Diagnosis of Chest X-Ray images

using Deep Convolution Network

Project Mentor - Prof. Hari Om

Presented By:

Avi Sahney (17JE003050)
Devashish Gupta (17JE002834)
Aditya Sharma (17JE002941)
Rohit Kumar (17JE003215)

Abstract

A chest X-ray is a fast and painless imaging test that uses certain electromagnetic waves to create pictures of the structures in and around your chest.

This test can help diagnose and monitor conditions such as pneumonia, heart failure, lung cancer, tuberculosis, sarcoidosis, and lung tissue scarring, called fibrosis.

The NIH recently released a chest X-Ray dataset comprising 112,120 frontal-view X-ray images of 30,805 unique patients with fourteen disease image labels, mined using NLP (natural language processing).

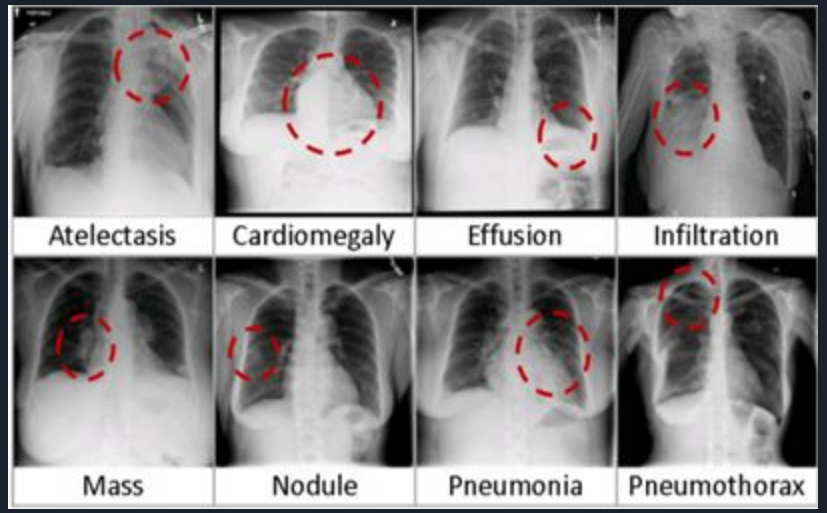
Why Deep Convolutional Networks?

Deep Convolutional neural networks have been widely used to solve computer vision problems such as image recognition and disease predictions in image data. We used a wide array of deep neural network architectures to train models for highly accurate predictions of diseases in the Chest X-Ray.

Deep learning techniques are state-of-the-art learning algorithms and perform quite well in medical image analysis. Our objective is to train models on the provided dataset to generate accurate predictions on new data and localizing the diseases in the Xray images.

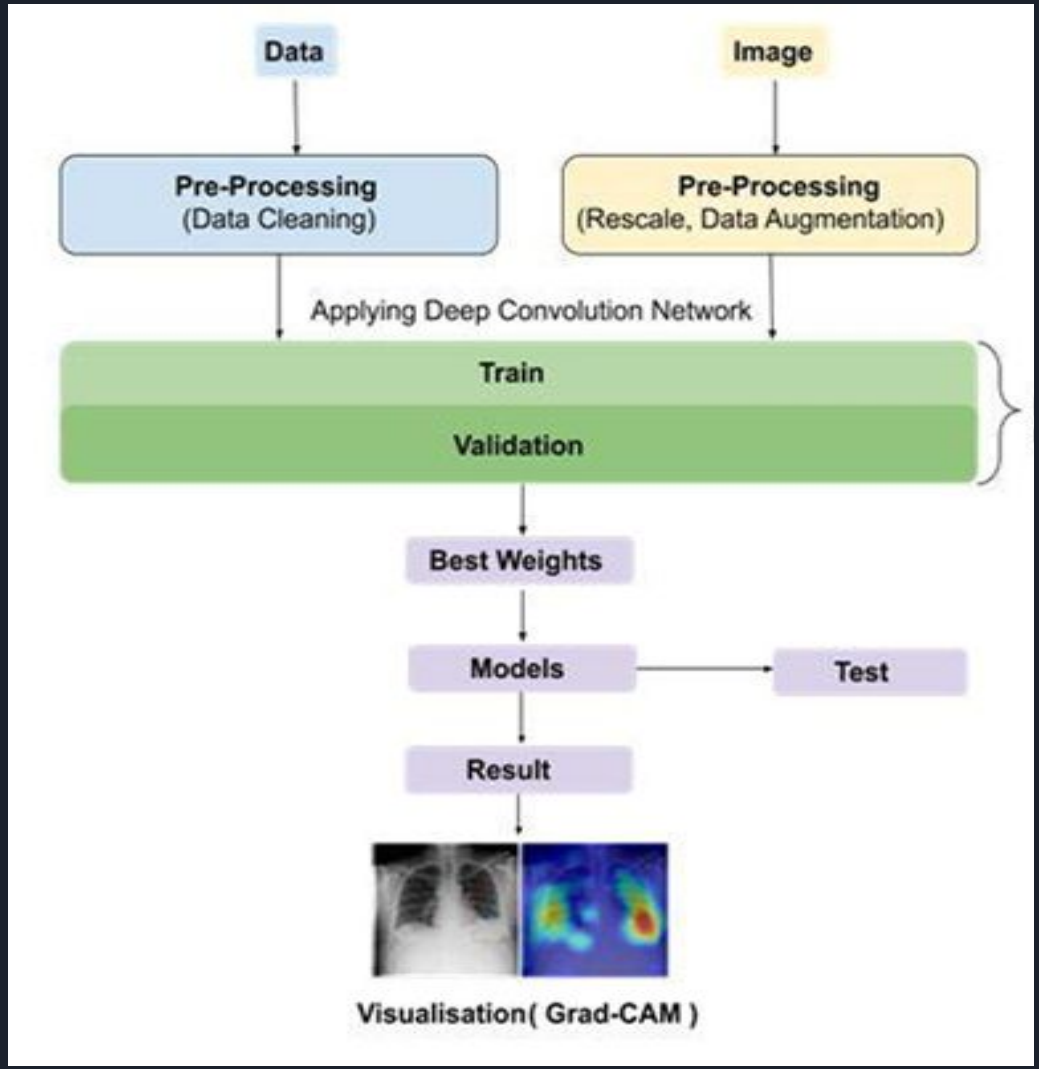
DataSet

1) Dataset for this project has been taken from National Institute of health.
2) The NIH recently released a chest X-Ray dataset comprising 112,120 frontal-view X-ray images of 30,805 unique patients.
3) It consists of 14 different diseases . Fourteen common pathologies include Cardiomegaly, Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural_thickening, Nodule, Mass and Hernia.



Flow-Diagram:

Timeline of our project



The project can be broken down into two major parts. These include :

1. Classification of diseases
2. Localization of diseases

Steps involved in the classification of diseases

- Data preprocessing and augmentation
- Training various models and finding the best fit for the data
- Visualizing the predictions made by the model on test dataset through measurements such as AUC ROC , and plotting the confusion matrix

Steps involved in the localization of diseases

- Using Grad CAM or Grad CAM++ algorithms to plot the activation maps for generating heatmaps
- Validating the generated heatmaps using the bounding boxes provided.

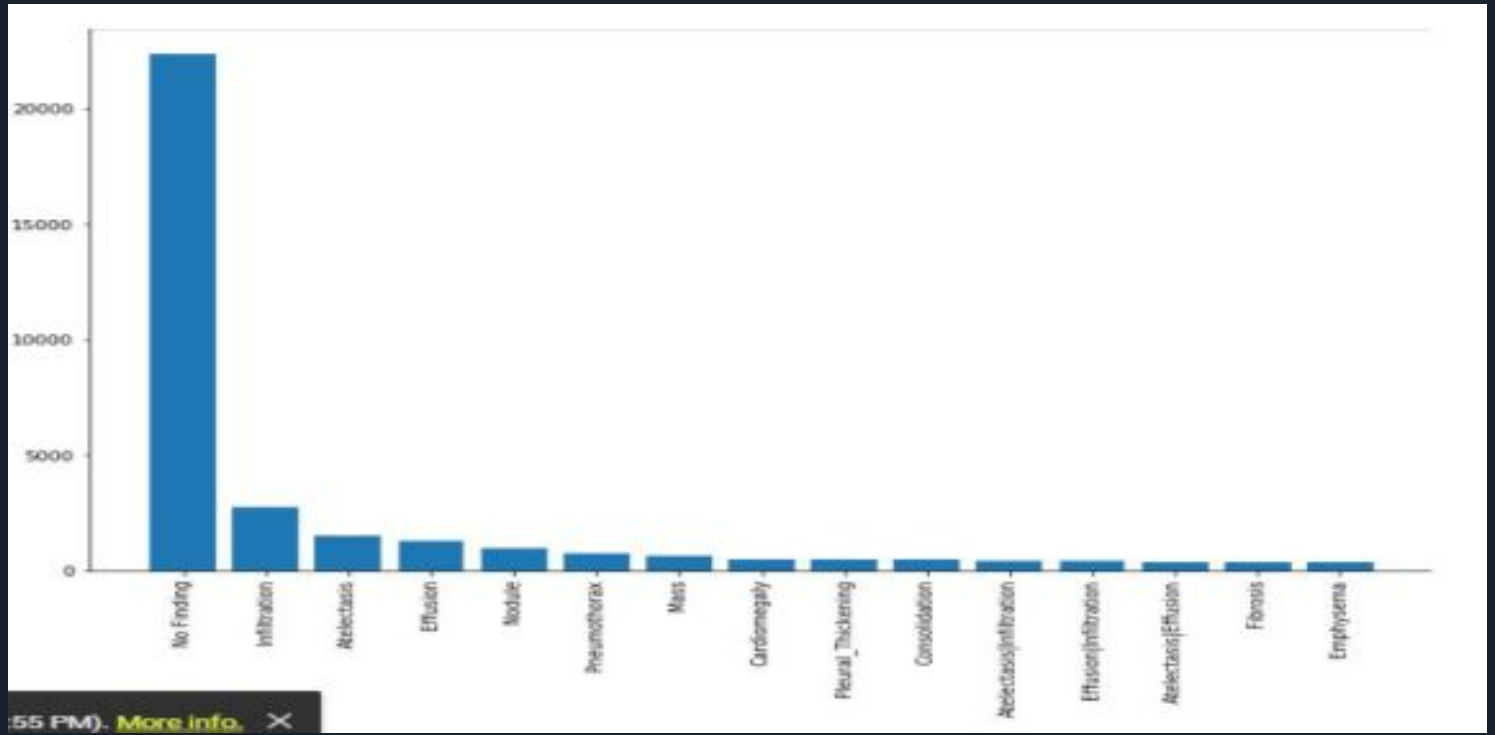
Data preprocessing and augmentation

1) The labels of the data entries were converted to categorical labels for each disease. We are using all 14 diseases in our project.
2) The images provided in the dataset were rescaled to images of resolution 100*100 pixels. The data augmentation was applied to the images in training data which included techniques such as horizontal flipping and shifting of images.
3) DownSampling of the DataSet was done.

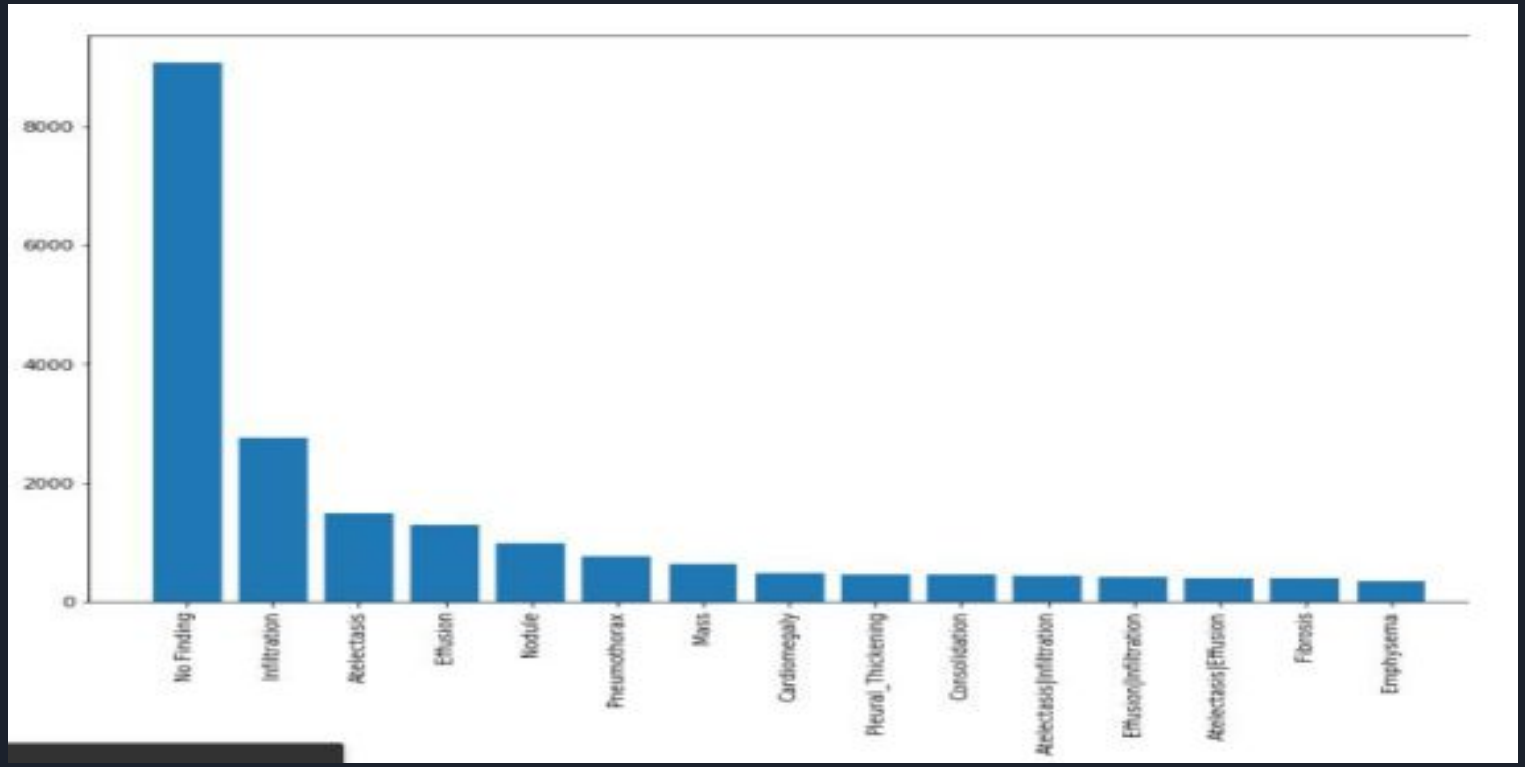
Why DownSampling?

1) The Dataset we were using was having more than 50% images with No Findings.
2) We downsampled dataset by 50% using concept of class weightage.
3) Total number of images were reduced to 50000 from 1 lakh.

The below two histograms show the frequency of each class in the dataset before downsampling of the dataset



The below two histograms show the frequency of each class in the dataset after downsampling of the dataset



Several neural network architectures were employed for the task, these include

- ResNet50 (23 million parameters), ResNet18 (11 million parameters), DenseNet121, MobileNet

- We are also looking forward to train models such as GoogleNet and Inception networks as well as simple networks such as VGG16.

AUC ROC curve

In the metrics, we are using accuracy and a custom metric loss which evaluates the AUC score i.e area under the ROC curve.

We used the AUC ROC curve because the dataset provided is strongly biased towards the absence of diseases in the images.

This will help us to select models that have true positive and false positive rates that are significantly above random chances that are not guaranteed by accuracy alone.

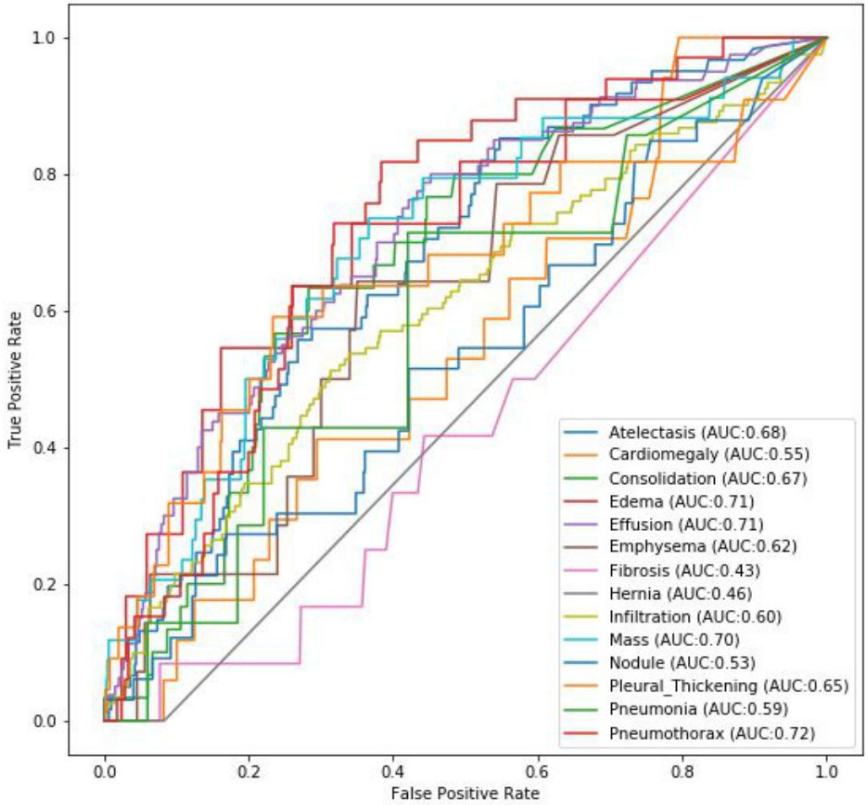
Result - ResNet50

The first model we trained was **ResNet50**. We used the [illegible] was used to provide the output. And a dense layer of 14 units was used to provide the output. Learning rate = 0.001, optimizer = rmsprop (decay = 1e-6) , loss = binary_cross_entropy, epochs = 20

Result cont.

Model	Training Loss	Val. loss	Mean AUC ROC score
ResNet50	0.1009	0.22928	0.76

Result cont.


True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Atelectasis (AUC:0.68)
Cardiomegaly (AUC:0.55)
Consolidation (AUC:0.67)
Edema (AUC:0.71)
Effusion (AUC:0.71)
Emphysema (AUC:0.62)
Fibrosis (AUC:0.43)
Hernia (AUC:0.46)
Infiltration (AUC:0.60)
Mass (AUC:0.70)
Nodule (AUC:0.53)
Pleural_Thickening (AUC:0.65)
Pneumonia (AUC:0.59)
Pneumothorax (AUC:0.72)

Result - DenseNet121

The result we got was not satisfactory on the test dataset. We got random predictions on the test dataset. The ratio of TP(True positive) to FP(False Positive) was very low. We suspected we might be overfitting the data on such a large neural network with almost 25 million parameters. The next model we trained was **DenseNet121.**

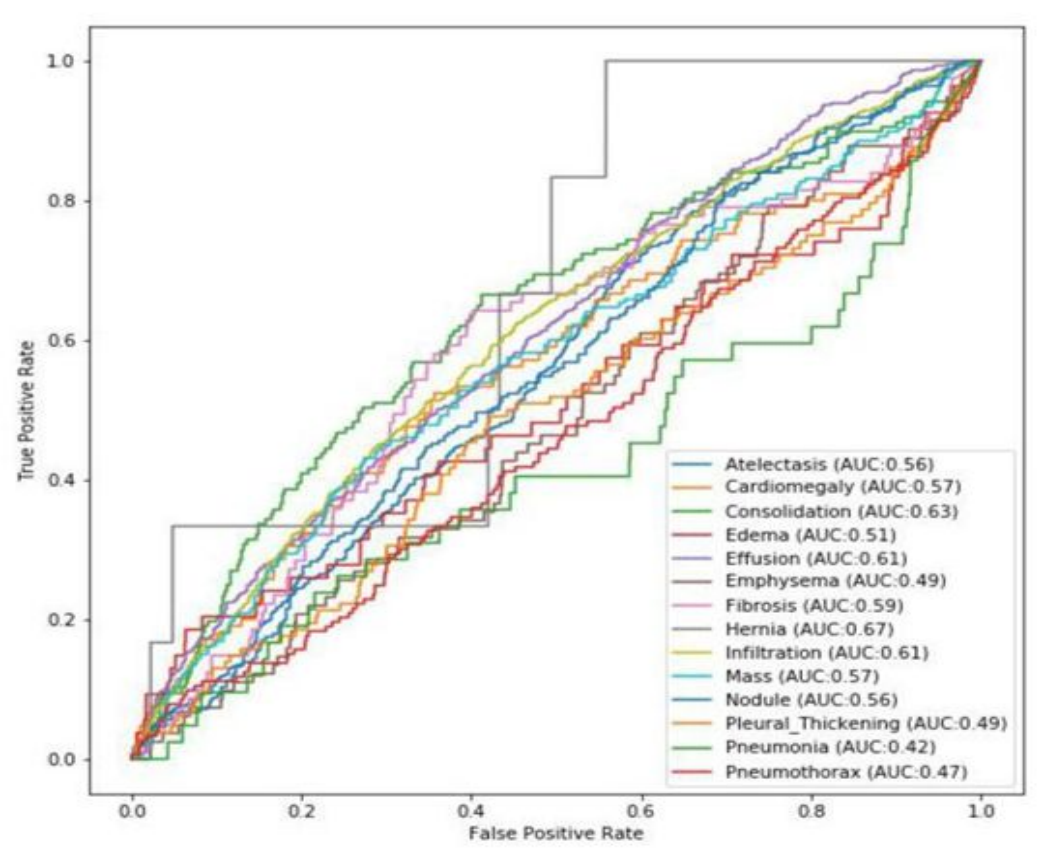
The pre trained weight were loaded from imagenet and the model was trained using :

Learning_rate = 0.001, optimizer = Adam, loss = binary_crossentropy , epochs = 20

Result contd

Model	Training Loss	Val. loss	Mean AUC ROC score
DenseNet121	0.1854	0.18556	0.752

Result contd - DenseNet



Result cont.

The model generated similar results as ResNet50 and a very poor the solution by correcting few problems as given below: approach to

1. We increased the size of the dataset by including more images from the dataset to stop our model from overfitting.
2. We trained our model with more number of epochs to train longer on the data.
3. We used relatively smaller models than ResNet50 but still large enough to get better results on the dataset.

Result - ResNet18

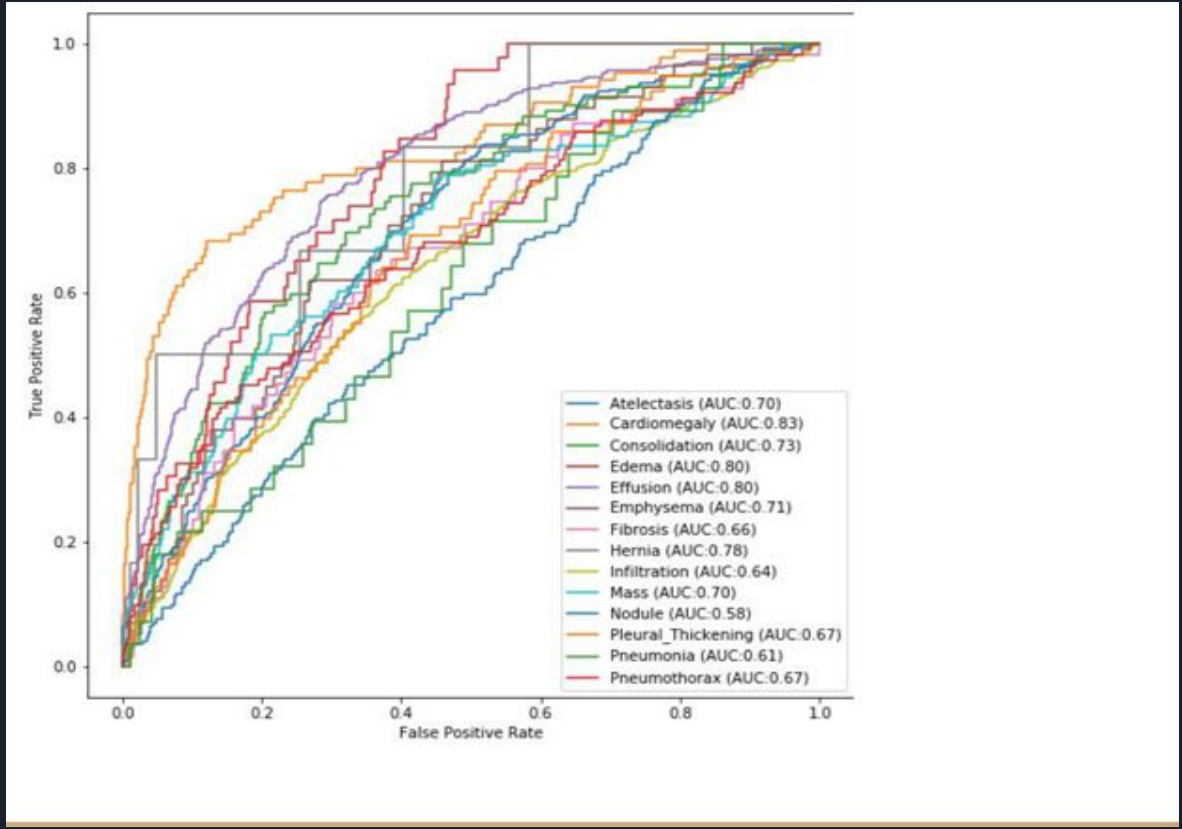
The next model we trained was **ResNet18** using:

Learning rate = 0.001 (decreasing learning rate every time the validation loss
as 0.0001, the same as before by a factor of 2 with lower bound on learning rate
Optimizer = Adam, number of epochs = 40, pretrained weights imported from
ImageNet .
The outcome of the training was better than the previous 2 models
suggests that increase in dataset with more epochs were the right
engines for the training.

Result cont.

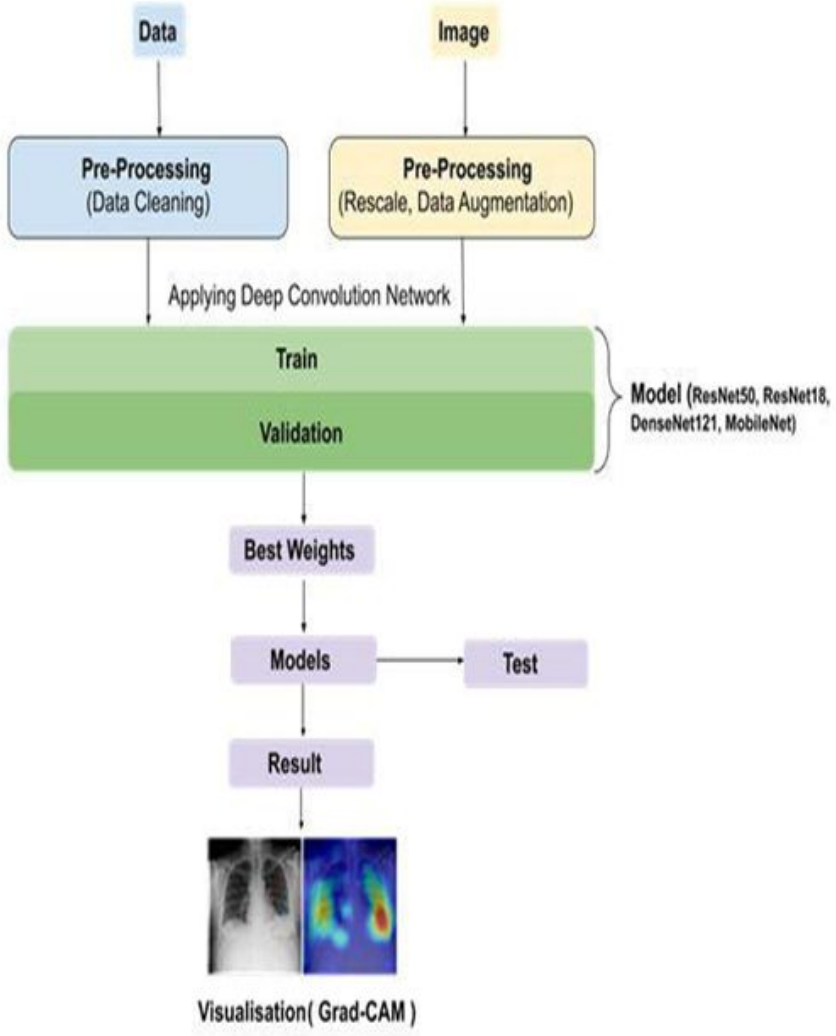
Model	Training loss	Val. loss	Mean AUC ROC score
ResNet18	0.1606	0.19857	0.83

Result cont.


True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Atelectasis (AUC:0.70)
Cardiomegaly (AUC:0.83)
Consolidation (AUC:0.73)
Edema (AUC:0.80)
Effusion (AUC:0.80)
Emphysema (AUC:0.71)
Fibrosis (AUC:0.66)
Hernia (AUC:0.78)
Infiltration (AUC:0.64)
Mass (AUC:0.70)
Nodule (AUC:0.58)
Pleural_Thickening (AUC:0.67)
Pneumonia (AUC:0.61)
Pneumothorax (AUC:0.67)

What was achieved?

1) Data Pre-Processing and Augmentation
2) We used a wide area of deep neural network architectures to train models for highly accurate predictions of diseases in the Chest X-Ray.
3) Visualizing the predictions made by the model on test dataset through measurements such as AUC ROC.



What is left?

1) We need to try more learning models on this data for even better results.
2) We haven’t done localization of disease. Will be generating heatmaps and compare them with the bounding box provided in NIH dataset.
3) Make a model with even high binary and categorical accuracy.

References

1. Rajpurkar, P., et al.: Chexnet: radiologist-level pneumonia detection on chest xrays with deep learning. arXiv preprint arXiv:1711.05225 (2017)

2. Focal loss for dense object detection. arXiv 2017 TY Lin, P Goyal, R Girshick, K He, P Dollár - arXiv preprint arXiv:1708.02002, 2000

3. Wang, Xiaosong, et al. "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.