PROJECT REPORT

Diagnosis of Chest X-Ray images using Deep Convolution Network

Project Mentor – Prof. Hari Om

Submitted By - Avi Sahney (17JE003050)
Rohit kumar (17JE003215)
Devashish gupta (17JE002834)
Aditya sharma (17JE002941)

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). Most of the radiographs are labelled with more than one disease making it a typical multi-label classification problem. Apart from the inherent challenges posed by multi-label classification, the dataset has a heavy imbalance in the number of instances of individual classes. Fourteen common pathologies include Cardiomegaly, Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural_thickening, Nodule, Mass and Hernia.

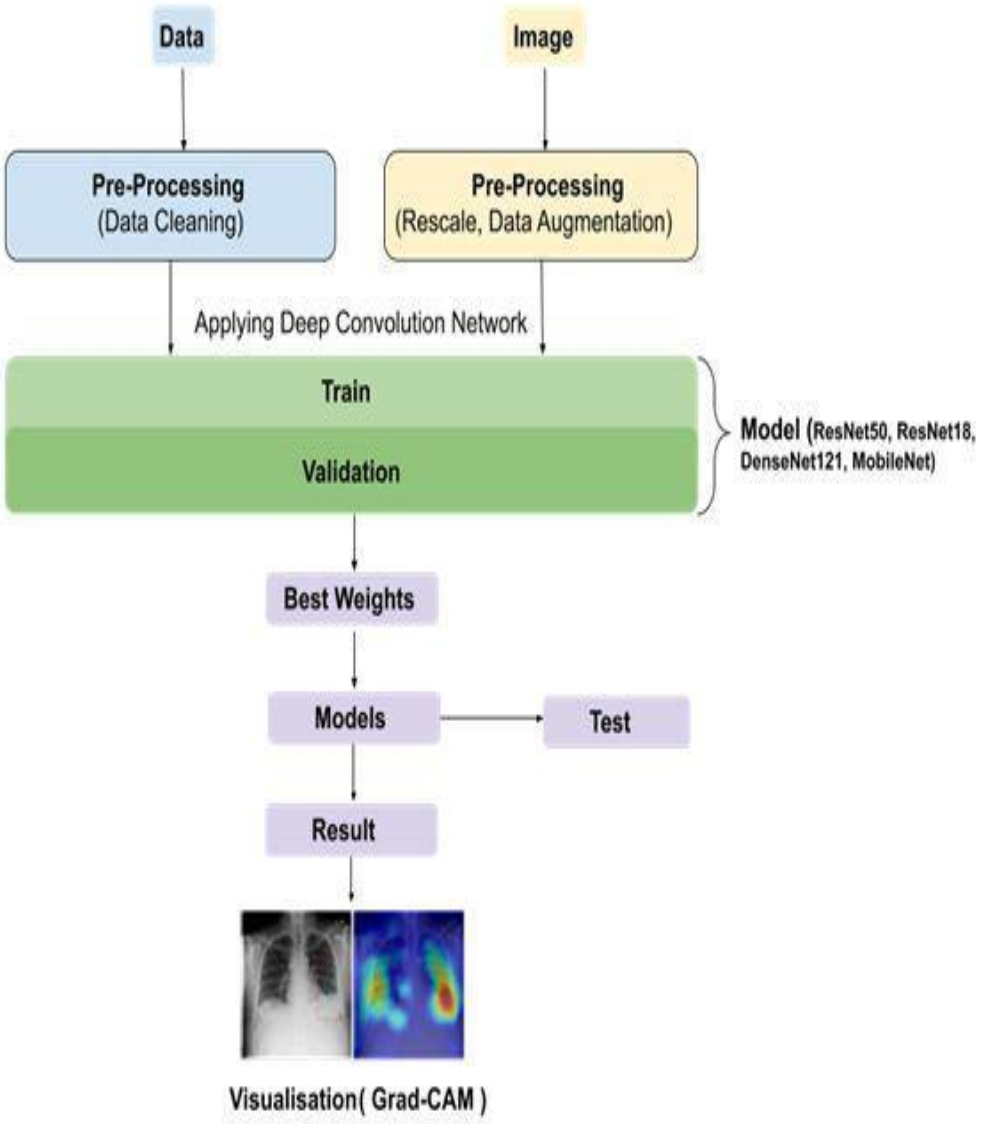
Deep learning techniques are state-of-the-art learning algorithms and perform quite well in medical image analysis. Our objective is to train models on the provided dataset to generate accurate predictions on new data and localizing the diseases in the Xray images.

Introduction

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.

Methodology

Flow diagram



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

We collected the data from the NIH dataset from Kaggle. Then the labels of the data entries were converted to categorical labels for each disease. We are using all 14 diseases in our project. 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.

Horizontal flipping is performed to augment the training set so as to simulate images where the abnormality may be present in any of the lungs, since lung diseases may take place in either lung. This step is important as it makes the model robust to abnormalities in CXR images. The images are then normalized by rescaling each pixel to 1/255 of its value. This set us ready for the training phase.

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.

All the trained models were loaded with pre-trained weights from imagenet dataset for faster learning.

The learning rate is kept very low to 0.001 with a decrease in the learning rate every time the validation loss didn't improve

AUC ROC curve

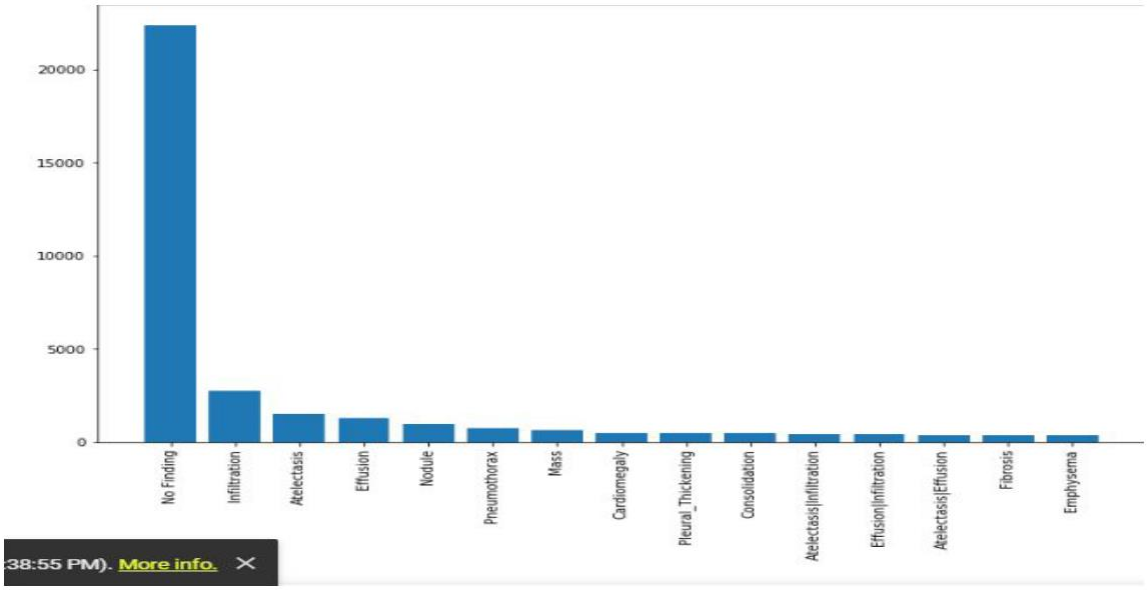
In the metrics, we are using accuracy and a custom metric loss which evaluates the AUC score i.e area under the ROC curve.

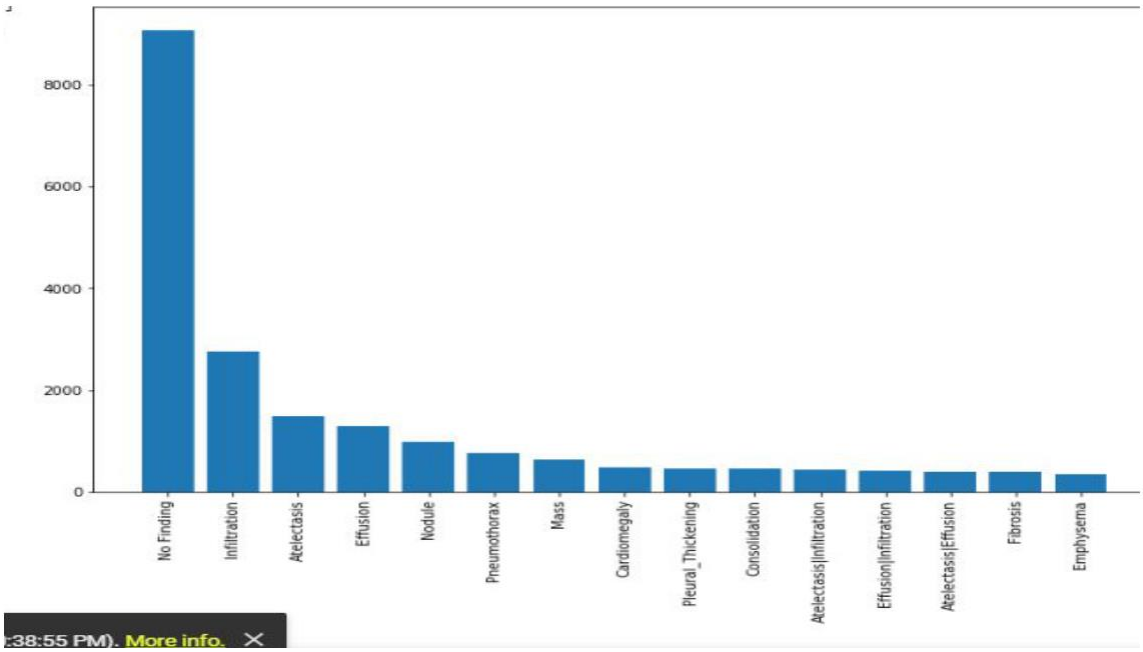
AUC - ROC curve is a performance measurement for classification problems at various threshold settings. ROC is a probability curve and AUC represents the degree or measure of separability. It tells how much model is capable of distinguishing between classes. Higher the AUC, better the model is at predicting 0s as 0s and 1s as 1s.

We used the AUC ROC curve because the dataset provided is strongly biased towards the absence of diseases in the images. So even if we predict the absence of all diseases on all test dataset we can get accuracy as high as 94%, which led us to the idea of using ROC curve to measure performance of the dataset. 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.

Downsampling the dataset

The total number of samples having "No Findings" label were very large in proportion to the rest of the 14 disease labels. So we down-sampled the dataset by randomly erasing 50% of the "No Finding " labels to generate a fairly balanced dataset. The below two histograms show the frequency of each class in the dataset before and after downsampling of the dataset.


20000
15000
10000
5000
0
No Finding
Infiltration
Atelectasis
Effusion
Nodule
Pneumothorax
Mass
Cardiomegaly
Pleural_Thickening
Consolidation
Atelectasis|Infiltration
Effusion|Infiltration
Atelectasis|Effusion
Fibrosis
Emphysema
:38:55 PM). More info.

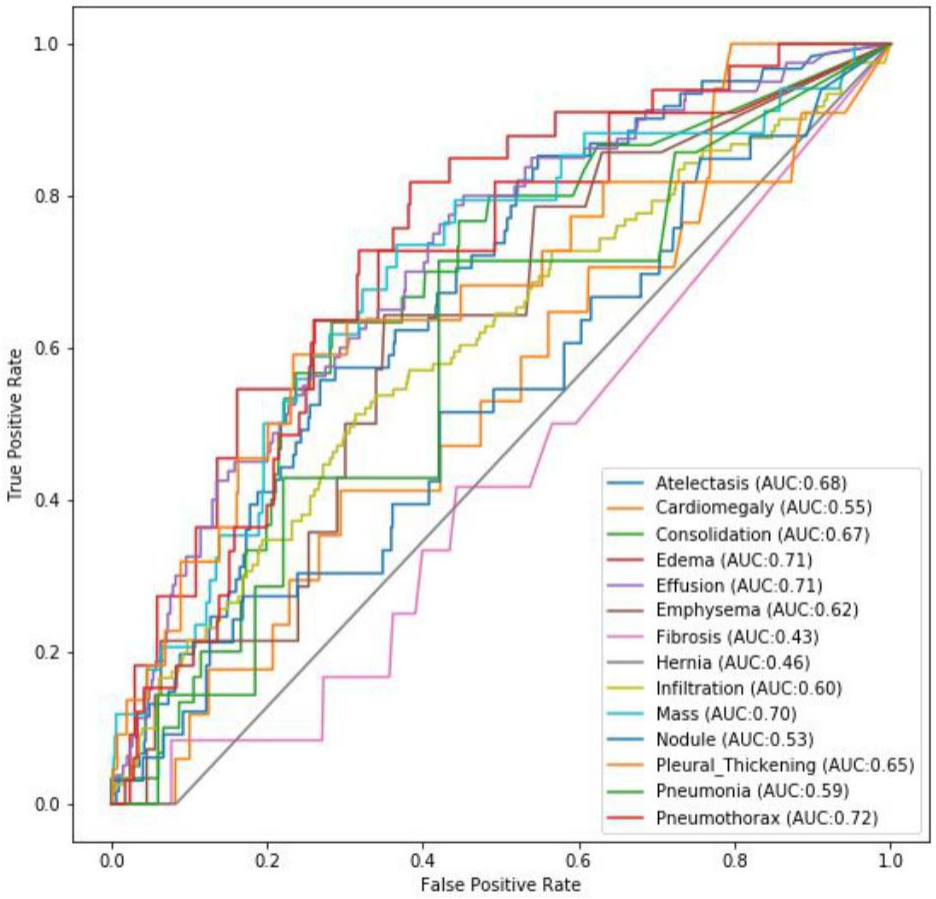

8000
6000
4000
2000
0
No Finding
Infiltration
Atelectasis
Effusion
Nodule
Pneumothorax
Mass
Cardiomegaly
Pleural_Thickening
Consolidation
Atelectasis|Infiltration
Effusion|Infiltration
Atelectasis|Effusion
Fibrosis
Emphysema
:38:55 PM). More info.

Result

The first model we trained was **ResNet-50**.We used the pre-trained weights from the imagenet dataset. The top layer was removed from the model 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

Model	Train-loss	Val-loss	Mean AUC ROC score
ResNet50	0.1009	0.22928	0.76

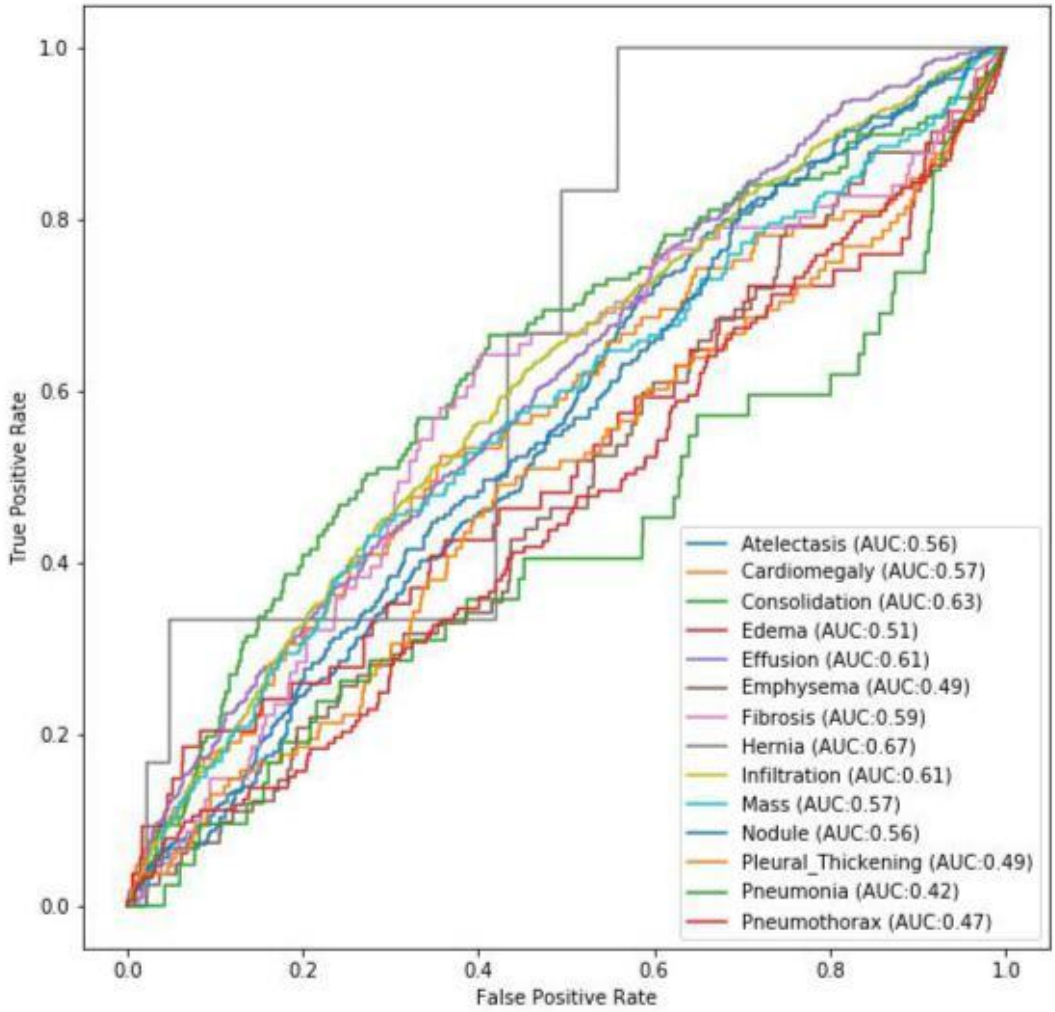


The result we got was not satisfactory on the test dataset. We got random predictions on the test dataset. The ratio of TP(True positive) to FP(False Positive) was very low. We suspected we might be overfitting the data on such a large neural network with almost 25 million parameters. The next model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

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



The model generated similar results as ResNet50 and a very poor prediction on the test dataset . We employed a different approach to the solution by correcting few problems as given below:

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.

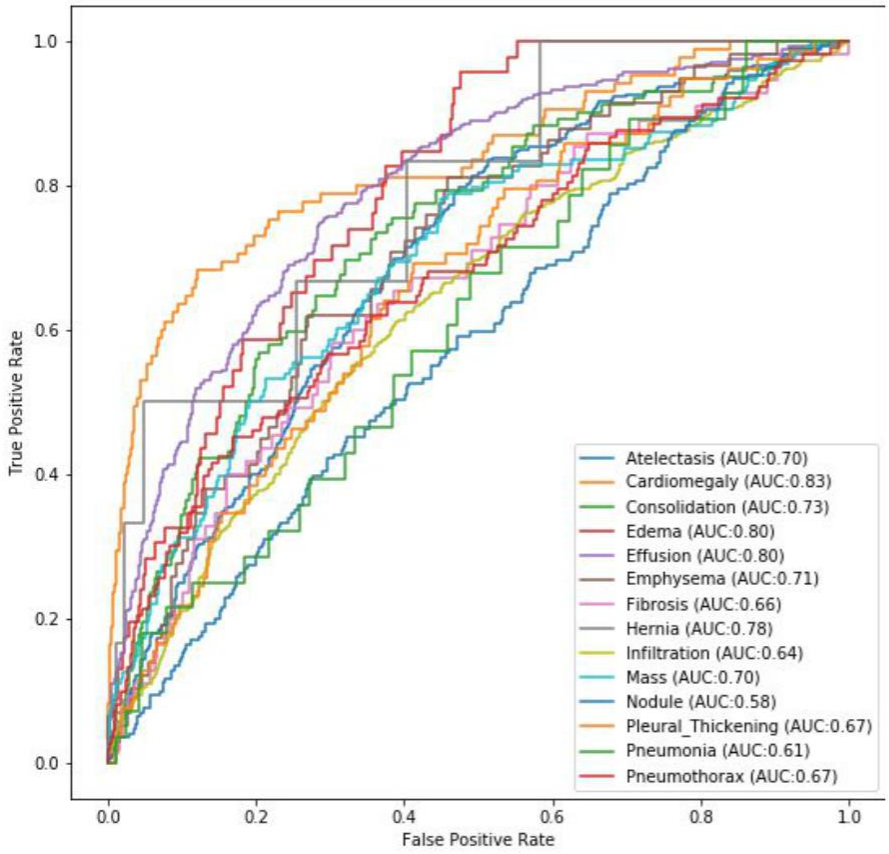
The next model we trained was **ResNet18**using:

Learning_rate = 0.001 (decreasing learning rate every time the validation loss remained the same as before by a factor of 2 with lower bound on learning rate as 0.0001), Optimizer = Adam, number of epochs = 40, pretrained weights imported from imagenet .

The outcome of the training was better than the previous 2 models suggesting that an increase in dataset with more epochs were the right choices for the train .

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

The below graph of AUC ROC shows a significant improvement in the results of the predictions. At the same time using a threshold of 0.3 on the predictions gave us better results with improvement in TP/FP ratio for many diseases.



Till Now we have trained models on the dataset and derived important results for better predictions , but there are few objectives that we are currently working on, and concepts we are going to integrate with our project, These include:

1. Class weighted loss: The dataset is imbalanced with varying frequency of each class, which biases the result to some extent. To counter this , we are currently working on a method to incorporate the weight of class depending upon its frequency in the loss calculated in the output layer. This , in theory, can help in reducing the biases of the data.
2. Localization of diseases: We are working on localizing the diseases with the help of Grad CAM and Grad CAM++ algorithms. We have successfully applied these algorithms to the dog vs cat binary classification problem. The algorithm is simple to apply by using the penultimate layer of our convolution neural network and applying a layer of global average pooling followed by a layer of a fully connected network. Then we train the model so that the new weights learn the features and then we calculate the gradients with respect to the input image to get the area of high interest (i.e the area of the input having a high activation). Then superimposing the Class Activation Mapping image with the original image we generate a visualization of disease affected regions.

3. One of the most important steps of our project is to validate the results of the Grad CAM or Grad CAM++ visualization on images already having bounding boxes for the diseases. We acquired around 800 images in the dataset which have the bounding box for the disease infected areas. We will be working on comparing the results obtained after localization with the given bounding boxes.

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.