TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN   
VÀ TRUYỀN THÔNG VIỆT - HÀN

**KHOA KHOA HỌC MÁY TÍNH**



BÁO CÁO KỸ THUẬT

**ỨNG DỤNG DEEPLEARNING  
PHÁT HIỆN BỆNH VỀ PHỔI**

Học phần: Thị giác máy tính (1)

Sinh viên thực hiện: Nguyễn Đức Gia Phúc – 22AD037

Nguyễn Quang Tuấn – 22AD053

**Đà Nẵng, tháng 05 năm 2025**

# MỤC LỤC

[MỤC LỤC i](#_Toc198126226)

[1. Giới thiệu 1](#_Toc198126227)

[2. Kiến trúc tổng thể của Hệ thống 1](#_Toc198126228)

[3. Dữ liệu và tiền xử lý 1](#_Toc198126229)

[3.1. Bộ dữ liệu 1](#_Toc198126230)

[3.2. Tiền xử lý dữ liệu 2](#_Toc198126231)

[4. Kiến trúc mô hình 3](#_Toc198126232)

[4.1. Mô hình CNN: 3](#_Toc198126233)

[4.2 Mô hình Resnet 50: 4](#_Toc198126234)

[4.3 Mô hình Efficient Net v2: 5](#_Toc198126235)

[5.Huấn luyện mô hình 6](#_Toc198126236)

[5.1. Huấn luyện mô hình CNN: 6](#_Toc198126237)

[5.2. Huấn luyện mô hình ResNet 50: 6](#_Toc198126238)

[5.3. Huấn luyện mô hình EfficientNetV2: 6](#_Toc198126239)

[6. Kết quả thử nghiệm, so sánh, đánh giá 7](#_Toc198126240)

[6.1. Mô hình CNN 7](#_Toc198126241)

[6.2. Mô hình ResNet50 8](#_Toc198126242)

[6.3. Mô hình EfficientNetV2 9](#_Toc198126243)

[7. Kết luận 10](#_Toc198126244)

# 1. Giới thiệu

* Mục tiêu chính là hỗ trợ phát triển các mô hình học sâu (Deep learning), đặc biệt là các mạng nơ-ron tích chập (CNN), để tự động phân loại hình ảnh X-quang ngực thành hai nhóm chính:
* NORMAL : Phổi bình thường
* PNEUMONIA: Phổi bị viêm (do vi khuẩn hoặc virus)
* Giúp tăng cường khả năng chẩn đoán sớm và chính xác các bệnh về viêm phổi, đặc biệt trong các bối cảnh thiếu hụt bác sĩ chẩn đoán hình ảnh tại các khu vực đặc biệt, có nguồn lực y tế hạn chế.

# 2. Kiến trúc tổng thể của Hệ thống



**Hình 1. Kiến trúc tổng thể hệ thống**

* Ban đầu hệ thống nhận đầu vào là ảnh X-quang phổi.
* Sau đó ảnh được tiền xử lý bằng cách thay đổi kích thước ảnh về một kích thước chuẩn (224x224). Ảnh được áp dụng các kỹ thuật tăng cường dữ liệu (Data Augmentation) như: Xoay, lật ảnh, thay đổi độ sáng/tương phản. Chuẩn hóa dữ liệu ảnh (Normalization) theo giá trị trung bình và độ lệch chuẩn theo bộ ImageNet.
* Sau bước tiền xử lý, ảnh đã xử lý được đưa vào một mô hình học sâu để trích xuất đặc trưng và phân loại: CNN (Mạng nơ-ron tích chập cơ bản), ResNet50 (Mạng sâu có kết nối dư), Efficient Net V2. Mô hình dự đoán kết quả dựa trên các đặc trưng trích xuất từ ảnh. Mô hình đưa ra một trong hai nhãn: Phổi bình thường hoặc Phổi bị tổn thương (viêm phổi).
* Hiển thị kết quả dự đoán cuối cùng.

# 3. Dữ liệu và tiền xử lý

## 3.1. Bộ dữ liệu

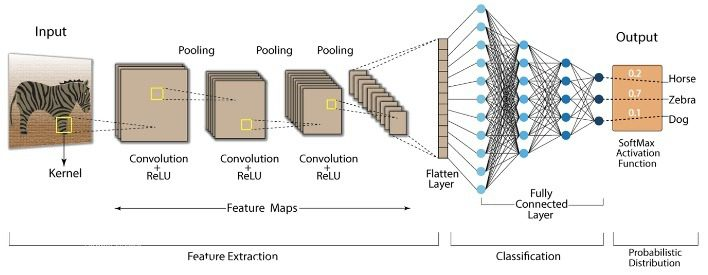
* Trong đề tài này sử dụng bộ dữ liệu hình ảnh X-quang ngực (Chest X-Ray Images) được công khai trên nền tảng Kaggle.
* Nguồn gốc bộ dữ liệu:
* Bộ dữ liệu được cung cấp bởi Paul Timothy Mooney, có nguồn từ nghiên cứu của NIH (National Institutes of Health).
* Bao gồm các hình ảnh X-quang ngực được gán nhãn bởi chuyên gia chẩn đoán.
* Đặc điểm, số lượng mẫu, phân bố số lượng giữa các lớp
* Đặc điểm: Dữ liệu bao gồm các ảnh X-quang ngực (ảnh grayscale, định dạng JPEG) của bệnh nhân nhi (trẻ em) ở độ tuổi 1–5.
* Số lượng mẫu: 5856 ảnh
* Tập Train có 5216 ảnh, trong đó NORMAL có 1341 ảnh và PNEUMONIA có 1341 ảnh
* Tập Validation có 16 ảnh, trong đó NORMAL có 8 ảnh và PNEUMONIA có 8 ảnh
* Tập Test có 624 ảnh, trong đó NORMAL có 234 ảnh và PNEUMONIA có 390 ảnh
* Số lượng ảnh PNEUMONIA chiếm hơn 70% tổng số liệu, nên tỉ lệ giữa NORMAL và PNEUMONIA khoảng 1/3

## 3.2. Tiền xử lý dữ liệu

* Mã hóa theo bệnh nhân: Với ảnh viêm phổi (PNEUMONIA), tên file có chứa thông tin ID bệnh nhân (ví dụ: person01\_bacteria\_...). Ta trích xuất ID bệnh nhân để phân chia dữ liệu theo nhóm bệnh nhân, nhằm tránh hiện tượng một bệnh nhân xuất hiện ở cả tập huấn luyện và tập validation → tránh rò rỉ dữ liệu
* Random phân chia tập dữ liệu
* 80% ảnh được dùng để huấn luyện
* 20% ảnh được dùng để xác thực
* Tập Test được giữ nguyên
* Áp dụng các kỹ thuật xử lý ảnh:
  + Chuyển ảnh grayscale sang 3 kênh để phù hợp với mô hình pretrained
* Xoay ảnh ngẫu nhiên trong khoảng ±30 độ
* Lật ảnh ngang với xác suất 50%
* Thay đổi nhẹ độ sáng, độ tương phản, độ bão hòa, sắc độ
* Cắt và phóng ảnh ngẫu nhiên để mô phỏng nhiều góc chụp khác nhau
* Biến dạng nhẹ hình học
* Chuẩn hóa ảnh theo mean và std theo ImageNet
* Giải quyết mất cân bằng lớp
* Tạo DataLoader

# 4. Kiến trúc mô hình

## 4.1. Mô hình CNN:



***Hình 2. Kiến trúc mô hình CNN***

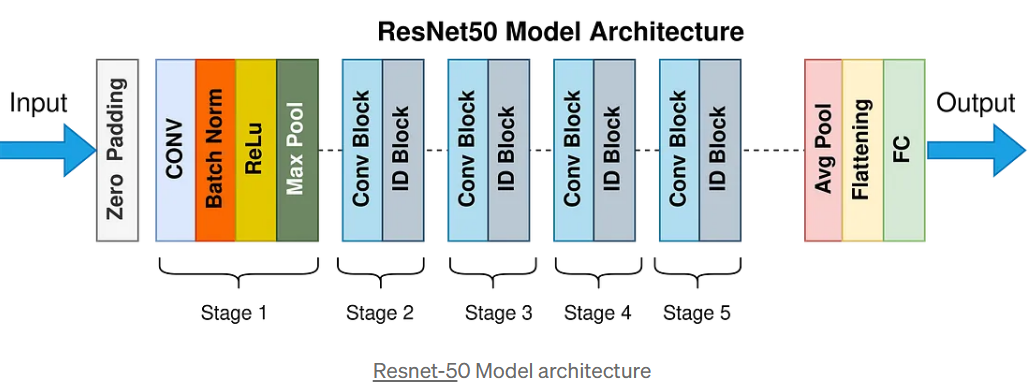
Cấu trúc tổng thể của CNN gồm 5 thành phần chính:

* 1. Lớp đầu vào (Input Layer)
* Nhận đầu vào là ảnh có kích thước cố định, ví dụ: 224 × 224 × 3 (cao × rộng × số kênh màu).
* 2. Lớp tích chập (Convolutional Layer)
* Mục đích: trích xuất các đặc trưng như cạnh, góc, kết cấu... từ ảnh.
* Sử dụng các kernel/filter trượt qua ảnh và tính toán tích chập.
* Output: bản đồ đặc trưng (feature maps)
* Có thể có nhiều lớp liên tiếp nhau.

Sau mỗi lớp tích chập, thường có:

* Batch Normalization (BN):Chuẩn hóa đầu ra để tăng tốc quá trình huấn luyện.
* Hàm kích hoạt (Activation Function): Thường dùng ReLU: f(x) = max(0, x) giúp mô hình học phi tuyến.
* 3. Lớp gộp (Pooling Layer)
* Mục đích: giảm kích thước không gian (chiều cao và chiều rộng), giúp giảm số lượng tham số và tính toán, chống overfitting.
* 4. Lớp kết nối đầy đủ (Fully Connected Layer - FC)
* Nhận vector đặc trưng từ các lớp trước (thường sau Global Average Pooling). Dùng để phân loại, đưa ra xác suất các lớp.
* Softmax Layer (cho classification): Biến đầu ra thành xác suất (tổng các xác suất bằng 1).
* 5. Lớp đầu ra (Output Layer)

## 4.2 Mô hình Resnet 50:

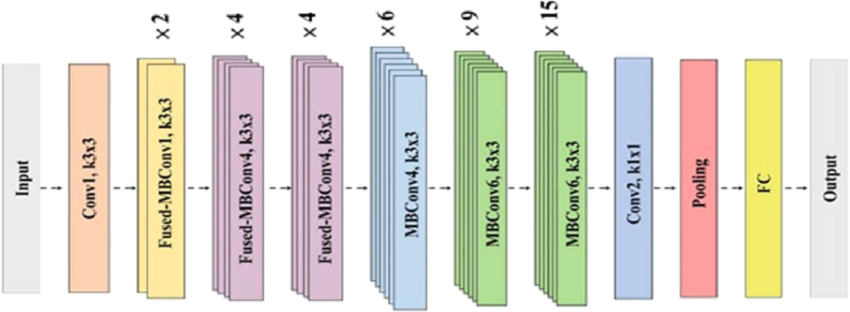


***Hình 3. Kiến trúc mô hình ResNet 50***

Cấu trúc tổng thể của ResNet-50: ResNet-50 bao gồm 5 giai đoạn chính:

* Giai đoạn đầu (Stem):
* Lớp tích chập 7×7 với 64 bộ lọc, stride 2.
* Lớp chuẩn hóa theo lô (Batch Normalization) và hàm kích hoạt ReLU.
* Lớp gộp cực đại (Max Pooling) 3×3 với stride 2.
* Giai đoạn 1:
* 3 khối dư với số lượng bộ lọc lần lượt là 64, 64, và 256.
* Giai đoạn 2:
* 4 khối dư với số lượng bộ lọc lần lượt là 128, 128, và 512.
* Giai đoạn 3:
* 6 khối dư với số lượng bộ lọc lần lượt là 256, 256, và 1024.
* Giai đoạn 4:
* 3 khối dư với số lượng bộ lọc lần lượt là 512, 512, và 2048.
* Lớp kết thúc:
* Lớp gộp trung bình toàn cục (Global Average Pooling).
* Lớp kết nối đầy đủ (Fully Connected) với số lượng đầu ra tùy thuộc vào số lớp cần phân loại.
* Hàm kích hoạt Softmax để đưa ra xác suất phân loại

## 4.3 Mô hình Efficient Net v2:

****

**Hình 4. Kiến trúc mô hình Efficient Net v2**

EfficientNetV2 bao gồm 6 giai đoạn chính:

* Giai đoạn đầu (Stem):
  + Conv Layer: 3×3, 32 bộ lọc, stride 2.
  + Batch Normalization và Swish Activation.
* Giai đoạn 1:
  + 1 block Fused-MBConv với cấu trúc 1x1 Conv -> 3x3 DWConv -> 1x1 Conv.
  + Số lượng bộ lọc lần lượt là 16, 16.
  + Kích thước đầu ra không thay đổi (stride 1).
* Giai đoạn 2:
  + 2 blocks Fused-MBConv.
  + Số lượng bộ lọc lần lượt là 32, 32.
  + Kích thước đầu ra giảm (stride 2).
* Giai đoạn 3:
  + 3 blocks Fused-MBConv.
  + Số lượng bộ lọc lần lượt là 48, 48.
  + Kích thước đầu ra giảm (stride 2).
* Giai đoạn 4:
  + 4 blocks MBConv (với Skip Connection).
  + Số lượng bộ lọc lần lượt là 96, 112, 160, 192.
  + Sử dụng 1x1 Conv -> 3x3 DWConv -> 1x1 Conv với Swish Activation.
  + Kích thước đầu ra giảm (stride 2).
* Lớp kết thúc:
  + Global Average Pooling (GAP).
  + Fully Connected Layer (FC) với số lượng đầu ra tùy thuộc vào số lớp cần phân loại.
  + Softmax Activation để đưa ra xác suất phân loại.

# 5.Huấn luyện mô hình

## 5.1. Huấn luyện mô hình CNN:

- Mô hình sử dụng**:** Mạng CNN tự xây dựng gồm 3 tầng tích chập (Conv2d) và 2 tầng ẩn (Linear)

- Ảnh đầu vào: được resize về kích thước 224x224

- Các lớp đều được huấn luyện (không đóng băng)

- Tối ưu: Adam, CrossEntropyLoss, scheduler CosineAnnealingLR

## 5.2. Huấn luyện mô hình ResNet 50:

- Mô hình sử dụng**:** resnet50 với trọng số pretrained từ ImageNet (ResNet50\_Weights.IMAGENET1K\_V1)

- Ảnh đầu vào: resize về 224x224

- Thay đổi cấu trúc:

+ Đóng băng toàn bộ các lớp trừ layer4[-1] và fc

+ Tầng fc (fully connected) được thay bằng: Linear → Dropout → BatchNorm → ReLU → Linear (ra 2 lớp)

- Tối ưu: Adam, CrossEntropyLoss, scheduler CosineAnnealingLR

## 5.3. Huấn luyện mô hình EfficientNetV2:

- Mô hình sử dụng: tf\_efficientnetv2\_s.in21k từ thư viện timm, pretrained từ ImageNet21k

- Ảnh đầu vào: resize về 224x224, chuyển sang grayscale

- Chỉnh sửa cấu trúc:

+ Đóng băng toàn bộ mô hình

+ Chỉ fine-tune block cuối (blocks[-1]) và tầng phân loại (classifier)

+ Thay tầng classifier bằng: Dropout → Linear → BatchNorm → ReLU → Dropout → Linear (ra 2 lớp)

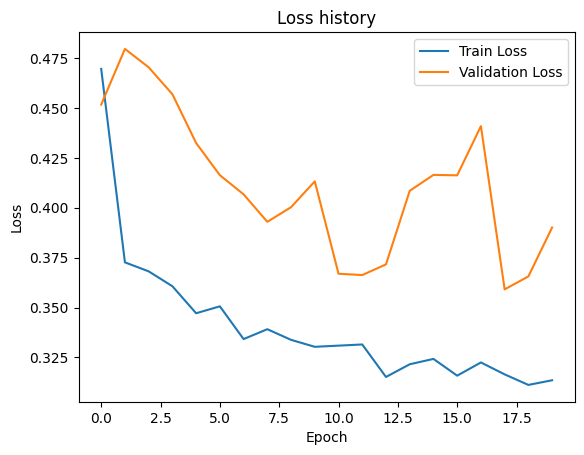
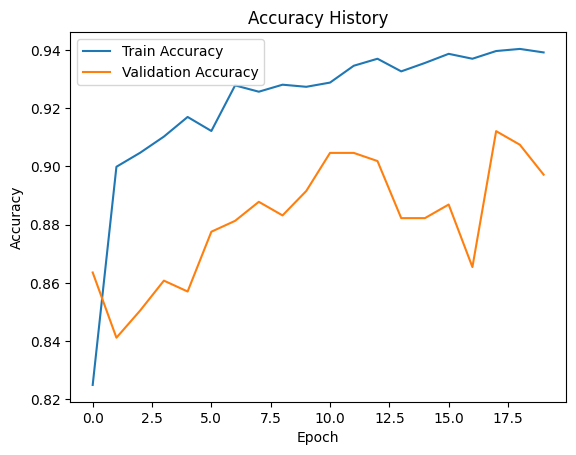
- Tối ưu: Adam, CrossEntropyLoss, scheduler CosineAnnealingLR

# 6. Kết quả thử nghiệm, so sánh, đánh giá

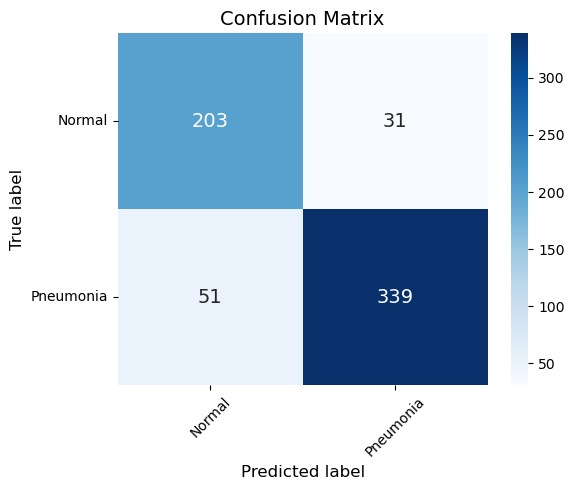
Trong quá trình huấn luyện, cả ba mô hình CNN, ResNet18 và EfficientNetv2 đều được huấn luyện với cùng tập dữ liệu và cấu hình tương đương. Kết quả cụ thể như sau:

|  |  |  |
| --- | --- | --- |
| Mô hình | Accuracy (Train) | Accuracy (Test) |
| CNN | 93.92% | 86.86% |
| ResNet50 | 97.3% | 93.11% |
| EfficientNetV2 | 96.56% | 92.95% |

## 6.1. Mô hình CNN

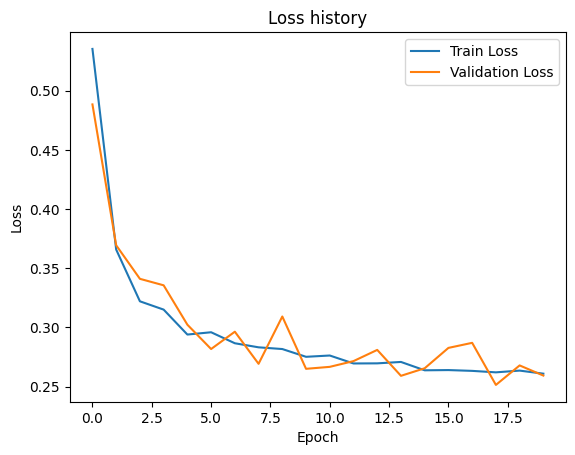
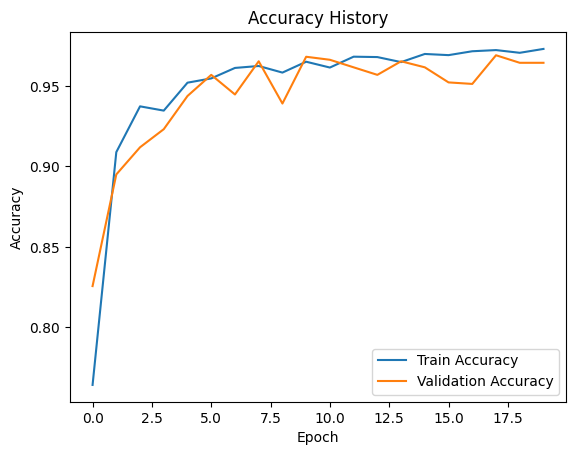


**Hình 5. Đồ thị huấn luyện CNN**

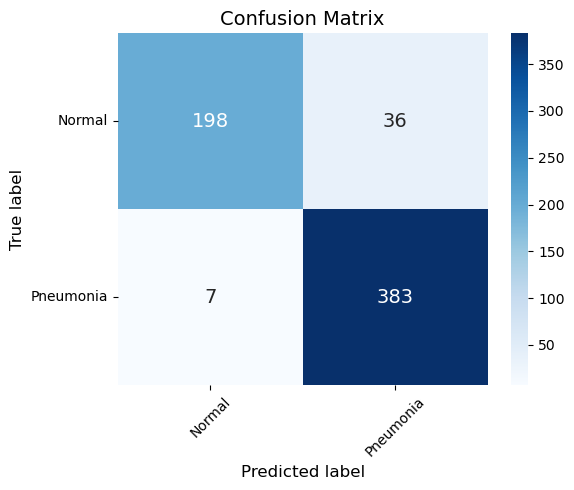


**Hình 6. Confusion matrix của CNN**

## 6.2. Mô hình ResNet50

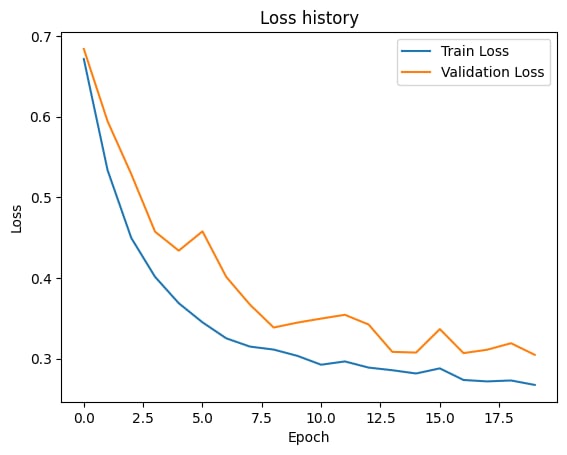
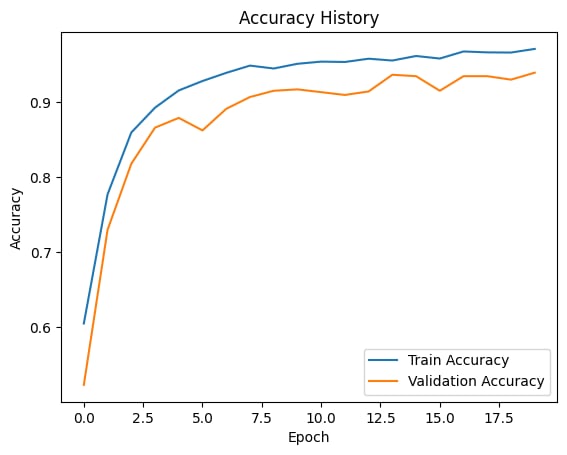


**Hình 7. Đồ thị huấn luyện ResNet50**

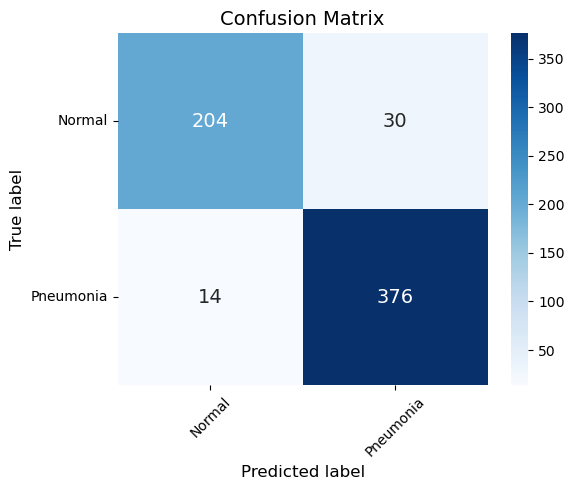


**Hình 8. Confusion matrix của Efficient Net v2**

## 6.3. Mô hình EfficientNetV2



**Hình 9. Đồ thị huấn luyện Efficient Net v2**



**Hình 10. Confusion matrix của Efficient Net v2**

# 7. Kết luận

Trong quá trình thử nghiệm trên bộ dữ liệu pneumonia, ba mô hình được so sánh bao gồm ResNet50, EfficientNetV2 và một mạng CNN được thiết kế và huấn luyện từ đầu.

Kết quả cho thấy:

* ResNet50 đạt hiệu suất vượt trội hơn hẳn so với hai mô hình còn lại, cả về độ chính xác và tốc độ hội tụ. Điều này nhấn mạnh rằng việc tận dụng mô hình tiền huấn luyện (pre-trained) với kiến trúc sâu rộng và khả năng nhận diện đặc trưng mạnh mẽ như ResNet50 có thể mang lại lợi ích rõ rệt trong các bài toán phân loại ảnh y tế.
* Ngược lại, mặc dù EfficientNetV2 được thiết kế để cân bằng giữa độ chính xác và tốc độ tính toán, nhưng trong trường hợp ảnh grayscale và dữ liệu hạn chế, ResNet50 vẫn chứng minh được khả năng phát huy hiệu quả cao hơn khi được tinh chỉnh đúng cách.