

NÂNG CAO HIỆU QUẢ PHÂN LỚP ẢNH X-QUANG VỚI MÔ HÌNH HỌC KẾT HỢP BỎ PHIẾU MỀM CÓ TRỌNG SỐ ĐƯỢC TỐI ƯU HÓA BẰNG TÌM KIẾM LƯỚI

Phạm Thanh An^{1,2}, Lê Văn Tường Lân³, Hoàng Văn Dũng^{4*}

¹ Trường Đại học Khoa học, Đại học Huế, Huế, Việt Nam

² Trường Đại học Ngân hàng TP. Hồ Chí Minh, Hồ Chí Minh, Việt Nam

³ Đại học Huế, Huế, Việt Nam

⁴ Trường Đại học Công nghệ Kỹ thuật TP. Hồ Chí Minh, Hồ Chí Minh, Việt Nam

* Tác giả liên hệ Hoàng Văn Dũng <dunghv@hcmute.edu.vn>

(Ngày nhận bài: 30-03-2026; Hoàn thành phản biện: 10-06-2026; Ngày chấp nhận đăng: 20-06-2026)

Tóm tắt. Các mô hình học sâu độc lập thường bị hạn chế do thiên kiến quy nạp vốn có của chúng: mạng nơ-ron tích chập (CNNs) vượt trội trong việc tự động trích xuất các kết cấu, đặc trưng cục bộ, nhưng thiếu nhận thức ngữ cảnh toàn cục. Vision Transformer (ViT) có thể nắm bắt các phụ thuộc tầm xa nhưng có độ phức tạp cao, cần lượng dữ liệu lớn. Bài báo này đề xuất một khung kết hợp lai tích hợp các thể mạnh khác nhau của các mô hình CNN (ResNet-50, DenseNet-201) và Transformer (Swin-T). Thay vì áp dụng phương pháp bỏ phiếu mềm tiêu chuẩn, trong đó tất cả các mô hình được xem xét như nhau, chúng tôi giới thiệu chiến lược bỏ phiếu mềm có trọng số dựa trên tìm kiếm lưới. Đóng góp tối ưu của mỗi mô hình cơ sở được tinh chỉnh thông qua tìm kiếm lưới toàn diện trên một tập dữ liệu xác thực, và đánh giá cuối cùng trên tập dữ liệu kiểm thử độc lập nhằm tránh rò rỉ dữ liệu. Các thực nghiệm được thực hiện trên hai bộ dữ liệu ảnh X-quang ngực, mỗi bộ gồm bốn lớp bệnh lý. Phương pháp cải tiến dựa trên bỏ phiếu mềm của chúng tôi đạt hiệu quả phân lớp tốt hơn so với các mô hình cơ sở riêng lẻ và các phương pháp học kết hợp truyền thống. Trên tập dữ liệu ChestXray-ImageDataSet, độ chính xác tăng 0,23% và 3,22% so với mô hình tốt nhất Soft Voting và mô hình kém nhất ResNet-50. Trên tập dữ liệu COVID-19, phương pháp đề xuất tăng 0,54% về độ chính xác so với mô hình tốt nhất ResNet-50 và 4,76% độ chính xác so với mô hình cơ sở kém nhất Swin-T với cùng một phương pháp huấn luyện.

Từ khóa: tìm kiếm lưới, bỏ phiếu mềm có trọng số, học kết hợp, phân lớp ảnh X-quang

Improving Chest X-ray Image Classification Efficiency through a Grid-Search Optimized Weighted Soft Voting Ensemble

Phạm Thanh An^{1,2}, Lê Văn Tường Lân³, Hoàng Văn Dũng^{4*}

¹ University of Sciences, Hue University, Hue, Vietnam

² Ho Chi Minh University of Banking, Hochiminh, Vietnam

³ Hue University, Hue, Vietnam

⁴ HCMC University of Technology and Engineering, Hochiminh, Vietnam

* Correspondence to Hoàng Văn Dũng <dunghv@hcmute.edu.vn>

(Received: 30 March 2026; Revised: 10 June 2026; Accepted: 20 June 2026)

Abstract. Independent deep learning models are often limited by their inherent inductive bias: convolutional neural networks (CNNs) excel at automatically extracting local textures and features but lack global contextual awareness. Vision Transformers (ViTs) can capture long-range dependencies but are highly complex and require large datasets. This paper proposes a hybrid framework that integrates the different strengths of CNNs (ResNet-50, DenseNet-201) and Transformer (Swin-T) models. Instead of using the standard soft voting method, where all models are treated equally, we introduce a weighted soft voting strategy based on a grid search. The optimal contribution of each baseline model is refined via a comprehensive grid search on a validation dataset, and the final evaluation is performed on an independent test dataset to avoid data leakage. Experiments were conducted on two chest X-ray datasets, each comprising four pathological classes. Our proposed soft voting-based improvement achieved better classification performance compared to individual baseline models and traditional ensemble methods. On the ChestXray-ImageDataSet dataset, accuracy increased by 0.23% and 3.22%, respectively, compared with the best-performing model, Soft Voting, and the worst-performing model, ResNet-50. On the COVID-19 dataset, the proposed method increased accuracy by 0.54% compared to the best-performing ResNet-50 model and 4.76% compared to the worst-performing baseline model Swin-T, using the same training method.

Keywords: Grid search, Weighted soft voting, ensemble learning, X-ray classification

1 Mở đầu

Học kết hợp (Ensemble Learning) là một phương pháp trong học máy, được xây dựng dựa trên triết lý "Trí tuệ đám đông". Mục tiêu cốt lõi của phương pháp này là kết hợp sức mạnh dự đoán của nhiều mô hình học máy cơ sở (base learners) để tạo ra một mô hình tổng hợp (strong learner) có độ chính xác, tính ổn định và khả năng tổng quát hóa vượt trội so với bất kỳ mô hình đơn lẻ nào trong tập hợp [1]. Học kết hợp đã được nghiên cứu và phát triển mạnh mẽ từ những năm 1990. Học kết hợp xây dựng nhiều bộ phân loại hoặc mô hình học cơ sở, sau đó kết hợp các kết quả dự đoán của chúng để cải thiện hiệu quả phân loại. Các mô hình cơ sở được gọi là các mô hình yếu [2]. Các phương pháp kết hợp toán học cố gắng xây dựng mô hình và suy ra các quy tắc kết hợp bằng cách sử dụng logic và thống kê [3].

Fumera và cộng sự [4] đã tiến hành một nghiên cứu toàn diện, kết hợp giữa phân tích lý thuyết và thực nghiệm, nhằm đánh giá hiệu quả của các mô hình phân loại, trong đó trọng tâm đặt vào các phương pháp kết hợp tuyến tính. Các tác giả đã tập trung khảo sát hai cơ chế kết hợp phổ biến: trung bình đơn giản và trung bình có trọng số. Kết quả phân tích cho thấy hiệu năng của các

mô hình kết hợp tuyến tính không chỉ phụ thuộc vào độ chính xác riêng lẻ của từng bộ phân loại thành phần, mà còn chịu ảnh hưởng đáng kể bởi mức độ tương quan giữa các đầu ra của chúng. Khi các bộ phân loại có sai số độc lập hoặc có tương quan thấp, việc kết hợp sẽ mang lại lợi ích rõ rệt hơn. Đáng chú ý, các tác giả chứng minh rằng phương pháp trung bình có trọng số thường đạt hiệu quả vượt trội so với trung bình đơn giản. Điều này được lý giải bởi khả năng phân bổ mức độ đóng góp khác nhau cho từng bộ phân loại, qua đó khai thác tốt hơn thông tin từ các mô hình có hiệu suất cao và hạn chế ảnh hưởng của các mô hình kém chính xác. Kittler và cộng sự [5] đã đề xuất một khung lý thuyết tổng quát cho bài toán kết hợp các bộ phân loại. Các tác giả đặc biệt tập trung vào các mô hình tổ hợp, trong đó mỗi bộ phân loại được xây dựng trên những tập đặc trưng riêng biệt, qua đó khai thác tính bổ sung thông tin giữa các không gian đặc trưng khác nhau nhằm nâng cao hiệu quả phân loại tổng thể [5].

Học sâu đã nổi lên như một nhánh quan trọng của học máy và được ứng dụng rộng rãi trong các bài toán thị giác máy tính. Đó là một phương pháp xử lý dữ liệu sử dụng nhiều lớp cấu trúc phức tạp hoặc nhiều lớp xử lý bao gồm nhiều

phép biến đổi phi tuyến tính [6]. Từ năm 2012, sau thành công của AlexNet [7] trong bài toán phân loại ảnh tự nhiên, học sâu được tập trung nghiên cứu và ứng dụng rộng rãi. Học sâu được ứng dụng trong phân loại hình ảnh là một chủ đề thu hút sự quan tâm nghiên cứu. Trong số đó, mạng nơ-ron tích chập (Convolutional Neural Networks - CNNs) được sử dụng rộng rãi nhất.

Hình ảnh y tế, là một loại hình ảnh trực quan đặc biệt, đã thu hút sự chú ý của nhiều nhà nghiên cứu. Trong những năm gần đây, hệ thống chẩn đoán có sự hỗ trợ của máy tính (Computer-Aided Diagnosis - CAD) dựa trên học sâu (Deep Learning) đã tạo ra một bước ngoặt lớn trong lĩnh vực phân tích hình ảnh y khoa [8]. Nhờ khả năng trích xuất đặc trưng tự động với độ chính xác cao, các mô hình CNNs đã trở thành công cụ đắc lực giúp giảm tải áp lực cho các chuyên gia y tế và hạn chế tối đa các sai sót chẩn đoán mang tính chủ quan. Trong đó, mạng nơ-ron tích chập từ lâu đã thiết lập tiêu chuẩn vàng trong các tác vụ phân loại ảnh X-quang nhờ đặc tính bất biến không gian và thiên vị quy nạp (inductive bias) mạnh mẽ, cho phép nhận diện xuất sắc các vi tổn thương cục bộ.

Tuy nhiên, cấu trúc tích chập cục bộ của kiến trúc CNNs lại bộc lộ giới hạn trong việc nắm bắt các mối liên hệ khoảng cách xa và bối cảnh giải phẫu toàn cục. Để giải quyết rào cản này, Vision Transformer (ViT) [9] đã được giới thiệu và nhanh chóng chứng minh sự ưu việt thông qua cơ chế tự chú ý (self-attention), giúp mô hình hóa xuất sắc các đặc trưng ngữ cảnh toàn cục (global context). Dù vậy, ViT thường đòi hỏi khối lượng dữ liệu huấn luyện khổng lồ và có xu hướng bỏ qua các chi tiết vật lý tinh vi ở cấp độ pixel nếu không được thiết kế tối ưu cho hình ảnh y tế. Do đó, xu hướng nghiên cứu hiện đại đang dịch chuyển mạnh mẽ sang việc xây dựng các hệ thống lai, các mô hình học kết hợp nhằm tạo sức mạnh cho sự kết hợp giữa kiến trúc CNNs và Transformer. Mặc dù sự kết hợp các kiến trúc

không đồng nhất mang lại tiềm năng to lớn, các chiến lược kết hợp hiện tại vẫn còn những điểm nghẽn nghiêm trọng. Các phương pháp phổ biến như Bỏ phiếu cứng (Hard Voting) hay Bỏ phiếu mềm tiêu chuẩn (Soft Voting) thường áp dụng chiến lược "cào bằng" quyền lực giữa các mô hình, vô tình làm loãng các dự đoán có độ tin cậy cao và không khai thác triệt để thế mạnh riêng biệt của từng kiến trúc. Trên các tập dữ liệu y tế thường bị giới hạn về quy mô và có sự mất cân bằng lớp sâu sắc, cách tiếp cận này tiêu tốn lượng lớn tài nguyên tính toán và cực kỳ dễ dẫn đến hiện tượng quá khớp (overfitting), làm suy giảm nghiêm trọng khả năng khái quát hóa khi triển khai trên lâm sàng.

Học tập kết hợp là một phương pháp toán học tiêu chuẩn nhằm giảm thiểu những điểm yếu của các mô hình riêng lẻ. Bằng cách tổng hợp các dự đoán từ nhiều mô hình độc lập, các phương pháp kết hợp làm giảm phương sai và sự không chắc chắn trong dự đoán. Kỹ thuật kết hợp phổ biến nhất, bỏ phiếu mềm tiêu chuẩn, tính toán trung bình cộng không trọng số của xác suất lớp [10]. Mặc dù đơn giản về mặt tính toán, phương pháp này giả định rằng tất cả các mô hình thành phần đều có độ tin cậy dự đoán như nhau. Trong các tập hợp không đồng nhất chứa cả CNNs và Transformer, giả định này là sai lầm. Một sơ đồ trọng số đồng nhất thường làm giảm độ tin cậy của các dự đoán cao của một mô hình vượt trội bằng cách lấy trung bình chúng với các đầu ra của các mô hình kém chính xác hơn.

Với những rào cản nêu trên, nghiên cứu này đề xuất một khung hệ thống học kết hợp, kết hợp hai mạng CNNs tiên tiến (gồm ResNet-50, DenseNet-201) và một kiến trúc Swin Transformer (Swin-T). Để tối ưu hóa quá trình kết hợp, chúng tôi giới thiệu cơ chế bỏ phiếu mềm có trọng số (Weighted Soft Voting) được tinh chỉnh thông qua thuật toán tìm kiếm lưới (Grid Search) toàn diện. Bằng cách thực hiện hiệu chuẩn hậu kỳ trên một tập kiểm định độc lập, hệ thống có khả năng tự

động dò tìm tỷ trọng phân bố trọng số tối ưu cho từng mạng cơ sở mà không cần bất kỳ chi phí lan truyền ngược (backpropagation) bổ sung nào, đồng thời tránh được hoàn toàn rủi ro rò rỉ dữ liệu (data leakage). Ý tưởng kết hợp các mô hình học sâu cơ sở không phải là mới, nhưng vẫn còn rất nhiều tiềm năng để phát triển các phương pháp kết hợp mới, các loại đặc trưng và bộ phân loại được sử dụng, cũng như các ứng dụng mới trong lĩnh vực y tế.

Những đóng góp chính của nghiên cứu này có thể được tóm tắt như sau:

- **Đề xuất mô hình kết hợp các kiến trúc không đồng nhất:** Đề xuất một khung học kết hợp lai ghép hiệu quả giữa khả năng trích xuất vùng tổn thương cục bộ của CNNs (gồm ResNet-50, DenseNet-201) và khả năng nắm bắt bối cảnh toàn cục của Transformer (Swin-T), giúp bù đắp các khiếm khuyết về khả năng biểu diễn đặc trưng của hình ảnh y khoa.

- **Tối ưu hóa trọng số hiệu quả:** Triển khai chiến lược tối ưu hóa trọng số thông qua thuật toán tìm kiếm lưới. Phương pháp của chúng tôi nhằm tối đa hóa trực tiếp các chỉ số chẩn đoán (F1-Score) trên tập xác thực, đồng thời vẫn duy trì khả năng diễn giải trên tập kiểm thử mà không phát sinh thêm chi phí huấn luyện. Đây là một cải tiến của phương pháp đề xuất so với phương pháp bỏ phiếu mềm sử dụng trung bình xác suất của ba mô hình cơ sở.

2 Các nghiên cứu liên quan

Trong thập kỷ qua, sự bùng nổ của học sâu (Deep Learning - DL) đã tạo ra một bước ngoặt lớn trong chẩn đoán hình ảnh y khoa. Việc áp dụng kỹ thuật học chuyển giao (Transfer Learning) [11] từ các bộ trọng số được tiền huấn luyện trên ImageNet đã trở thành phương pháp tiêu chuẩn nhằm khắc phục tình trạng khan hiếm dữ liệu y tế được gán nhãn. Nhiều nghiên cứu đã chứng minh tính hiệu quả của các kiến trúc mạng

neuron tích chập như ResNet [12] và DenseNet [13] trong việc nhận diện các bệnh lý ở lồng ngực. Với cấu trúc kết nối tắt và kết nối dày đặc (dense connectivity), các mạng này giải quyết thành công vấn đề triệt tiêu đạo hàm, cho phép trích xuất các đặc trưng vi mô cực kỳ rõ nét như các nốt mờ hay vùng tổn thương trên phim X-quang. Manaee và cộng sự [14] đã sử dụng phương pháp học chuyển giao trên mô hình ResNet, DenseNet để tự động phát hiện COVID-19. Phương pháp của họ đạt tỷ lệ độ nhạy là 98% ($\pm 3\%$), trong khi tỷ lệ độ đặc hiệu khoảng 90%.

Gần đây với sự xuất hiện của Vision Transformer (ViT) [9] và những mô hình cải tiến dựa trên Transformer tiêu biểu là kiến trúc Swin-Transformer [15] đã mang đến một cách tiếp cận mới. Thay vì phụ thuộc vào các bộ lọc tích chập cục bộ, Swin Transformer sử dụng cơ chế tự chú ý dạng cửa sổ dịch chuyển (shifted-window self-attention) để mô hình hóa các mối quan hệ ngữ cảnh toàn cục (global context) của các cơ quan giải phẫu. Mặc dù cả CNNs và Transformer đều đạt được những kết quả ấn tượng, những nghiên cứu gần đây chỉ ra rằng: mô hình CNNs thường có ưu thế trong việc nắm bắt, trích xuất các kết cấu cục bộ (local textures), trong khi Transformer lại vượt trội ở khả năng nhận diện hình thái cấu trúc tổng thể. Sự bù trừ này chính là tiền đề cho các mô hình kết hợp không đồng nhất [16]. Phạm và Hoàng [17] đã cải tiến mô hình Transformer để phát hiện các bệnh phổi sử dụng phim X-quang, đối với phát hiện lớp COVID-19 tỷ lệ đạt 99.78%.

Để khai thác sức mạnh của nhiều mô hình khác nhau, nhiều giải pháp đề xuất sử dụng kỹ thuật học tập kết hợp. Các phương pháp như bỏ phiếu cứng (Majority/Hard Voting) và bỏ phiếu mềm (Soft Voting) [5], [10]. Trong bỏ phiếu cứng, mỗi bộ phân loại đưa ra một nhãn lớp riêng. Dự đoán cuối cùng được xác định bằng bỏ phiếu đa số, trong đó lớp nhận được nhiều phiếu bầu nhất sẽ được chọn. Với phương pháp bỏ phiếu mềm, mỗi bộ phân loại cung cấp một phân phối xác suất

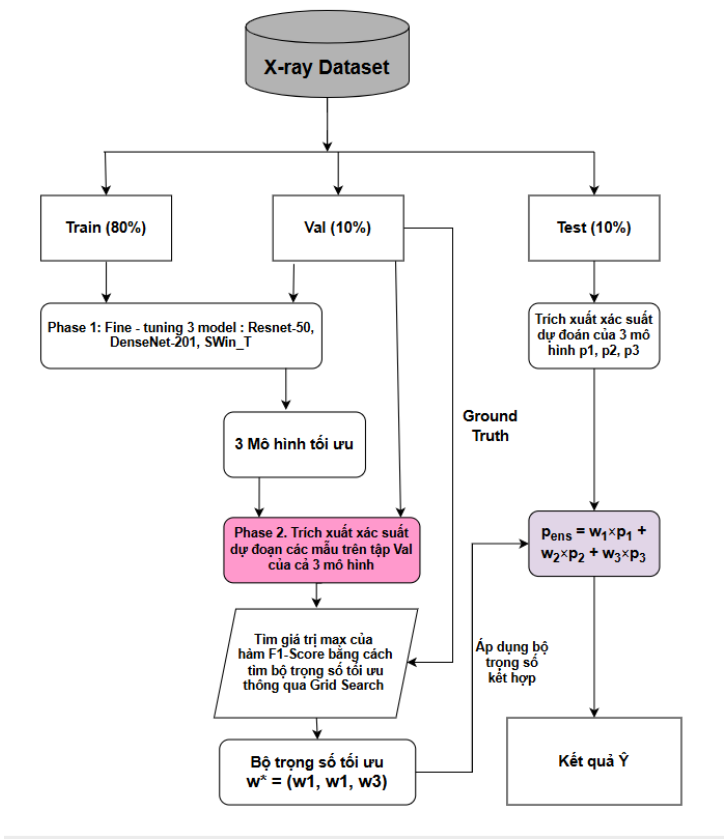
trên tất cả các lớp. Các xác suất này được tính trung bình, và lớp có xác suất kết hợp cao nhất được chọn làm dự đoán cuối cùng. Mặc dù cách tiếp cận học kết hợp cải thiện độ chính xác so với các mô hình đơn lẻ, các phương pháp truyền thống này vẫn tồn tại 2 khiếm khuyết mang tính hệ thống khi áp dụng vào môi trường y tế lâm sàng: Thứ nhất là giả định về mức độ tin cậy ngang nhau giữa các mô hình. Phương pháp bỏ phiếu mềm giả định rằng mọi mô hình cơ sở đều đóng góp vai trò ngang nhau, điều này là một thiếu sót lớn về mặt toán học. Trên thực tế, một mạng DenseNet có thể cực kỳ nhạy bén với bệnh lý A, trong khi Swin-T lại chẩn đoán chính xác hơn với bệnh lý B. Việc giả định về xác suất ngang nhau sẽ làm giảm mức độ đóng góp của mô hình có khả năng dự đoán tốt hơn và khuếch đại sai lầm của mô hình dự đoán kém, dẫn đến giảm hiệu suất tổng thể. Bỏ qua hiện tượng mất cân bằng lớp. Một số phương pháp khác gán trọng số

(weighted voting) thường sử dụng thuật toán tìm kiếm ngẫu nhiên (Random Search) hoặc Heuristic (dựa trên kinh nghiệm) để tối ưu hóa độ chính xác tổng thể (Global Accuracy). Như đã phân tích, trong các bộ dữ liệu y khoa thường xảy ra mất cân bằng (giữa ca bị bệnh và ca bình thường), việc tối ưu hóa Accuracy sẽ đẩy hệ thống rơi vào "Nghịch lý độ chính xác" - nơi mô hình có xu hướng thiên vị khi dự đoán cho các ca bình thường (lớp đa số) và bỏ sót nghiêm trọng các ca bệnh lý thực sự (lớp thiểu số).

3 Phương pháp đề xuất

3.1 Tổng quan về phương pháp

Trong nghiên cứu này, chúng tôi đề xuất phương pháp kết hợp các mô hình sử dụng bộ trọng số tối ưu thông qua phương pháp tìm kiếm lưới được thực hiện qua ba giai đoạn, được tóm tắt như sau (Xem Hình 1):



Hình 1. Sơ đồ phương pháp đề xuất

Giai đoạn 1: Tập dữ liệu huấn luyện (tập Train) và tập xác thực (tập Val) được sử dụng để thực hiện huấn luyện tinh chỉnh (fine-tuning) nhằm tạo ra các mô hình cơ sở tối ưu. Mô hình có độ chính xác cao nhất trên tập xác thực trong quá trình fine-tuning được lưu lại để đánh giá và thực hiện việc tìm kiếm lưới các bộ trọng số tối ưu.

Giai đoạn 2: Sử dụng ba mô hình tối ưu đã tìm ra ở giai đoạn 1, áp dụng các mô hình này vào tập dữ liệu xác thực để trích xuất xác suất dự đoán của tất cả các mẫu trong tập xác thực. Từ các xác suất dự đoán của ba mô hình và nhãn của tập xác thực, sử dụng thuật toán tìm kiếm lưới để tìm bộ trọng số kết hợp tối ưu của ba mô hình.

Giai đoạn 3: Sử dụng ba mô hình tối ưu đã tìm, tính xác suất dự đoán trên tập dữ liệu đánh giá độc lập (tập Test) của ba mô hình cơ sở, áp

dụng bộ trọng số tối ưu đã tìm ra ở giai đoạn 2, tính xác suất dự đoán cuối cùng của mô hình kết hợp trên tập Test, để xác định nhãn dự đoán cho từng mẫu.

3.2 Thuật toán tìm kiếm lưới trọng số tối ưu và suy luận tối ưu

Giả sử tập xác thực (tập Val) có N mẫu và bài toán phân loại gồm C lớp bệnh lý. Sau khi thực hiện quá trình suy luận bằng ba mô hình cơ sở độc lập trên tập Val, ta thu được các bộ tham số đầu vào sau:

$P_1, P_2, P_3 \in \mathbb{R}^{N \times C}$: Là 3 ma trận xác suất đầu ra (đã qua hàm Softmax) của ba mô hình ResNet-50, DenseNet-201 và Swin-T. Mỗi hàng trong ma trận biểu diễn phân phối xác suất dự đoán cho một mẫu.

Thuật toán 1: Tìm bộ trọng số tối ưu (Grid Search)

Input:

$Y_{Val} \in \mathbb{R}^N$: Là vector chứa N nhãn thực tế (Ground Truth) của tập Val.

$P_1, P_2, P_3 \in \mathbb{R}^{N \times C}$ Là ba ma trận xác suất dự đoán từ ba mô hình

Δw : Là bước nhảy tìm kiếm trong không gian lưới

F : Hàm mục tiêu để tìm giá trị tối ưu (F1-Score)

Output:

$w^* = [w_1, w_2, w_3]$ là véc tơ trọng số tối ưu

S_{max} : Giá trị max hàm mục tiêu trên tập Val

1. Initialization:

2. $S_{max} \leftarrow 0$

3. $w^* = [1/3, 1/3, 1/3]$

4. Grid Traversal

5. for $w_1 = 0$ to 1 step Δw do

6. for $w_2 = 0$ to $(1 - w_1)$ step Δw do

7. $w_3 \leftarrow 1 - (w_1 + w_2)$

8. if $w_3 < 0$ then

9. continue

10. end if

11. $P_{ens} = w_1 \times P_1 + w_2 \times P_2 + w_3 \times P_3$

12. $\hat{Y}_{val} = \text{argmax}(P_{ens})$

13. $S_{current} = F(Y_{Val}, \hat{Y}_{val})$

14. if ($S_{current} > S_{max}$)

```
15.            $S_{\max} \leftarrow S_{\text{current}}$ 
16.            $w^* \leftarrow [w_1, w_2, w_3]$ 
17.       end if
18.   end for
19. end for
20. Return  $w^*, S_{\max}$ 
```

Thuật toán 2: Suy luận áp dụng trọng số tối ưu trên tập Test

Input:

$D_{\text{test}} = \{x_1, x_2, \dots, x_{N_{\text{test}}}\}$: Tập test có N_{test} mẫu

M_1, M_2, M_3 : các mô hình tốt nhất trên tập xác thực thu được sau khi thực hiện Fine-tuning

$w^* = [w_1, w_2, w_3]$ là bộ trọng số tối ưu tìm được từ thuật toán 1

C: Số lượng lớp bệnh lý mục tiêu

Output:

Ŷ: Nhãn dự đoán của tập mẫu Test

```
1. Khởi tạo 3 ma trận xác suất dự đoán  $P_1, P_2, P_3 \in \mathbb{R}^{N_{\text{test}} \times C}$ 
2. for  $i = 1$  to  $N_{\text{test}}$  do
3.    $P_1[i] \leftarrow \text{Softmax}(M_1(x_i))$ 
4.    $P_2[i] \leftarrow \text{Softmax}(M_2(x_i))$ 
5.    $P_3[i] \leftarrow \text{Softmax}(M_3(x_i))$ 
6. end for
7. Khởi tạo mảng Ŷ kích thước  $N_{\text{test}}$ 
8. for  $i = 1$  to  $N_{\text{test}}$  do
9.    $P_{\text{ens}} \leftarrow w_1 \times P_1[i] + w_2 \times P_2[i] + w_3 \times P_3[i]$ 
10.   $\hat{Y}[i] \leftarrow \text{argmax}(P_{\text{ens}})$ 
11. end for
12. return Ŷ
```

3.3 Phân tích độ phức tạp thuật toán

Độ phức tạp thuật toán tìm kiếm lưới: Giả sử chúng ta có các hằng số sau:

- N: Số lượng mẫu trong tập Val.
- C: Số lượng lớp bệnh lý (lớp).
- Δw : Bước nhảy của lưới (0.05).
- K: Số lượng mốc phân chia trên trục trọng số. $K = 1/\Delta w$

Vòng lặp ngoài duyệt w_1 chạy $(K + 1)$ lần.

Vòng lặp trong duyệt w_2 chạy trung bình $(K + 1)/2$ lần.

Tổng số kích bản trọng số (tổ hợp) cần quét là: $M = (K+1) \times (K+2)/2$.

Bên trong mỗi lần lặp, hệ thống thực hiện phép cộng ma trận kích thước $N \times C$, tìm argmax và tính F1-Score. Các phép toán này tốn $O(N \times C)$.

Độ phức tạp thuật toán tìm kiếm lưới: $O(K^2 \times N \times C)$, vì K và C là hằng số. Độ phức tạp là $O(N)$.

Phân tích độ phức tạp thuật toán suy luận tối ưu. Gọi

- N_{test} : Tổng số lượng mẫu đầu vào trong tập kiểm thử.
- M : Tổng số mô hình cơ sở trong hệ thống ($M = 3$).
- F_m : Độ phức tạp của mô hình cơ sở thứ m ($m \in \{1, 2, 3\}$).

Chi phí Tính toán Xác suất Dự đoán (Base Model Inference): Vì thuật toán chạy tuần tự qua M mô hình để tính toán ra M vector xác suất phân phối dự đoán cho mỗi bức ảnh, chi phí tiệm cận cho N_{test} mẫu là $O(N_{test} \times \max_{i=1}^m F_m)$.

Chi phí kết hợp dự đoán: Chi phí thực hiện vòng lặp N_{test} lần áp dụng bộ trọng số tối ưu lên các vector xác suất, trong mỗi lần lặp thực $M \times C$ phép nhân, $(M-1) \times C$ phép cộng giá trị và $(C-1)$ phép so sánh argmax . Tổng độ phức tạp chi phí kết hợp dự đoán là $O(N_{test} \times M \times C)$. M, C là các hằng số, vì vậy chi phí kết hợp dự đoán lớp là $O(N_{test})$.

Một cách tổng quát: Độ phức tạp thuật toán suy luận là $O(N \times \max_{i=1}^m F_m)$.

4 Thực nghiệm và đánh giá

4.1 Tập dữ liệu thực nghiệm

Trong nghiên cứu này, chúng tôi tiến hành thực nghiệm trên 2 tập dữ liệu ảnh X-quang ngực. Tập dữ liệu ChestXray-ImageDataset [18], gồm bốn lớp phân loại bệnh lý phổi, được lựa chọn để đánh giá phương pháp đề xuất. Bộ dữ liệu hình ảnh X-quang ngực ChestXray-ImageDataset là một bộ dữ liệu công khai được sử dụng để chẩn đoán và nghiên cứu các bệnh về phổi, bao gồm cả COVID-19. Bộ dữ liệu này chứa các hình ảnh được thu thập từ nhiều bệnh viện khác nhau ở Bangladesh, nơi các chuyên gia chăm sóc sức khỏe đã theo dõi và thu thập cẩn thận các hình ảnh X-quang cho mục đích nghiên cứu và chẩn đoán. Bộ dữ liệu cung cấp tổng cộng 4.350 hình ảnh chất lượng cao, được phân loại thành bốn lớp riêng biệt. Đây là tập dữ liệu khá cân bằng, phân bố các lớp khá đồng đều, phân bố chi tiết được mô tả ở Bảng 1:

Bảng 1. Phân bố tập dữ liệu ChestXray-ImageDataset

Lớp	Số mẫu	Tỷ lệ (%)
COVID-19	1.100	25,3
LUNG OPACITY	1.200	27,6
NORMAL	1.000	23,0
VIRAL PNEUMONIA	1.050	24,1
Tổng	4.350	100

Trong tập dữ liệu ChestXray-ImageDataset, lớp COVID-19 (phân bố gồm 1100 hình ảnh): Nhóm này chứa các hình ảnh X-quang của bệnh nhân được xác nhận hoặc nghi ngờ mắc COVID-19, hỗ trợ phát hiện sớm và theo dõi bệnh. Lớp bệnh lý mờ phổi (Phân bố nhóm: 1200 hình ảnh): Những hình ảnh này thể hiện các trường hợp mờ phổi hoặc bất thường cần được kiểm tra và chẩn đoán thêm. Lớp bình thường (Phân bố nhóm:

1000 hình ảnh): Nhóm này bao gồm các hình ảnh X-quang của phổi khỏe mạnh, được sử dụng làm tài liệu tham khảo để so sánh với phổi bị bệnh hoặc nhiễm trùng. Lớp bệnh lý viêm phổi do virus (Phân bố nhóm: 1050 hình ảnh): Các hình ảnh X-quang trong nhóm này liên quan đến các trường hợp viêm phổi do virus, giúp xác định và hiểu rõ hơn về loại nhiễm trùng phổi đặc biệt này.

Tập dữ liệu COVID-19 [19], đây là tập dữ liệu mất cân bằng, được tổng hợp và chọn lọc gồm các hình ảnh X-quang ngực COVID-19 thu được bằng cách đối chiếu 15 tập dữ liệu công khai. Tập dữ liệu chứa 1281 hình ảnh X-quang

bệnh lý COVID-19, 3270 hình ảnh X-quang bình thường, 1656 hình ảnh X-quang viêm phổi do virus và 3001 hình ảnh X-quang viêm phổi do vi khuẩn. Đây là tập dữ liệu mất cân bằng. Phân bố tập dữ liệu được trình bày ở Bảng 2.

Bảng 2. Phân bố tập dữ liệu COVID-19

Lớp	Số mẫu	Tỷ lệ (%)
COVID-19	1.281	13,9
Normal	3.271	35,5
Bacterial Pneumonia	3.001	32,6
Viral Pneumonia	1.656	18,0
Tổng	9.208	100

Chúng tôi thực hiện phân chia ngẫu nhiên, các tập dữ liệu nghiên cứu thành tập huấn luyện, tập xác thực và tập kiểm thử theo tỷ lệ 80:10:10. Các tập huấn luyện, tập xác thực và tập kiểm thử được lưu trữ trong 3 thư mục vật lý độc lập để đảm bảo tính nhất quán khi so sánh đánh giá các mô hình khác nhau và giữa các lần chạy thuật toán.

Quá trình chia tập dữ liệu hình ảnh gốc thành ba tập con riêng biệt: tập huấn luyện (Train), tập xác thực (Val) và tập kiểm thử (Test) được thực hiện qua hai bước chính bằng thư viện NumPy như sau:

Xáo trộn ngẫu nhiên dữ liệu (Data Shuffling). Trước tiên, toàn bộ danh sách các đường dẫn file ảnh được xào trộn ngẫu nhiên bằng hàm `np.random.seed(42)` và `np.random.shuffle()`. Bước này là bắt buộc để triệt tiêu tính thứ tự trong tập dữ liệu gốc. Việc xào trộn giúp đảm bảo phân phối dữ liệu đồng đều ở

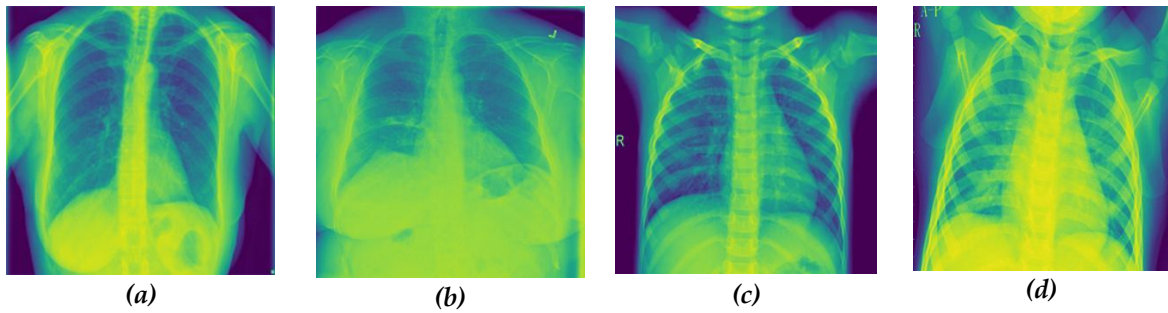
cả ba tập, từ đó giúp mô hình học được các đặc trưng tổng quát và tránh tình trạng học vẹt (overfitting).

Phân chia tập dữ liệu (Data Splitting). Sau khi dữ liệu đã được xào trộn, hàm `np.split` được sử dụng để cắt danh sách này thành 3 phần theo tỷ lệ được thiết lập (80% Train, 10% Val, 10% Test). Chúng ta sẽ tính toán các chỉ số (index) làm điểm cắt dựa trên tổng số lượng mẫu. Hàm `np.split` sẽ nhận mảng dữ liệu và các điểm cắt này để tách dữ liệu thành 3 mảng con hoàn toàn độc lập. Các mẫu dữ liệu được đưa vào ba tập huấn luyện, xác thực và kiểm thử, được lưu trữ trên 3 thư mục vật lý độc lập phục vụ cho quá trình nghiên cứu.

Chi tiết số lượng mẫu trong các tập huấn luyện, tập xác thực và tập kiểm tra của hai bộ dữ liệu nghiên cứu được mô tả trong Bảng 3 và Bảng 4. Hình ảnh minh họa của tập dữ liệu ChestXray-ImageDataset thể hiện ở Hình 2.

Bảng 3. Phân chia tập ChestXray-ImageDataset thành tập huấn luyện, tập xác thực và tập kiểm thử

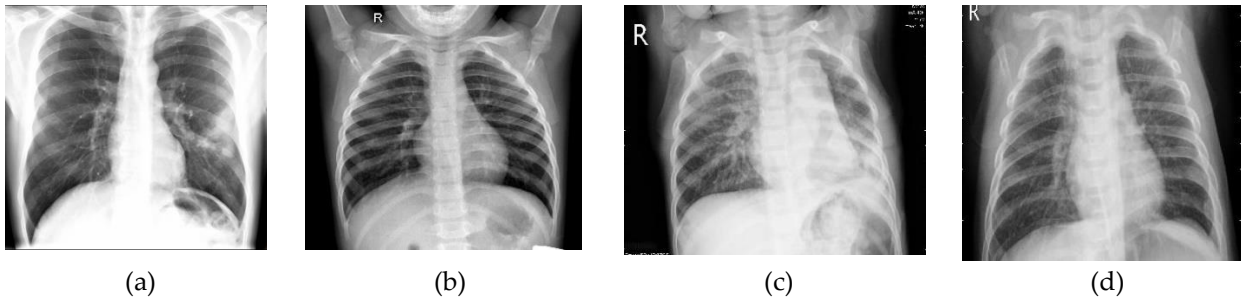
Lớp	Số mẫu	Tập Train	Tập Val	Tập Test
COVID-19	1.100	880	110	110
Lung-Opacity	1.200	960	120	120
Normal	1.000	800	100	100
Viral Pneumonia	1.050	840	105	105
Tổng	4.350			



Hình 2. Hình ảnh từ tập ChestXray-ImageDataset: (a) COVID-19, (b) Lung-Opacity, (c) Normal, (d) Viral Pneumonia

Bảng 4. Phân chia tập COVID-19 thành tập huấn luyện, tập xác thực và tập kiểm thử

Lớp	Số mẫu	Tập Train	Tập Val	Tập Test
COVID-19	1.281	1.024	128	129
Normal	3.271	2.616	327	328
Bacterial Pneumonia	3.001	2.400	300	301
Viral Pneumonia	1.656	1.324	166	166
Tổng	9.208			



Hình 3. Hình ảnh từ tập COVID-19 (a) COVID-19, (b) Normal, (c) Bacterial Pneumonia, (d) Viral Pneumonia

4.2 Phương pháp tăng cường dữ liệu

Để gia tăng khả năng tổng quát hóa của mô hình và hạn chế hiện tượng quá khớp (overfitting), trong quá trình huấn luyện các mô hình, chúng tôi triển khai chiến lược tăng cường dữ liệu động (Online Data Augmentation) thông qua thư viện torchvision.transforms của nền tảng PyTorch. Thay vì thực hiện tăng cường dữ liệu theo phương pháp tạo và lưu trữ các biến thể của các mẫu dữ liệu trên bộ nhớ ngoài, các kỹ thuật biến đổi hình học và cường độ sáng: bao gồm cắt ngẫu nhiên (Random Resized Crop), lật ngang (Horizontal Flip) và điều chỉnh màu sắc (Color Jitter) được thực thi trực tiếp trong quá trình nạp

batch dữ liệu. Phương pháp tiếp cận này đảm bảo rằng tại mỗi kỳ nguyên (epoch) trong quá trình huấn luyện, mô hình được tiếp cận với các biến thể dữ liệu khác biệt về mặt cấu trúc không gian nhưng vẫn bảo toàn đặc trưng ngữ nghĩa. Phương pháp tiếp cận này không chỉ giúp giảm thiểu đáng kể chi phí lưu trữ vật lý mà còn tối ưu hóa hiệu suất tính toán thông qua cơ chế nạp dữ liệu đa luồng, giúp duy trì sự cân bằng giữa tốc độ xử lý của CPU và khả năng huấn luyện của GPU.

4.3 Tham số huấn luyện mô hình.

Các mô hình cơ sở ResNet-50, DenseNet-201, Swin-T được sử dụng từ thư viện

Torchvision, sử dụng bộ trọng số ImageNet-1K. Lớp kết nối đầy đủ (Fully Connected - FC) của các mô hình cơ sở được sửa đổi số lớp đầu ra tương ứng với số nhãn bệnh lý của tập dữ liệu y tế lựa chọn nghiên cứu.

Bảng 5. Tham số huấn luyện tinh chỉnh

Số epoch huấn luyện	100
Batch size	64
Learning rate	1e-04
Trình tối ưu	AdamW, weight_decay=1e-2
Môi trường thực nghiệm	Ubuntu 22.04, RTX 4090, PyTorch
Kích thước ảnh đầu vào	224×224
Hàm mất mát	Cross-Entropy Loss
Giá trị bước lưới Δw	0.05

Hiệu suất các mô hình được giám sát liên tục trên tập xác thực. Mô hình có độ chính xác cao nhất trên tập xác thực, được ghi lại trong suốt quá trình huấn luyện. Để kiểm soát sự hội tụ trong quá trình tinh chỉnh, chúng tôi áp dụng chiến lược

giữ ổn định tốc độ học, giúp quá trình tinh chỉnh trọng số mạng ổn định trong suốt huấn luyện.

4.4 Các độ đo đánh giá hiệu năng

Trong phân tích hình ảnh y tế, sự phân bố dữ liệu giữa các lớp bệnh lý và lớp bình thường (normal class) thường xuyên xảy ra tình trạng mất cân bằng. Vì vậy, trong nghiên cứu này để đảm bảo sự so sánh công bằng và đánh giá chính xác năng lực chẩn đoán lâm sàng của ba mô hình cơ sở cũng như mô hình kết hợp, chúng tôi áp dụng một hệ thống các độ đo toàn diện. Đặc biệt, để triệt tiêu hoàn toàn vấn đề thiên vị dữ liệu (data bias), ngoại trừ Accuracy được tính trên toàn cục, tất cả các độ đo đánh giá ở cấp độ lớp (class-level metrics) đều được tính toán tuân thủ nghiêm ngặt theo chiến lược trung bình vĩ mô không có trọng số (Unweighted Macro-averaging).

Giả sử bài toán có C lớp, ma trận nhầm lẫn (Confusion Matrix) cho lớp thứ i cung cấp các giá trị: Dương tính thực sự (TP_i), âm tính thực sự (TN_i), dương tính giả (FP_i), và âm tính giả (FN_i). Các độ đo được định nghĩa cụ thể như sau:

$$\text{Accuracy (ACC)} = \frac{\sum_{i=1}^k TP_i}{N} \tag{1}$$

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \tag{2}$$

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \tag{3}$$

$$\text{Specificity}_i = \frac{TN_i}{TN_i + FP_i} \tag{4}$$

$$\text{F1-Score}_i (F1_i) = 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \tag{5}$$

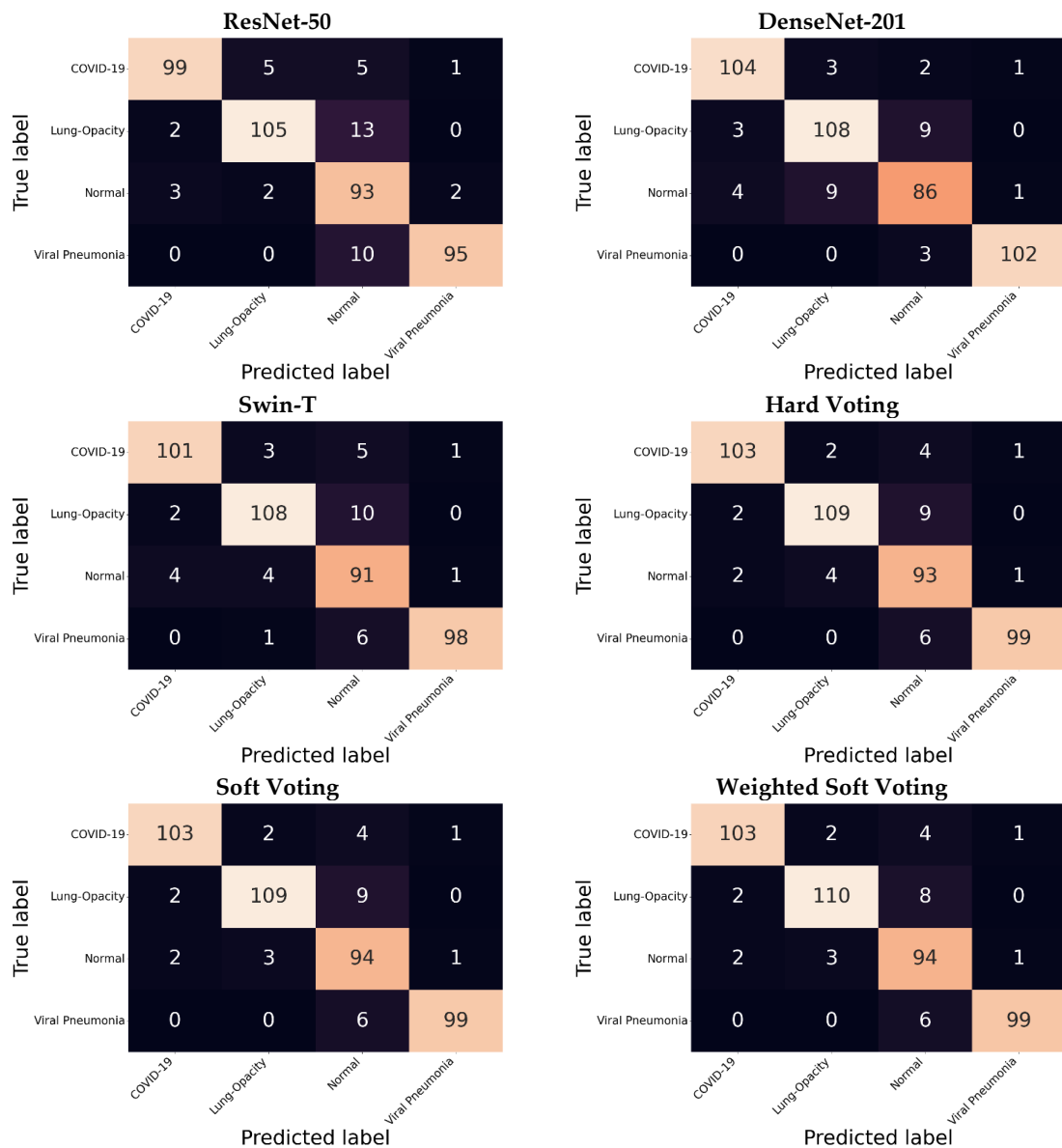
$$\text{Macro-M} = \frac{1}{C} \sum_{i=1}^C M_i \tag{6}$$

Các độ đo Precision, Recall, Specificity, F1-Score của các mô hình cơ sở và mô hình kết hợp được tính bằng công thức trung bình không có trọng số của độ đo các lớp, theo công thức (6).

4.5 Kết quả thực nghiệm

Ma trận nhầm lẫn trên tập kiểm thử bộ dữ liệu ChestXray-ImageDataset thể hiện ở Hình 4. Bảng 6 thể hiện hiệu suất phân lớp của các

phương pháp trên tập dữ liệu nghiên cứu ChestXray-ImageDataset. Độ chính xác của phương pháp đề xuất có sự gia tăng, trong khoảng từ 0,23% so với mô hình tốt nhất là Soft Voting, và 3,22% so với mô hình kém nhất là ResNet-50. Các mô hình kết hợp có thời gian suy luận xấp xỉ tổng thời gian suy luận của các mô hình cơ sở, chi tiết thể hiện ở Bảng 7.



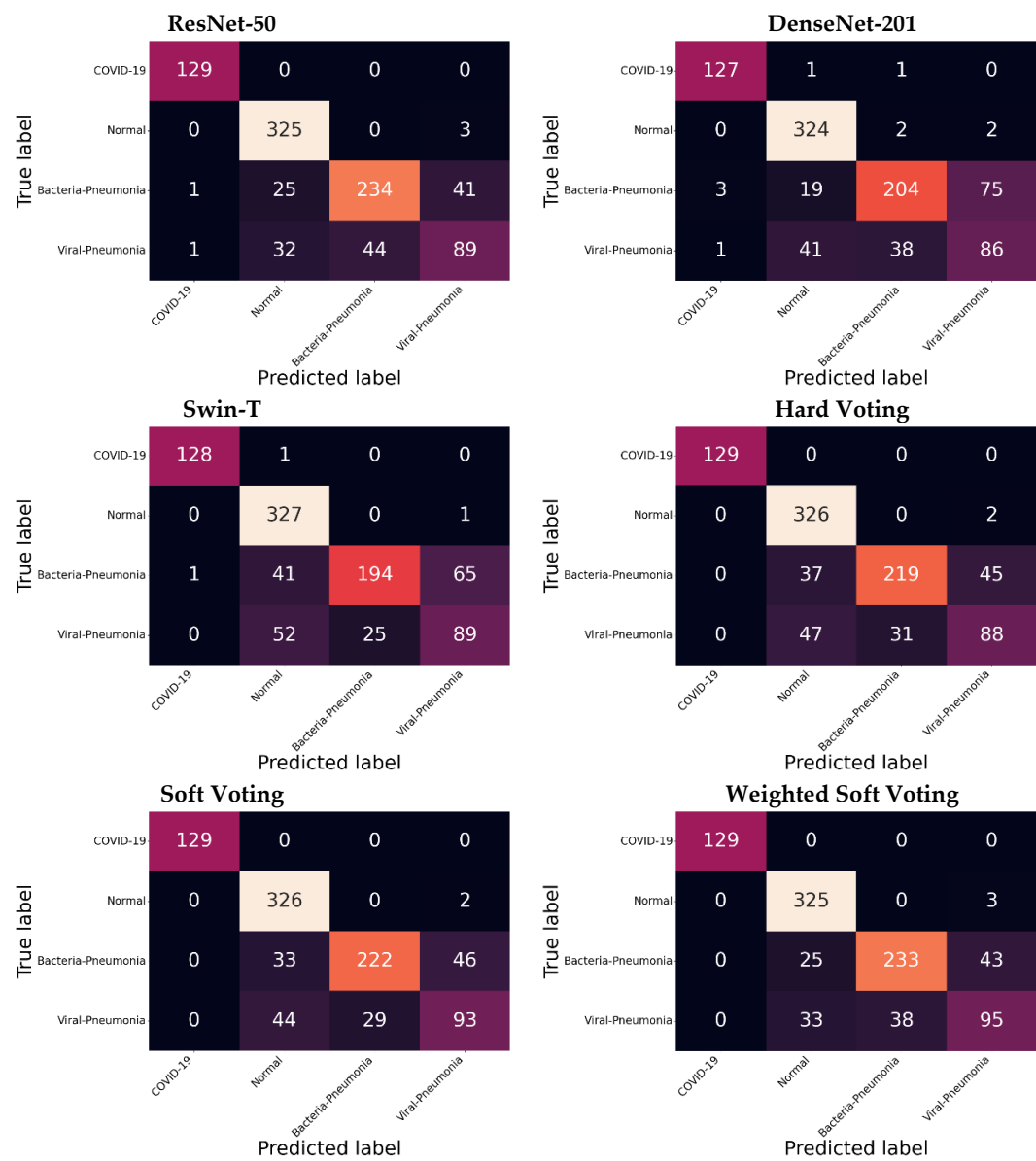
Hình 4. Ma trận nhầm lẫn trên tập Test của bộ ChestXray-ImageDataset

Bảng 6. Kết quả phân lớp các mô hình trên tập ChestXray-ImageDataSet

Model	Accuracy (%)	Precision (%)	Recall/ Sensitivity (%)	Specificity (%)	F1-Score (%)
ResNet-50	90,11	90,69	90,24	96,74	90,20
DenseNet-201	91,95	91,94	91,92	97,31	91,93
Swin-T	91,49	91,69	91,54	97,18	91,52
Hard Voting (3 Model)	92,87	93,02	92,94	97,65	92,89
Soft Voting (3 model)	93,10	93,27	93,19	97,73	93,12
Weighted Soft Voting (Phương pháp đề xuất)	93,33	93,47	93,40	97,80	93,34

Bảng 7. So sánh hiệu năng các mô hình trên tập ChestXray-ImageDataset

Model	Thời gian suy luận (s)	Thông lượng (FPS, img/s)	Độ trễ (ms/img)
ResNet-50	0,3354	1296,88	0,77
DenseNet-201	0,3739	1163,57	0,86
Swin-T	0,4453	976,84	1,02
HardVoting	1,1553	376,54	2,66
SoftVoting	1,1546	376,75	2,65
Weighted Soft Voting (Phương pháp đề xuất)	1,1546	376,75	2,65



Hình 5. Ma trận nhầm lẫn trên tập kiểm thử bộ dữ liệu COVID-19

Bảng 8. Kết quả phân lớp trên tập test dữ liệu COVID-19

Model	Accuracy (%)	Precision (%)	Recall/ Sensitivity (%)	Specificity (%)	F1-Score (%)
ResNet-50	84,09	83,66	82,61	94,33	82,79
DenseNet-201	80,19	79,28	79,20	93,13	78,90
Swin-T	79,87	80,73	79,25	92,85	79,15
Hard Voting	82,47	83,07	81,29	93,68	81,58
Soft Voting	83,33	83,82	82,29	94,02	82,55
Weighted Soft Voting (Phương pháp đề xuất)	84,63	84,55	83,43	94,53	83,69

Trên tập dữ liệu mất cân bằng COVID-19, hiệu quả phân lớp của các mô hình có sự khác biệt rõ rệt, phương pháp chúng tôi đề xuất đạt hiệu quả cách biệt, vượt xa so với các mô hình cơ sở và các phương pháp kết hợp truyền thống. Độ chính xác, độ nhạy của phương pháp đề xuất vượt các mô hình cơ sở tốt nhất và kém nhất, dao động trong khoảng từ 0,54 đến 4,76%, chi tiết thể hiện ở Bảng 8. Ma trận nhầm lẫn ở Hình 5 cho thấy phương pháp đề xuất có sự cách biệt hiệu quả phân lớp ở lớp bệnh lý viêm phổi do vi khuẩn (Bacterial-Pneumonia).

5 Thảo luận

Trong các hệ thống chẩn đoán hình ảnh y tế, sự phân bố dữ liệu thường xuyên đối mặt với tình trạng mất cân bằng nghiêm trọng giữa các lớp, trong đó số lượng ca bệnh lý thường ít hơn rất nhiều so với các ca bình thường. Do đó, việc lựa chọn hàm mục tiêu để đánh giá và định tuyến bộ trọng số trong thuật toán tìm kiếm lưới (Grid Search) mang tính quyết định đến hiệu quả phân lớp tổng thể của mô hình.

Về mặt toán học, giá trị F1-Score là trung bình điều hòa giữa độ chính xác dự đoán (Precision) và độ nhạy (Recall). Đặc tính của trung bình điều hòa là nó sẽ trừng phạt rất nặng mô hình nếu một trong hai giá trị trên tụt xuống mức thấp. Nhờ đó, thuật toán Grid Search hướng tới việc tìm bộ trọng số w^* tạo ra sự cân bằng tối ưu giữa việc không bỏ sót bệnh nhân và không chẩn đoán sai người khỏe mạnh. Hơn nữa, việc áp dụng chiến lược tính toán trung bình vĩ mô

(Macro-averaging) đảm bảo rằng hiệu suất chẩn đoán trên các lớp thiểu số (minority classes) được đánh giá ngang bằng với các lớp đa số, mang lại một thước đo công bằng, khách khe và sát với yêu cầu thực tiễn của chẩn đoán. Nếu sử dụng độ chính xác tổng thể làm hàm mục tiêu, thuật toán tối ưu hóa sẽ dễ dàng ưu tiên lớp đa số để nâng cao độ chính xác. Cụ thể, hệ thống có xu hướng dồn trọng số để dự đoán thiên vị cho lớp đa số (ca bình thường) nhằm tối đa hóa hàm mục tiêu. Ngược lại, nếu chỉ tập trung tối đa hóa độ nhạy (Recall/Sensitivity) nhằm không bỏ sót ca bệnh nào, thuật toán sẽ sinh ra một mô hình dán nhãn bệnh lý cho cả những người khỏe mạnh.

Trên 2 tập dữ liệu chúng tôi thực hiện nghiên cứu, lớp bệnh lý COVID-19 là lớp chiếm số mẫu ít nhất. Kết quả dự đoán của phương pháp đề xuất luôn đạt hiệu suất phân lớp ngang với mô hình cơ sở tốt nhất, chi tiết được thể hiện ở ma trận nhầm lẫn hình 4, hình 5.

6 Kết luận và hướng phát triển

Trong nghiên cứu này, chúng tôi đã đề xuất phương pháp kết hợp Weighted Soft Voting được cải tiến dựa trên phương pháp Soft Voting, sử dụng cách tiếp cận tìm kiếm lưới để tìm ra bộ trọng số kết hợp tối ưu. Kết quả thực nghiệm trên hai tập dữ liệu công khai ảnh X-quang ngực để phân lớp các bệnh lý về phổi đã đạt được hiệu suất vượt trội so với các mô hình cơ sở và các phương pháp kết hợp truyền thống. Mặc dù phương pháp đề xuất (Grid-Searched Weighted Soft Voting) sử dụng đồng thời 3 kiến trúc học

sâu kích thước lớn (ResNet-50, DenseNet-201, Swin-T), thời gian suy luận vẫn ở mức phù hợp với yêu cầu của các ứng dụng thực tế. Trên hệ thống thử nghiệm phần cứng NVIDIA RTX 4090, tổng thời gian suy luận tuần tự trung bình chỉ mất 2,65 mili giây cho mỗi bức ảnh X-quang, tương đương thông lượng 376,75 FPS.

Trong phần thực nghiệm, bước lưới được lựa chọn là 0,05. Kết quả thực nghiệm trên 2 tập dữ liệu cho thấy phương pháp đề xuất cải thiện hiệu quả phân lớp so với các mô hình và các phương pháp kết hợp truyền thống. Tuy nhiên bộ trọng số này cố định cho toàn bộ các mẫu dữ liệu trong tập kiểm thử, đây là một hạn chế của phương pháp đề xuất. Phương pháp tiếp cận theo trọng số thích ứng phụ thuộc đặc trưng mẫu dữ liệu đầu vào hoặc các phương pháp hỗn hợp chuyên gia (MoE) cần phải được đánh giá trong những nghiên cứu sắp tới. Việc lựa chọn, khảo sát giá trị bước nhảy trong không gian tìm kiếm ảnh hưởng đến kết quả mô hình, vì vậy cần các phương pháp tối ưu hóa để lựa chọn bước lưới. Đây là những hướng phát triển trong những nghiên cứu tiếp theo nhằm tối ưu bước nhảy, thiết kế bộ trọng số thích ứng. Ngoài ra, kết quả phương pháp đề xuất chưa được kiểm chứng trên những tập dữ liệu ngoài khác, cũng như thực hiện xác thực chéo và so sánh toàn diện với nhiều phương pháp kết hợp hiện đại khác gần đây. Những hạn chế trên mở ra những hướng nghiên cứu sắp tới để đánh giá một cách trọn vẹn hơn về những cải tiến được đề xuất.

Tài liệu tham khảo

1. Yang Y, Lv H, Chen N. A survey on ensemble learning under the era of deep learning. *Artificial Intelligence Review*. 2023;56(6):5545-89.
2. Rincy TN, Gupta R, editors. *Ensemble Learning Techniques and its Efficiency in Machine Learning: A Survey*. 2nd International Conference on Data, Engineering and Applications (IDEA); 2020 28-29 Feb. 2020.
3. Tulyakov S, Jaeger S, Govindaraju V, Doermann D. Review of Classifier Combination Methods. In: Marinai S, Fujisawa H, editors. *Machine Learning in Document Analysis and Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 361-86.
4. Fumera G, Roli F. A theoretical and experimental analysis of linear combiners for multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2005;27(6):942-56.
5. Kittler J, Hatef M, Duin RP, Matas J. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2002;20(3):226-39.
6. LeCun Y, Bengio Y, Hinton G. Deep learning. *nature communications*. 2015;521(7553):436-44.
7. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*; Lake Tahoe, Nevada: Curran Associates Inc.; 2012. p. 1097-105.
8. Cai L, Gao J, Zhao D. A review of the application of deep learning in medical image classification and segmentation. *Annals of translational medicine*. 2020;8(11):713.
9. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. 2020.
10. Dietterich TG, editor *Ensemble methods in machine learning*. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; 2000. Berlin: Springer Verlag.
11. Torrey L, Shavlik J. Transfer learning. *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*: IGI Global; 2009. p. 242-64.
12. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society. 2016:770-778.
13. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ, editors. *Densely connected convolutional networks*. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*; 2017: Institute of Electrical and Electronics Engineers Inc.

14. Minaee S, Kafieh R, Sonka M, Yazdani S, Jamalipour Soufi G. Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning. *Medical Image Analysis*. 2020;65:101794.
15. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, et al., editors. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *Proceedings of the IEEE International Conference on Computer Vision*; 2021: Institute of Electrical and Electronics Engineers Inc.
16. He K, Gan C, Li Z, Rekik I, Yin Z, Ji W, et al. Transformers in medical image analysis. *Intelligent Medicine*. 2023;3(1):59-78.
17. Pham T-A, Hoang V-D. Chest X-ray image classification using transfer learning and hyperparameter customization for lung disease diagnosis. *Journal of Information and Telecommunication*. 2024:1-15.
18. Talukder MA. Chest X-Ray Image Dataset. Mendeley Data, V1. Mendeley Data; 2023.
19. Sait U, Lal K, Prajapati S, Bhaumik R, Kumar T, Sanjana S, et al. Curated dataset for COVID-19 posterior-anterior chest radiography images (X-Rays). Mendeley Data; 2020.