

CHẨN ĐOÁN COVID-19 QUA X-QUANG LỒNG NGỰC BẰNG MÔ HÌNH HỌC SÂU

**Võ Trí Nhân^{1*}, Nguyễn Thành Đạt¹, Nguyễn Phương Nam¹,
PGS. TS. Phạm Thị Thu Hiền¹, TS. Lê Thanh Hải²**

¹Trường Đại học Quốc tế - Đại học Quốc gia Thành Phố Hồ Chí Minh, Việt Nam

²Trường Đại học Bách khoa - Đại học Quốc gia Thành Phố Hồ Chí Minh, Việt Nam

*Tác giả liên hệ: votrinhan88@gmail.com

TÓM TẮT

Do đại dịch Coronavirus 2019 (COVID-19) đã để lại những hậu quả tàn khốc trên toàn thế giới, một quy trình chẩn đoán sàng lọc hiệu quả đóng vai trò rất quan trọng trong việc hỗ trợ cho ngành y tế cộng đồng và ngăn chặn dịch bệnh. Nghiên cứu này giới thiệu một quy trình tinh chỉnh cho việc huấn luyện các mô hình dự đoán học sâu để phát hiện COVID-19 từ ảnh chụp X-quang lồng ngực, cũng như chính các kết quả mô hình thu được. Quy trình này bao gồm nhiều kỹ thuật để tránh tình trạng mô hình quá khớp và tối ưu hóa các kết quả dự đoán, chẳng hạn như tăng dữ liệu, tối ưu hóa Bayes để điều chỉnh siêu thông số, chọn trị số hiệu suất phù hợp và dừng sớm trong quá trình huấn luyện mô hình. Trên tập dữ liệu COVID-XRay-5K v3, ba kiểu mô hình ResNet50, NASNet-A-Mobile và Xception đã đạt được chỉ số AU-PRC là 0,9773, 0,9633 và 0,9003; và AU-ROC lần lượt là 0,9940, 0,9964 và 0,9812. Ở độ nhạy 98%, chúng duy trì độ đặc hiệu cao là 97,53%, 97,60% và 86,00%. Với hiệu suất như vậy, các mô hình học sâu này là những công cụ đầy hứa hẹn để hỗ trợ trong cuộc chiến chống lại đại dịch.

Từ khóa: COVID-19, X-quang lồng ngực, Học sâu, Thử nghiệm sàng lọc, Phân loại nhị phân

DETECTION OF COVID-19 FROM CHEST X-RAY USING DEEP LEARNING

**Vo Tri Nhan^{1*}, Nguyen Thanh Dat¹, Nguyen Phuong Nam¹,
Assoc. Dr. Pham Thi Thu Hien¹, Dr. Le Thanh Hai¹**

¹International University - Vietnam National University, Ho Chi Minh City, Vietnam

²University of Science and Technology - Vietnam National University, Ho Chi Minh City, Vietnam

* Corresponding Author: votrinhan88@gmail.com

ABSTRACT

As the Coronavirus Disease 2019 (COVID-19) have left devastating consequences over the world, an effective screening test procedure is crucial to strengthen public health and put the disease to a halt. This study introduces a refined pipeline to train deep learning predictive models for the detection of COVID-19 from chest X-rays, as well as the resulting models themselves. The pipeline involves multiple techniques to combat overfitting and optimize predictive results, such as data augmentation, Bayesian optimization for hyperparameter tuning, selecting the appropriate performance metric, and early stopping during model training. On the COVID-XRay-5K v3 dataset, the three models, ResNet50, NASNet-A-Mobile, and Xception, achieved the areas under the precision-recall curves of 0.9773, 0.9633, and 0.9003; and the areas under the receiver operating characteristic curves of 0.9940, 0.9964, and 0.9812, respectively. At 98% recall (sensitivity), they sustained high specificity of 97.53%, 97.60%, and 86.00%. With such performance, these deep learning models are promising tools to aid in the combat against the pandemic.

Keywords: COVID-19, Chest X-rays, Deep Learning, Screening Test, Binary Classification

1. Giới thiệu

Đại dịch Corona 2019 (COVID-19) gây ra đã làm suy yếu thể giới của chúng ta trong hai năm qua bằng cách cướp đi sinh mạng của hàng triệu người và để lại những biến chứng vĩnh viễn cho hàng trăm triệu người mắc bệnh. Hàng triệu doanh nghiệp đang trên bờ vực sụp đổ và gần 1,65 tỷ lực lượng lao động toàn cầu đang gặp khó khăn để kiếm sống (Chriscaden, 2020).

Dù sẽ mất khá lâu để không chế dịch hoàn toàn nhưng củng cố y tế dự phòng bằng cách thường xuyên thực hiện các xét nghiệm sàng lọc cộng đồng được chứng minh là một cách hiệu quả để kiểm soát sự tiến triển của dịch bệnh. Hiện tại, một số kỹ thuật chẩn đoán đã được thực hành, phổ biến nhất là phản ứng chuỗi polymerase phiên mã ngược (RT-PCR), PCR thời gian thực và bộ xét nghiệm nhanh, chụp cắt lớp vi tính (CT) và chụp X quang lồng ngực.

Bên cạnh đó, học sâu là công nghệ ngày càng trở nên phổ biến và hữu ích hơn trong cuộc sống hàng ngày của chúng ta. Sự kết hợp giữa học sâu và phân tích X-quang lồng ngực nhằm chẩn đoán COVID-19 đã đạt được những thành tựu mới và cung cấp một giải pháp thay thế khả thi để chẩn đoán bệnh.

Nghiên cứu này cố gắng triển khai các kỹ thuật học sâu để tạo ra các mô hình dự đoán có thể phát hiện COVID-19 từ ảnh X-quang lồng ngực. Điều này thể cắt giảm chi phí chẩn đoán, giảm nhân lực thực hiện chẩn đoán cũng như nguy cơ lây nhiễm và phơi nhiễm trong phạm vi gần, đồng thời vẫn duy trì độ đặc hiệu đáng tin cậy.

Mục tiêu của nghiên cứu này là thiết kế một quy trình hoàn chỉnh để xây dựng các mô hình học sâu nhằm phát hiện COVID-19 từ chụp X-quang ngực sau - trước, khảo sát đầy đủ và tối ưu hóa từng giai đoạn của đường ống, đồng thời cung cấp các mô hình có thể sử dụng được đã qua huấn luyện.

Về phạm vi, nghiên cứu này nhằm thực hiện phân loại nhị phân COVID-19 trên tập dữ liệu COVID-XRay-5K v3 bằng cách sử dụng các kỹ thuật học sâu hiện đại. Ba mô hình được đào tạo trước được thay thế cho nhiệm vụ này với sự trợ giúp của việc học chuyển giao. Điều chỉnh siêu tham số đã được thực hiện với kiểm tra chéo k-lần và tối ưu hóa Bayes cho các điều kiện đào tạo tối ưu cho mỗi mô hình. Chiến lược dừng sớm được áp dụng và cải tiến. Ngoài ra, một phân tích chi tiết về một số chỉ số đánh giá hiệu suất được trình bày để giải thích lựa chọn chính của AU-PRC và cách nó có thể chống lại các vấn đề mất cân bằng tập dữ liệu và ngưỡng quyết định di động. Cuối cùng, sau khi đào tạo, các mô hình được đánh giá trên mẫu đối chứng và so sánh với các công trình khác. Kết quả có thể bị hạn chế bởi các nguồn tài nguyên tính toán phát huy hết khả năng của nó. Một số sửa đổi và chiến lược đã được áp dụng để vượt qua khó khăn này.

2. Cơ sở lý thuyết

ResNet (Residual Network) là mô hình học sâu được giới thiệu vào năm 2015 (He, Zhang, Ren, & Sun, 2016). Vào lúc đó, xây dựng mô hình sâu bằng cách thêm nhiều lớp chập để đạt được kết quả tốt hơn đang là xu hướng chung của thế giới. Tuy nhiên, việc tăng các lớp chập đã xảy ra hiện tượng “Vanishing Gradient”, dẫn đến quá trình học tập của mô hình không tốt. Mạng ResNet ra đời để giải quyết hiện tượng này bằng cách đưa ra giải pháp sử dụng kết nối “tắt” (skip connection) đồng nhất để xuyên qua một hay nhiều lớp. Một khối như vậy được gọi là một Residual Block.

Xception được giới thiệu vào năm 2017 (Chollet, 2017) và dựa trên cấu trúc của mô hình Inception (Szegedy et al., 2015). Điểm đặc biệt của mô hình này là sử dụng kiến trúc tích chập tách biệt theo chiều sâu (Depthwise Seperable Convolution) có tác dụng giảm thiểu số lượng tham số và khối lượng tính toán. Tuy nhiên, khác với người tiền nhiệm, Xception sử dụng lớp tích chập điểm (Pointwise Convolution) trước lớp tích chập chiều sâu (Depthwise Convolution). Kết quả cho thấy mô hình Xception đạt được kết quả tốt hơn Inception với cùng một tập dữ liệu.

NASNet (Neural Architecture Search neural network) được phát triển bởi Google và giới thiệu vào năm 2018 (Zoph, Vasudevan, Shlens, & Le, 2018). Mô hình được xây dựng bằng cách học tăng cường (reinforcement learning) dựa trên các kiến trúc của mô hình khác để hoàn thiện kiến trúc của mình. Kiến trúc mô hình NASNet cho phân loại hình ảnh bao gồm hai thành phần chính là Normal Cell và Reduction Cell. Số lượng của hai khối này có thể thay đổi như là một tham số để điều chỉnh mô hình hoạt động tốt hơn. Nhờ khả năng tự điều chỉnh của các khối, NASNet có nhiều phiên bản khác nhau, trong bài nghiên cứu này, NASNet-A-mobile được chọn vì sự gọn nhẹ của nó.

Ba mô hình trên đều có những ưu điểm khác nhau trong việc phân loại hình ảnh. Mô hình ResNet tập trung vào giảm nhẹ khối lượng tính toán và giải quyết vấn đề “Gradient Vanishing” khi các lớp trong mô hình được tăng lên bằng cách sử dụng các “skip connection” nhưng vẫn đảm bảo độ chính xác trong quá trình phân loại ảnh. Xception tách riêng các dữ liệu đầu vào thành từng chiều (dimension) khác nhau và xử lý các chiều này độc lập. Việc áp dụng kiến trúc Depthwise Seperable Convolution đã giúp mô hình có kết quả tốt với khối lượng tính toán không quá cao. NASNet-A-Mobile có thể đạt được kết quả tốt nhất với mô hình nhỏ với độ phức tạp ít nhất bằng các khối được xây dựng thông qua quá trình học tăng cường.

3. Phương pháp nghiên cứu

Nghiên cứu áp dụng một quy trình tổng quát gồm có bốn chặng tuần tự: (1) Chuẩn bị bộ dữ liệu, (2) Tinh chỉnh trọng số, (3) Huấn luyện mô hình, (4) Đánh giá kết quả. Kết quả từ mỗi chặng đều đóng vai trò tiền đề để thực hiện công việc của chặng liên sau.

3.1. Chuẩn bị bộ dữ liệu

Bộ dữ liệu được dùng trong nghiên cứu này là Covid-Xray-5K v3, chuẩn bị bởi Minaee và các cộng sự. Covid-Xray-5K v3 gồm các hình chụp X-quang thẳng lồng ngực, được giám định bởi một chuyên viên đã được hội đồng chứng nhận (Minaee, Kafieh, Sonka, Yazdani, & Jamalipour Soufi, 2020). Trong bộ dữ liệu, 5000 mẫu âm tính được tổng hợp từ 13 phân loại: xẹp phổi, tim to, đông đặc, phù nề, mở rộng trung thất tim, gãy xương, tổn thương phổi, màng phổi, màng phổi, viêm phổi, tràn khí màng phổi, dùng thiết bị hỗ trợ, và không có phát hiện nào. 184 mẫu dương tính được lấy từ các nghiên cứu khác, và chỉ bao gồm các mẫu thể hiện rõ dấu hiệu của COVID-19. Bộ dữ liệu gốc đã được chia sẵn thành tập train có 2000 mẫu âm tính, 84 mẫu dương tính và tập test có 3000 mẫu âm tính, 100 mẫu dương tính. Trong nghiên cứu này, tập train đã được chia lại thành một tập train nhỏ hơn và tập validation với tỉ lệ 4:1 (Tương đương với 1600 mẫu âm tính, 67 mẫu âm tính ở tập train và 400 mẫu âm tính, 17 mẫu dương tính ở tập validation) Để đối phó với vấn đề overfit, bộ dữ liệu được làm giàu bằng cách áp dụng các phép biến hình. Đầu tiên, các mẫu ảnh sẽ được thay đổi kích thước về 256×256 điểm ảnh. Tiếp đến, đối với tập train, các biến thể ngẫu nhiên sẽ tạo ra thông qua các phép biến hình như cắt hình ngẫu nhiên (224×224), xoay hình ngẫu nhiên (tối đa 10°), lật ngẫu nhiên theo phương ngang (50%), để mô phỏng những thay đổi trong các tình huống thực tế. Đối với tập validation và test, các mẫu sẽ được cắt 224×224 về trung tâm, đảm bảo phép biến hình sẽ luôn giống nhau và cho kết quả cố định trong quá trình đánh giá. Cuối cùng, mọi mẫu ảnh sẽ được chuẩn hóa về các tensor $3 \times 224 \times 224$ để đưa vào mô hình.

3.2. Tinh chỉnh trọng số

Tinh chỉnh trọng số là quá trình điều chỉnh các trọng số để tìm được một bộ trọng số cho ra quá trình huấn luyện tối ưu hoặc đáp ứng được kỳ vọng. Mỗi mô hình sẽ hoạt động tốt nhất với một bộ trọng số nhất định, và hiện tượng nhiễu cũng sẽ ảnh hưởng đến sự dao động của việc huấn luyện. Hiểu một cách đơn giản, tinh chỉnh trọng số là bài toán tìm giá trị cực đại của hàm $f(x)$ từ tập A, với $f(x)$ là một hàm hợp đen phức tạp về mặt tính toán, còn A là tập hợp các bộ tham số đầu vào để xác định.

Trong trường hợp tinh chỉnh trọng số, tập A là tập chứa tất cả các trường hợp của các bộ trọng số, với giá trị của mỗi trọng số nằm trong miền cho trước, còn hàm $f(x)$ là quá trình huấn luyện thử mô hình với một bộ trọng số đầu vào. Bằng cách thử lặp đi lặp lại nhiều bộ trọng số khác nhau, đến khi đạt được kết quả tốt trên tập validation, bộ trọng số phù hợp sẽ được tìm thấy. Tuy vậy, để quá trình này diễn ra một cách hiệu quả, việc đánh giá kết quả (phần $f(x)$) và một thuật toán tối ưu trên toàn miền (phần max) là cần thiết. Kiểm chứng chéo k-lần (k-fold cross-validation) và tối ưu hoá Bayes (Bayesian optimization) (Mockus, 1989) đã được áp dụng một cách tương ứng. Đặc biệt, tối ưu hoá Bayes cho kết quả nhanh hơn và tiết kiệm tính toán hơn so với thuật toán tìm theo lưới hay tìm ngẫu nhiên.

3.3. Huấn luyện mô hình

Sau khi được tinh chỉnh, các bộ trọng số sẽ được dùng để cấu hình vòng lặp huấn luyện. Trong vòng lặp này, dữ liệu từ tập train sẽ được dùng để huấn luyện, còn dữ liệu từ tập validation sẽ được dùng để đánh giá quá trình huấn luyện đó. Mô hình học được kiến thức từ dữ liệu dựa trên thuật toán gradient descent, bằng cách cập nhật các trị số trong mạng thần kinh của mô hình. Khi kết quả trên tập validation bắt đầu có dấu hiệu overfit, quá trình huấn luyện sẽ được dừng một cách tự động bằng thuật toán Early Stopping, dựa trên nghiên cứu của Prechelt (Prechelt, 2012). Đồng thời, mỗi lần mô hình đạt được kết quả tốt nhất trong suốt quá trình huấn luyện, các trị số của mạng thần kinh sẽ được lưu lại và xem như là kết quả cuối cùng của mô hình. Ứng với từng kiểu mô hình ResNet50, NASNet-A-Mobile, Xception, có tổng cộng 05 mô hình được huấn luyện. Các mô hình cho ra kết quả tốt nhất trên tập validation từ mỗi kiểu sẽ được chọn để dự đoán trên tập test.

3.4. Đánh giá kết quả

Để đánh giá kết quả của một mô hình dự đoán, đặc biệt là dự đoán nhị phân, các chỉ số sau thường được sử dụng phổ biến: độ chính xác (accuracy), độ phủ/nhảy (recall), độ đặc hiệu (specificity), độ chuẩn xác (precision). Đặc biệt, chỉ số điểm F1 (F1-score) sẽ hiệu quả đối với trường hợp bộ dữ liệu bị mất cân bằng (số lượng mẫu của các nhãn không đồng đều), chỉ số diện tích bên dưới đường Receiver Operating Characteristic (AU-ROC) hiệu quả trong bài toán ngưỡng quyết định di động (thay đổi ngưỡng quyết định để đề phòng các trường hợp âm tính giả hoặc dương tính giả). Đặc biệt, chỉ số diện tích bên dưới đường Precision-Recall (AU-PRC) có thể giải quyết cả hai vấn đề này (Chou; Saito & Rehmsmeier, 2015). Vì lý do đó, trong nghiên cứu này, AU-PRC là chỉ số tối quan trọng trong việc quyết định lựa chọn bộ trọng số ở và mô hình ở các bước trước trong quy trình tổng quát đã giới thiệu ở đầu mục này, các chỉ số còn lại chỉ được dùng để đánh giá. Ngoài ra, với bài toán ngưỡng quyết định di động, nghiên cứu này còn tính toán chỉ số độ đặc hiệu ứng với với độ nhạy 98% - đây là độ nhạy tương đối cao để đảm bảo tầm soát dịch một cách hiệu quả.

4. Kết quả nghiên cứu và thảo luận

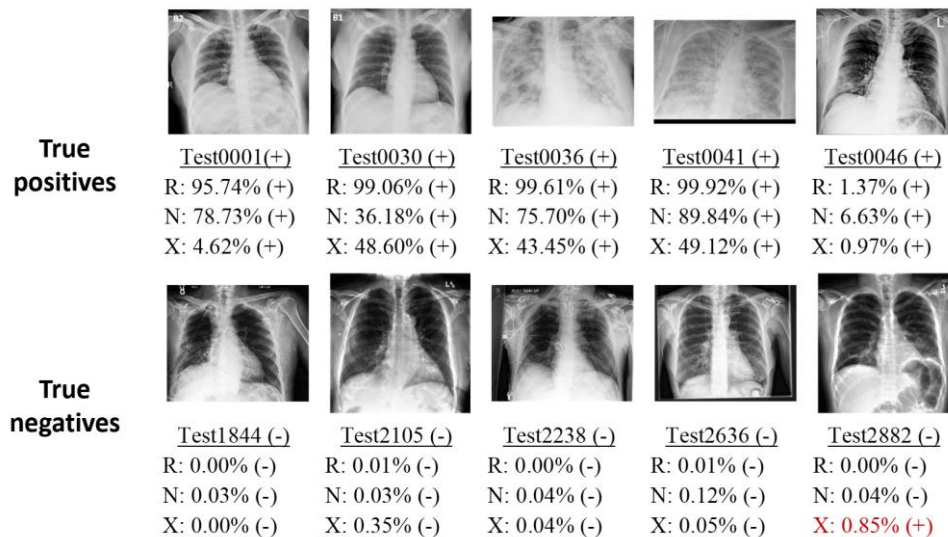
4.1. Kết quả nghiên cứu

Cả ba mô hình được chuyển tiếp qua bộ thử nghiệm để trả về kết quả đánh giá. Do kích thước lớn của tập dữ liệu, chỉ một số ví dụ từ tất cả các danh mục: dương tính thật, dương tính giả, âm tính thật, âm tính giả được trích xuất để dễ hình dung. Dấu (+) và (-) được kí hiệu lần lượt cho kết quả chẩn đoán là dương tính hay âm tính. Các dự đoán được đưa ra bởi ResNet50, NASNet-A-Mobile và Xception được ký hiệu lần lượt là R, N và X. Ngoài ra, các dự đoán sai được đánh dấu màu đỏ. Các ngưỡng quyết định tương đương với độ nhạy 98% đã được sử dụng cho mỗi mô hình (xem Hình 1).

Mặc dù độ chính xác không phải là thước đo hiệu suất chính của nghiên cứu này, nhưng ba mô hình đã đạt được những kết quả đáng chú ý. Các giá trị độ chính xác đã chọn của ba mô hình để

so sánh, đối với các nghiên cứu khác, chúng được giữ nguyên như trong tài liệu gốc. Sự khác biệt về độ chính xác giữa các nghiên cứu khác nhau chỉ có thể so sánh đưa ra một góc nhìn hạn chế. Một lần nữa, điều này ngụ ý nhu cầu về một số liệu có khả năng phân biệt hiệu suất trong tập dữ liệu không cân bằng, chẳng hạn như AU-PRC. Vì điều quan trọng là phải duy trì một độ nhạy cao để phát hiện COVID-19, độ nhạy ở mức 98% nhớ lại đã được sửa đổi để đưa ra một so sánh rõ ràng hơn. Cuối cùng, ResNet50 và NASNet-A-Mobile được đề xuất trong nghiên cứu này đã hoạt động tốt hơn những mô hình khác (xem Bảng 1).

4.1.1. Kết quả chẩn đoán



Hình 1: Kết quả thu được sau khi huấn luyện các mô hình

4.1.2. So sánh kết quả so với các công trình nghiên cứu khác

Bảng 1

Bảng so sánh kết quả so với các công trình nghiên cứu khác

Mô hình nghiên cứu	Độ chính xác	Độ đặc hiệu (độ nhạy 98%)	AU-ROC	AU-PRC
ResNet50	99.61%	97.53%	0.9940	0.9773
NASNet-A-Mobile	99.45%	97.60%	0.9964	0.9633
Xception	98.77%	86.00%	0.9812	0.9003
ResNet50*	—	89.6%	0.990	0.899
SqueezeNet*	—	92.9%	0.992	0.897
CNN by Chimp optimization**	98.25%	—	—	—
CNN by hybrid sine-cosine***	98.83%	93.2%	—	—
ResNet18 ****	—	—	0.994	—
AlexNet ****	—	—	0.995	—

*(Minaee et al., 2020)

** (Hu et al., 2021)

*** (Wu, Khishe, Mohammadi, Taher Karim, & Rashid, 2021)

**** (Naviwala & Qureshi, 2021)

4.2. Thảo luận

Nhờ có dung lượng nhẹ mà mô hình NASNet-A-Mobile có khả năng được tiếp tục phát triển để trở thành ứng dụng chẩn đoán trên điện thoại thông minh (xem Hình. 2). Dự kiến ứng dụng có thể giúp người dùng tự tham khảo chẩn đoán của mình một cách nhanh chóng nhất nếu họ có ảnh chụp X-quang lồng ngực của mình. Điều này giúp giảm áp lực lên hệ thống y tế và phổ cập nhanh hơn việc sử dụng trí tuệ nhân tạo vào chẩn đoán y khoa tại những cơ sở y tế cấp địa phương và những nơi chưa có thiết bị y tế tích hợp hệ thống trí tuệ nhân tạo.



Hình 2: Giao diện dự kiến của ứng dụng chẩn đoán Covid-19 bằng X-quang phổi

5. Kết luận & Gợi ý

Tóm lại, nghiên cứu này đã tạo ra một số kết quả đầy hứa hẹn một cách hiệu quả. Một quy trình tinh chỉnh về việc phát triển các mô hình học sâu để phát hiện COVID-19 đã được hình thành. Quy trình này xem xét và khắc phục hai cài đặt phổ biến thường xuyên phát sinh trong các vấn đề chẩn đoán y học: mất cân bằng tập dữ liệu và ngưỡng quyết định di động. Quy trình này cũng có thể là nền tảng để phát triển các mô hình học sâu để dự đoán các tình trạng sức khỏe khác. Với quy trình này, việc học chuyển giao của ba loại mô hình học sâu được đào tạo trước đã được thực hiện để tạo ra ba mô hình dự đoán COVID-19 với kết quả hiện đại.

Trong khi đó, ba mô hình này có thể hỗ trợ trong cuộc chiến khẩn cấp chống lại đại dịch toàn cầu. Mô hình hiệu suất tốt nhất, ResNet50, có thể được tích hợp vào một ứng dụng máy tính để bàn nhằm hỗ trợ các xét nghiệm sàng lọc COVID-19 bên cạnh các phương pháp test nhanh và PCR thời gian thực. Ứng dụng có thể được cài đặt trên hệ thống máy tính của bệnh viện để đẩy nhanh việc phân loại và nhập viện, trên một thiết bị y tế di động có kèm X-quang để thực hiện xét nghiệm cục bộ cho một vùng dân cư với tốc độ nhanh chóng và chi phí hợp lý. Mô hình nhẹ nhất, NASNet-A-Mobile, có thể được phát triển thành một ứng dụng di động để mọi người có thể dễ dàng theo dõi khả năng mắc bệnh nếu họ có thể chụp X-quang phổi.

Nghiên cứu này có tiềm năng tốt để cải tiến như có thể là mở rộng khả năng dự đoán bao gồm cả chẩn đoán X-quang ngực bên. Ngoài ra, các mô hình dự đoán có thể phân loại nhiều loại tình trạng hô hấp hơn trong một bài toán phân loại nhiều lớp. Cách tiếp cận phân loại có thể được thay đổi từ phân loại nhị phân sang phân đoạn ngữ nghĩa hoặc một thuật toán để trực quan hóa các hình ảnh dự đoán dưới dạng bản đồ nhiệt có thể được phát triển. Cuối cùng, với nhiều tài

nguyên tính toán hơn, các mô hình có kích thước lớn hơn có thể được đào tạo để cho kết quả tốt hơn.

LỜI CẢM ƠN

Nhóm tác giả xin chân thành gửi lời cảm ơn đến PGS. TS. Phạm Thị Thu Hiền và TS. Lê Thanh Hải vì sự hướng dẫn tận tình từ quý thầy cô trong quá trình thực hiện đề tài cũng như những góp ý hữu ích trong quá trình viết bài tham luận.

TÀI LIỆU THAM KHẢO

- Chollet, F. (2017). *Xception: Deep learning with depthwise separable convolutions*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Chou, S.-Y. Precision - Recall Curve, a Different View of Imbalanced Classifiers. Retrieved from <https://sinyi-chou.github.io/classification-pr-curve/>
- Chriscaden, K. (2020). Impact of COVID-19 on people's livelihoods, their health and our food systems. Retrieved from <https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people's-livelihoods-their-health-and-our-food-systems>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Hu, T., Khishe, M., Mohammadi, M., Parvizi, G.-R., Taher Karim, S. H., & Rashid, T. A. (2021). Real-time COVID-19 diagnosis from X-Ray images using deep CNN and extreme learning machines stabilized by chimp optimization algorithm. *Biomedical Signal Processing and Control*, 68, 102764. doi:10.1016/j.bspc.2021.102764
- Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., & Jamalipour Soufi, G. (2020). Deep-COVID: Predicting COVID-19 From Chest X-Ray Images Using Deep Transfer Learning. *Medical Image Analysis*, 65, 101794. doi:10.1016/j.media.2020.101794
- Mockus, J. (1989). *Bayesian Approach to Global Optimization: Theory and Applications* (Vol. 37). Dordrecht: Springer Netherlands.
- Naviwala, M. H., & Qureshi, R. (2021, 2021-05). *Performance Analysis of Deep Learning Frameworks for COVID 19 Detection*. Paper presented at the 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2).
- Prechelt, L. (2012). Early Stopping — But When? In G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Neural Networks: Tricks of the Trade: Second Edition* (Vol. 7700, pp. 53-67). Berlin, Heidelberg: Springer.
- Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS One*, 10(3), e0118432. doi:10.1371/journal.pone.0118432
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). *Going deeper with convolutions*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Wu, C., Khishe, M., Mohammadi, M., Taher Karim, S. H., & Rashid, T. A. (2021). Evolving deep convolutional neural network by hybrid sine-cosine and extreme learning machine for real-time COVID19 diagnosis from X-ray images. *Soft Computing*. doi:10.1007/s00500-021-05839-6
- Zoph, B., Vasudevan, V., Shlens, J., & Le, Q. V. (2018). *Learning transferable architectures for scalable image recognition*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.