

ỨNG DỤNG KỸ THUẬT HỌC SÂU TRONG HỖ TRỢ CHẨN ĐOÁN BỆNH VIÊM PHỔI THÔNG QUA ẢNH CHỤP X-QUANG

Võ Đức Quang ^{(1), (5)}, Nguyễn Hải Yến ⁽²⁾,
Mai Hồng Mận ⁽³⁾, Nguyễn Thị Nhã ⁽⁴⁾

¹ Viện Kỹ thuật và Công nghệ, Trường Đại học Vinh

² Lớp 58K2-CNTT, Viện Kỹ thuật và Công nghệ, Trường Đại học Vinh

³ Lớp 58K4-CNTT, Viện Kỹ thuật và Công nghệ, Trường Đại học Vinh

⁴ Trung tâm Thực hành Thí nghiệm, Trường Đại học Vinh

⁵ Nghiên cứu sinh, Trường Đại học Bách Khoa Hà Nội

Ngày nhận bài 07/10/2021, ngày nhận đăng 10/12/2021

Tóm tắt: Trong khung cảnh dịch bệnh Covid-19 đang bùng phát tại Việt Nam và trên toàn thế giới, một ứng dụng học máy hỗ trợ chẩn đoán bệnh viêm phổi có độ chính xác cao sẽ giúp tiết kiệm chi phí về thời gian, nhân lực cho ngành y tế, giúp các bệnh nhân được chữa trị kịp thời, giảm nguy cơ tăng nặng. Bài báo này trình bày về đặc trưng của các kiến trúc mạng học sâu hiện đại dựa trên mạng nơ-ron tích chập như ResNet50, VGG16, Inception, DenseNet và thử nghiệm đánh giá các mô hình này trong bài toán chẩn đoán viêm phổi sử dụng bộ dữ liệu Chest-Xray. Kết quả thử nghiệm chỉ ra rằng mô hình học sâu sử dụng kiến trúc mạng học sâu VGG16 cho tỷ lệ chính xác cao nhất. Đây là cơ sở để đề xuất xây dựng ứng dụng hỗ trợ chẩn đoán viêm phổi dựa trên ảnh chụp X-quang hiệu quả.

Từ khóa: Học sâu; CNN; VGG16; Covid19; X-quang lồng ngực.

1. Mở đầu

Hiện nay, dịch bệnh COVID-19 đang bùng phát lan rộng, cướp đi sinh mạng của hơn 5 triệu người ở các quốc gia trên thế giới. Cơ chế bệnh lý của virus này thường gây ra triệu chứng viêm phổi, suy hô hấp. Đặc biệt, bệnh tiến triển chuyển nặng diễn ra rất nhanh, dẫn đến nguy cơ tử vong rất cao, đặc biệt ở những người có bệnh nền hoặc người cao tuổi. Việc phát hiện sớm và chính xác bệnh viêm phổi có ý nghĩa rất quan trọng, giúp bệnh nhân được điều trị kịp thời, giảm bớt nguy cơ tử vong. Trong thực tế khám chữa bệnh, các bác sĩ thực hiện nghiệp vụ đọc và phân tích các ảnh chụp X-quang lồng ngực để chẩn đoán chính xác tình trạng viêm phổi. Khi số lượng bệnh nhân nhiều, điều này tạo nên áp lực về thời gian và hiệu suất công việc. Với những tiến bộ công nghệ, đặc biệt là trong lĩnh vực trí tuệ nhân tạo, chúng ta hoàn toàn có thể tạo ra các ứng dụng dựa trên học máy giúp tự động phát hiện viêm phổi dựa trên hình ảnh X-quang lồng ngực đạt độ chính xác cao. Một hệ thống ứng dụng như thế sẽ giúp giảm bớt thời gian cho các bác sĩ, chuyên gia y tế trong đây nhanh tiến trình khám bệnh, giúp người bệnh được chữa trị kịp thời, giảm bớt nguy cơ tăng nặng.

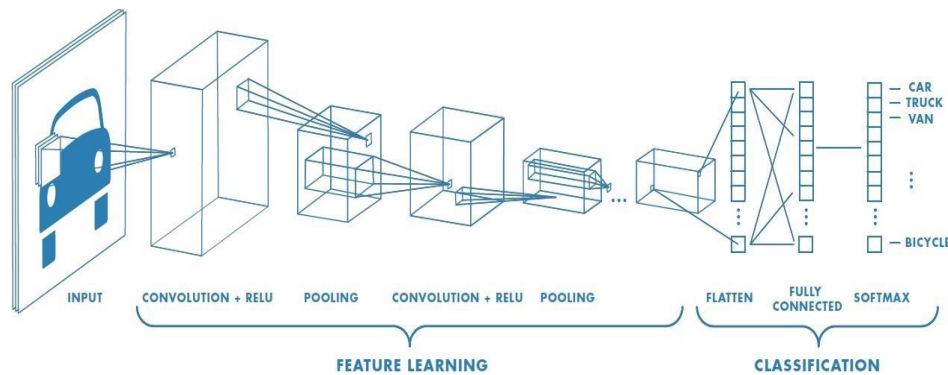
Các kỹ thuật học sâu (deep learning) đã được nghiên cứu rộng rãi và thử nghiệm chứng minh đạt hiệu quả rất cao trong các bài toán phân lớp, nhận dạng ảnh. Các kiến trúc mạng học sâu dựa trên mạng nơ-ron tích chập (CNN - convolutional neural network) đã liên tục được nghiên cứu cải tiến đem đến kết quả thử nghiệm ngày càng ấn tượng như VGG16 [1], ResNet50 [2], Inception [7]. Trong bài báo này, chúng tôi thực hiện thử nghiệm các giải thuật học sâu hiện đại trong phân tích chẩn đoán bệnh viêm phổi thông

qua 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. Trong phần tiếp theo, bài báo sẽ giới thiệu và trình bày các kiến trúc mạng học sâu hiện đại dựa trên CNN; Mô hình học sâu hỗ trợ chẩn đoán viêm phổi được đề xuất ở Phần 3, cuối cùng là các thử nghiệm và kết luận ở Phần 4.

2. Mô hình kiến trúc Deep Learning

2.1. Mạng nơ-ron tích chập (CNN-Convolutional Neural Network)

CNN là một loại mạng nơ-ron bao gồm nhiều lớp nơ-ron nhân tạo được kết nối với nhau giúp mô phỏng gần như hệ thống thị giác của con người trong việc phân biệt các hình ảnh. Về kiến trúc, CNN có cấu tạo gồm 3 tầng: Tầng tích chập (convolutional layer), Tầng gộp (pooling layer) và Tầng kết nối đầy đủ (fully connected layer) được mô tả như Hình 1.



Hình 1: Kiến trúc CNN

Một mô hình nhận dạng/phân loại ảnh sử dụng kiến trúc CNN được thực hiện trong 3 giai đoạn chính như sau:

- Tích chập (convolution): là quá trình tính toán trong đó sử dụng nhiều bộ lọc đặc trưng trượt trên các ma trận điểm ảnh để trích xuất các đặc trưng tương ứng của dữ liệu ảnh. Với mỗi tầng, giá trị tích chập cho một điểm ảnh đầu ra được tính theo công thức (1), trong đó $Y[m, n]$ là giá trị đầu ra cho tầng tiếp theo, m và n là chỉ số của hàng và cột của ma trận đầu ra, f biểu thị cho ma trận điểm ảnh đầu vào, ma trận lọc được biểu diễn là h và $*$ là phép tích chập.

$$Y[m, n] = (f * h)[m, n] = \sum_j \sum_k h[j, k] f[m - j, n - k]. \quad (1)$$

Sau mỗi quá trình tích chập, ta thu được ma trận các giá trị đặc trưng. Sau đó, cần sử dụng một hàm kích hoạt (activation function) để phi tuyến hóa các giá trị đó. Trong hầu hết các nghiên cứu trước đây, **sigmoid** và **tanh** là những hàm kích hoạt phi tuyến được sử dụng phổ biến nhất. Tuy nhiên, gần đây ReLU (rectified linear unit) [9] đã được sử dụng rộng rãi và phổ biến hơn. ReLU có tốc độ tính toán nhanh hơn nhờ đạo hàm chỉ có 2 giá trị $\{0, 1\}$ và không có lũy thừa cơ số e như hàm sigmoid. Công thức hàm ReLU được tính theo (2) và (3).

$$ReLU(x) = \max(0, x) \quad (2)$$

với điều kiện

$$\frac{d}{dx} ReLU(x) = \{1 \text{ if } x > 0; 0 \text{ otherwise}\}. \quad (3)$$

- Tổng hợp (pooling): còn được gọi là quá trình lấy mẫu con (subsampling) được sử dụng để giảm số chiều của bản đồ đặc trưng thu được từ phép toán tích chập. Max pooling và Average pooling là hai phép toán tổng hợp phổ biến nhất được sử dụng trong CNN.

- Kết nối đầy đủ (fully connected): là quá trình làm phẳng (flatten) ma trận đặc trưng thành một vector đặc trưng kết hợp việc sử dụng các kết nối đầy đủ giữa các tầng. Tầng kết nối đầy đủ cuối cùng sẽ có số lượng đơn vị bằng với số nhân ảnh và áp dụng hàm kích hoạt là Softmax. Đây là một hàm logistic để quyết định nhãn lớp dựa trên các giá trị thuộc vector đặc trưng. Đầu vào của hàm là một vector chứa các giá trị $x \in R^n$ và đầu ra là một vector gồm các xác suất $p \in R^n$. Nó được định nghĩa theo công thức (4):

$$p = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \text{ với } p_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}. \quad (4)$$

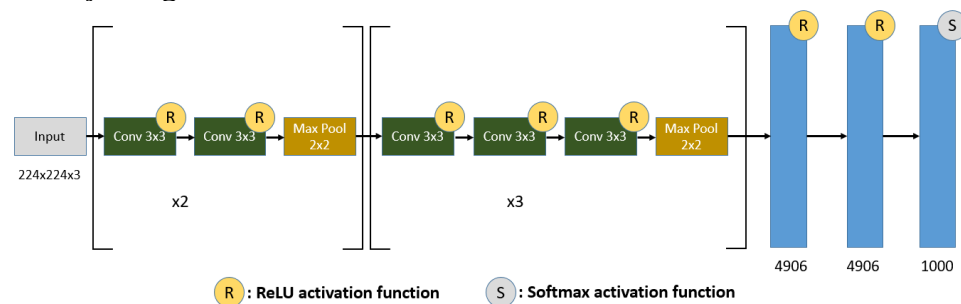
Mô hình phân lớp sử dụng kiến trúc CNN có thể được coi là sự kết hợp của hai giai đoạn: trích xuất đặc trưng và phân loại. Các tầng tích chập và tầng gộp thực hiện chức năng trích xuất các đặc trưng của ảnh. Sau đó, các tầng được kết nối đầy đủ hoạt động như một bộ phân loại trên các đặc trưng này và chỉ định xác suất cho hình ảnh đầu vào thuộc nhãn nào. Đặc điểm có lợi nhất của CNN là giảm số lượng tham số trong mạng nơ-ron nhân tạo thông thường. Một khía cạnh quan trọng khác của CNN là độ phức tạp phát hiện hình ảnh được tăng dần, càng ở những tầng tích chập về sau càng có khả năng phát hiện các đường nét phức tạp, đã rõ ràng hình thù và thậm chí là cấu thành vật thể, đây được gọi là những đặc trưng bậc cao (high level) của dữ liệu ảnh. Các lớp được kết nối đầy đủ sẽ học cách sử dụng kết hợp những đặc trưng này để phân loại hình ảnh một cách chính xác.

Với sự phát triển không ngừng của lĩnh vực thị giác máy tính, dựa trên kiến trúc nền tảng là CNN, các nhà nghiên cứu về mạng học sâu trên thế giới đã khai phá ra nhiều biến thể của CNN mang lại độ chính xác và hiệu năng được cải thiện rõ rệt. Sau đây, bài báo sẽ trình bày một số kiến trúc CNN hiện đại kèm theo những đặc điểm nổi bật của chúng.

2.2. Các kiến trúc CNN hiện đại

2.2.1. VGG-16

Kiến trúc mô hình VGG-16 được Karen Simonyan và Andrew Zisserman giới thiệu vào năm 2014 trong [1]. So với kiến trúc CNN cổ điển, VGG-16 đã được cải tiến về độ sâu lên đến 16 tầng, trong đó gồm 13 tầng tích chập, 3 tầng được kết nối đầy đủ (với tầng cuối sử dụng hàm kích hoạt Softmax để phân loại). Mô tả kiến trúc các tầng của VGG-16 được trình bày trong Hình 2.



Hình 2: Kiến trúc các tầng trong VGG16

Kiến trúc VGG-16 gộp các tầng tích chập thành các modun gọi là khối tích chập (block). Trong mỗi khối, VGG-16 kết hợp 2 hoặc 3 tầng tích chập, theo sau là tầng gộp cực đại (max pooling) để giảm kích thước bản đồ đặc trưng. Bằng việc sử dụng các block dạng [Conv2D*n + Max Pooling], các mô hình mạng VGG-16 trở nên sâu hơn. Đồng thời, VGG-16 chỉ sử dụng các bộ lọc kích thước nhỏ 3x3 (với tầng tích chập) và 2x2 (với tầng gộp) giúp giảm số lượng tham số cho mô hình dẫn đến giảm khối lượng tính toán phải thực hiện. Các thử nghiệm trên tập dữ liệu xác thực ImageNet trên Keras framework cho thấy, độ chính xác của mô hình VGG-16 thuộc Top-1 và Top-5 có hiệu quả cao (với 71,3% và 90,01%). VGG-16 cho thấy độ sâu của mạng là một thành phần quan trọng để mô hình đạt được hiệu suất tốt.

2.2.2. GoogleNet-Inception-V3

Kiến trúc mô hình GoogleNet-Inception-V1 được Christian Szegedy và cộng sự giới thiệu vào năm 2014 trong bài báo [7]. Kiến trúc GoogleNet hướng đến việc cố gắng đạt được độ chính xác cao với chi phí tính toán giảm. Trước đây, các kiến trúc CNN chọn từng kích thước bộ lọc từ nhỏ như 1x1 tới lớn như 11x11. Trong kiến trúc này, GoogleNet-Inception sử dụng kết hợp các bộ lọc tích chập khác nhau gọi là khối Inception. Trong khi mô hình kiến trúc VGG-16 có 16 tầng với 138 triệu tham số thì GoogleNet-Inception-V1 chỉ dùng 5 triệu tham số với 22 tầng. Bằng việc kết hợp những điểm mạnh của mô hình Network in Network [3] và mô hình chứa các khối Inception lặp lại, GoogleNet-Inception-V1 đã giải quyết được vấn đề tồn kém chi phí tính toán trong khi độ sâu của mạng tăng lên. Một kiến trúc cải tiến đáng chú ý của GoogleNet-Inception là GoogleNet-Inception-V3 được đề xuất vào năm 2015 trong bài báo [4]. Kiến trúc này sử dụng 24 triệu tham số, trong đó sau các tầng tích chập của Inception-V3 được bổ sung thêm một tầng Batch normalization (BN) [5] và một ReLU activation. Tại mỗi tầng, kỹ thuật chuẩn hóa đầu vào BN được sử dụng trong từng minibatch theo phân phối chuẩn hóa $N(0,1)$, giúp cho quá trình huấn luyện thuật toán nhanh hơn. BN chuyển đổi những giá trị kích hoạt tại mỗi tầng (x) theo công thức sau:

$$BN(x) = \gamma \odot \frac{x - \hat{\mu}}{\hat{\sigma}} + \beta. \quad (5)$$

Ở đây, $\hat{\mu}$ là giá trị trung bình, $\hat{\sigma}$ là độ lệch chuẩn của các mẫu trong minibatch, γ là hệ số tỷ lệ, β là độ chệch. Ký hiệu một minibatch là $\mathcal{B}$, $\epsilon > 0$ là hằng số rất nhỏ, ta tính $\hat{\mu}_{\mathcal{B}}$ và $\hat{\sigma}_{\mathcal{B}}$ theo công thức sau:

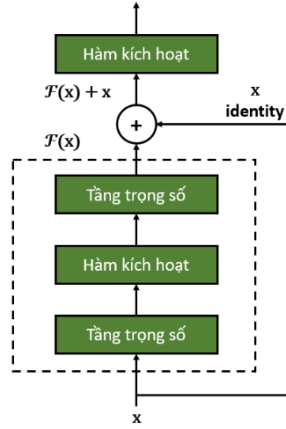
$$\hat{\mu}_{\mathcal{B}} \leftarrow \frac{1}{|\mathcal{B}|} \sum_{x \in \mathcal{B}} x \text{ và } \hat{\sigma}_{\mathcal{B}}^2 \leftarrow \frac{1}{|\mathcal{B}|} \sum_{x \in \mathcal{B}} (x - \hat{\mu}_{\mathcal{B}})^2 + \epsilon. \quad (6)$$

Inception-V3 sử dụng 3 kiểu Inception module: Inception-A, Inception-B và Inception-C và 2 kiến trúc giảm chiều dữ liệu là Reduction-A và Reduction-B. Với việc kích thước của các tầng không bị giảm một cách đột ngột, Inception-V3 giải quyết được vấn đề thắt cổ chai (representational bottlenecks). Thêm vào đó, Inception-V3 sử dụng phương pháp nhân tố (factorisation methods) giúp việc tính toán hiệu quả hơn. Độ chính xác Top-1 và Top-5 của mô hình này trên tập dữ liệu xác thực ImageNet lần lượt là 77,9% và 93,7%.

2.2.3. ResNet-50 (Residual Nets)

Khi thiết kế các mạng ngày càng sâu thì độ phức tạp ở các tầng sẽ tăng lên, khả năng biểu diễn mạng cũng trở nên khó khăn hơn và dễ xảy ra hiện tượng *Vanishing*

Gradient [8]. Thêm vào đó, các đặc trưng gốc ban đầu cũng dễ bị mất mát khi đi qua quá nhiều tầng. Để giải quyết vấn đề này, kiến trúc mô hình ResNet được Kaiming He và cộng sự giới thiệu vào năm 2015 trong bài báo [2]. ResNet sử dụng cấu trúc gồm các khối phần dư (residual blocks) chứa các kết nối tắt (shortcut connections) và áp dụng Batch Normalization sau mỗi tầng tích chập. Kết nối tắt ở đây là bỏ qua việc đào tạo một số tầng trung gian. Điều này giúp dễ dàng hơn trong việc huấn luyện mạng và giữ được các đặc trưng ban đầu khi mạng có độ sâu lớn. Khối phần dư định danh được mô tả trong Hình 3.



Hình 3: Khối phần dư định danh

Về tính toán, các khối phần dư định danh ResNet được thực hiện như sau:

$$y = \mathcal{F}(x, \{w_i\}) + x. \quad (7)$$

Trong đó x và y là các vector đầu vào và đầu ra của các lớp được xem xét; hàm $\mathcal{F}(x, \{w_i\})$ biểu diễn ánh xạ theo phần dư sẽ được học. Phép toán $\mathcal{F} + x$ được thực hiện bằng một kết nối tắt và phép cộng phần tử.

Nhìn chung, chúng ta không thể xác định được số lượng tầng hoặc khối phần dư tối ưu cần thiết cho một mạng nơ-ron bởi nó phụ thuộc vào độ phức tạp của dữ liệu được huấn luyện. Việc sử dụng các kết nối tắt trong mạng cho phép chúng ta bỏ qua việc huấn luyện trên một số tầng trung gian mà không làm giảm độ chính xác tổng thể. Điều này giúp cho mạng nơ-ron trở nên linh hoạt để có thể điều chỉnh số tầng một cách tối ưu trong quá trình đào tạo. Độ chính xác Top-1 và Top-5 của mô hình ResNet-50 trên tập dữ liệu xác thực ImageNet lần lượt là 74,9% và 92,1%.

2.2.4. DenseNet (Dense Convolutional Network)

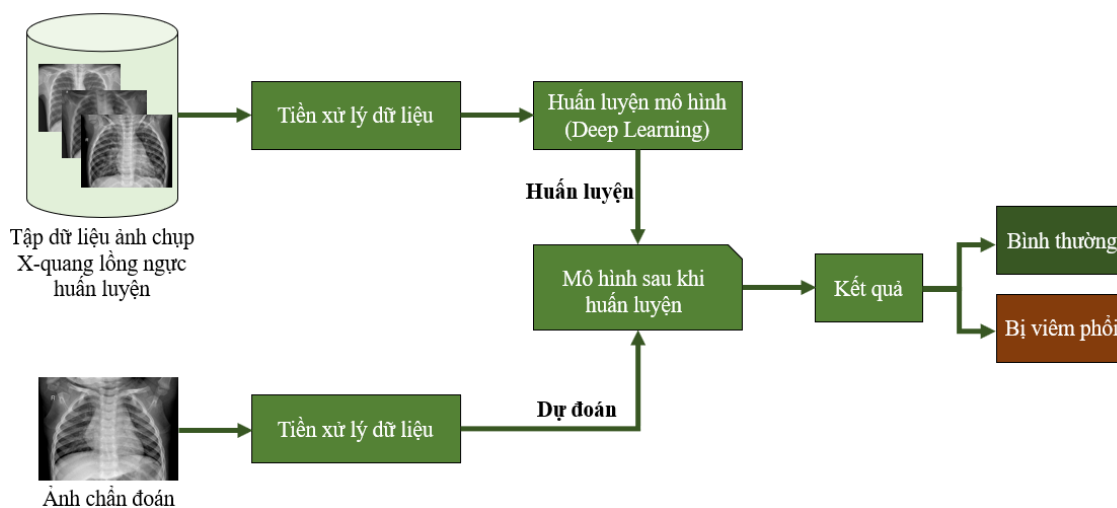
Kiến trúc mô hình DenseNet được Gao Huang và cộng sự giới thiệu vào năm 2016 trong bài báo [6]. DenseNet được coi là phiên bản mở rộng của ResNet khi kế thừa kiến trúc khối và phát triển kết nối tắt theo một mạng dày đặc. Nhằm mục đích cải thiện hơn nữa luồng thông tin giữa các tầng, các kết nối trực tiếp từ một tầng bất kỳ đến tất cả các tầng tiếp theo đã được đề xuất trong kiến trúc DenseNet. Theo đó, tầng thứ l nhận các bản đồ đặc trưng của tất cả các tầng trước đó $x_0, \dots, x_{l-1}$ làm đầu vào. Trong đó, $[X_0, X_1, \dots, X_{l-1}]$ biểu thị phép nối các bản đồ đặc trưng được tạo trong các tầng từ 0 đến $l-1$. Để dễ thực hiện, nhiều đầu vào của H_l theo công thức (8) được nối với nhau bằng một tensor duy nhất.

$$X_l = H_l([X_0, X_1, \dots, X_{l-1}]). \quad (8)$$

Điểm khác biệt chính của DenseNet so với ResNet là DenseNet nối các đầu ra lại với nhau thay vì cộng lại như ở ResNet. Sau đó ta áp dụng tầng chuyển tiếp (transition layer) để giảm chiều dữ liệu. DenseNet đồng thời cũng áp dụng Batch Normalization trước khi thực hiện tích chập ở các tầng chuyển tiếp để tránh hiện tượng *Vanishing Gradient* [8]. Kết quả là DenseNet-121 chỉ với 8 triệu tham số nhưng có độ chính xác cao hơn so với ResNet-50 với gần 26 triệu tham số trên bộ dữ liệu ImageNet. Cụ thể, độ chính xác Top-1 và Top-5 của mô hình DenseNet-121 trên tập dữ liệu xác thực ImageNet lần lượt là 75% và 92,3%.

3. Đề xuất mô hình học sâu hỗ trợ chẩn đoán viêm phổi

Trong phần 2, chúng tôi đã trình bày các kiến trúc mô hình học sâu hiện đại dựa trên mạng nơ-ron tích chập CNN. Dựa trên các phân tích các đặc trưng của các kiến trúc đó, bài báo đề xuất một mô hình ứng dụng học máy hỗ trợ chẩn đoán bệnh viêm phổi dựa trên ảnh chụp X-quang. Theo đó, mô hình phân lớp nhận dạng này được thực hiện qua 2 giai đoạn: Training (huấn luyện) và Validation (xác thực). Các bước trong mô hình ứng dụng được mô tả trong sơ đồ dưới đây.



Hình 4: Mô hình ứng dụng chẩn đoán bệnh viêm phổi

Quá trình huấn luyện:

Input: Tập dữ liệu huấn luyện

Bước 1: Tiền xử lý dữ liệu huấn luyện (thực hiện các thao tác dịch ảnh, zoom ảnh, lật ảnh để tạo sự đa dạng cho các mẫu dữ liệu ảnh).

Bước 2: Sử dụng các kiến trúc tích chập và pooling để trích xuất đặc trưng.

Bước 3: Huấn luyện các tham số mô hình.

Output: Mô hình phân lớp đã huấn luyện

Quá trình nhận dạng chẩn đoán:

Input: Ảnh cần chẩn đoán, các tham số mô hình đã huấn luyện được

Bước 1: Tiền xử lý dữ liệu ảnh cần nhận dạng (thực hiện thao tác resize kích thước của ảnh cần nhận dạng nhằm mục đích đồng nhất kích thước với các ảnh trong tập huấn luyện).

Bước 2: Phân lớp/chẩn đoán dựa vào mô hình đã được huấn luyện.

Output: Kết quả phân lớp/chẩn đoán

4. Thử nghiệm và kết quả

4.1. Bộ dữ liệu Chest X-Ray

Mô hình ứng dụng học sâu hỗ trợ chẩn đoán viêm phổi được huấn luyện thử nghiệm với các kiến trúc CNN hiện đại đã trình bày ở Phần 2 trên bộ dữ liệu Chest X-Ray. Bộ dữ liệu này có 5.863 ảnh X-Quang lồng ngực được Paul Mooney xây dựng, cập nhật năm 2018 và lưu trữ trên Kaggle. Hình ảnh X-quang lồng ngực được thu thập từ các bệnh nhi từ một đến năm tuổi tại Trung tâm Y tế Phụ nữ và Trẻ em Quảng Châu, Trung Quốc. Một số ảnh trong bộ dữ liệu được mô tả trong Hình 5.



Hình 5: Ví dụ về X-quang ngực bệnh nhân viêm phổi

Trong Hình 5, ảnh chụp X-quang ngực bình thường (ảnh bên trái) có hình ảnh phổi rõ ràng mà không có bất kỳ vùng mờ bất thường nào trong hình ảnh. Hình ảnh của bệnh nhân viêm phổi do vi khuẩn (ảnh giữa) thường xuất hiện các thùy khu trú (chỉ ra bởi các mũi tên trắng), trong khi viêm phổi do vi rút (ảnh bên phải) có các “kẽ” lan tỏa ra ở cả hai phổi. Các hình ảnh X-quang được tổ chức thành 3 thư mục bao gồm: *train*, *test*, *val* tương ứng với dữ liệu huấn luyện, dữ liệu kiểm tra và dữ liệu xác thực.

Tập dữ liệu	Nhãn Viêm phổi	Nhãn Bình thường	Tổng số mẫu
Chest X-Ray	4.273	1.583	5.856
Training	3.875	1.341	5.216
Test	390	234	624
Validation	8	8	16

4.2. Các độ đo đánh giá

Bài toán chẩn đoán bệnh viêm phổi là bài toán phân lớp hai nhãn lớp. Để đánh giá hiệu quả của các mô hình trong bài toán nhận dạng ảnh này, chúng tôi sử dụng các độ đo Accuracy, Precision, Recall và F1-Score [10]. Giả sử gọi nhãn lớp của các ảnh được xác nhận viêm phổi là lớp Positive (+1, tích cực), nhãn lớp của ảnh không bị viêm phổi là Negative (-1, tiêu cực). Khi mô hình phân lớp tiến hành phân loại các ảnh sẽ xảy ra 4 trường hợp như trong Hình 6, trong đó:

		Thực tế	
		Positive	Negative
Mô hình phân loại	Positive	TP	FP
	Negative	FN	TN

Hình 6: Ma trận biểu diễn kết quả

TP (True Positive): mẫu được phân loại positive chính xác (thực tế là positive).

FP (False Positive): mẫu bị phân loại sai positive (thực tế là negative).

TN (True Negative): mẫu được phân loại negative chính xác (thực tế là negative).

FN (False Negative): mẫu bị phân loại sai negative (trong thực tế là positive).

Trong bài toán này, bộ dữ liệu huấn luyện có tỷ lệ số lượng mẫu dữ liệu Positive chiếm khoảng 30%, khá mất cân bằng, do vậy để xem xét kỹ hơn về chất lượng của mô hình phân lớp, ngoài sử dụng độ đo Accuracy, chúng tôi sử dụng thêm độ đo Precision và Recall và F1-Score. Các công thức tính toán như sau:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

4.3. Kết quả thực nghiệm

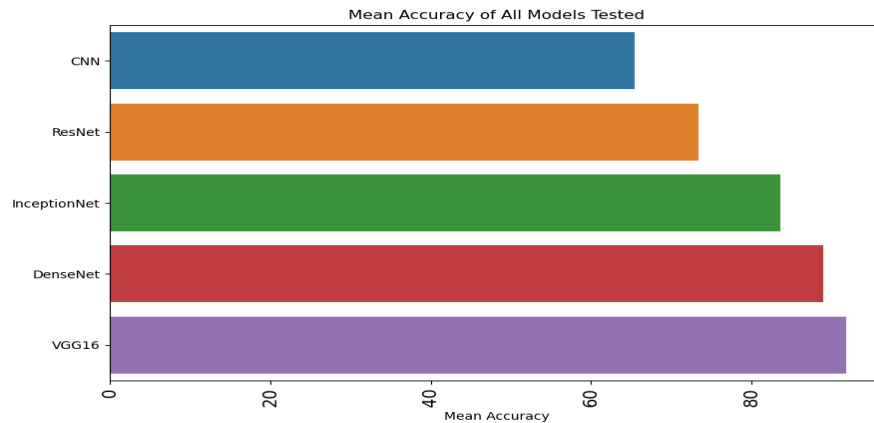
Chương trình huấn luyện và thử nghiệm chẩn đoán bệnh viêm phổi dựa trên ảnh chụp X-quang được lập trình với các công cụ và môi trường thực hiện như sau:

- Cấu hình máy tính: RAM 8GB; Ổ cứng SSD 500GB; CPU Core i7.
- Môi trường: Windows 10. Ngôn ngữ: Python 3.6, Javascript.
- Thư viện sử dụng: Numpy, Pandas, Matplotlib, Seaborn, Tensorflow, Keras

Mô hình học sâu huấn luyện các tham số và trích xuất đặc trưng của ảnh sử dụng các kiến trúc mạng lần lượt: CNN, ResNet, InceptionNet, DenseNet và VGG16. Mô hình kiến trúc các tầng mạng này đã trình bày chi tiết ở Phần 2. Chúng tôi thu nhận các kết quả, tính ra các giá trị độ đo tỷ lệ phân loại/ chẩn đoán chính xác Accuracy, Recall và Precision để đánh giá chất lượng của mô hình. Bảng 1 là kết quả tham số mô hình học được, tỷ lệ Accuracy trên bộ dữ liệu huấn luyện và bộ dữ liệu kiểm tra, các độ đo Precision, Recall, F1-score trên mẫu nhãn +1 (viêm phổi).

Bảng 1: Kết quả tham số và độ đo Accuracy

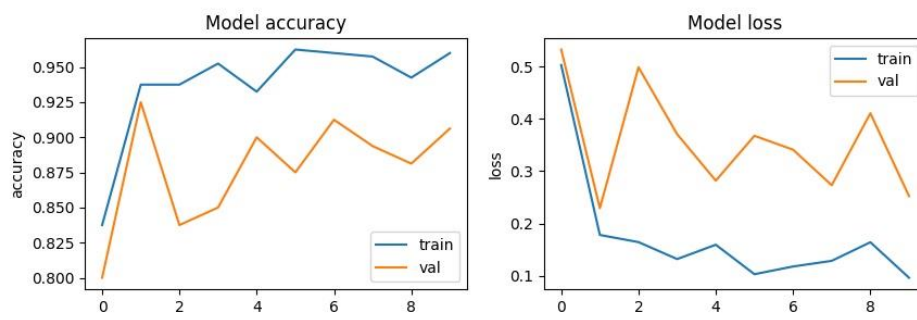
	CNN	ResNet	InceptionNet	DenseNet	VGG16
Total params	9.726.370	23.587.712	21.802.784	7.037.504	14.714.688
Trainable params	9.725.474	23.534.592	21.768.352	6.953.856	14.714.688
Test Accuracy	0,6550	0,7337	0,8365	0,8894	0,9183
Train Accuracy	0,8862	0,8556	0,8796	0,9473	0,9617
Precision (+)	0,7025	0,6250	0,8012	0,8264	0,8384
Recall (+)	0,9871	1,0000	0,9820	0,9769	0,9832
F1-score (+)	0,8208	0,7692	0,8824	0,8954	0,9051



Hình 7: Biểu đồ so sánh kết quả kiểm tra của các mô hình

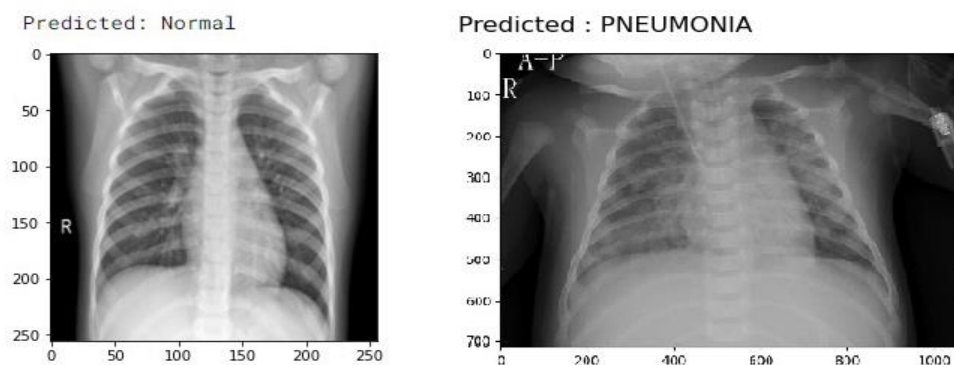
Các số liệu ở Bảng 1 và Hình 7 chỉ ra rằng kết quả mô hình phân lớp nhận dạng sử dụng kiến trúc mạng VGG16 cho kết quả độ chính xác cao nhất (91,83%) so với các mô hình sử dụng mạng CNN, ResNet, InceptionNet, DenseNet. Mặt khác, với bộ dữ liệu huấn luyện có tỷ lệ mất cân bằng ~1:3.3, (tỷ lệ mẫu nhãn +1 chiếm khoảng 31%) ta cần xem xét các độ đo Precision, Recall và đặc biệt là F1-Score để đánh giá chất lượng của mô hình. Kết quả trong bảng cũng chỉ ra rằng, mô hình kiến trúc VGG16 có độ đo Precision cao nhất 083,84%, độ đo Recall là 98,33% kém hơn CNN (98,72%) và ResNet (100%). Tuy nhiên, CNN và ResNet có kết quả đo Precision thấp hơn rất nhiều (70,25% và 62,50%), nghĩa là có khá nhiều mẫu bình thường bị phân loại nhầm sang viêm phổi. Độ đo trung bình điều hòa F1-Score của VGG16 có kết quả cao nhất 90,51% cho thấy đây là mô hình hiệu quả nhất trong chất lượng phân loại.

Về đánh giá kết quả lỗi của mô hình sử dụng kiến trúc VGG16 trên tập huấn luyện và tập xác thực trong Hình 8, chúng ta thấy các hàm đo độ chính xác và hàm 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).



Hình 8: Kết quả Accuracy VGG16

Các kết quả ấn tượng của mô hình học sâu sử dụng kiến trúc VGG16 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. Hình 9 mô phỏng một số kết quả chẩn đoán bệnh trong ứng dụng thực tế.



Hình 9: Ví dụ kết quả chẩn đoán

5. Kết luận

Trong bài báo này, chúng tôi đã thực hiện các thử nghiệm phân lớp/ nhận dạng chẩn đoán bệnh viêm phổi dựa trên các kiến trúc mạng nơ-ron học sâu hiện đại: CNN, ResNet, InceptionNet, DenseNet và VGG16. Kết quả chỉ ra rằng, mô hình kiến trúc mạng VGG16 có kết quả độ chính xác cao nhất (đạt xấp xỉ đến 91% trên bộ dữ liệu Chest-Xray). Đây là một con số ấn tượng, đem đến tiềm năng cho việc xây dựng một ứng dụng hỗ trợ chẩn đoán bệnh viêm phổi thông qua ảnh chụp X-quang.

TÀI LIỆU THAM KHẢO

- [1] Karen Simonyan, Andrew Zisserman, “Very Deep Convolutional Network for Large Scale Image Recognition,” *arXiv:1409.1556*, 2014.
- [2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition,” *arXiv:1512.03385*, 2015.
- [3] Min Lin, Qiang Chen, Shuicheng Yan, “Network in Network,” *arXiv:1312.4400*, 2013.
- [4] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna, “Rethinking the Inception Architecture for Computer Vision,” *arXiv:1512.00567*, 2015.
- [5] Vignesh Thakkar, Suman Tewary, Chakraborty, “Batch Normalization in Convolutional Neural Networks - A comparative study with CIFAR-10 data,” *2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT)*, 2018.
- [6] Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger, “Densely Connected Convolutional Networks,” *arXiv:1608.06993*, 2016.
- [7] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru, Vincent Vanhoucke, Andrew Rabinovich, “Going Deeper with Convolutions,” *arXiv:1409.4842*, 2014.
- [8] Hong Hui Tan, King Hann Lim, “Vanishing Gradient Mitigation with Deep Learning Neural Network Optimization,” *In: 2019 7th International Conference on Smart Computing & Communications (ICSCC)*, 2019.

- [9] Abien Fred, M. Agarap, “Deep Learning using Rectified Linear Units (ReLU),” *arXiv:1803.08375*, 2018.
- [10] Cyril Goutte, Eric Gaussier, “A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation”, In: *European Conference on Information Retrieval (ECIR)*, 2005.

SUMMARY

APPLICATION OF DEEP LEARNING TECHNIQUES IN SUPPORTING THE DIAGNOSIS OF PNEUMONIA THROUGH X-RAY IMAGES

Vo Duc Quang ^{(1), (5)}, **Nguyen Hai Yen** ⁽²⁾,
Mai Hong Man ⁽³⁾, **Nguyen Thi Nha** ⁽⁴⁾

¹ *Institute of Engineering and Technology, Vinh University*

² *Class 58K2-IT, Institute of Engineering and Technology, Vinh University*

³ *Class 58K4-IT, Institute of Engineering and Technology, Vinh University*

⁴ *Center for Practices and Experiments, Vinh University*

⁵ *PhD Student, Hanoi University of Science and Technology*

Received on 07/10/2021, accepted for publication on 10/12/2021

In the context of the outbreak of the COVID-19 epidemic in Vietnam and around the world, an artificial intelligence application that accurately diagnoses pneumonia will help reduce time and human resources for medical examination and treatment. This helps patients receive timely treatment, reducing the risk of aggravation and death. This paper presents the characteristics of modern deep learning network architectures based on convolutional neural networks such as ResNet50, VGG16, Inception, DenseNet. Thereby performing a test to evaluate these models in the diagnosis of pneumonia using the Chest-Xray dataset. The test results show that the deep learning model using VGG16 deep learning network architecture gives the highest accuracy rate. This is the basis for proposing to build an application to support effective pneumonia diagnosis based on X-ray images.

Keywords: Deep learning; Covid19; VGG16; CNN; X-ray images.