



CY Tech
Big Data Deep Learning

Lung X-ray Covid Classification

A presentaion by Ayush Bajracharya

18 December, 2024



Introduction

The data aims to build an image classification model based on Deep Learning to detect COVID-19 infections using chest X-ray images or CT scans. The model is used to classify images into multiple categories, such as:

1. COVID-19 (cases of coronavirus infection)
2. Normal (healthy individuals)
3. Pneumonia (pneumonia unrelated to COVID-19).

Data Preparation

Dataset Overview:

- Dataset organization (train/test split).
- Labels: Covid, Normal, Viral Pneumonia.

Preprocessing:

- Image resizing to 224x224.
- Class distribution visualization.
- Use of techniques like SMOTE for balancing. (worse results)



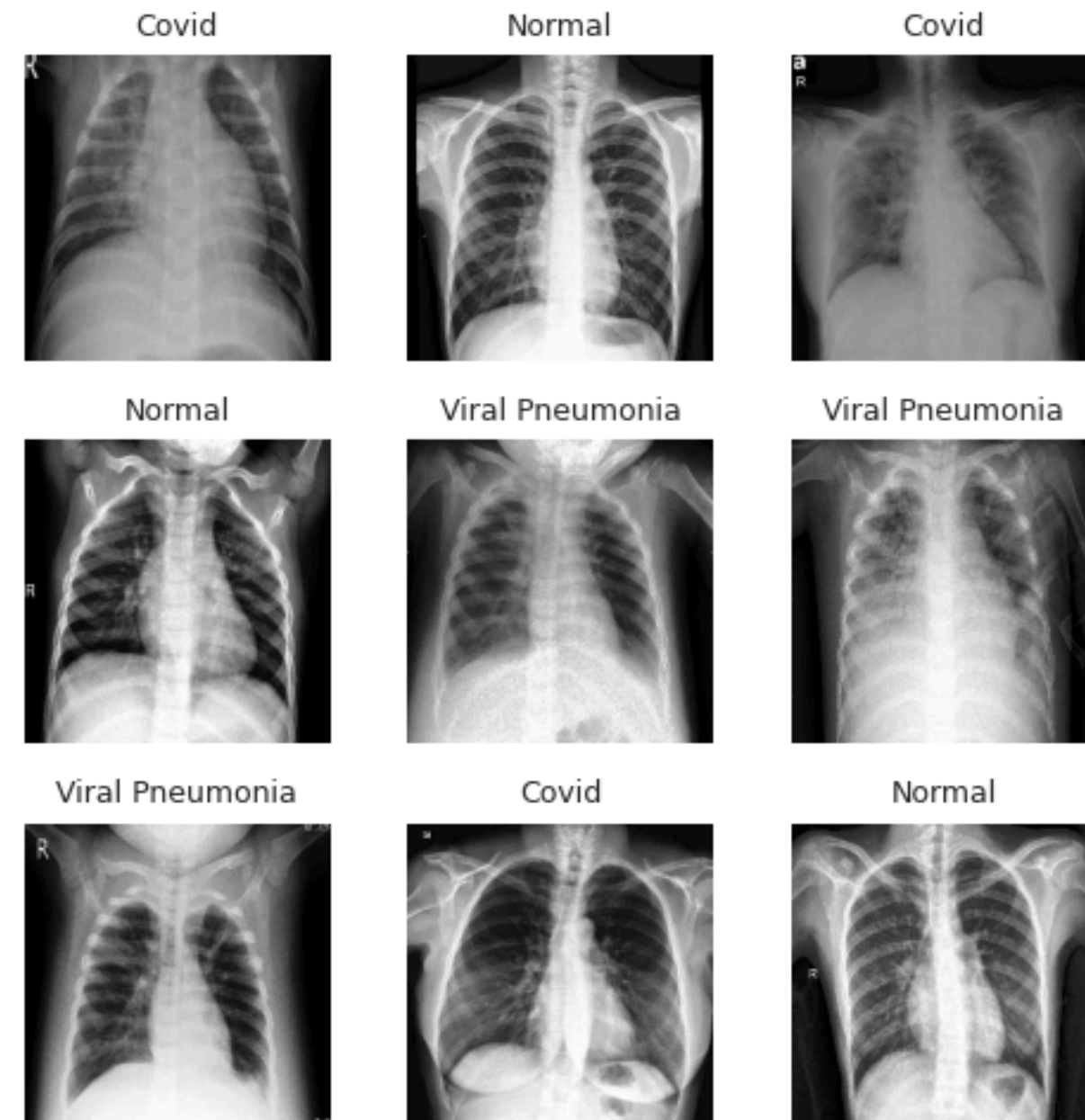
Data Augmentation and Exploration

Image Augmentation:

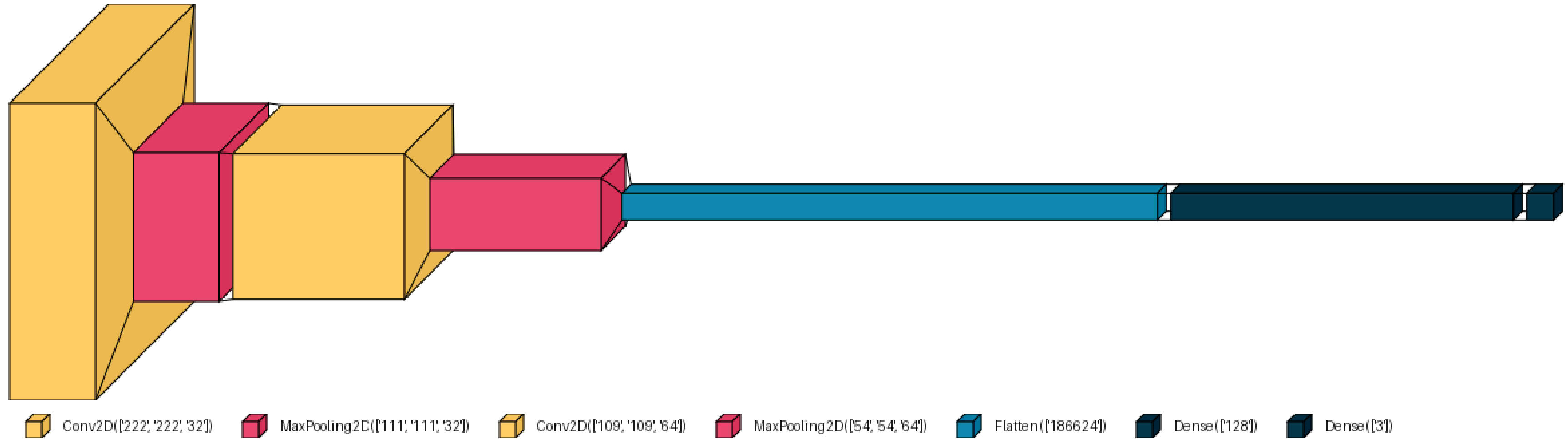
- Display sample images with labels.

Data Dimensions:

- Training data dimensions and reshaping process.



Model Architecture



Model Design



- Sequential CNN model with:
 - 2 Convolutional layers.
 - 2 MaxPooling layers.
 - Fully connected dense layer.
- Activation functions:
 - Relu and Softmax
- Loss:
 - Categorical Crossentropy
- Metrics:
 - Accuracy

Training Setup

```
# Learning rate reduction on plateau
learning_rate_reduction = ReduceLROnPlateau(
    monitor='val_accuracy', # Monitor validation accuracy
    patience=2, # Number of epochs without improvement before reducing the learning rate
    verbose=1, # Display a message when reducing the learning rate
    factor=0.3, # Reduce learning rate by 70%
    min_lr=0.000001 # Minimum learning rate
)

# Early stopping to prevent overfitting
early_stopping = EarlyStopping(
    monitor='val_loss', # Monitor validation loss
    patience=5, # Number of epochs without improvement before stopping
    restore_best_weights=True # Restore the best weights after training stops
)
```

- Epochs:
 - 50
- Batch Size:
 - 64
- Callbacks:
 - Learning Rate Reduction
 - Early stopping

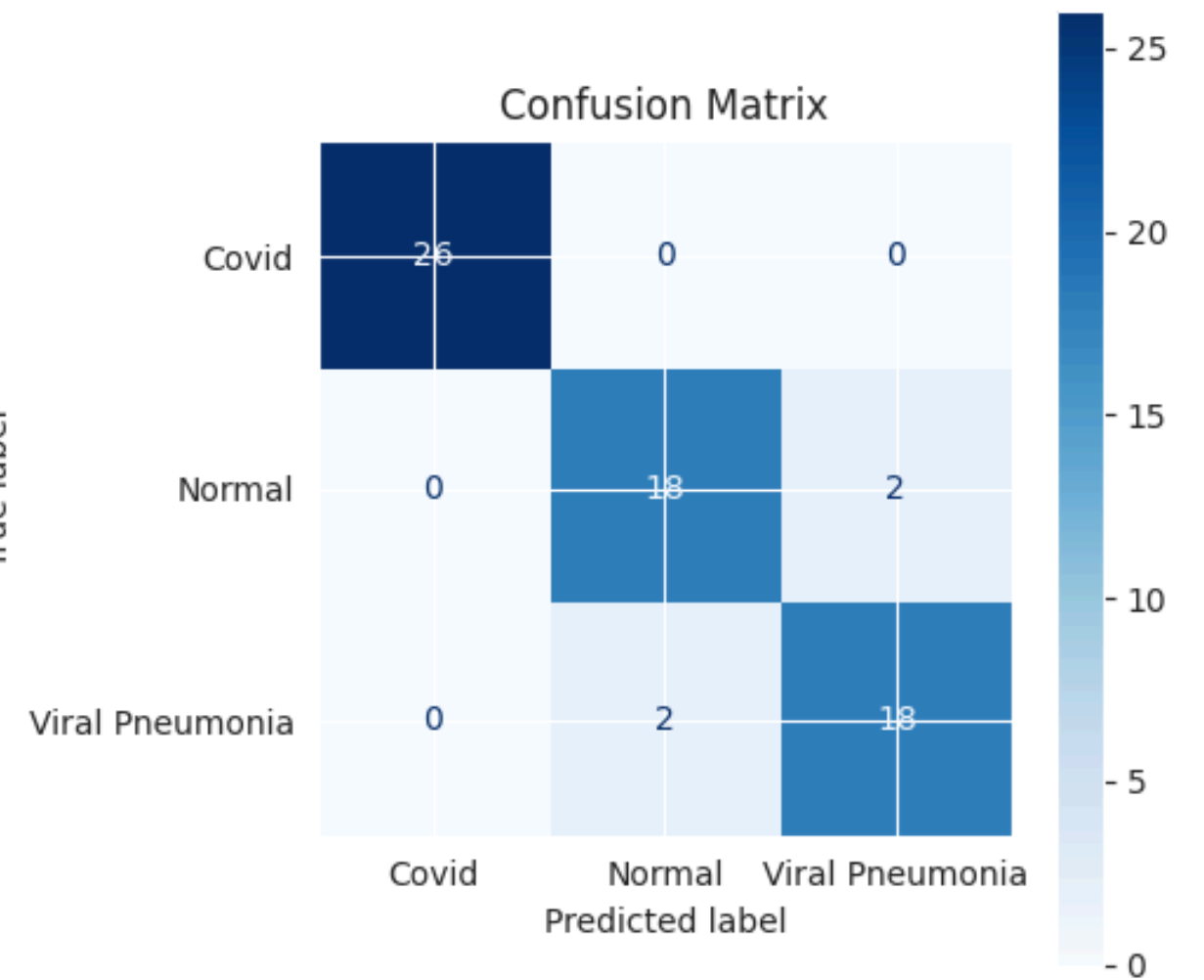
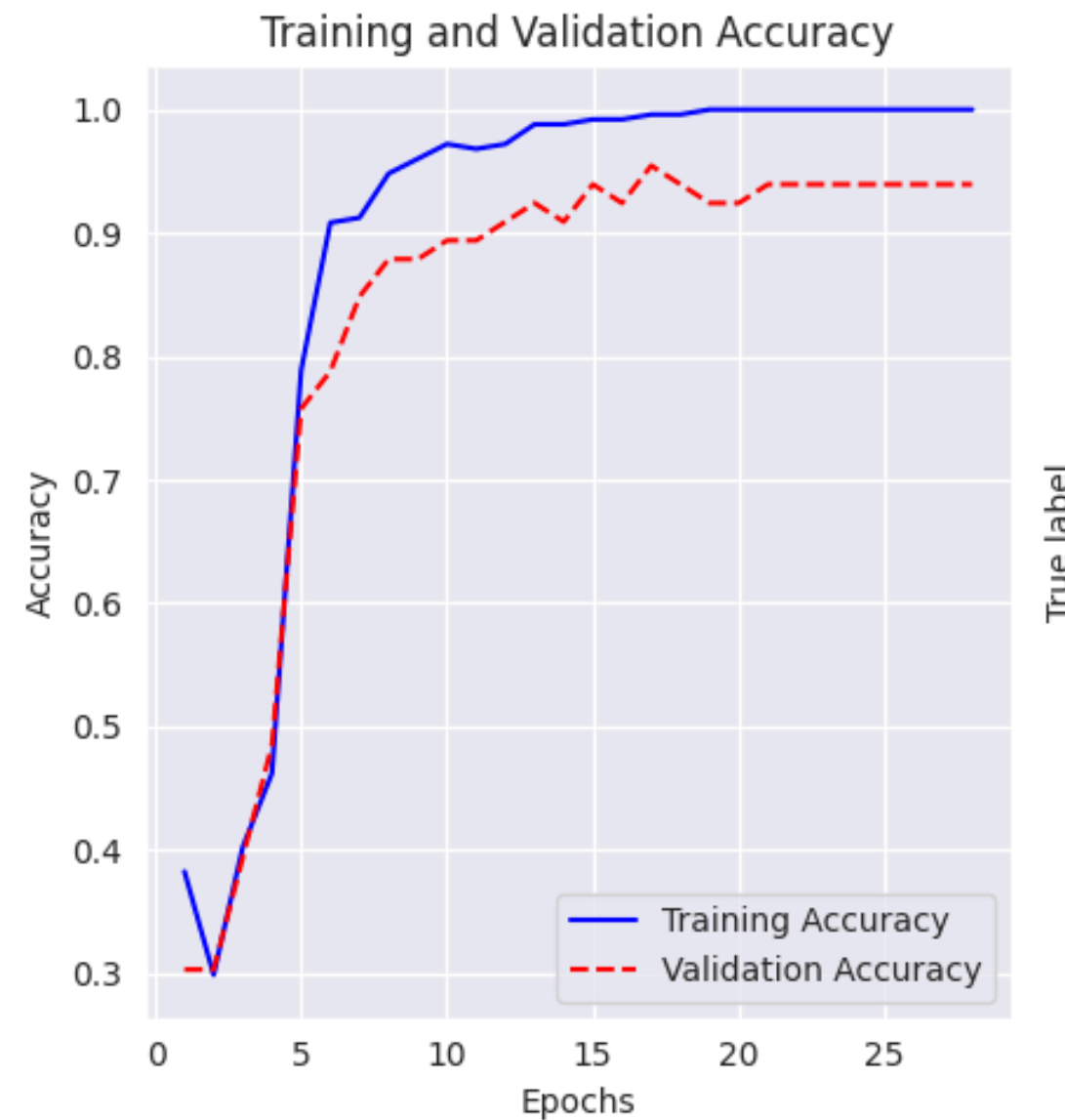
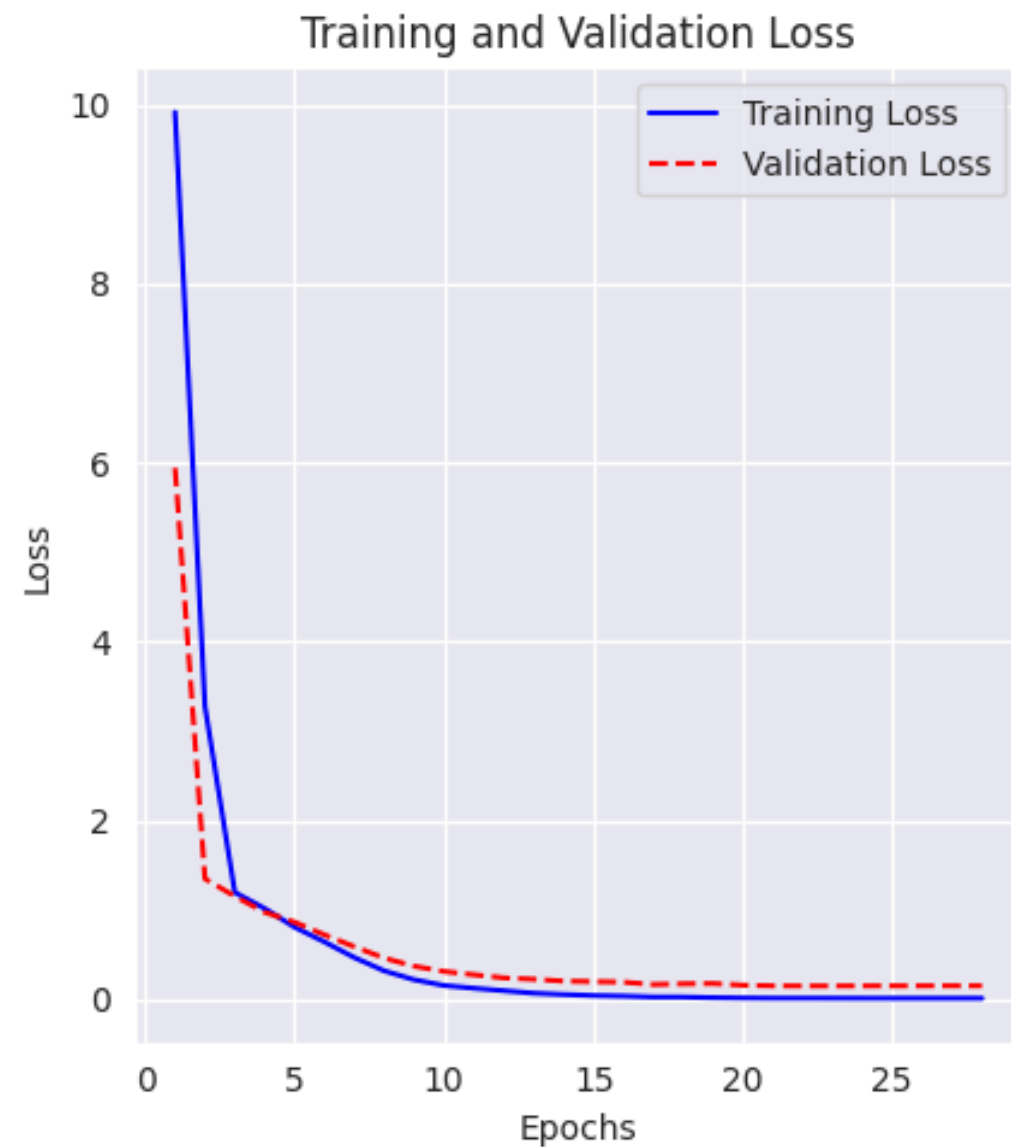
Model Comparision

	First Model	Second Model	Third Model	Fourth Model	Fifth Model
Changes	Image Data Generator (IDG) 150 Input 1 Output (GreyScale)	Remove IDG SMOTE 5 CN Layers	224 Input 3 CN Layers 3 Output	Remove SMOTE Learning Rate Adam Optimizer	2 CN Layers AdamX Optimizer
Accuracy	50%	77%	82%	89%	94%

Best Model Metrics

Runtime: ~10 mins
(with early stop)

	precision	recall	f1-score	support
Covid	1.00	1.00	1.00	26
Normal	0.90	0.90	0.90	20
Viral Pneumonia	0.90	0.90	0.90	20
accuracy			0.94	66
macro avg	0.93	0.93	0.93	66
weighted avg	0.94	0.94	0.94	66





Thank You

18 December, 2024