Pneumonia detection in chest X-ray images using deep learning models

Presenter : Yiang Yuet Meng

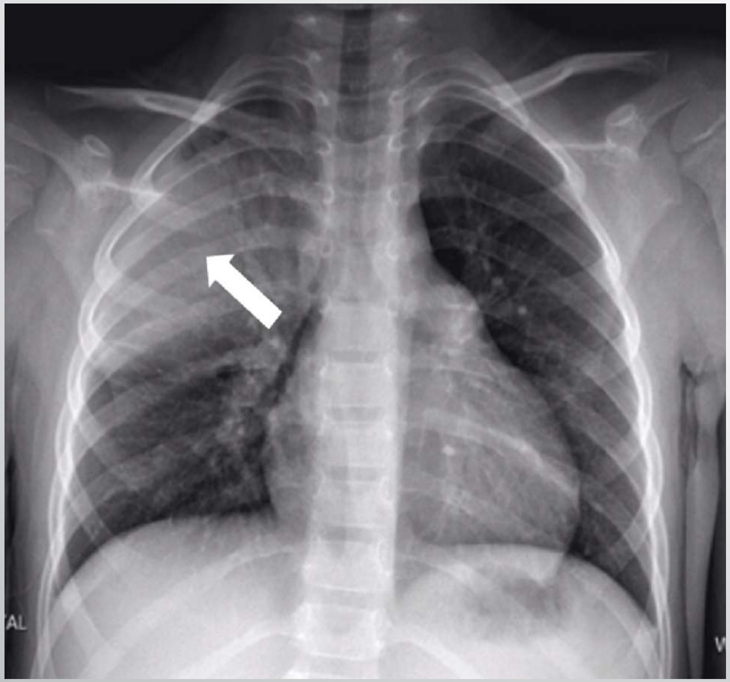
Agenda

- Background
- Data Cleaning/Exploratory Data Analysis
- Feature Engineering
- Evaluation Metrics
- CNN Model
- Modelling
- Error Analysis
- Summary

Background

- There is an outbreak of bacteria in the community that is affecting children and elderly. The infection can be detected via Chest X ray.
- The normal chest X-ray depicts clear lungs without any areas of abnormal opacification in the image.
- Bacterial pneumonia typically exhibits a focal lobar consolidation
- whereas viral pneumonia manifests with a more diffuse “interstitial” pattern in both lungs.
- We are establishing a diagnostic tool based on a deep-learning framework for the screening of patients that are affected by this bacteria so that doctors can focus on the treatment

Background

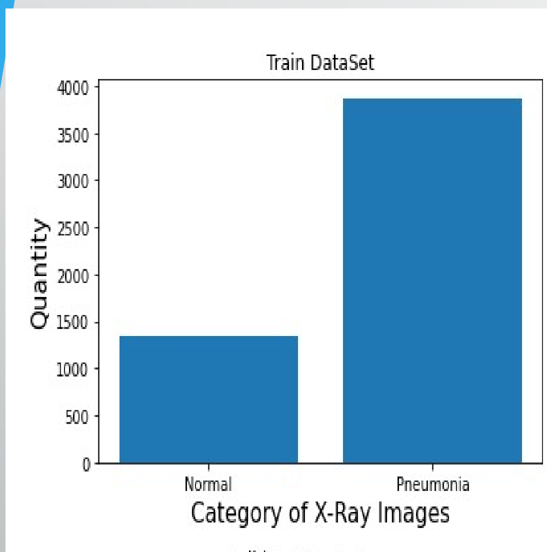


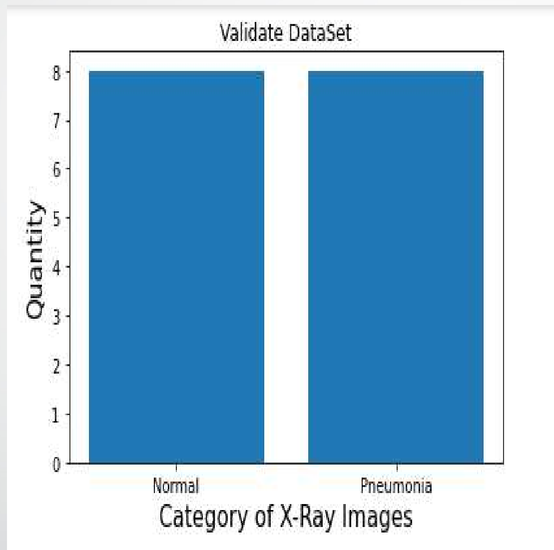
focal lobar consolidation

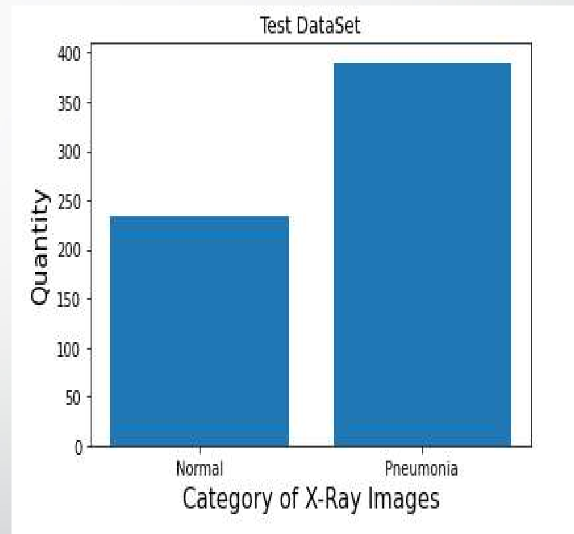
diffuse “interstitial” in both lungs

Data Cleaning/Exploratory Data Analysis

- Datasets:





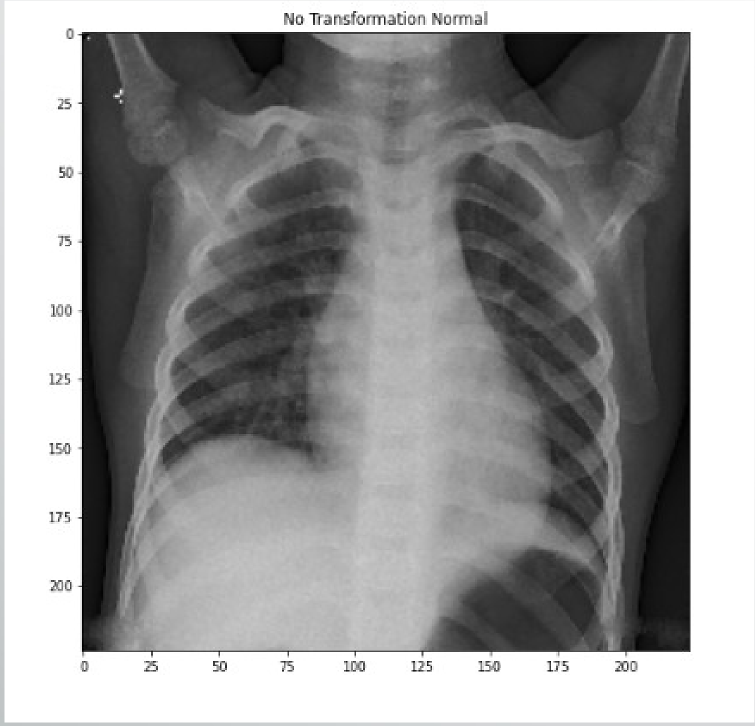


Normal Train: B&W 1341 pictures
Pneumonia Train: B&W 3875 pictures

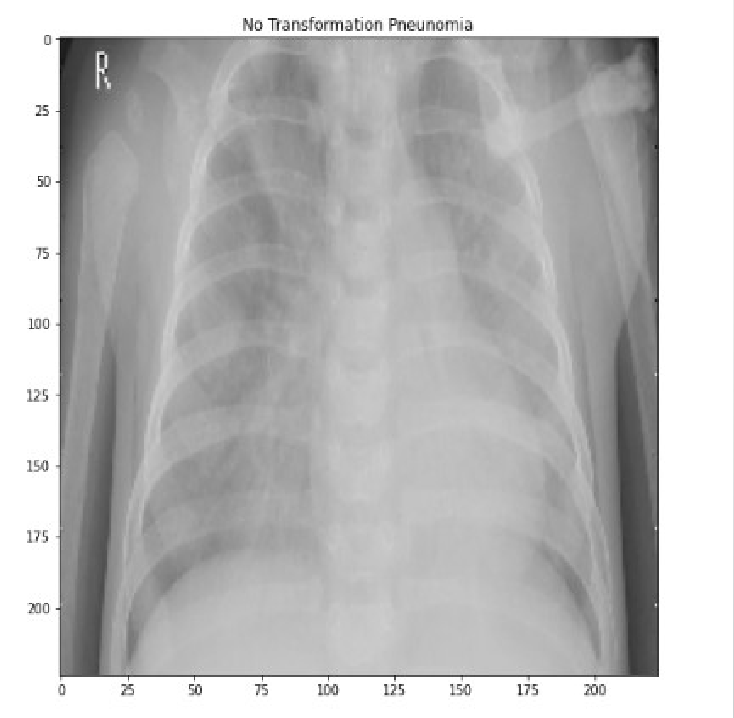
Normal Val: B&W 8 pictures
Pneumonia Val: B&W 8 pictures

Normal Test: B&W 234 pictures
Pneumonia Test: B&W 390 pictures

Data Cleaning/Exploratory Data Analysis



Normal Chest X-Ray (label as 0)



Pneumonia Chest X-Ray (label as 1)

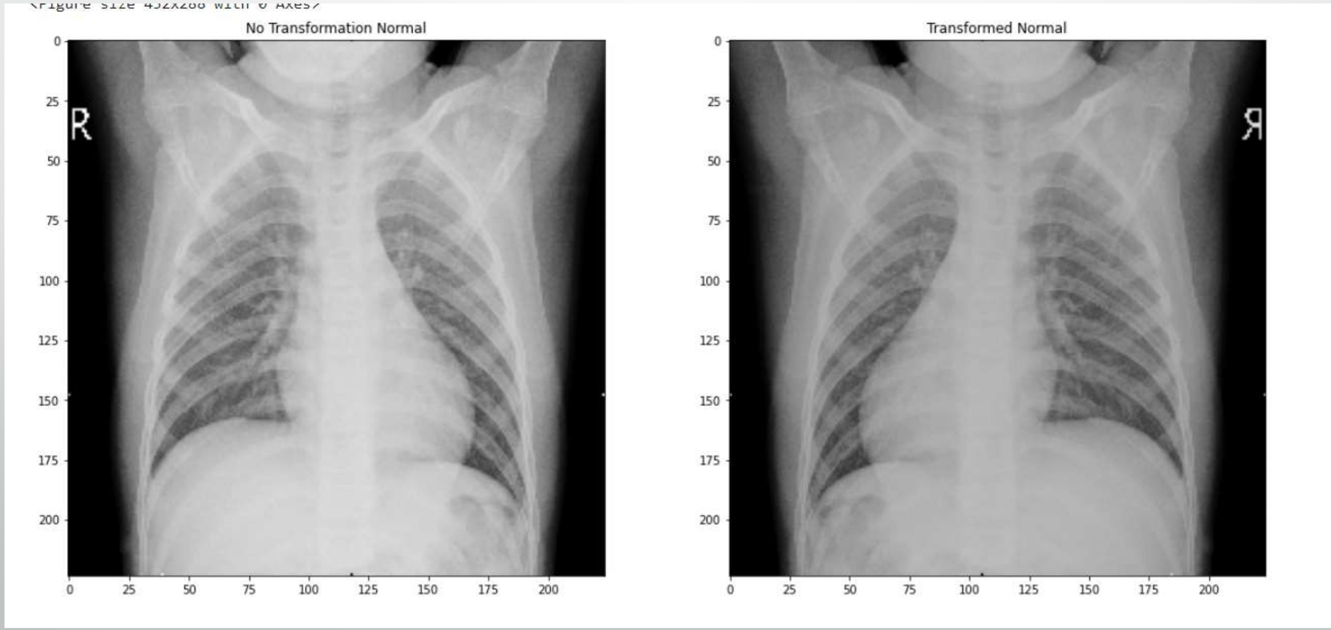
Above plot confirms that for Pneumonia cases, there are white patches in the lung area as compared to Normal case. In order to make a stronger contrast between the 2 types of images, data augmentation (such as tune brightness, horizonal and vertical flip) is recommended

Feature Engineering

Data augmentation will increase the diversity of the training set. This is done by

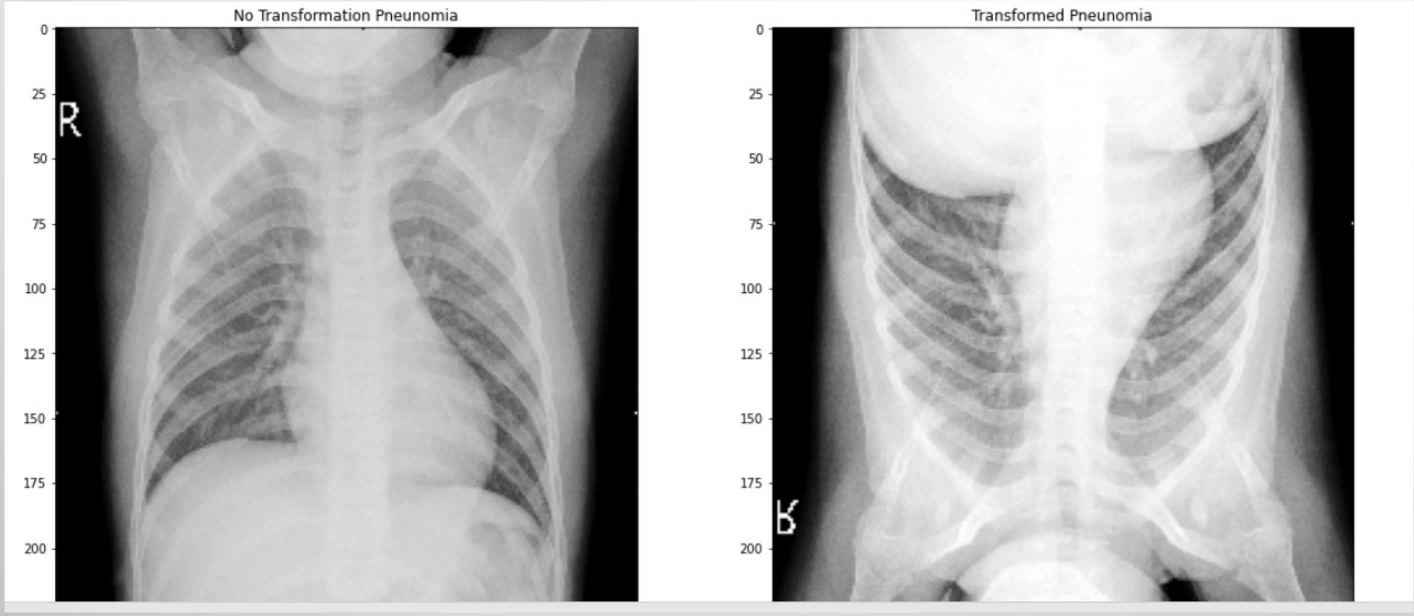
- image rotation,
- tune the brightness of the images,
- flip the images vertically and horizontally.

- Below transformed image for a Normal case is darker and horizonal flipped.



Feature Engineering

- This is a Pneumonia case. Transformed image brighter, horizontal/vertical flip.



Evaluation Metrics

In the selection of the appropriate model, the metrics used will be the accuracy and recall. Objective is to increase accuracy yet with low level of false negative

Accuracy = (tp+tn)/(tp+fp+tn+fn)

Recall= tp/(tp+fn)

On top of it, it needs to be supported with a plot of accuracy and loss of training and validation dataset

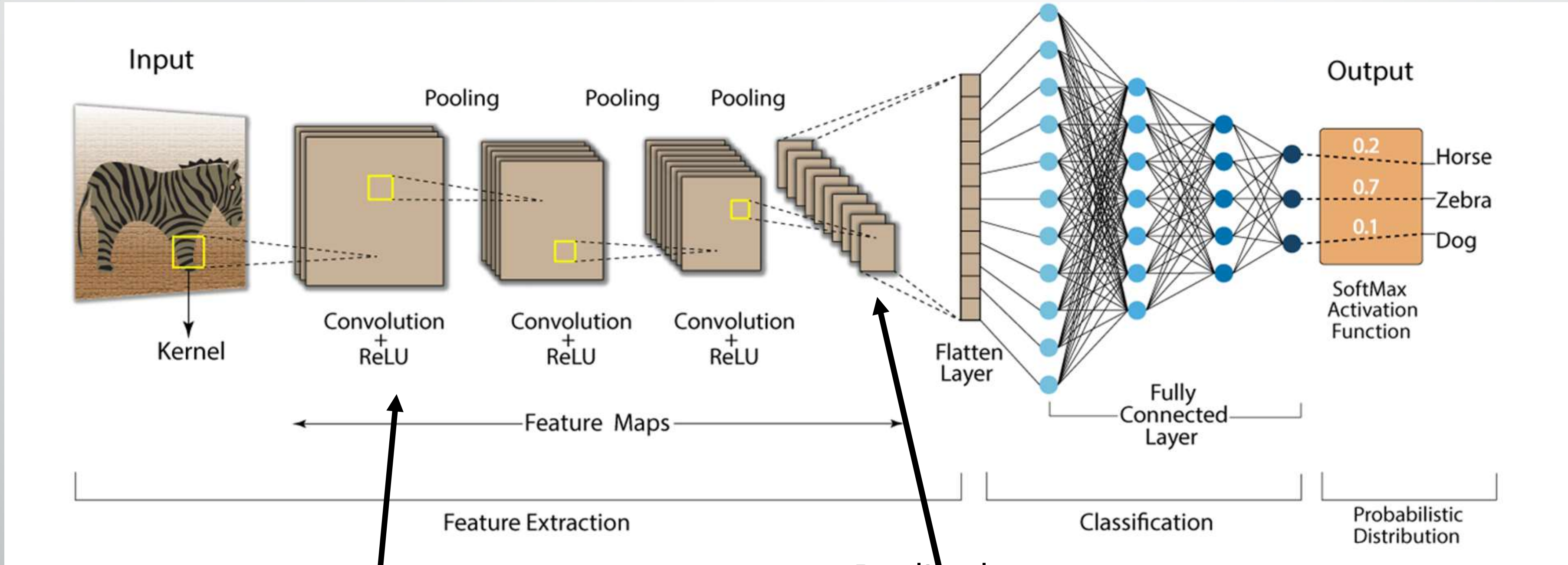
tp :True Positive

fn : False Negative

tn : True Negative

fn : False Negative

CNN Model

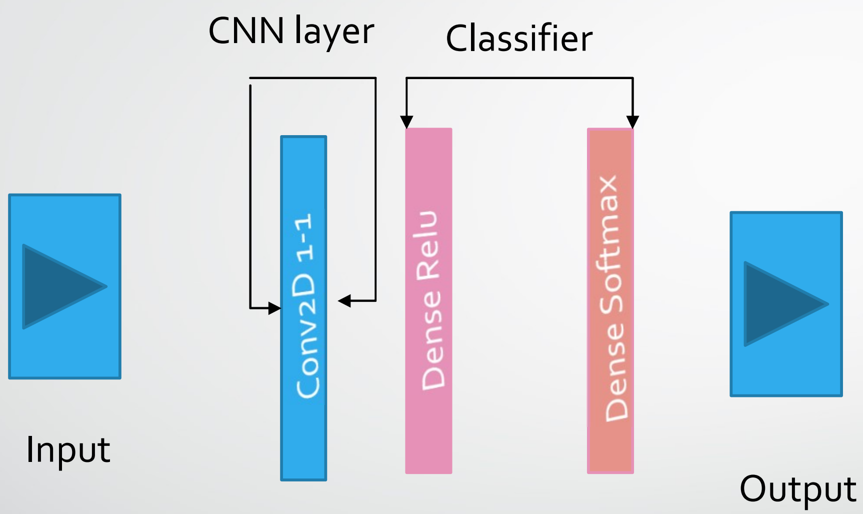


Convolution layers are used to detect features in images

Pooling layers are used to compress the output of Convolution layers

Modelling

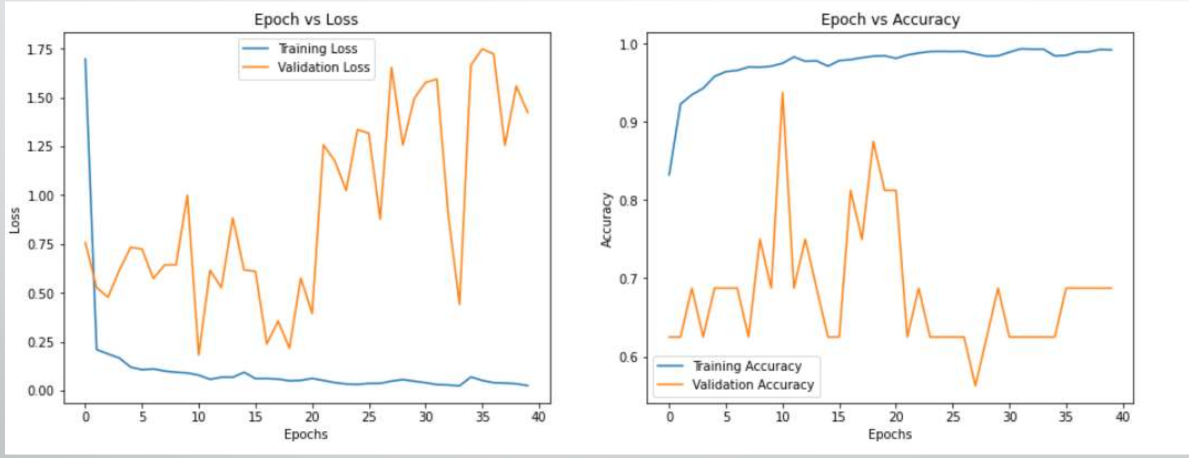
Base Model- CNN architecture



This is architecture for Base Model

Modelling

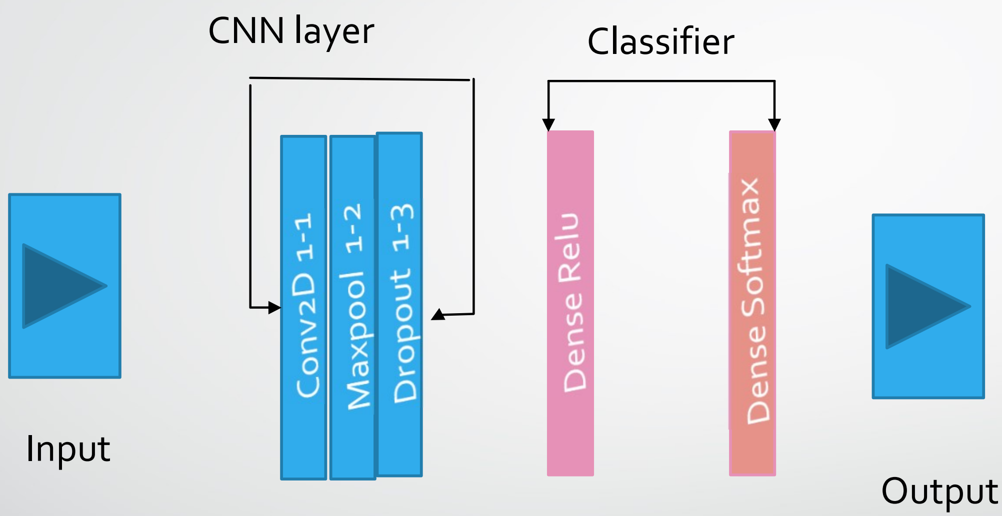
Model	HyperParameters	Epoch	Accuracy	Misclass	Precision	Recall	Specificity
Base CNN- 3 layers	learning_rate=0.000001, beta_1=0.9, beta_2=0.999, epsilon=1e-08, decay=0.000005	40	0.7163	0.2837	0.6969	0.9667	0.2991



The graphs seemed to suggest that overfitting seemed to occur since Training Accuracy/Loss improves while Validation Accuracy/Loss got worst as epoch increases

Modelling

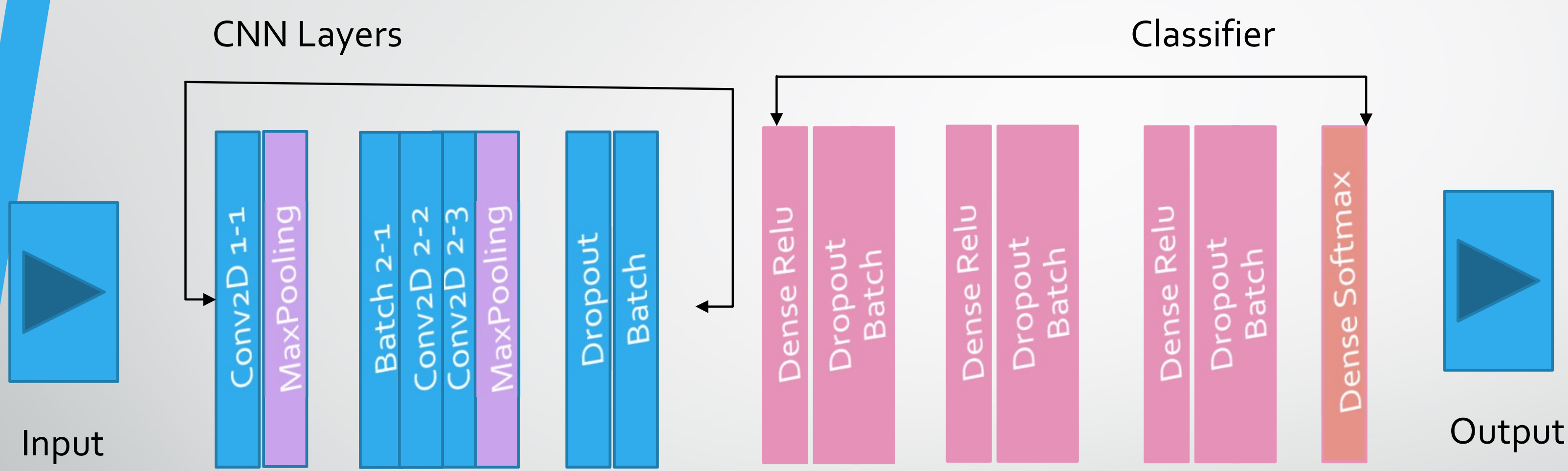
Model 1- CNN architecture



This is architecture for Model 1

Modelling

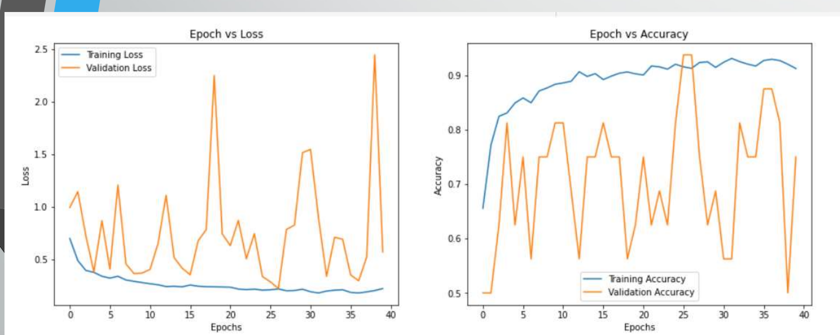
Model 2- CNN architecture



This is model 2 architecture

Modelling

Model	HyperParameters	Epoch	Accuracy	Misclass	Precision	Recall	Specificity
Model 1 - CNN- 7 layers	learning_rate=0.000001, beta_1=0.9, beta_2=0.999, epsilon=1e-08, decay=0.000005	40	0.7115	0.2885	0.6959	0.9564	0.3034
Model 2 - CNN- 19 layers	learning_rate=0.000001, beta_1=0.9, beta_2=0.999, epsilon=1e-08, decay=0.000005	40	0.7308	0.2692	0.6996	0.9974	0.2863



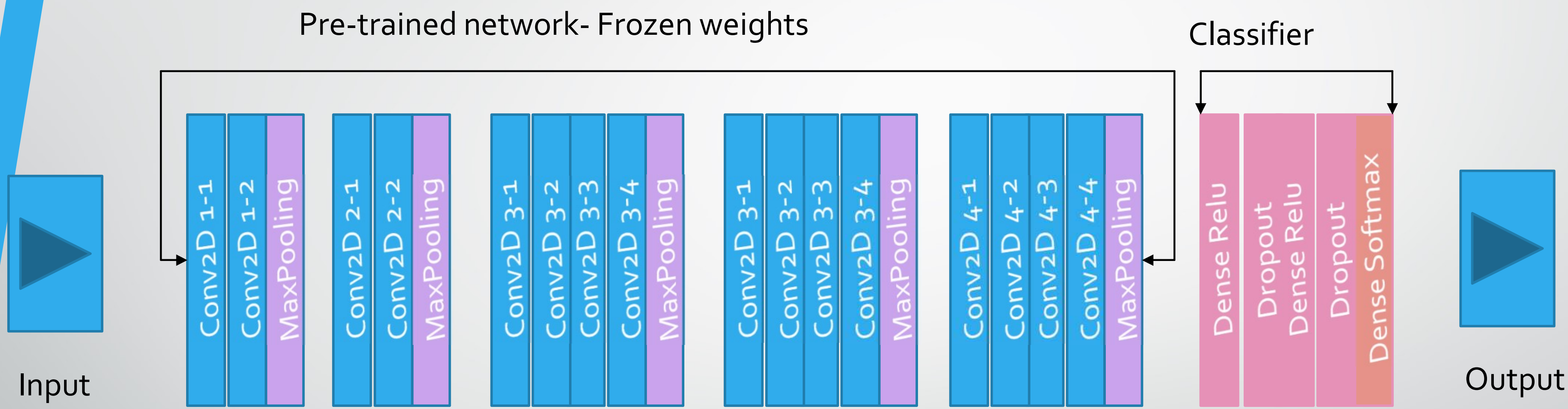
The plot on accuracy and loss of train and validation improves. However, the validation data accuracy and loss is fluctuating.

CNN model built from scratch may not be a desirable model as large amount of data is necessary to acquire the desired accuracy. Next is to try out Transfer Learning.

Transfer learning is a work-around to surmount this obstacle. A pre-trained model on a large dataset is re-used and the network weights determined are applied

Modelling

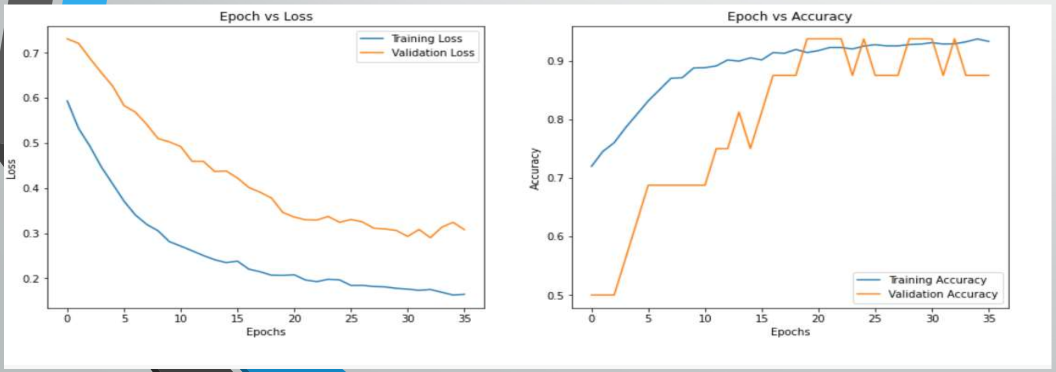
VGG-19 architecture



For model 3 and model 4, I will be using the above pre-trained VGG-19 architecture with no change in classifier layers

Modelling

Model	HyperParameters	Epoch	Accuracy	Misclass	Precision	Recall	Specificity
Model 3 - VGG 19	decay=0.000005, amsgrad: False, betas: (0.9, 0.999) epsilon: 1e-08, learning rate: 1e-06,maximize: False weight_decay: 0	40	0.851	0.149	0.8293	0.959	0.6709
Model 4 - VGG 19	decay=0.000005, amsgrad: False, betas: (0.9, 0.999) epsilon: 1e-08, learning rate: 1e-06, maximize: False weight_decay: 0, Earlystopping	35	0.8542	0.1458	0.8286	0.9667	0.6667

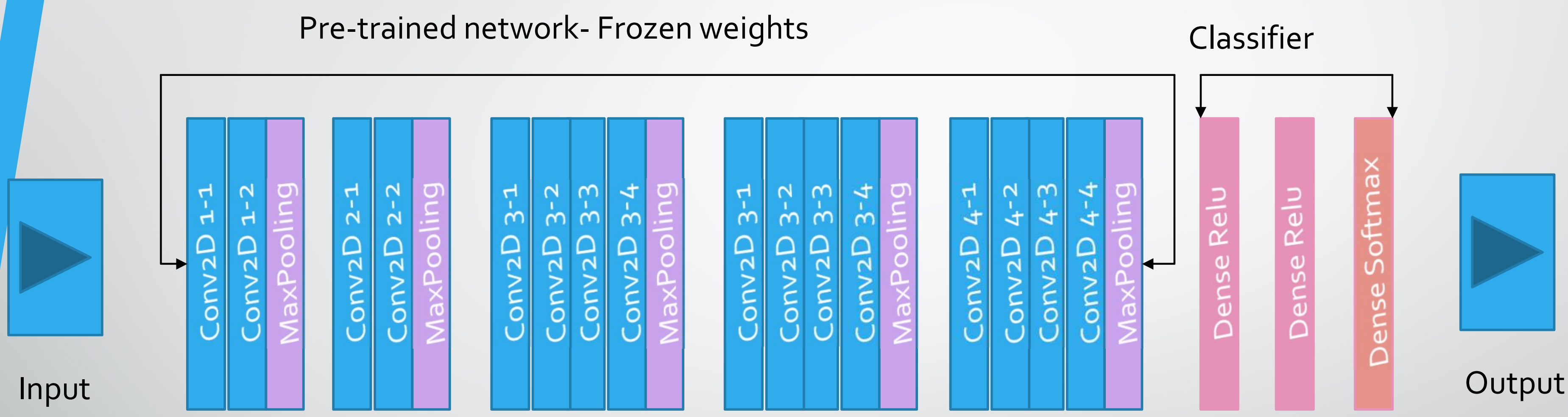


The accuracy and recall both improved using Transfer Learning.

The plot on loss and accuracy of train and validation dataset also improves.

Modelling

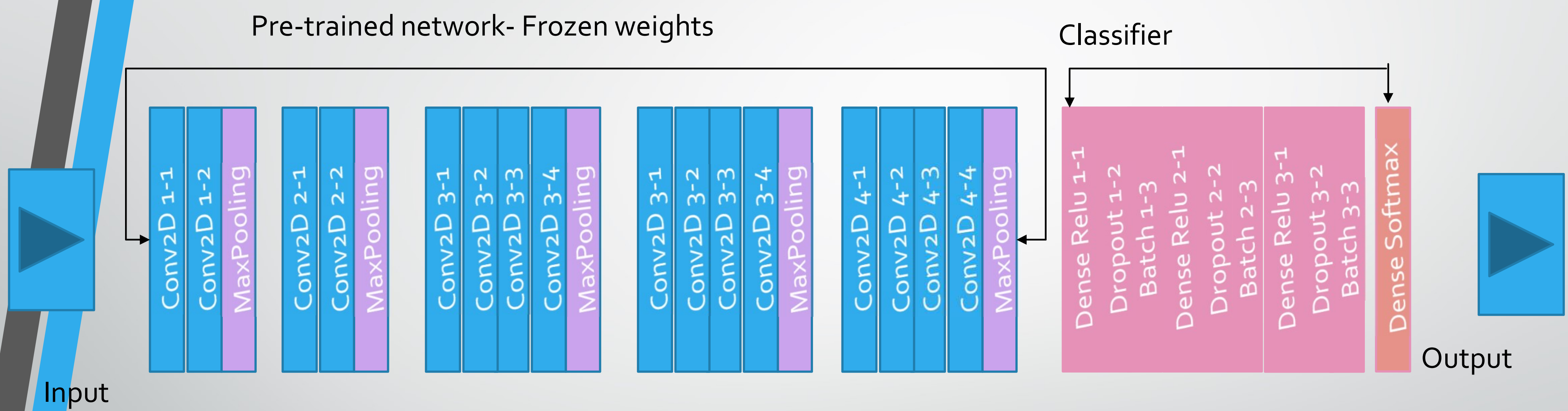
Model 5- VGG-19 architecture



In model 5, I will remove some layers in the Classifier layers.

Modelling

Model 6- VGG-19 architecture

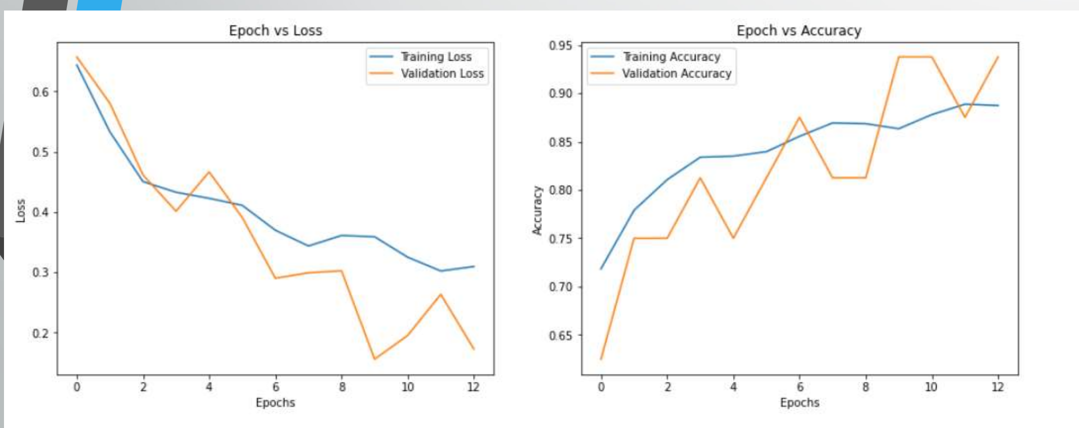


In model 6, I will add in more layers to the Classifier layers. Batch Normalisation layers will be introduced.

It is a technique to train deep neural networks to standardise the inputs to a layer for every mini-batch to stabilise the learning process and reduce the number of training epochs in deep network training

Modelling

Model	HyperParameters	Epoch	Accuracy	Misclass	Precision	Recall	Specificity
Model 5 - VGG 19	initial_learning_rate=1e-6, decay_steps=100000, decay_rate=0.000005 earlystopping	6	0.7981	0.2019	0.7568	0.9974	0.4658
Model 6 - VGG 19	initial_learning_rate=1e-6, decay_steps=100000, decay_rate=0.000005 earlystopping	13	0.8814	0.1186	0.878	0.941	0.7821

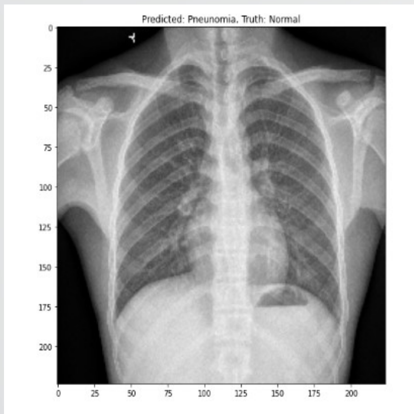


The accuracy and recall both improved using Transfer Learning.

The plot on loss and accuracy of train and validation dataset also improves.

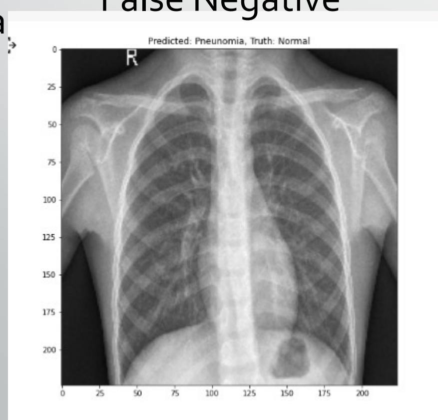
Error Analysis

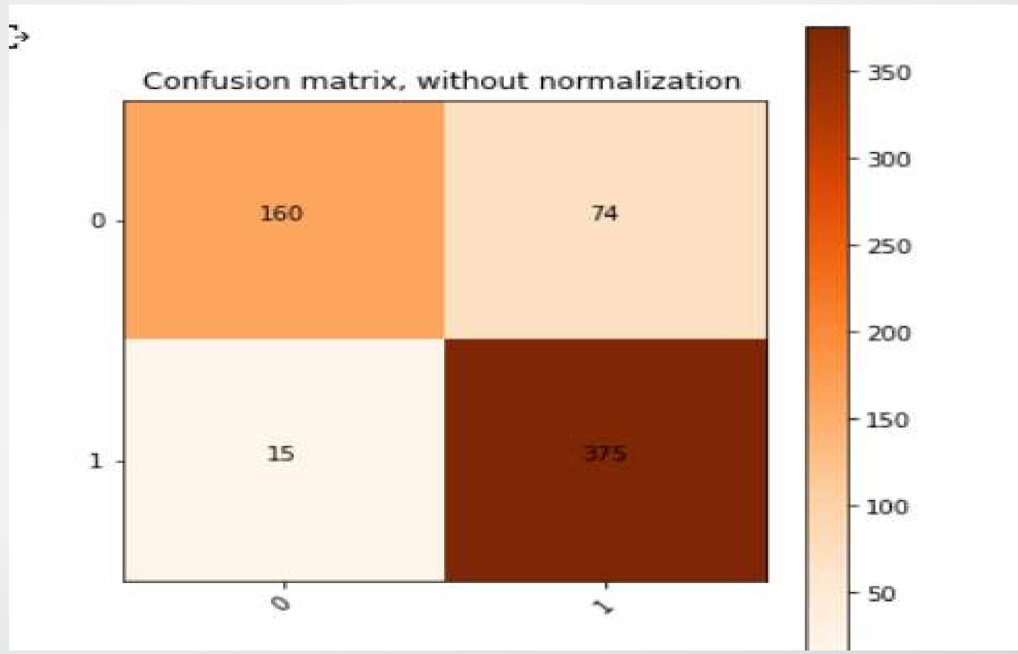
False Positive



Legend
0: Normal
1: Pneumonia

False Negative





- This may be due to the poor image quality, where the contrast of the image is not adequate, resulting in the base learners classifying the sample incorrectly.
- It is plausible that such an early stage of pneumonia, where the white infiltrates have just started to appear sparingly in the lungs, was not captured by the CNN models.

Summary

A vast amount of labeled data is difficult to acquire because it requires that expert doctors classify each image, which is an expensive and time-consuming task.

- Transfer learning is a work-around to surmount this obstacle. A model trained on a large dataset is re-used and the network weights determined in this model are applied.

- The final model is taking weights from the pre-trained model. I introduced additional layers after the Flatten layer so as to customise the weights to the X-Ray images.

There is still room for improvement. In the future, we may investigate techniques such as

- contrast enhancement of the images or other pre-processing steps to improve the image quality.

- segmentation of the lung image before classification to enable the CNN models to achieve improved feature extraction.

- ensemble of pre-trained models