

PHÂN LOẠI VÀ KHOANH VÙNG 14 LOẠI TÔN THƯƠNG PHỔI SỬ DỤNG MẠNG NƠI RON CLASSIFICATION AND LOCALIZATION OF 14 TYPES OF LUNG LESIONS USING NEURAL NETWORKS

TRẦN QUỐC THỊNH

Lớp Cao học CNTT K3. Khoa Công nghệ thông tin,

Trường Đại học SPKT Vĩnh Long

Email: thinhqt@vlute.edu.vn

Nhận bài (Received): 25/3/2022; Phản biện (Reviewed): 15/5/2022; Chấp nhận (Accepted): 28/7/2022

TÓM TẮT:

Nguy cơ mắc bệnh phổi là vô cùng lớn đối với nhiều người, đặc biệt là ở các quốc gia đang phát triển. Các bác sĩ gặp khó khăn trong việc giải thích tia X ở ngực có thể dẫn đến chẩn đoán sai. Phương pháp học sâu giúp xác định vị trí các bất thường của phổi. Sử dụng tập dữ liệu gồm 112.000 ảnh được gán nhãn bởi các bác sĩ có kinh nghiệm. Chúng tôi đào tạo một mạng lưới thần kinh dựa trên khuôn khổ Faster-RCNN và Retina theo hướng tiếp cận dữ liệu lớn làm giảm thời gian huấn luyện. Hơn nữa, chúng tôi đã sử dụng Grad-CAM để làm nổi bật vùng tổn thương trong việc đưa ra dự đoán. Kết quả chứng minh phương pháp sử dụng kỹ thuật học sâu có thể đạt được độ chính xác lên đến 95.5%.

Từ khóa: Tổn thương phổi, Faster-RCNN, Retina, transfer learning, bigdata.

ABSTRACT:

The risk of lung disease is immense for many people, especially those in developing countries. Doctors have difficulty interpreting chest X-rays, which can lead to misdiagnosis. Deep learning methods help locate lung abnormalities. A dataset of 112,000 images is labeled by experienced physicians. A neural network is trained based on Faster-RCNN and Retina framework towards big data approach to reduce the training time. Furthermore, Grad-CAM is used to highlight the lesion region in making predictions. The results demonstrate that the method using deep learning techniques can achieve the accuracy of up to 95.5%.

Keywords: Lung lesion, Faster-RCNN, Retina, transfer learning, bigdata.

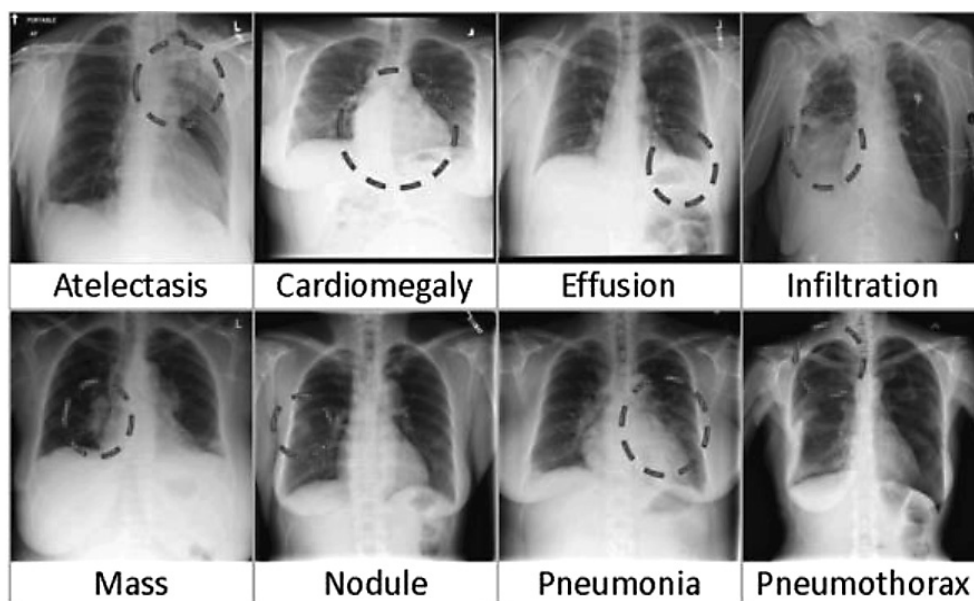
1. MỞ ĐẦU

1.1. Giới thiệu các đặc điểm tổn thương phổi, các nghiên cứu liên quan

1.1.1. Các đặc điểm tổn thương trên phổi

Hình ảnh X-quang phổi có giá trị trong việc phát hiện, chẩn đoán bệnh, đánh giá

mức độ nặng, đánh giá các biến chứng hô hấp. Tuy nhiên, nếu chỉ dựa trên hình ảnh X-quang, rất khó để phân biệt giữa viêm phổi do virus với một số căn nguyên khác, do vậy phải kết hợp với đặc điểm dịch tễ, biểu hiện lâm sàng để đưa ra chẩn đoán phù hợp.



Hình 1. Hình ảnh X-quang ngực các loại tổn thương phổi.

1.1.2. Các nghiên cứu có liên quan

Nhóm tác giả Rachna Jain, Preeti Nagrath trong nghiên cứu “Phân loại viêm phổi trong hình ảnh X-quang ngực bằng cách sử dụng mạng nơ-ron tích chập và transfer learning” [4]. Mô hình thứ nhất và thứ hai bao gồm hai và ba lớp chập. Bốn mô hình còn lại là các mô hình được đào tạo trước đó là VGG16, VGG19, ResNet50 và Inception-v3. Mô hình thứ nhất và thứ hai đạt được độ chính xác lần lượt là 85,26% và 92,31%. Độ chính xác của VGG16, VGG19, ResNet50 và Inception-v3 lần lượt là 87,28%, 88,46%, 77,56% và 70,99%. Nghiên cứu tiếp theo là của tác giả Alex Alemi, nghiên cứu “Improving Inception and Image Classification in TensorFlow” vào năm 2016 [5] đã đào tạo lại bốn mô hình Inception-ResNet-v2, Inception V3, ResNet 152 và ResNet V2 200 sử dụng tập dữ liệu ImageNet tỷ lệ chính xác lần lượt là 80.4%, 78%, 76.8% và 79.9%. Kết quả cho thấy mô hình có

thể phát hiện đối tượng tốt trên Inception-ResNet-V2. Một cách tiếp cận khác sử dụng CheXNet của nhóm tác giả Laboni Sarker, Md. Mohaiminul Islam với nghiên cứu “COVID-DenseNet: A deep learning architecture to detect COVID-19 from chest radiology images” vào năm 2021 [6]. Trong nghiên cứu này, nhóm tác giả đề xuất sử dụng CheXNet để phát hiện bệnh nhân COVID-19. Để kiểm tra độ chắc chắn của mô hình tác giả đã thực hiện cả phân loại hai lớp và ba lớp và đạt độ chính xác lần lượt là 96,49% và 93,71%.

1.2. Cơ sở lý thuyết

Như chúng ta đã biết, xét nghiệm tổn thương phổi bằng các cách như chụp CT, nội soi phế quản, xét nghiệm máu và chụp X-Quang ngực. Trong đó, cách chụp X-quang ngực cho được cái nhìn toàn thể hai lá phổi, nhìn thấy các tổn thương đủ lớn, không bị che lấp trên hai lá phổi và cuối cùng là giảm được chi phí khi xét nghiệm. Gần đây, một số nhà nghiên cứu

đã đề xuất các trí thông minh nhân tạo (AI) khác nhau dựa trên giải pháp cho các vấn đề y tế khác nhau. Mạng nơ-ron (CNN) đã cho phép các nhà nghiên cứu thu được kết quả thành công trong các vấn đề y tế rộng rãi như phát hiện ung thư vú, não phát hiện và phân đoạn khối u, phân loại bệnh trên hình ảnh X-quang.

1.2.1 Các mạng nơ-ron trích xuất đặc trưng

ResNet trở thành kiến trúc được sử dụng phổ biến nhất ở thời điểm hiện tại, với ít tham số, huấn luyện dễ dàng mà không làm tăng tỷ lệ lỗi huấn luyện. ResNet bao gồm bốn kiến trúc khác nhau lần lượt là 34, 50, 101 và 152 layer [1]. Trong đó, ResNet-50 đạt được hiệu suất khá tốt.

CheXnet bao gồm 121 lớp CNN [11]. Nó tạo ra bản đồ nhiệt bao gồm các khu vực cục bộ có thể xác định các khu vực bị ảnh hưởng bởi bệnh trong hình ảnh cùng với xác suất dự đoán. Phương pháp này được phát triển để dự đoán bệnh viêm phổi khi chụp X-quang phổi.

Inception Resnet V2 cải thiện thành công độ chính xác và tốc độ hội tụ [12] của mô hình ban đầu. Khi điều chỉnh trọng số bằng phương pháp lan truyền ngược, gradient từ lớp trên có thể bỏ qua lớp giữa cho đến lớp thấp nhất, đảm bảo rằng tất cả các trọng số có thể được điều chỉnh một cách hiệu quả.

1.2.2 Các mạng CNNs dùng trong huấn luyện

Faster R-CNN: huấn luyện một mô hình hiệu quả hơn và thay thế vai trò của các thuật toán như Selective Search vốn rất chậm chạp [7]. Faster R-CNN được đánh giá nhanh hơn hẳn các dòng

R-CNN trước đó. Ngoài ra, mạng này còn được sử dụng phổ biến cho việc phát hiện và phân loại.

Retina: ở các mạng YOLO, SSD.... cho thấy rằng các mạng đó có một điểm yếu [8] là rất dễ bỏ qua các vật thể nhỏ trong ảnh. Và theo yêu cầu của bài toàn là tìm các tổn thương nhỏ trên phổi. RetinaNet được đánh giá là khá ổn trong việc phát hiện các đối tượng nhỏ.

1.2.3 Các độ đo dùng để đánh giá mô hình

Độ đo Loss: Hàm loss sẽ chỉ ra mô hình đoán sai bao nhiêu so với giá trị thực tế. Mục đích của quá trình huấn luyện là tìm ra bộ số để độ lớn hàm mất mát là nhỏ nhất (cực tiểu).

Độ đo accuracy: Accuracy (độ chính xác) cách đánh giá này đơn giản tính tỉ lệ giữa số điểm được dự đoán đúng và tổng số điểm trong tập dữ liệu kiểm thử. Được tính bằng công thức sau:

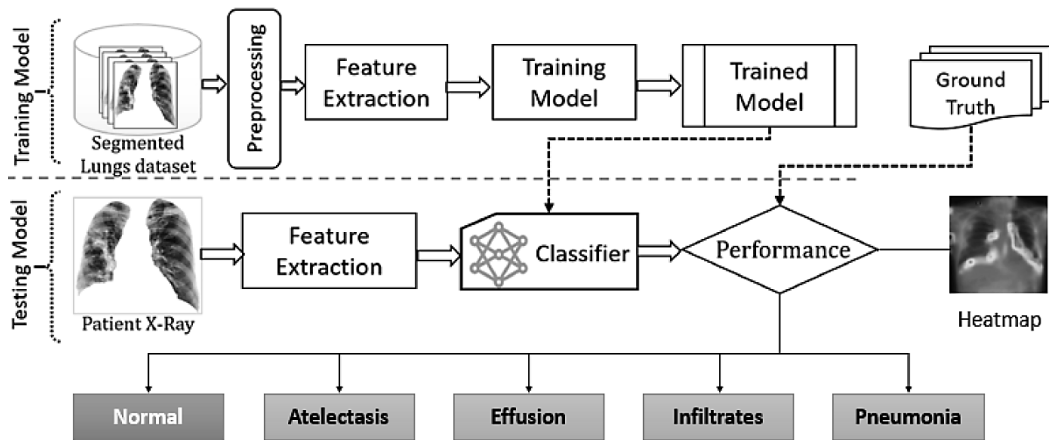
$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

Độ đo AUC: đường cong AUC - ROC là một phép đo hiệu suất cho phân loại ở các cài đặt ngưỡng khác nhau. ROC là một đường cong xác suất và AUC đại diện cho mức độ. Nó cho biết mô hình có khả năng phân biệt giữa các lớp như thế nào. AUC càng cao, mô hình càng tốt trong việc dự đoán là 1.

2. KẾT QUẢ NGHIÊN CỨU

2.1. Phương pháp đề xuất

Để giải quyết bài toán, chúng tôi đề xuất sử dụng mô hình tổng quát như Hình 2. Mô hình tổng quát gồm có hai pha: Pha huấn luyện và pha kiểm thử được mô tả cụ thể như hình sau:



Hình 2. Mô hình giải pháp đề xuất

A. Pha huấn luyện

Ở giai đoạn tiền xử lý dữ liệu các hình ảnh được thay đổi kích thước thành 256×256 và thay đổi tỷ lệ thành phạm vi 0–1 để phù hợp với đầu vào của mô hình. Để tăng thêm dữ liệu đầu vào chúng tôi áp dụng các phép như cắt ngẫu nhiên các vùng trên ảnh để tạo sự đa dạng. Lật ngẫu nhiên tạo ra một ảnh phản chiếu từ điểm gốc. Rút trích đặc trưng chúng tôi đề xuất phương pháp với ba mô hình. Sau khi rút trích đặc trưng chúng tôi sử dụng lại mạng Faster R-CNN và RetinaNet để có thể sử dụng được transfer learning [9] từ bộ trọng số (weights) được trích xuất sẵn của mô hình.

Điều này giúp cho việc học các đặc trưng mới nhanh hơn, rút ngắn được thời gian huấn luyện.

B. Pha kiểm thử

Ảnh sau khi đã rút trích đặc trưng sẽ được đưa qua mô hình đã được huấn luyện trước đó để cho ra kết quả dò tìm tự động gồm bounding box chứa tổn thương và nhãn của tổn thương. Một điểm mới của bài báo chúng tôi là áp dụng thêm bản đồ nhiệt GradCAM để xác định chính xác vùng tổn thương nặng trên phổi.

2.2. Kết quả thực nghiệm

2.2.1 Các kịch bản áp dụng

Bảng 1. Các kịch bản được đề xuất và các tham số huấn luyện

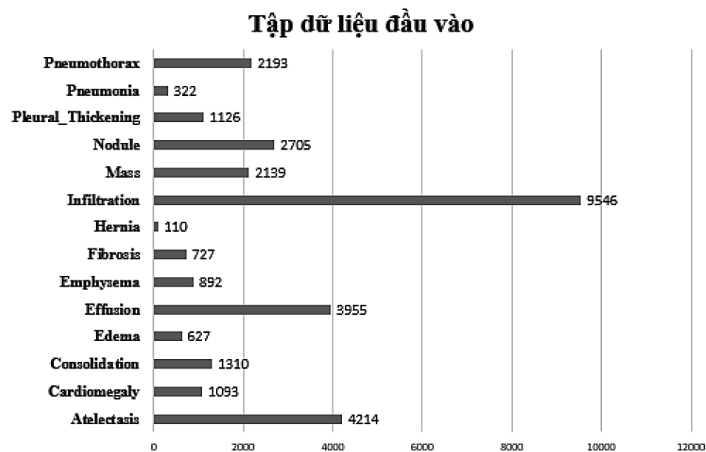
Kịch bản	Mạng huấn luyện	Mạng rút trích đặc trưng	Learning rate	Num_steps	mAP @IoU	Num classes
1	Faster R-CNN	ResNet-50	0.001	20	0.5	14
2		CheXNet	0.001	20	0.5	14
3		Inception ResNet V2	0.001	20	0.5	14
4	Retina	ResNet-50	0.001	20	0.5	14

2.2.2 Môi trường cài đặt

Ngôn ngữ lập trình Python 3.7. Thư viện, công cụ: OpenCV, Tensorflow, kaggle. Cấu hình máy tính sử dụng GPU: GPU Nvidia K80, Ram 13Gb.

2.2.3 Tập dữ liệu thử nghiệm

Tập dữ liệu ChestX-ray14 [10] gồm 112.000 ảnh trong đó 89600 ảnh được dùng làm tập huấn luyện và 22400 ảnh cho tập kiểm thử, tương ứng với tỉ lệ là 80:20.



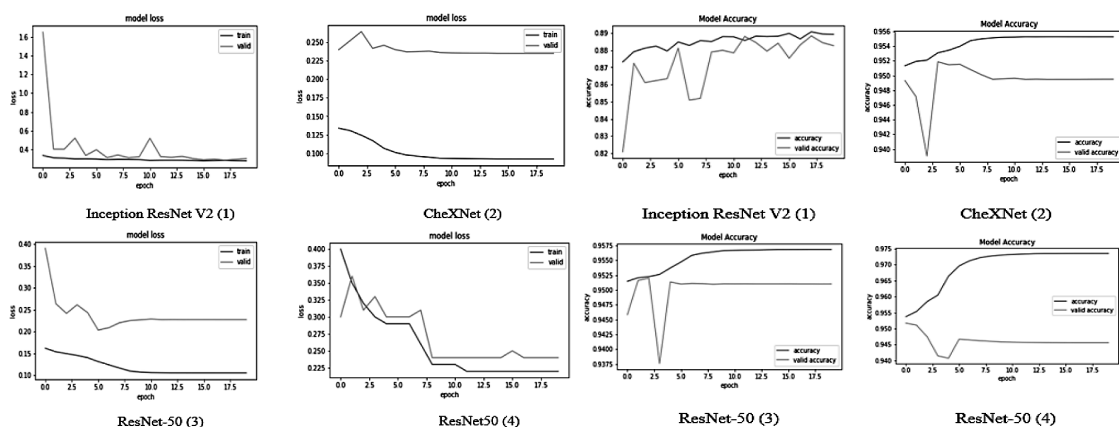
Hình 3. Số lượng tập ảnh của từng class.

2.2.4 Các kết quả thu được

a) Độ đo Loss

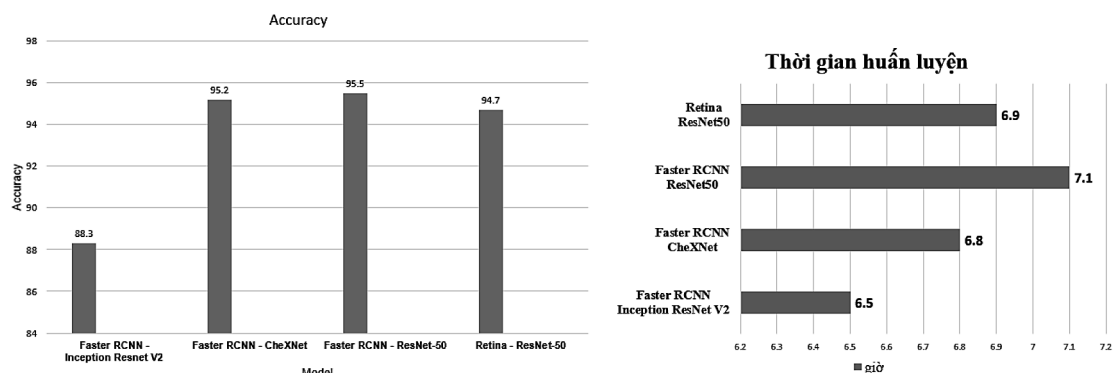
Khi giá trị Loss không còn giảm chúng tôi quyết định dừng huấn luyện, khi đó giá trị Loss trung bình của 3 kịch

bản lần lượt là 0.24, 0.096, 0.112. Mạng Retina – Resnet 50 cho kết quả rút trích đặc trưng cũng khá tốt nhưng chỉ số Loss vẫn còn khá cao so với các kịch bản trước lên đến 0.21



Hình 4. Biểu đồ về độ đo Loss và Accuracy khi huấn luyện.

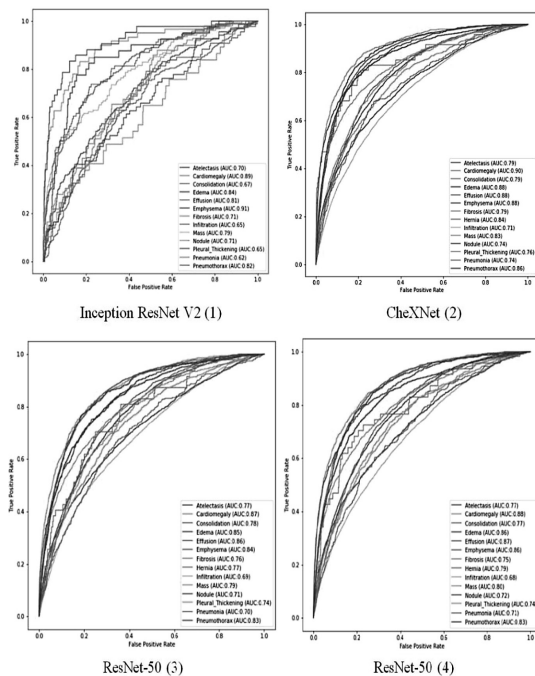
b) Độ đo Accuracy và thời gian huấn luyện



Hình 5. Biểu đồ về Accuracy và thời gian huấn luyện.

Kết quả ở Hình 5 cho thấy đối với mô hình Inception ResNet V2 cho ra độ chính xác thấp hơn các mô hình còn lại và không ổn định giữa các lần huấn luyện. Hai mô hình còn lại cho kết quả khá tương đồng và ổn định ở mức chính xác 95% so với 88% ở mô hình Inception ResNet V2.

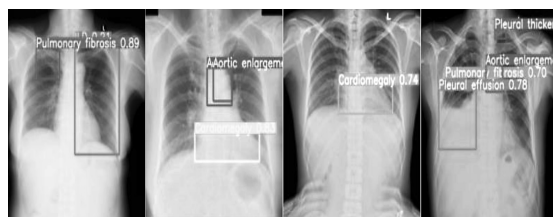
c) Độ đo AUC-ROC:



Hình 6. Biểu đồ về độ đo AUC khi kiểm thử.

2.3. Kết quả dò tìm và phân loại tổn thương

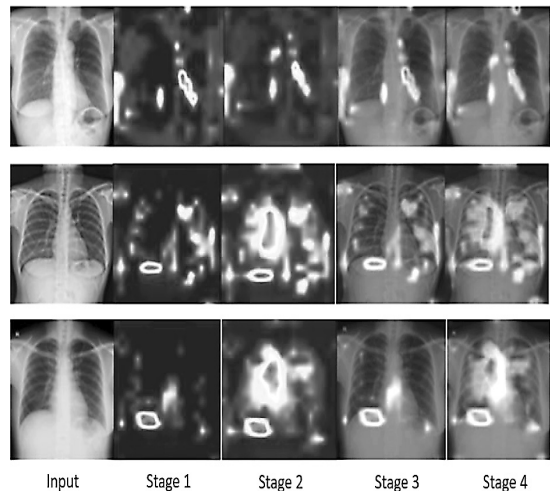
Một số hình ảnh thực nghiệm khi dò tìm và phân loại tự động tổn thương trên phổi của 4 kịch bản đề xuất được trình bày trong bảng sau:



Hình 7. Kết quả thực nghiệm trên các kịch bản đề xuất.

Ở những kết quả trên có thể khoanh vùng được vị trí của những nơi bị tổn

thương của phổi nhưng không thể xác định được vùng tổn thương nặng nhất hay vùng trung tâm mà bác sĩ cần phải chữa trị. Để giải quyết bài toán đó, chúng tôi đã áp dụng thêm bản đồ nhiệt (GradCam). Ở giai đoạn đầu sẽ rút trích các đặc trưng cấp thấp như noise hoặc kết cấu của lớp background, sau đó thu hẹp không gian rút trích và dần tập trung vào mục tiêu chính. Cuối cùng vị trí tổn thương chính xác có màu đỏ cho tổn thương nặng.



Hình 8. Ảnh minh họa áp dụng bản đồ nhiệt lên ảnh x-quang

3. KẾT LUẬN

Trong bài báo này, chúng tôi nghiên cứu cách xây dựng các mô hình học sâu để chẩn đoán các tổn thương của phổi bằng cách chọn 3 kiến trúc CNN khác nhau để thử nghiệm hiệu suất của chúng trong bài toán phân loại đã cho và áp dụng thêm các mạng Faster R-CNN và Retina. Tôi cũng thực hiện tăng cường dữ liệu như lật và cắt ngẫu nhiên để tăng sự đa dạng cho tập dữ liệu đầu vào. Ngoài việc đào tạo mô hình bằng cách sử dụng ngẫu nhiên trọng số khởi tạo, chúng tôi áp dụng phương pháp transfer learning để cải thiện hơn nữa hiệu suất kết quả cho thấy mô hình Faster R-CNN backbone ResNet 50 cho ra

kết quả tốt nhất với độ chính xác 95.5%. Công việc trong tương lai của chúng tôi sẽ áp dụng phương pháp transfer learning trên các bộ dữ liệu khác, sử dụng nhiều mô hình để có thể đánh giá và so sánh các mô hình.

TÀI LIỆU THAM KHẢO

- [1] Kaiming He, X. Zhang, S. Ren, J. Sun (2015), “*Deep Residual Learning for Image Recognition*”.
- [2] S. Ren, K. He, R. Girshick, and J. Sun (2017), “*Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*”.
- [3] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár (2018), “*Focal Loss for Dense Object Detection*”.
- [4] Rachna Jain, Preeti Nagrath, Gaurav Kataria, V. Sirish Kaushik, D. Jude Hemanth (2020), “*Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning*”.
- [5] Alex Alemi (2016), “*Improving Inception and Image Classification in TensorFlow*”.
- [6] Laboni Sarker , Md. Mohaiminul Islam, Tanveer Hannan và Zakaria Ahmed (2021), “*COVID-DenseNet: A deep learning architecture to detect COVID-19 from chest radiology images*”.
- [7] S. Ren, K. He, R. Girshick, and J. Sun (2017), “*Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*”.
- [8] Ivan William Harsono, Suryadiputra Liawatimena, Tjeng Wawan Cenggoro (2020), “*Lung nodule detection and classification from Thorax CT-scan using RetinaNet with transfer learning*”.
- [9] S. J. Pan and Q. Yang (2010), “*A survey on transfer learning*”.
- [10] <https://paperswithcode.com/dataset/chestx-ray14> (14/02/2022)
- [11] Mawanda Almuhayar, Henry Horng-Shing Lu (2019). “*Classification of Abnormality in Chest X-Ray Images by Transfer Learning of CheXNet*”.
- [12] Szegedy, Vanhoucke(2017). “*Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*”.