# Comparative Evaluation of Pretrained Models for Lung Disease Classification from Chest X-Rays

DATA 586 Project - Group 13

Craig Adlam, Kulaphong Jitareerat, Nijiati Abulizi

## Summary

This project investigates the comparative efficacy of various pretrained convolutional neural network (CNN) models for the classification of lung diseases from chest X-rays. Building upon an original study which used VGG16, this project extends the comparison to include EfficientNetV2 models and evaluates the impact of different optimizers, specifically Adam and AdamW. The aim is to identify optimal model configurations that enhance diagnostic accuracy in a multiclass setting, which includes ten lung disease classes alongside a normal class.

## Introduction

Lung diseases, notably worsened by environmental factors and global health trends, pose a significant burden on healthcare systems. Traditional diagnostic methods, heavily reliant on human expertise, face challenges like variability in diagnosis and limited accessibility. Advanced deep learning techniques, particularly CNNs, offer a promising solution due to their ability to learn complex patterns in imaging data. Previous research demonstrated the potential of the VGG16 model and achieved significant multi-class classification accuracy; this study expands on that by exploring additional architectures and training optimizations to further enhance classification performance.

## Methodology

The original study involved a comprehensive dataset of 80,000 chest X-rays, preprocessed and augmented to address class imbalance and enhance model training. Initial benchmarks were established using a range of pretrained models including AlexNet, GoogLeNet, InceptionV3, MobileNetV2, VGG16, ResNet 50, DenseNet121, and EfficientNetB7, with VGG16 initially outperforming others in terms of accuracy.

For this project, we conducted further experiments with EfficientNetV2 variants with different numbers of parameters, using Adam and AdamW optimizers. Models were fine-tuned on the lung disease dataset, with emphasis on evaluating their performance across different configurations. Metrics such as accuracy, precision, recall and f1 values were employed to assess model effectiveness.

Specifically, out of the original 80000 X-ray images, we used 5000 images, ensuring balanced representation across ten disease classes. These images underwent preprocessing to meet the input standards required for deep learning models, which included normalization and augmentation. We then split this dataset into a training set of 4,000 images and a testing set of 1,000 images, maintaining an 80:20 ratio.

For the training phase, we chose six distinct pre-trained CNN models, each undergoing two training sessions with different optimizers: Adam and AdamW, totaling twelve trained models. The validation performance was monitored to apply early stopping criteria, thereby preventing overfitting. Upon completion of the training, we evaluated each model against the testing dataset to measure generalizability and classification capability, recording performance metrics such as accuracy and other performance values.

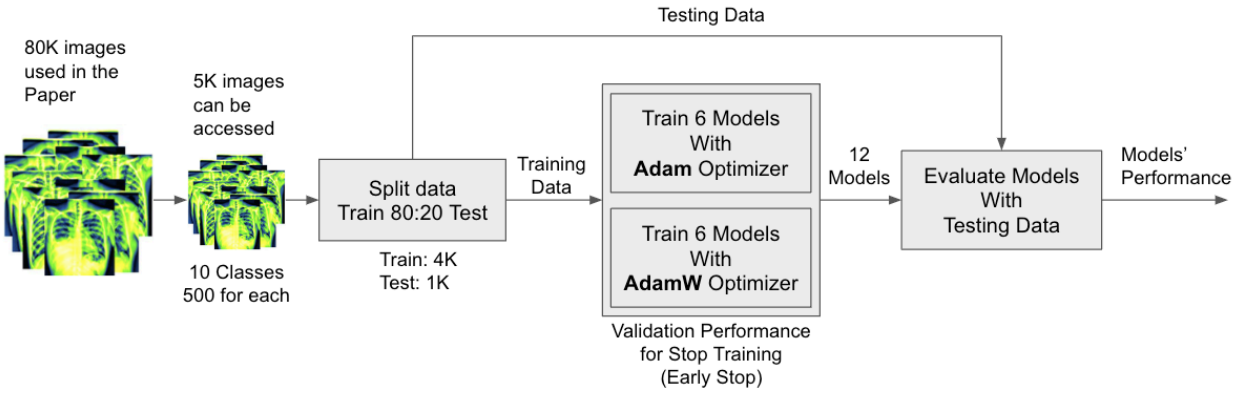


Figure1. Workflow

In our analysis, we compared the performance of all twelve model configurations to find out the most effective model-optimizer combination. These findings were documented to capture trends and efficiency of the optimizers, particularly highlighting their performance across varied architectures. Finally, we synthesized our research into a comprehensive report that not only showed the comparative effectiveness of each model and optimizer but also outlined potential future research avenues and the practical implications of our study for medical diagnostics

**Results**

We were able to replicate the original study's methodology and its findings with reduced sample images, establishing a robust baseline for our research. With the successful reproduction of results using the VGG16 model, as reported in the original study, we set the stage for an expanded investigation. Our approach was to do a comprehensive comparative analysis,

incorporating a suite of six state-of-the-art pretrained models of EfficientNetV2 and VGG19. Each model was trained with both the Adam and AdamW optimizers, resulting in twelve unique model-optimizer configurations. This extensive exploration allowed us to evaluate the effects of various architectures and optimization techniques on the classification accuracy for lung diseases from chest X-ray images. By doing this, we were able to identify the most effective combinations and provide deeper insights into the optimization of deep learning models for medical imaging diagnostics as shown in Figure 2.

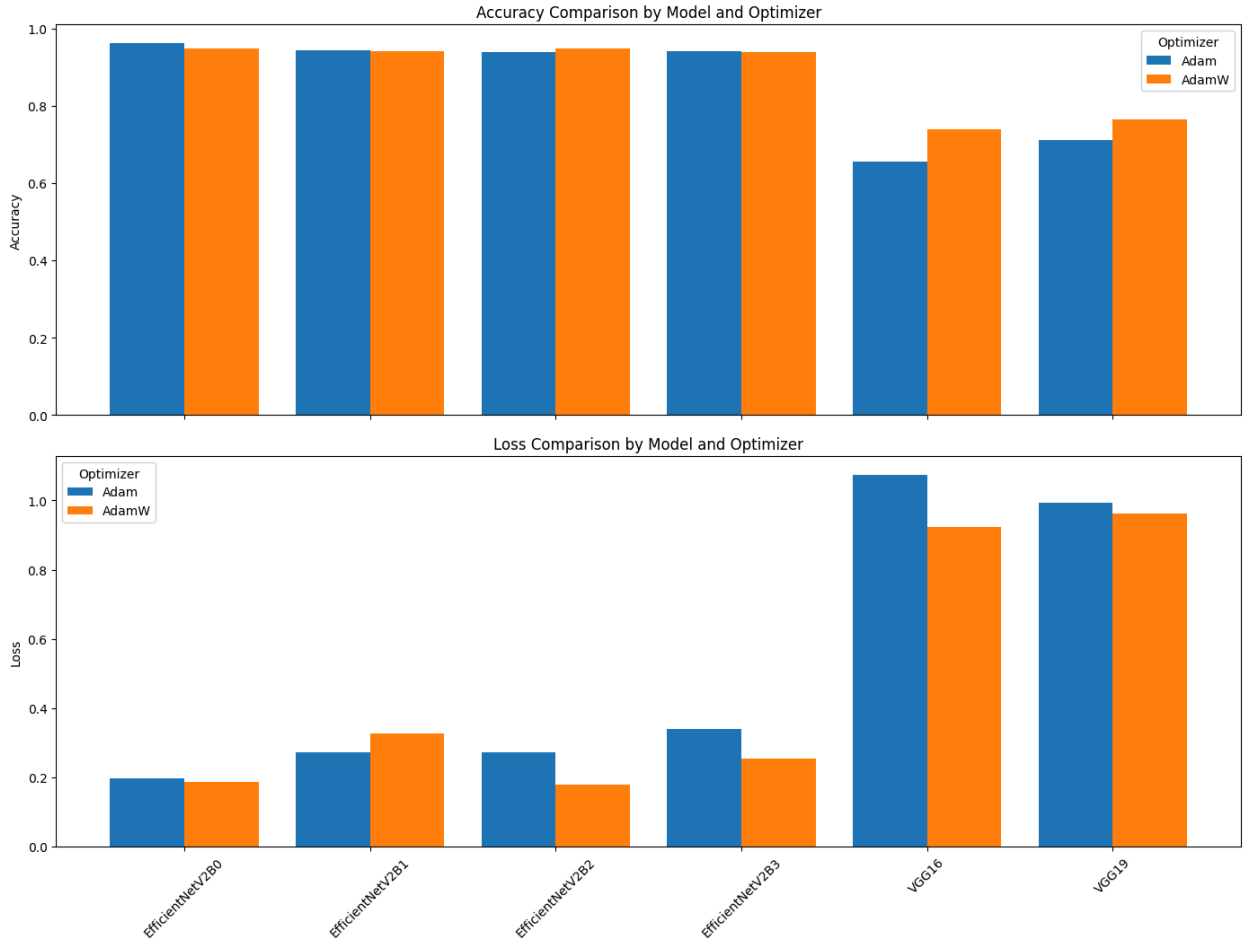


Figure 2. Testing Dataset Performance (Accuracy)

From Figure 2, we can see that EfficientNetV2 models generally provided better accuracy (around 94%) than VGG16 and VGG19. The AdamW optimizer led to significant performance improvements in VGG16 models compared to Adam. For EfficientNetV2 models, both Adam

and AdamW optimizers performed comparably, suggesting robustness in optimizer choice for these architectures.

For completeness, here we also provided the precision, recall and f1-scores plots (Figure 3) of our 12 models with the two different optimizers.

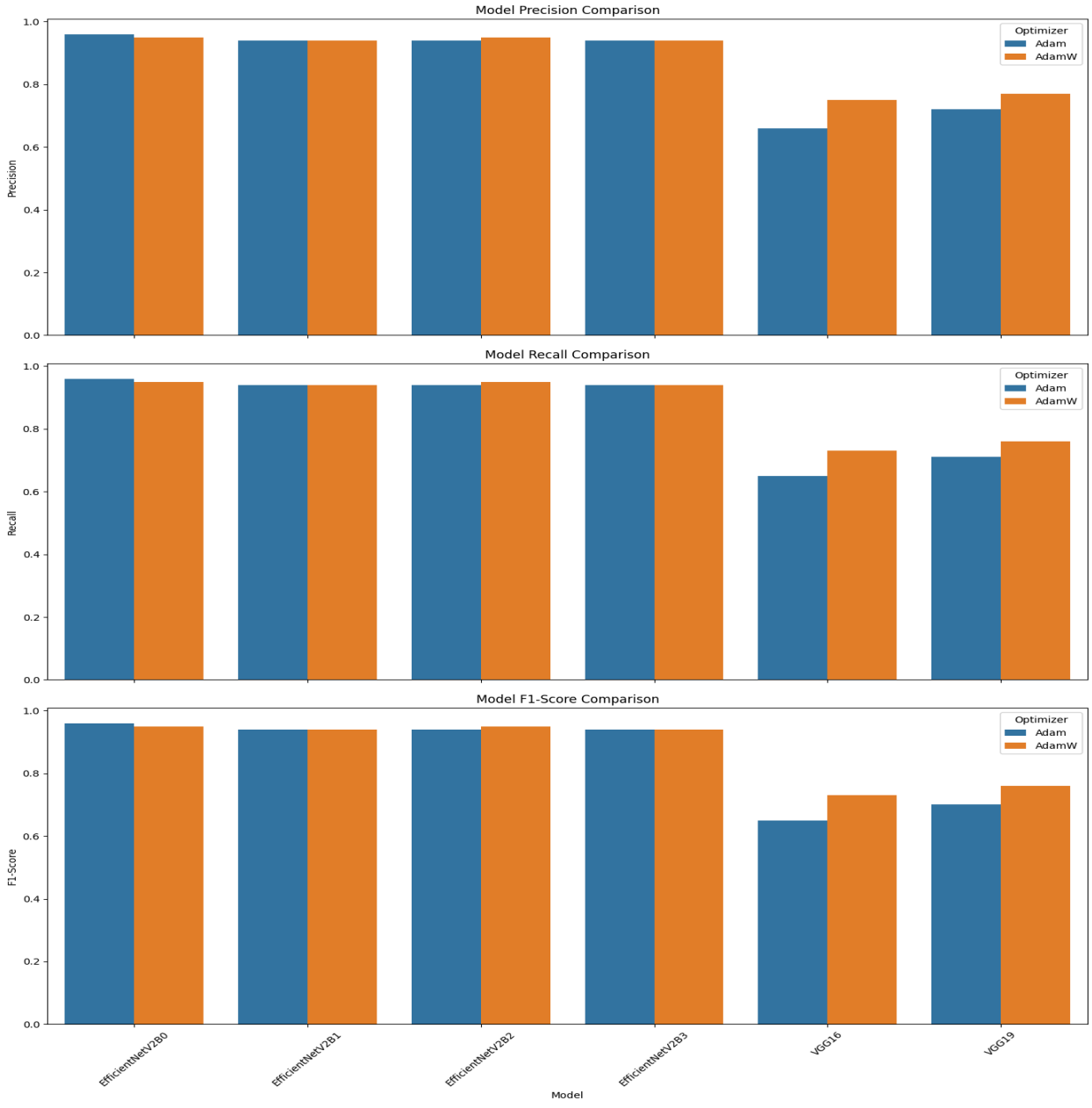


Figure 3. Testing Dataset Performance (Precision, Recall, F1-Score)

## Conclusion and Future Work

The findings indicate that newer architectures like EfficientNetV2 may offer advantages over traditional models like VGG16 in specific medical imaging tasks due to their advanced design principles. Additionally, optimizer choice can significantly influence model performance, particularly in how weight adjustments are handled during training.

Future research could explore:

- Integration of more complex data augmentation and preprocessing techniques to further enhance model robustness.
- Expansion of the dataset size and diversity to improve the generalizability of the models.
- Examination of additional emerging CNN architectures and their potential in medical image analysis.
- Deployment of these models in real-world clinical settings to evaluate practical utility and user acceptance.

**Limitations**

The study encountered several limitations:

- The preprocessing of images was crucial for our model performance but also could introduce potential biases depending on the chosen techniques.
- Dataset size could be expanded to include more varied pathological cases.
- Hardware constraints, particularly GPU availability, limited the extent of experimentation and model complexity.

**References**

LungNet22: A Fine-Tuned Model for Multiclass Classification and Prediction of Lung Disease Using X-ray Images J. Pers. Med. 2022, 12(5), 680; https://doi.org/10.3390/jpm12050680 https://www.mdpi.com/2075-4426/12/5/680

**Contributions of Group Members**

All members of the group contributed equally to the project.

GitHub: https://github.com/Nijat27/CNN_Lung_Xray/blob/main/Combined.ipynb, https://github.com/kulaphongj/cnn_lung_diseases/tree/main/data_lung (data)

**Appendix**

Confusion Matrix of Testing Dataset

EfficientNetV2B0 - Adam

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| control | 0.98 | 1.00 | 0.99 | 116 |
| covid | 0.98 | 0.97 | 0.98 | 101 |
| effusion | 0.90 | 0.95 | 0.93 | 100 |
| lung Opacity | 0.96 | 0.92 | 0.94 | 100 |
| mass | 0.98 | 1.00 | 0.99 | 101 |
| nodule | 0.98 | 0.92 | 0.95 | 101 |
| pneumonia | 0.99 | 0.95 | 0.97 | 100 |
| pneumothorax | 0.98 | 0.98 | 0.98 | 100 |
| pulmunary fibrosis | 0.88 | 0.97 | 0.92 | 99 |
| tuberculosis | 0.99 | 0.95 | 0.97 | 101 |
| accuracy | | | 0.96 | 1019 |
| macro avg | 0.96 | 0.96 | 0.96 | 1019 |
| weighted avg | 0.96 | 0.96 | 0.96 | 1019 |

EfficientNetV2B0 - AdamW

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| control | 0.99 | 1.00 | 1.00 | 116 |
| covid | 0.99 | 0.97 | 0.98 | 101 |
| effusion | 0.93 | 0.94 | 0.94 | 100 |
| lung Opacity | 0.93 | 0.98 | 0.96 | 100 |
| mass | 0.93 | 0.96 | 0.95 | 101 |
| nodule | 0.87 | 0.94 | 0.90 | 101 |
| pneumonia | 0.92 | 0.99 | 0.95 | 100 |
| pneumothorax | 0.96 | 0.80 | 0.87 | 100 |
| pulmunary fibrosis | 0.98 | 0.95 | 0.96 | 99 |
| tuberculosis | 0.99 | 0.95 | 0.97 | 101 |
| accuracy | | | 0.95 | 1019 |
| macro avg | 0.95 | 0.95 | 0.95 | 1019 |
| weighted avg | 0.95 | 0.95 | 0.95 | 1019 |

EfficientNetV2B1 - Adam

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| control | 1.00 | 1.00 | 1.00 | 116 |
| covid | 1.00 | 0.98 | 0.99 | 101 |
| effusion | 0.87 | 0.88 | 0.88 | 100 |
| lung Opacity | 0.95 | 0.94 | 0.94 | 100 |
| mass | 0.99 | 0.89 | 0.94 | 101 |
| nodule | 0.93 | 0.94 | 0.94 | 101 |
| pneumonia | 0.92 | 0.95 | 0.94 | 100 |
| pneumothorax | 0.93 | 0.88 | 0.90 | 100 |
| pulmunary fibrosis | 0.90 | 0.96 | 0.93 | 99 |
| tuberculosis | 0.93 | 0.99 | 0.96 | 101 |
| accuracy | | | 0.94 | 1019 |
| macro avg | 0.94 | 0.94 | 0.94 | 1019 |
| weighted avg | 0.94 | 0.94 | 0.94 | 1019 |

EfficientNetV2B1 - AdamW

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| control | 0.97 | 1.00 | 0.99 | 116 |
| covid | 0.97 | 0.99 | 0.98 | 101 |
| effusion | 0.90 | 0.94 | 0.92 | 100 |
| lung Opacity | 0.96 | 0.97 | 0.97 | 100 |
| mass | 0.95 | 0.95 | 0.95 | 101 |
| nodule | 0.90 | 0.94 | 0.92 | 101 |
| pneumonia | 0.93 | 0.95 | 0.94 | 100 |
| pneumothorax | 0.94 | 0.85 | 0.89 | 100 |
| pulmunary fibrosis | 0.92 | 0.93 | 0.92 | 99 |
| tuberculosis | 0.97 | 0.88 | 0.92 | 101 |
| accuracy | | | 0.94 | 1019 |
| macro avg | 0.94 | 0.94 | 0.94 | 1019 |
| weighted avg | 0.94 | 0.94 | 0.94 | 1019 |

EfficientNetV2B2 - Adam

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| control | 0.97 | 1.00 | 0.99 | 116 |
| covid | 0.97 | 1.00 | 0.99 | 101 |
| effusion | 0.95 | 0.91 | 0.93 | 100 |
| lung Opacity | 0.95 | 0.94 | 0.94 | 100 |
| mass | 0.93 | 0.92 | 0.93 | 101 |
| nodule | 0.91 | 0.91 | 0.91 | 101 |
| pneumonia | 0.86 | 1.00 | 0.93 | 100 |
| pneumothorax | 0.93 | 0.90 | 0.91 | 100 |
| pulmunary fibrosis | 0.93 | 0.92 | 0.92 | 99 |
| tuberculosis | 1.00 | 0.88 | 0.94 | 101 |
| accuracy | | | 0.94 | 1019 |
| macro avg | 0.94 | 0.94 | 0.94 | 1019 |
| weighted avg | 0.94 | 0.94 | 0.94 | 1019 |

EfficientNetV2B2 - AdamW

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| control | 0.98 | 1.00 | 0.99 | 116 |
| covid | 1.00 | 0.97 | 0.98 | 101 |
| effusion | 0.92 | 0.94 | 0.93 | 100 |
| lung Opacity | 0.97 | 0.92 | 0.94 | 100 |
| mass | 0.98 | 0.92 | 0.95 | 101 |
| nodule | 0.90 | 0.92 | 0.91 | 101 |
| pneumonia | 0.94 | 0.97 | 0.96 | 100 |
| pneumothorax | 0.90 | 0.96 | 0.93 | 100 |
| pulmunary fibrosis | 0.92 | 0.96 | 0.94 | 99 |
| tuberculosis | 0.97 | 0.91 | 0.94 | 101 |
| accuracy | | | 0.95 | 1019 |
| macro avg | 0.95 | 0.95 | 0.95 | 1019 |
| weighted avg | 0.95 | 0.95 | 0.95 | 1019 |

EfficientNetV2B3 - Adam

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| control | 0.96 | 1.00 | 0.98 | 116 |
| covid | 0.98 | 1.00 | 0.99 | 101 |
| effusion | 0.94 | 0.91 | 0.92 | 100 |
| lung Opacity | 0.95 | 0.92 | 0.93 | 100 |
| mass | 0.93 | 0.91 | 0.92 | 101 |
| nodule | 0.99 | 0.95 | 0.97 | 101 |
| pneumonia | 0.93 | 0.94 | 0.94 | 100 |
| pneumothorax | 0.83 | 0.92 | 0.87 | 100 |
| pulmunary fibrosis | 0.93 | 0.96 | 0.95 | 99 |
| tuberculosis | 0.99 | 0.89 | 0.94 | 101 |
| accuracy | | | 0.94 | 1019 |
| macro avg | 0.94 | 0.94 | 0.94 | 1019 |
| weighted avg | 0.94 | 0.94 | 0.94 | 1019 |

EfficientNetV2B3 - AdamW

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| control | 0.98 | 1.00 | 0.99 | 116 |
| covid | 0.98 | 0.98 | 0.98 | 101 |
| effusion | 0.90 | 0.90 | 0.90 | 100 |
| lung Opacity | 0.97 | 0.91 | 0.94 | 100 |
| mass | 0.90 | 0.94 | 0.92 | 101 |
| nodule | 0.94 | 0.93 | 0.94 | 101 |
| pneumonia | 0.96 | 0.91 | 0.93 | 100 |
| pneumothorax | 0.93 | 0.91 | 0.92 | 100 |
| pulmunary fibrosis | 0.93 | 0.94 | 0.93 | 99 |
| tuberculosis | 0.90 | 0.95 | 0.92 | 101 |
| accuracy | | | 0.94 | 1019 |
| macro avg | 0.94 | 0.94 | 0.94 | 1019 |
| weighted avg | 0.94 | 0.94 | 0.94 | 1019 |

VGG16 - Adam

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| control | 0.89 | 1.00 | 0.94 | 116 |
| covid | 0.94 | 0.92 | 0.93 | 101 |
| effusion | 0.76 | 0.65 | 0.70 | 100 |
| lung Opacity | 0.38 | 0.48 | 0.42 | 100 |
| mass | 0.64 | 0.64 | 0.64 | 101 |
| nodule | 0.66 | 0.60 | 0.63 | 101 |
| pneumonia | 0.61 | 0.47 | 0.53 | 100 |
| pneumothorax | 0.58 | 0.51 | 0.54 | 100 |
| pulmunary fibrosis | 0.59 | 0.69 | 0.63 | 99 |
| tuberculosis | 0.52 | 0.52 | 0.52 | 101 |
| accuracy | | | 0.65 | 1019 |
| macro avg | 0.66 | 0.65 | 0.65 | 1019 |
| weighted avg | 0.66 | 0.65 | 0.65 | 1019 |

VGG16 - AdamW

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| control | 0.98 | 1.00 | 0.99 | 116 |
| covid | 0.94 | 0.94 | 0.94 | 101 |
| effusion | 0.83 | 0.50 | 0.62 | 100 |
| lung Opacity | 0.84 | 0.76 | 0.80 | 100 |
| mass | 0.60 | 0.57 | 0.59 | 101 |
| nodule | 0.76 | 0.61 | 0.68 | 101 |
| pneumonia | 0.59 | 0.92 | 0.72 | 100 |
| pneumothorax | 0.65 | 0.57 | 0.61 | 100 |
| pulmunary fibrosis | 0.57 | 0.78 | 0.66 | 99 |
| tuberculosis | 0.75 | 0.68 | 0.72 | 101 |
| accuracy | | | 0.74 | 1019 |
| macro avg | 0.75 | 0.73 | 0.73 | 1019 |
| weighted avg | 0.76 | 0.74 | 0.74 | 1019 |

VGG19 - Adam

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| control | 0.95 | 1.00 | 0.97 | 116 |
| covid | 0.86 | 0.94 | 0.90 | 101 |
| effusion | 0.65 | 0.51 | 0.57 | 100 |
| lung Opacity | 0.64 | 0.78 | 0.70 | 100 |
| mass | 0.54 | 0.70 | 0.61 | 101 |
| nodule | 0.75 | 0.72 | 0.74 | 101 |
| pneumonia | 0.70 | 0.57 | 0.63 | 100 |
| pneumothorax | 0.84 | 0.49 | 0.62 | 100 |
| pulmunary fibrosis | 0.66 | 0.68 | 0.67 | 99 |
| tuberculosis | 0.58 | 0.67 | 0.62 | 101 |
| accuracy | | | 0.71 | 1019 |
| macro avg | 0.72 | 0.71 | 0.70 | 1019 |
| weighted avg | 0.72 | 0.71 | 0.71 | 1019 |

VGG19 - AdamW

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| control | 0.94 | 1.00 | 0.97 | 116 |
| covid | 0.88 | 0.92 | 0.90 | 101 |
| effusion | 0.77 | 0.61 | 0.68 | 100 |
| lung Opacity | 0.65 | 0.68 | 0.66 | 100 |
| mass | 0.79 | 0.69 | 0.74 | 101 |
| nodule | 0.74 | 0.76 | 0.75 | 101 |
| pneumonia | 0.70 | 0.76 | 0.73 | 100 |
| pneumothorax | 0.81 | 0.66 | 0.73 | 100 |
| pulmunary fibrosis | 0.72 | 0.80 | 0.76 | 99 |
| tuberculosis | 0.65 | 0.73 | 0.69 | 101 |
| accuracy | | | 0.77 | 1019 |
| macro avg | 0.77 | 0.76 | 0.76 | 1019 |
| weighted avg | 0.77 | 0.77 | 0.76 | 1019 |