn luyện sau khi chuẩn hóa, kết quả thể hiện qua Bảng 4.

Bảng 4. Kết quả huấn luyện mô hình (Theo Toan và các cộng sự [2]) trên tập VRTBCombineTrain đã chuẩn hóa

	Giá trị AUC trung bình	Độ nhạy trung bình	Độ đặc hiệu trung bình	Độ chính xác trung bình
Theo Toan và các cộng sự [2] (Chưa chuẩn hóa bộ dữ liệu)	0,9929	0,8745	0,992	0,9333
Kỹ thuật đề xuất	0,9953	0,9748	0,9993	0,9821

Kết quả trên cho thấy việc áp dụng Score-CAM giúp làm rõ các vùng ảnh hưởng đến quyết định phân loại, qua đó giúp các bác sĩ dễ dàng hiểu nguyên nhân dẫn đến kết quả chẩn đoán của mô hình. Điều này không chỉ tăng cường khả năng tin cậy mà còn cải thiện độ chính xác của mô hình bằng cách cho phép điều chỉnh và chuẩn hóa dữ liệu huấn luyện dựa trên những phát hiện thực tế từ trực quan hóa.

4. Kết luận

Bệnh lao là nguyên nhân gây tử vong đứng hàng thứ hai trong các bệnh nhiễm trùng, hàng năm có khoảng 1,3 triệu người tử vong do lao, và có thêm khoảng 300.000 ca tử vong do đồng nhiễm lao/HIV. Việc ứng dụng công nghệ trí tuệ nhân tạo vào việc hỗ trợ phát hiện sớm và chính xác lao phổi thông qua ảnh X-quang là một bước quan trọng trong việc kiểm soát và điều trị bệnh, giảm thiểu nguy cơ lây nhiễm và cải thiện kết quả điều trị cho bệnh nhân.

Bài báo này đề xuất một giải pháp thông qua việc trực quan hóa trong chẩn đoán bệnh lao từ ảnh X-quang lồng ngực để làm nổi bật các vùng trong ảnh mà mô hình học sâu sử dụng, để phát hiện sự không ăn khớp của các ảnh trong tập dữ liệu huấn luyện. Nhờ đó có thể chuẩn hóa dữ liệu và áp dụng các kỹ thuật nâng cao độ chính xác của mô hình. Đề xuất đã được cài đặt thử nghiệm và tỏ ra hiệu quả với mô hình học sâu đối với ảnh X-quang lồng ngực nói chung và hỗ trợ chẩn đoán bệnh lao nói riêng.

Lời cảm ơn

Nghiên cứu này được tài trợ bởi đề tài KC-4.0.14/19-25, “Nghiên cứu xây dựng hệ thống hỗ trợ chẩn đoán và dự báo dịch tễ địa không gian (geo-spatial epidemiology) bệnh lao phổi bằng ảnh X-quang ngực ở Việt Nam”.

TÀI LIỆU THAM KHẢO/REFERENCES

- [1] VRPACS Software, 2022. [Online]. Available: <http://vnpcacs.vn>. [Accessed April 03, 2024].
- [2] T. M. Ha, H. T. Lam, V. D. Ngo, and T. N. Do, "Detecting Tuberculosis from Vietnamese X-Ray Imaging Using Transfer Learning Approach," *Computers, Materials & Continua*, vol.74, pp. 5001-5016, 2022.
- [3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.
- [4] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
- [5] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2921-2929.
- [6] T. Rahman, A. Khandakar, M. A. Kadir, K. R. Islam, K. F. Islam, R. Mazhar, T. Hamid, M. T. 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 Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 191586–191601.
- [7] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *CoRR*, 2013, doi: 10.48550/arXiv.1312.6034.
- [8] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *ECCV*, 2014, pp. 818–833.
- [9] Z. C. Lipton, "The Mythos of Model Interpretability," *Communications of the ACM*, vol. 61, no. 10, June 2016, doi: 10.1145/3233231.
- [10] R. R. Selvaraju *et al.*, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," *International Conference on Computer Vision (ICCV)*, 2017, doi: 10.1109/ICCV.2017.74.
- [11] R. R. Selvaraju, S. Lee, Y. Shen, H. Jin, S. Ghosh, L. Heck, D. Batra, and D. Parikh, "Taking a hint: Leveraging explanations to make vision and language models more grounded," in *Proceedings of the International Conference on Computer Vision (ICCV)*, 2019, pp. 2591-2600.
- [12] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding *et al.*, "Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 24-25.
- [13] H. Jung and Y. Oh, "Towards Better Explanations of Class Activation Mapping," *International Conference on Computer Vision (ICCV)*, 2021, pp. 1336 - 1344.
- [14] N. Pandey, "Chest Xray Masks and Labels," 2020. [Online]. Available: <https://www.kaggle.com/nikhilpandey360/chestxray-masks-and-labels>. [Accessed April 03, 2024].