Transfer Learning for COVID-19 Chest X-ray Classification - Lab 4 -

Abdelillah Serghine
a.serghine@esi-sba.dz

April 19, 2025

1 Introduction

This report presents the application of transfer learning for classifying chest X-ray images using the COVID-19 Radiography Dataset. The goal is to distinguish between four categories: COVID-19, Lung Opacity, Normal, and Viral Pneumonia using pretrained CNN architectures.

2 Dataset Description

The dataset was obtained from Kaggle:
`https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database`
It contains the following classes:

- COVID
- Lung Opacity
- Normal
- Viral Pneumonia

Each image was resized to 224×224 pixels and normalized by dividing pixel values by 255.

3 Data Preprocessing

- Images were loaded from structured folders and converted to NumPy arrays.
- Pixel values were scaled to [0, 1].
- Labels were encoded using one-hot encoding.
- Dataset was split into 70% training and 30% validation.

4 Model Architecture

The base model used is **MobileNetV2** pretrained on ImageNet. The original classification head was removed and replaced with:

- Global Average Pooling
- Dense(1024, ReLU)
- Dense(512, ReLU)
- Dense(4, Softmax)

Layer (type)	Output Shape	Param #
mobilenetv2_1.00_224 (Functional)	(None, 7, 7, 1280)	2,257,984
global_average_pooling2d_1 (GlobalAveragePooling2D)	(None, 1280)	0
dense_3 (Dense)	(None, 1024)	1,311,744
dense_4 (Dense)	(None, 512)	524,800
dense_5 (Dense)	(None, 4)	2,052

Total params: 4,096,580 (15.63 MB)

Trainable params: 1,838,596 (7.01 MB)

Non-trainable params: 2,257,984 (8.61 MB)

Total parameters: 4096580

Figure 1: Model Architecture

Three training scenarios were evaluated:

- **Scenario 1**: Train last two convolutional layers + FC layers
- **Scenario 2**: Train last one convolutional layer + FC layers
- **Scenario 3**: Train only FC layers

5 Evaluation Results

Training with 20 epochs, a batch size of 64 and the Adam optimizer with a learning rate of 0.001.

Scenario 1: Train Last Two Conv Layers + FC Layers

- Accuracy: 86%

- Macro F1-score: 87%

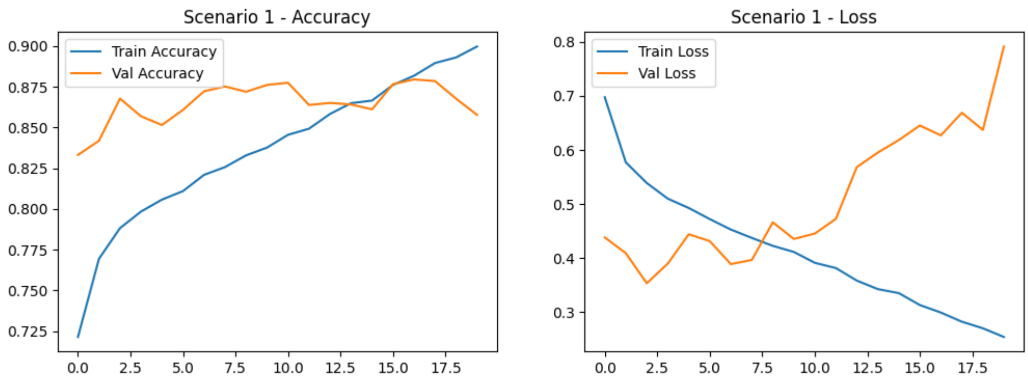

Figure 2: Scenario 1: Accuracy and Loss Curves

Table 1: Classification Report - Scenario 1

Class	Precision	Recall	F1-score	Support
COVID	0.84	0.88	0.86	2169
Lung Opacity	0.78	0.85	0.82	3607
Normal	0.91	0.85	0.87	6115
Viral Pneumonia	0.94	0.91	0.93	807
Accuracy		0.86 (12,698 samples)		
Macro Avg	0.87	0.87	0.87	
Weighted Avg	0.86	0.86	0.86	

Scenario 2: Train Last One Conv Layer + FC Layers

- Accuracy: 86%
- Macro F1-score: 86%

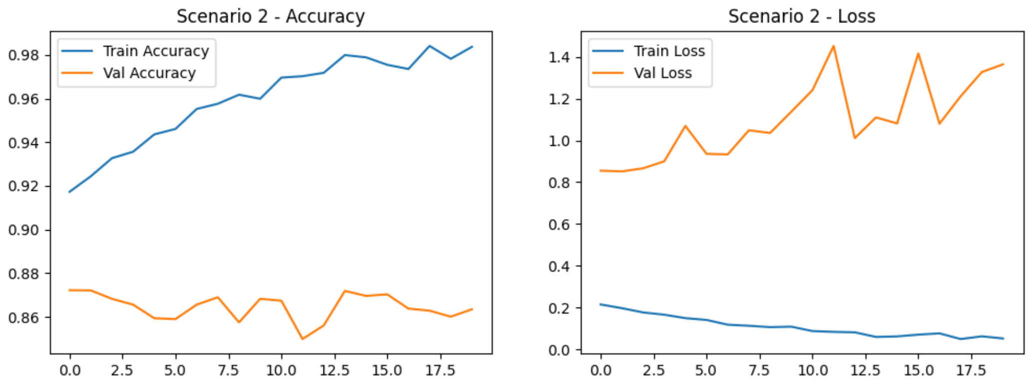

Figure 3: Scenario 2: Accuracy and Loss Curves

Table 2: Classification Report - Scenario 2

Class	Precision	Recall	F1-score	Support
COVID	0.89	0.83	0.86	2169
Lung Opacity	0.88	0.77	0.82	3607
Normal	0.84	0.94	0.89	6115
Viral Pneumonia	0.99	0.80	0.88	807
Accuracy		0.86 (12,698 samples)		
Macro Avg	0.90	0.83	0.86	
Weighted Avg	0.87	0.86	0.86	

Scenario 3: Train Only FC Layers

- Accuracy: 87%
- Macro F1-score: 87%

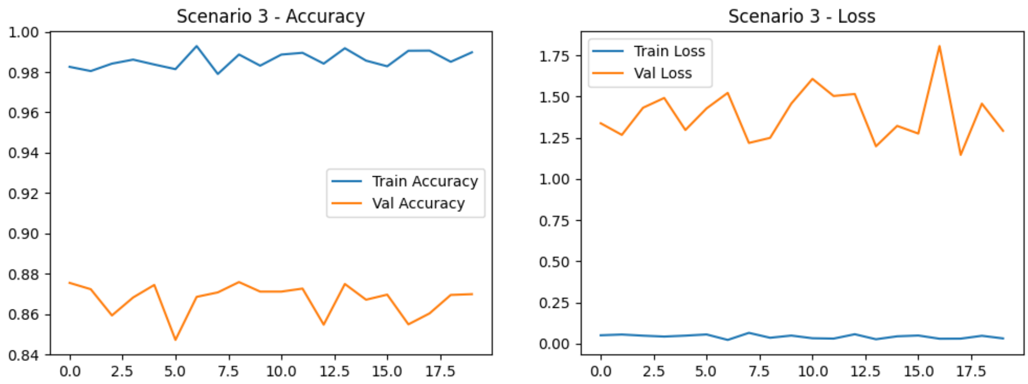

Figure 4: Scenario 3: Accuracy and Loss Curves

Table 3: Classification Report - Scenario 3

Class	Precision	Recall	F1-score	Support
COVID	0.90	0.81	0.85	2169
Lung Opacity	0.84	0.83	0.84	3607
Normal	0.87	0.91	0.89	6115
Viral Pneumonia	0.89	0.92	0.91	807
Accuracy		0.87 (12,698 samples)		
Macro Avg	0.88	0.87	0.87	
Weighted Avg	0.87	0.87	0.87	

6 Training Time Comparison

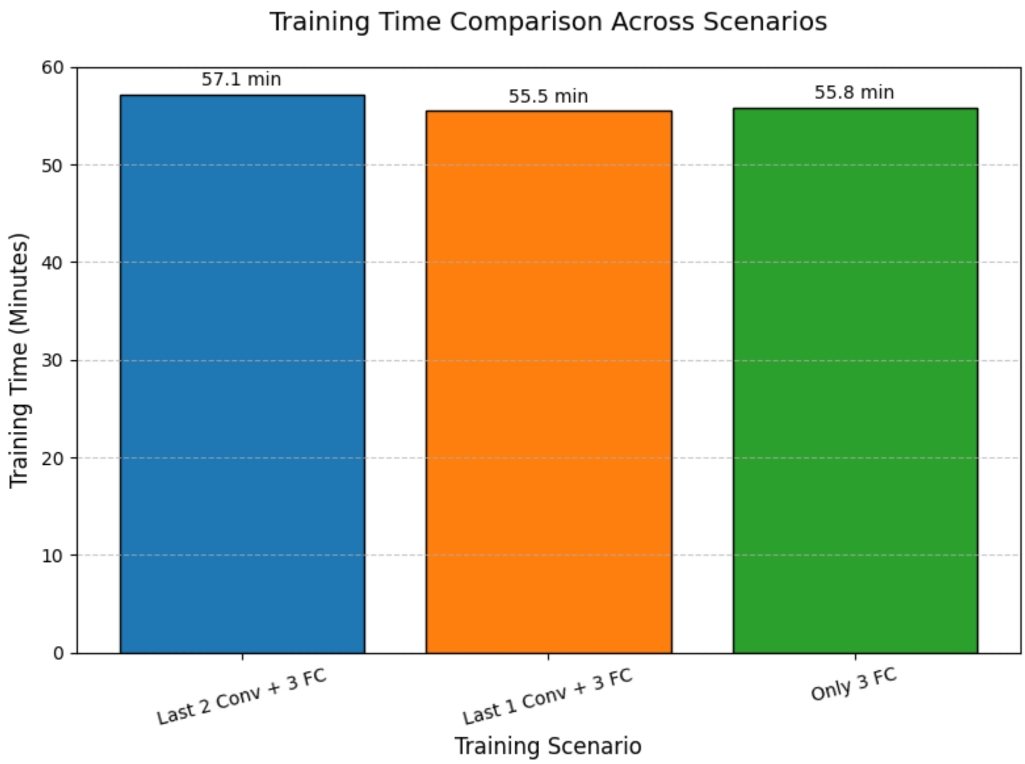

Figure 5: Training Time Comparison Across Scenarios

7 Discussion

All three training configurations yielded comparable results, with Scenario 3 slightly outperforming the others in terms of overall accuracy (87%) and F1-score.

- **Scenario 3 (FC only)** achieved the highest balance, indicating that MobileNetV2 already learns powerful features from ImageNet, requiring minimal fine-tuning.
- **Scenario 2** had a high precision for Viral Pneumonia (0.99), but slightly lower recall.
- **Scenario 1** achieved strong balance across all classes, with COVID and Viral Pneumonia being classified accurately.

- Training time (Figure 5) was shortest for Scenario 3, making it the most efficient approach.

Overall, training only the fully connected layers resulted in good performance, while saving computational cost and reducing overfitting risks, but there is an oscillations which may caused by :
- Small Validation Set or Batch Variability.
- Overfitting and Model Sensitivity.
- Class Imbalance and Dataset Noise.
- Randomness in Training.

8 Conclusion

Transfer learning using MobileNetV2 on chest X-ray images can effectively distinguish between multiple lung conditions. With minimal fine-tuning, the model achieves up to 87% accuracy.