

CHẨN ĐOÁN VIÊM PHỔI BẰNG HÌNH ẢNH X-QUANG NGỰC SỬ DỤNG KỸ THUẬT HỌC SÂU

DIAGNOSING PNEUMONIA USING CHEST X-RAY IMAGES USING DEEP LEARNING TECHNIQUE

Trần Thanh Hùng^{1,*}

DOI: <https://doi.org/10.57001/huih5804.2023.248>

TÓM TẮT

Hiện nay, bệnh viêm phổi đang là một trong những bệnh phổ biến nhất trên thế giới và cũng là một trong những tác nhân gây chết người hàng đầu theo nghiên cứu của tổ chức Y tế thế giới (WHO). Bài báo này nhằm nghiên cứu về bệnh viêm phổi dựa trên hình ảnh X-quang, và nghiên cứu về mô hình học sâu VGG-19, sử dụng mô hình học sâu này trong bài toán chẩn đoán viêm phổi bằng hình ảnh X-quang, qua đó có sự so sánh giữa các mô hình học sâu và để cho các nhà nghiên cứu có sự lựa chọn mô hình trong việc xây dựng các ứng dụng về chẩn đoán hình ảnh y tế.

Từ khóa: Viêm phổi, X-quang ngực, mô hình VGG-19, mạng nơ-ron tích chập.

ABSTRACT

Currently, pneumonia is one of the most common diseases in the world and also one of the leading causes of death according to research by the World Health Organization (WHO). This article aims to study pneumonia based on X-ray images, and study the deep learning model VGG-19, using this deep learning model in the problem of diagnosing pneumonia by X-ray images, thereby making a comparison between deep learning models and letting researchers have a choice of models in building medical imaging applications.

Keywords: Pneumonia, chest X-ray, VGG-19 model, Convolution Neural Network.

¹Khoa Công nghệ thông tin, Trường Đại học Công nghiệp Hà Nội

*Email: hungtt@fit-hau.edu.vn

Ngày nhận bài: 15/10/2023

Ngày nhận bài sửa sau phản biện: 10/12/2023

Ngày chấp nhận đăng: 25/12/2023

1. GIỚI THIỆU

Viêm phổi mắc phải cộng đồng (VPMPCĐ) là bệnh rất phổ biến trên toàn thế giới, có tỷ lệ mắc và tử vong cao đặc biệt ở các nước đang phát triển. Theo công bố của Tổ chức Y tế Thế giới (TCYTG) tại Sudan năm 2004 thì mỗi năm có 150,7 triệu trẻ mắc viêm phổi, trong đó có 11 - 20 triệu (7 - 13%) trẻ bị viêm phổi nặng cần phải nhập viện. Ước tính trên thế giới mỗi năm một đứa trẻ mắc viêm phổi 0,28 lần và 95% là trẻ em các nước đang phát triển. Tỷ lệ mới mắc viêm phổi ở trẻ < 5 tuổi hàng năm là 3% ở các nước đã phát triển và 7 - 18% ở các nước đang phát triển. Viêm phổi là nguyên nhân thứ 2 gây tử vong cho trẻ < 5 tuổi (18%) theo TCYTG năm 2010 [1].

Ở Việt Nam, viêm phổi chiếm khoảng 33% trong tổng số tử vong ở trẻ nhỏ do mọi nguyên nhân. Khoảng 2,8/1000 trẻ chết là do viêm phổi và với 7 triệu trẻ dưới 5 tuổi trong cả nước thì ước tính số chết do viêm phổi không dưới 20.000/ năm. Theo TCYTG năm 2004 thì ở Việt Nam trẻ em chết do viêm phổi là 4000 trẻ, chiếm 12% trong tổng số trẻ em chết < 5 tuổi [2].

Các kỹ thuật học sâu (Deep Learning) đã được nghiên cứu rộng rãi và thử nghiệm chứng minh đạt hiệu quả rất cao trong các bài toán phân lớp, nhận dạng ảnh. Các kiến trúc mạng học sâu dựa trên mạng nơ-ron tích chập (CNN - Convolutional Neural Network) đã liên tục được nghiên cứu cải tiến đem đến kết quả thử nghiệm ngày càng ấn tượng như VGG16, ResNet50, Inception. Trong bài báo này, tác giả thực hiện thử nghiệm với kiến trúc mạng VGG-19 (Một trong những kiến trúc tiên tiến nhất hiện nay) qua chẩn đoán hình ảnh X-quang lồng ngực. Các kết quả thử nghiệm được so sánh đánh giá để đề xuất một kiến trúc mạng học sâu phù hợp có chất lượng tốt nhất làm tiền đề cho việc xây dựng ứng dụng hỗ trợ chẩn đoán bệnh viêm phổi thông qua ảnh chụp X-quang.

2. HÌNH ẢNH X-QUANG PHỔI VÀ VIÊM PHỔI

2.1. X-quang phổi

X-quang là một loại kiểm tra rất phổ biến trong thăm khám sức khỏe thường quy và cũng là một trong những xét nghiệm đầu tiên mà bạn sẽ trải qua nếu bác sĩ nghi ngờ bạn bị bệnh lý về tim hoặc phổi. Đồng thời, một phim X-quang cũng có thể được sử dụng để kiểm tra xem bạn đang đáp ứng với điều trị như thế nào.

Chính vì thế, nhờ có phim X-quang phổi, bác sĩ sẽ biết được các vấn đề về phổi, bao gồm cả bệnh lý, bên trong cơ thể bạn, như là: Tình trạng của phổi: phát hiện khối u (lành tính hoặc ung thư), nhiễm trùng như bệnh viêm phổi hoặc bất thường trong khoang màng phổi như tràn dịch, tràn khí màng phổi. X-quang cũng có thể cho thấy tình trạng bệnh phổi mãn tính, chẳng hạn như khí phế thũng hoặc xơ nang, cũng như các biến chứng liên quan đến các tình trạng này. Ví dụ một phim X-quang viêm phổi cho thấy những đám mờ rải rác hay khu trú trong nhu mô phổi.

2.2. Đặc điểm chung của tổn thương phổi

Tổn thương X-quang phổi phụ thuộc vào một số yếu tố liên quan người bệnh như tuổi và tình trạng đáp ứng miễn dịch. Các yếu tố nguy cơ liên quan đến tần suất và mức độ tổn thương nặng bao gồm: người già, trẻ em, suy dinh dưỡng và suy giảm miễn dịch.

Hình ảnh X-quang có giá trị định hướng chẩn đoán bệnh và đánh giá mức độ nặng (dựa vào diện tổn thương), đánh giá các biến chứng hô hấp (tràn dịch, khí màng phổi), theo dõi đáp ứng điều trị và chẩn đoán phân biệt (viêm phổi do các nguyên nhân khác). Chỉ bằng hình ảnh X-quang, rất khó để chẩn đoán phân biệt giữa viêm phổi do virus với một số căn nguyên vi sinh khác (vi khuẩn không điển hình, nấm,...), do vậy khi phân tích hình ảnh X-quang dùng để chẩn đoán bệnh viêm phổi trong bài nghiên cứu này, các tác giả chỉ tập trung vào chẩn đoán hình ảnh viêm phổi do virus, không chẩn đoán căn nguyên của bệnh viêm phổi là do virus hay các căn nguyên vi sinh khác [3].

Tổn thương X-quang ngực của viêm phổi do virus có thể gồm các loại tổn thương:

Tổn thương tổ chức kẽ: hình ảnh lưới mờ, kính mờ (Ground-glass opacity GGO), đường kerley (phù tổ chức kẽ).

Nốt mờ, bóng mờ khu trú hay lan tỏa.

Đồng đặc nhu mô khu trú hay lan tỏa.

Đặc điểm chung của tổn thương X-quang của viêm phổi virus.

Tổn thương phụ thuộc vào giai đoạn bệnh: giai đoạn sớm hình ảnh X-quang có thể bình thường.

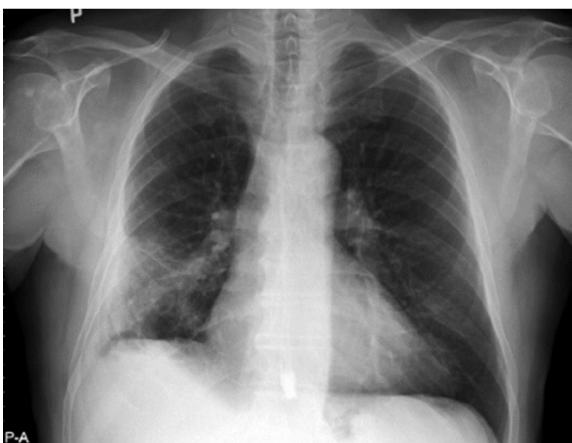
Tổn thương chủ yếu gồm tổn thương cả tổ chức kẽ kết hợp với đồng đặc nhu mô phổi.

Tổn thương thường lan tỏa, hay ở ngoại vi, thùy dưới hai bên, ít có phá hủy.

Tổn thương thường tiến triển nhanh, khi khỏi có thể để lại xơ phổi hai bên.

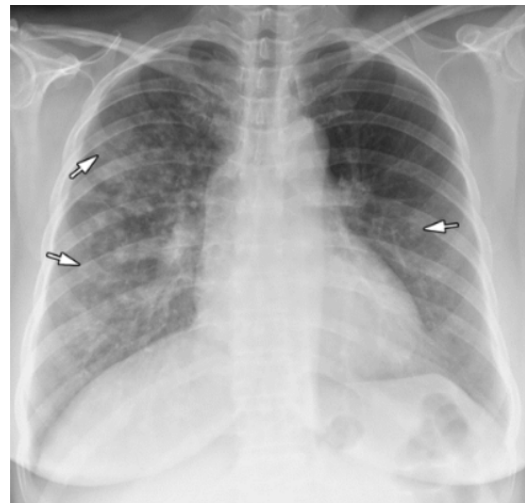
Một số hình ảnh tổn thương của viêm phổi virus [4]:

Hình ảnh kính mờ: là tổn thương đồng đặc không hoàn toàn, có tỷ trọng cao hơn nhu mô phổi xung quanh vẫn có thể thấy đường bờ các mạch máu hoặc phế quản bên trong tổn thương đó.



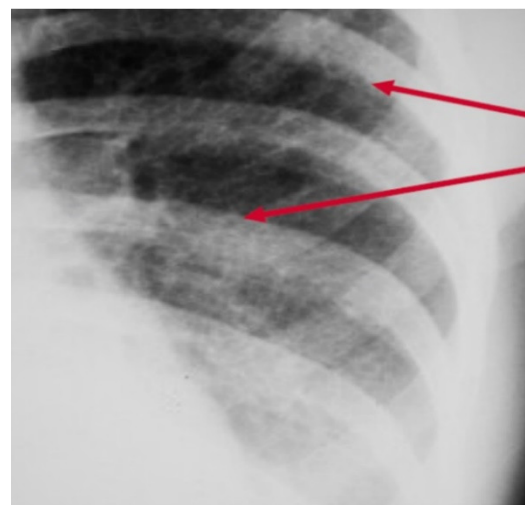
Hình 1. Hình mờ nhạt nền phổi phải (hình ảnh kính mờ)

Hình ảnh nốt mờ: là những hình mờ có đường kính dưới 3cm, dạng hình tròn, có thể đơn độc, có thể rải rác trong nhu mô phổi. Các nốt phổi thường có ranh giới rõ, được bao quanh bởi nhu mô phổi và không liên tục với rốn phổi hay trung thất.

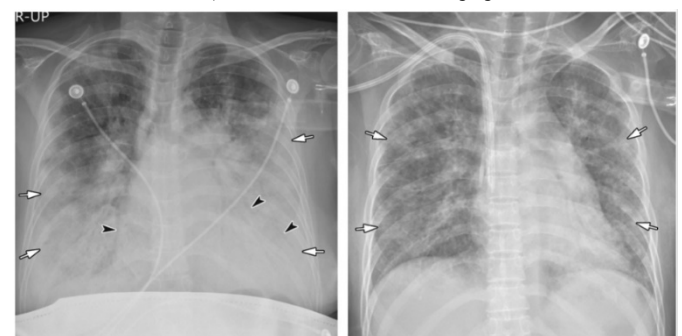


Hình 2. Hình nốt mờ trên X-quang phổi thẳng (mũi tên)

Hình ảnh dày thành phế quản: là những tổn thương thể hiện thành của phế quản dày lên, do có sự tích tụ dịch hay chất nhầy xung quanh thành phế quản, trong mô kẽ.



Hình 3. Hình ảnh dày thành phế quản trên X-quang ngực



Hình 4. Hình ảnh ARDS (bên trái) và sau khi điều trị

Hình ảnh hội chứng suy hô hấp cấp tính (ARDS): Dấu hiệu X-quang của ARDS thay đổi theo giai đoạn của bệnh. Dấu

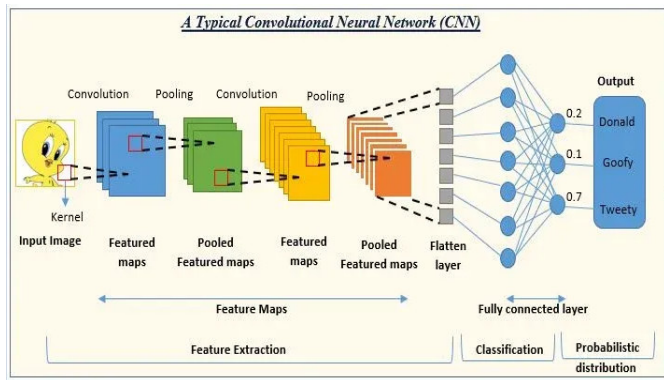
hiệu phổ biến nhất đó là tổn thương đồng đặc cả hai bên phổi, không cân xứng, chủ yếu ở ngoại vi của phổi.

3. PHÂN LỚP DỰA TRÊN MÔ HÌNH MẠNG HỌC SÂU VGG-19

3.1. Mạng nơ-ron tích chập (Convolutional neural network) [5]

Mạng nơ-ron tích chập (CNN) là sự kết hợp giữa sinh học, toán học và khoa học máy tính, nhưng những mạng này mạng là những đổi mới có ảnh hưởng nhất trong lĩnh vực thị giác máy tính và trí tuệ nhân tạo (AI). 2012 là năm đầu tiên CNN trở nên nổi tiếng khi Alex Krizhevsky sử dụng CNN 8 lớp (5 chuyển đổi, 3 kết nối đầy đủ) để giành chiến thắng trong cuộc thi ImageNet năm đó (sau này gọi là AlexNet), loại bỏ bản ghi lỗi phân loại khỏi 25,8% (năm 2011) lên 16,4% (năm 2012), một sự cải thiện đáng kinh ngạc vào thời điểm đó. Kể từ đó nhiều công ty đã sử dụng deep learning làm cốt lõi cho dịch vụ của họ. Ví dụ: Facebook sử dụng mạng lưới thần kinh để gắn thẻ tự động thuật toán, Google cho tìm kiếm ảnh của họ, Amazon cho các đề xuất sản phẩm của họ, Pinterest cho nguồn cấp dữ liệu trang chủ của họ cá nhân hóa và Instagram cho cơ sở hạ tầng tìm kiếm của họ.

Mạng CNN là một tập hợp các lớp tích chập chồng lên nhau và sử dụng các hàm nonlinear activation như ReLU và tanh để kích hoạt các trọng số trong các node. Mỗi một lớp sau khi thông qua các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo.



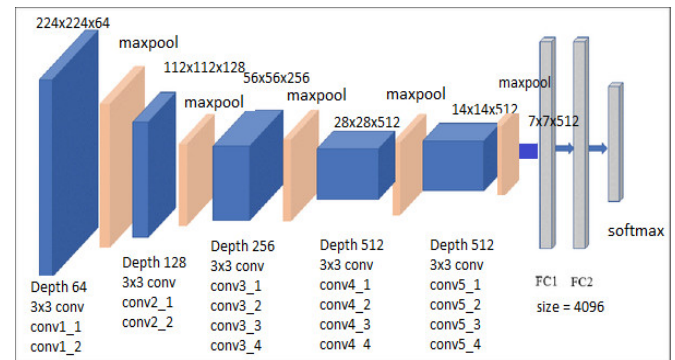
Hình 5. Mô hình mạng CNN

Mỗi một lớp sau khi thông qua các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo. Trong mô hình mạng truyền ngược (feedforward neural network) thì mỗi neural đầu vào (input node) cho mỗi neural đầu ra trong các lớp tiếp theo. Mô hình này gọi là mạng kết nối đầy đủ (fully connected layer). Còn trong mô hình CNNs thì ngược lại. Các lớp liên kết được với nhau thông qua cơ chế tích chập. Lớp tiếp theo là kết quả phép tính tích chập từ lớp trước đó, nhờ vậy mà ta có được các kết nối cục bộ. Như vậy mỗi neuron ở lớp kế tiếp sinh ra từ kết quả của filter áp đặt lên một vùng ảnh cục bộ của neuron trước đó. Mỗi một lớp được sử dụng các filter khác nhau thông thường có hàng trăm hàng nghìn filter như vậy và kết hợp kết quả của chúng lại. Ngoài ra có một số lớp khác như pooling/subsampling lớp dùng để chắt lọc lại các thông tin hữu ích hơn. Trong quá trình huấn luyện mạng (training) CNN tự động học các giá trị

qua các lớp filter. Ví dụ trong tác vụ phân lớp ảnh, CNNs sẽ cố gắng tìm ra thông số tối ưu cho các filter tương ứng theo thứ tự raw pixel > edges > shapes > facial > high-level features. Lớp cuối cùng được dùng để phân lớp ảnh [6].

3.2. Mô hình VGG-19 [7]

Simonyan và Zisserman của Đại học Oxford đã tạo ra một CNN 19 lớp (16 tích chập, 3 kết nối đầy đủ) đã sử dụng bộ lọc 3x3 với với mỗi giá trị trên ma trận lọc bằng 1, cùng với các maxpool kích thước 2x2 với số bước bằng 2, được gọi là mô hình VGG-19. So với AlexNet, VGG-19 (hình 6) là CNN sâu hơn với nhiều lớp hơn. Để giảm số lượng tham số trong các mạng sâu như vậy, nó sử dụng các bộ lọc tích chập kích thước nhỏ 3x3 trong tất cả các lớp chập và kết quả huấn luyện tốt nhất với tỷ lệ lỗi 7,3%.



Hình 6. Minh họa kiến trúc mô hình mạng VGG-19

3.3. Dữ liệu huấn luyện

Bộ dữ liệu huấn luyện mạng bao gồm các hình ảnh OCT và X-quang ngực đã được xác thực được mô tả và phân tích trong "Phân loại dựa trên học sâu và giới thiệu các bệnh có thể điều trị được ở người". Hình ảnh OCT được chia thành tập huấn luyện và tập kiểm tra các bệnh nhân độc lập. Hình ảnh OCT được gắn nhãn là (bệnh) - (ID bệnh nhân ngẫu nhiên) - (số hình ảnh của bệnh nhân này) và các hình ảnh gắn nhãn bình thường đối với hình ảnh không mắc bệnh [8].

Hình ảnh X-quang ngực (trước-sau) được chọn từ tập ảnh chụp X-quang của bệnh nhân nhi từ một đến năm tuổi từ Trung tâm Y tế Phụ nữ và Trẻ em Quảng Châu, Quảng Châu. Tất cả hình ảnh X-quang ngực được thực hiện như một phần của chăm sóc lâm sàng thông thường cho bệnh nhân.

Để phân tích hình ảnh X-quang ngực, tất cả các phim X-quang ngực ban đầu được sàng lọc để kiểm soát chất lượng bằng cách loại bỏ tất cả các bản quét chất lượng thấp hoặc không thể đọc được. Các chẩn đoán cho hình ảnh sau đó được hai bác sĩ chuyên môn phân loại trước khi được phép đào tạo hệ thống AI. Để giải quyết bất kỳ lỗi ảnh nào, bộ đánh giá cũng được chuyên gia thứ ba kiểm tra.

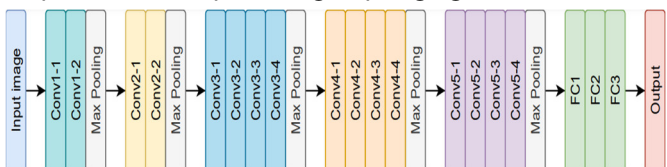
Bộ dữ liệu bao gồm 5.863 hình ảnh X-Ray (định dạng JPEG) trên hai loại (Viêm phổi/Bình thường). Tập dữ liệu được tổ chức thành ba thư mục: Train, Test và Validation. Mỗi thư mục chứa các thư mục con cho từng danh mục hình ảnh, cụ thể là Viêm phổi và Bình thường.

Bảng 1. Phân chia dữ liệu huấn luyện

Tập dữ liệu	Viêm phổi	Bình thường	Tổng số mẫu
X-Ray	4390	1473	5863
Train	3512	1178	4690
Test	868	285	1153
Validation	10	10	20

3.4. Kiến trúc mô hình VGG-19 trong bài toán phát hiện viêm phổi bằng X-quang ngực

Mô hình VGG19 được sử dụng trong bài toán huấn luyện để phát hiện viêm phổi bằng X-quang ngực:



Hình 7. Mô hình VGG-19 sử dụng trong bài toán chẩn đoán hình ảnh

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv4 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv4 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv4 (Conv2D)	(None, 14, 14, 512)	2359808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
Total params: 20.024.384		
Trainable params: 0		
Non-trainable params: 20.024.384		

Hình 8. Kiến trúc mô hình mạng VGG-19

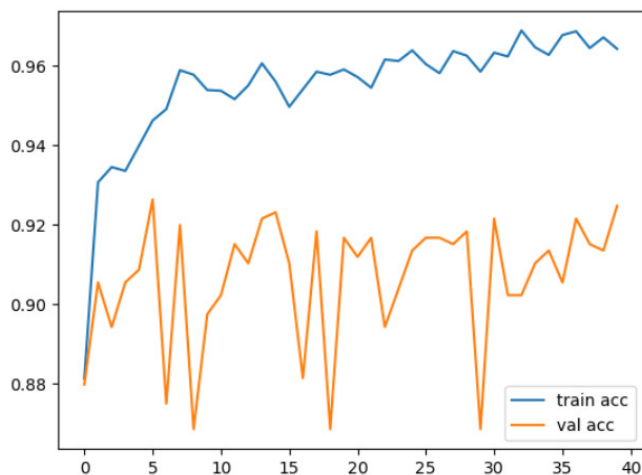
Mô hình mạng CNN dùng mô hình VGG-19 được sử dụng trong bài toán dự đoán viêm phổi do virus:

Layer (type)	Output Shape	Param #
vgg19 (Functional)	(None, 7, 7, 512)	20024384
flatten (Flatten)	(None, 25088)	0
dropout (Dropout)	(None, 25088)	0
dense (Dense)	(None, 16)	401424
dropout_1 (Dropout)	(None, 16)	0
dense_1 (Dense)	(None, 1)	17
Total params: 20.024.384		
Trainable params: 401.441		
Non-trainable params: 20.024.384		

Hình 9. Mạng CNN sử dụng VGG-19

4. KẾT QUẢ

Bài toán phát hiện viêm phổi bằng X-quang ngực sử dụng mô hình VGG-19 đưa ra kết quả dự đoán với độ chính xác khá cao accuracy = 0,96.



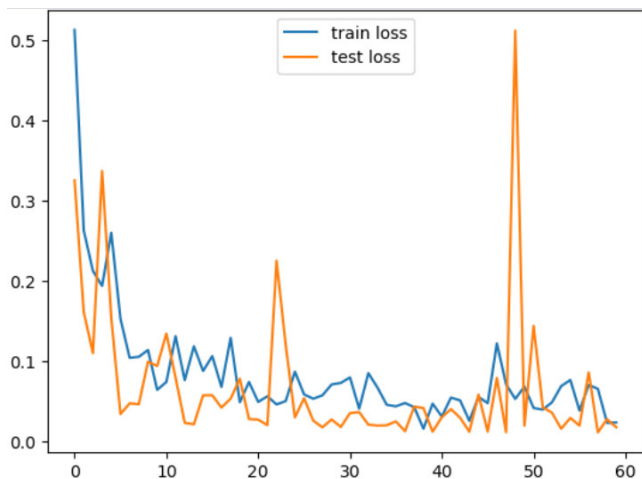
Hình 10. Kết quả độ chính xác của mô hình dự đoán

So sánh kết quả so với các mô hình deep learning khác trong bài toán chẩn đoán viêm phổi bằng X-quang ngực:

Bảng 2. Kết quả so sánh độ chính xác giữa các mô hình học sâu

Model	Accuracy
ResNet50	0,97
CNN	0,73
VGG-19	0,96

Các số liệu ở bảng 2 và hình 10 chỉ ra rằng kết quả mô hình phân lớp nhận dạng sử dụng kiến trúc mạng VGG19 cho kết quả độ chính xác cao nhất (~97%) so với các mô hình sử dụng mạng CNN, ResNet,...



Hình 11. Hàm Loss trong huấn luyện VGG-19

Về đánh giá kết quả lỗi của mô hình sử dụng kiến trúc VGG19 trên tập huấn luyện và tập xác thực trong hình 10, 11, chúng ta thấy các hàm đo độ chính xác và độ mất mát có độ ổn định. Điều này cho thấy khả năng tổng quát hóa của mô hình khá tốt, giảm được hiện tượng quá khớp (overfitting).

Các kết quả ấn tượng của mô hình học sâu sử dụng kiến trúc VGG19 trên đây đem đến tiềm năng cho việc xây dựng một ứng dụng hỗ trợ chẩn đoán bệnh viêm phổi thông qua ảnh chụp X-quang. Điều này đem lại hiệu quả cho việc rút ngắn thời gian khám bệnh, bệnh nhân được chữa trị kịp thời hơn, giảm nguy cơ tăng nặng.

5. KẾT LUẬN

Dựa trên kết quả thực nghiệm chẩn đoán viêm phổi virus bằng kỹ thuật học sâu sử dụng mô hình VGG-19, đây là một trong những mô hình tiên tiến hiện nay được sử dụng rộng rãi trong các bài toán thị giác máy tính. Tuy kết quả độ chính xác chưa vượt qua được mô hình ResNet50, nhưng kỹ thuật trong mô hình VGG-19 này lại là một trong những mô hình dễ hiểu và dễ triển khai nhất, nó khá quen thuộc với những bài toán sử dụng với các mô hình trong CNN.

TÀI LIỆU THAM KHẢO

- [1]. Philip Keith Pattemore, Lance C. Jennings, 2009. *Epidemiology of Respiratory Infections*. Pediatric Respiratory Medicine. 2008: 435–452. Published online 2009 May 18. doi: 10.1016/B978-032304048-8.50035-9.
- [2]. Nguyen Thi Ha, Doan Mai Thanh, Nguyen Thi Yen, 2020. *Real-time multiplex pcr technique in diagnosis the cause of community-acquired pneumonia in children*. Journal of Medical Research, Hanoi Medical University, Vol. 131 - No. 7.
- [3]. Tomas Franquet MD, 2011. *Imaging of Pulmonary Viral Pneumonia*. Radiology: Volume 260: Number 1.
- [4]. Do Quyet, Pham Minh Thong, Ta Ba Thang, Pham Ngoc Thach, Mai Xuan Khan, Nguyen Minh Hai, Dao Ngoc Bang, Nguyen Tuan Anh, 2019. *X-ray imaging characteristics of pneumonia caused by Corona virus*. Scientific conference of the Military and Police Lung Association.
- [5]. <https://www.mathworks.com/discovery/convolutional-neural-network.html>.
- [6]. <https://adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks/>
- [7]. Karen Simonyan, Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. <https://doi.org/10.48550/arXiv.1409.1556>
- [8]. Kermany Daniel, Zhang Kang, Goldbaum Michael, 2018. *Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification*. Mendeley Data, V2, doi: 10.17632/rscbjbr9sj.2
- [9]. Abien Fred, M. Agarap, 2018. *Deep Learning using Rectified Linear Units (ReLU)*. arXiv:1803.08375.
- [10]. Karen Simonyan, Andrew Zisserman, 2014. *Very Deep Convolutional Network for Large Scale Image Recognition*. arXiv:1409.1556.

AUTHOR INFORMATION

Tran Thanh Hung

Faculty of Information Technology, Hanoi University of Industry, Vietnam