

Interim Report of Capstone Project

(Pneumonia Detection Challenge)

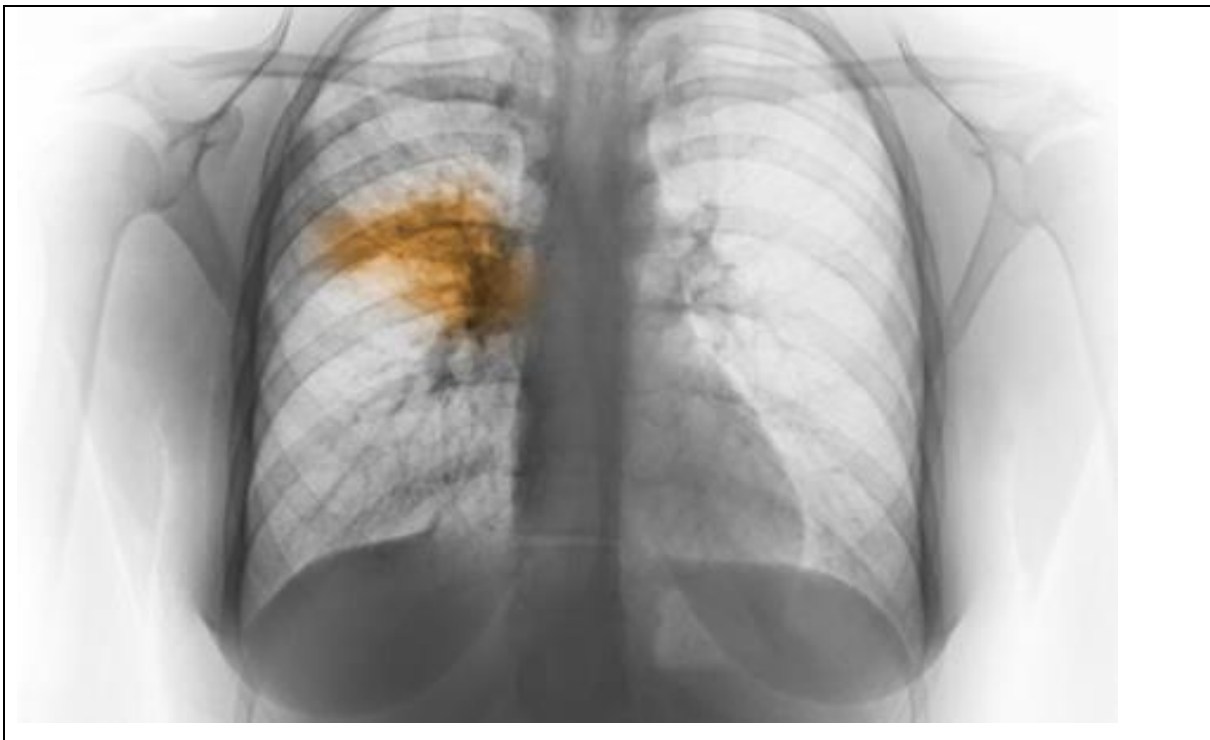
Essa Ibrahim Shariff, Duvvuru Lokesh, Debolina Dutta,
Bibhunandan Mohanta and Mohammed Hisham M

PROBLEM STATEMENT

In medicine, the next frontier for AI is anomaly localization in medical imaging. Localization of anomalies refers to both predicting anomalies and their boundaries. Automatic detection algorithms to locate inflammation in an image can help physicians make better clinical decisions. In this project, we analyze data with the knowledge of EDA. We build a detection model and present our findings based on the evaluations with the RSNA Pneumonia Detection Challenge dataset.

OVERVIEW OF PNEUMONIA

Pneumonia is a form of an acute respiratory infection that affects the lungs. The lungs comprise small sacs called alveoli that fill up with oxygen as a healthy person breathes. The alveoli are filled with pus and fluid when a person has pneumonia, making breathing difficult, and reducing oxygen intake.



The single most significant bacterial cause of death in children worldwide is pneumonia. In 2017, pneumonia killed 808,694 children under the age of 5, accounting for 15 percent of all deaths by children under five. Children and families worldwide are afflicted by pneumonia, but it is most common in South Asia and sub-Saharan Africa. It can be avoided with easy procedures and managed with low-cost, low-tech treatment and care.

In 2015 spending for maternal, infant, and child survival, the cost of antibiotic care for all children with pneumonia estimate at about US\$ 109 million per year among 66 countries. The expense requires antibiotics and diagnostics for the treatment of pneumonia.

CAUSES AND TRANSMISSION

According to WHO, pneumonia is caused by several infectious agents, including viruses, bacteria, and fungi. The most common are:

- *Streptococcus pneumoniae* – the most common cause of bacterial pneumonia in children;
- *Haemophilus influenzae* type b (Hib) – the second most common cause of bacterial pneumonia;
- the respiratory syncytial virus is the most common viral cause of pneumonia;
- in infants infected with HIV, *Pneumocystis jiroveci* is one of the most common reasons for pneumonia, responsible for at least one-quarter of all pneumonia deaths in HIV-infected infants.

Spreading of Pneumonia happens in many ways. The viruses and bacteria commonly found in a child's nose or throat can infect the lungs while breathing. They may also spread via air-borne droplets from a cough or sneeze. Besides, pneumonia may spread through blood, especially during and shortly after birth. More research needs to be done on the different pathogens causing pneumonia and how they are transmitted, as this is of critical importance for treatment and prevention.

TREATMENT AND PREVENTION

Pneumonia is treated with antibiotics. Amoxicillin-dispersible tablets are the antibiotic of choice. In most pneumonia cases, oral antibiotics are needed, which are mostly administered at a health clinic. These cases may also be diagnosed and treated at the neighborhood level by qualified community health professionals with affordable oral antibiotics. Only for severe cases of pneumonia is hospitalization recommended.

DIAGNOSTIC PROCEDURE

The doctor will diagnose pneumonia based on your medical history, a physical exam, and test results. Sometimes pneumonia is hard to analyze because symptoms may be the same as a cold or flu. The patient may not realize that his/her condition is more severe until it lasts longer than these other conditions.

If the doctor thinks the patient may have pneumonia, they may do one or more of the following tests.

- Chest X-ray to look for inflammation in the patient's lungs. A chest X-ray is often used to diagnose pneumonia.

- Blood tests, such as a complete blood count (CBC), determine whether the patient's immune system is fighting an infection.
- Pulse oximetry to measure how much oxygen is in his/her blood. Pneumonia can keep the patient's lungs from moving enough oxygen into his/her blood. A small sensor called a pulse oximeter is attached to the patient's finger or ear to calculate the levels.

DATA DESCRIPTION

In 2018, RSNA organized an AI challenge to detect pneumonia, one of the leading causes of mortality worldwide, as part of its efforts to help improve artificial intelligence (AI) instruments for radiology. RSNA Pneumonia dataset consists of 29684 thousand images. All the images are in Dicom format. There are 3000 images for testing and the remaining for training.

Dicom images: The images are in a particular format called DICOM files (*. dcm). They contain a mix of header metadata as well as pixel data underlying raw image arrays.

There are three classes in the dataset - Normal, Not normal/No opacity, and Lung opacity. Normal class indicates there is no anomaly in the lungs. Not normal/No opacity demonstrates to those who do not have pneumonia, but the image still has some abnormality. Sometimes, this finding could mimic the appearance of the right pneumonia. Lung opacity class indicates there is definite pneumonia in the lungs. Finally, these three classes are divided into two target variables, 0 and 1. The images with lung opacity come under target 1 and 0 for the other two classes. Along with the images, two csv files are provided. A detailed class info file consists of the image name and the class it belongs to. The train labels file consists of the bounding box coordinates belonging to each image. Bounding box coordinates are given in the following format:

- x -- the upper-left x coordinate of the bounding box.
- y -- the upper-left y coordinate of the bounding box.
- width -- the width of the bounding box.
- height -- the height of the bounding box.

With these bounding box coordinates, the target column is provided, which discriminates classes into categories of 0 and 1.

EDA AND PREPROCESSING

There are two datasets; each dataset has 30227 rows, and the trian_labels.csv has six columns while class_info.csv has two columns.

```
print("Bounding Box Shape:", bbox_df.shape)
print("Class info Shape:", class_df.shape)
```

```
Bounding Box Shape: (30227, 6)
Class info Shape: (30227, 2)
```




As we can see below, in train_labels.csv file contains patientId, which is a unique value per patient. Each patientId has one target column and four values: the corresponding abnormality bounding box defined by the upper-left-hand corner 'x' and 'y' coordinate and its corresponding width and height. The target column has two values 0 and 1. 0 is for No Lung Opacity / Not Normal, Normal, and 1 is for Lung opacity.

	patientId	x	y	width	height	Target
0	0004cfab-14fd-4e49-80ba-63a80b6bddd6	NaN	NaN	NaN	NaN	0
1	00313ee0-9eaa-42f4-b0ab-c148ed3241cd	NaN	NaN	NaN	NaN	0
2	00322d4d-1c29-4943-afc9-b6754be640eb	NaN	NaN	NaN	NaN	0
3	003d8fa0-6bf1-40ed-b54c-ac657f8495c5	NaN	NaN	NaN	NaN	0
4	00436515-870c-4b36-a041-de91049b9ab4	264.0	152.0	213.0	379.0	1

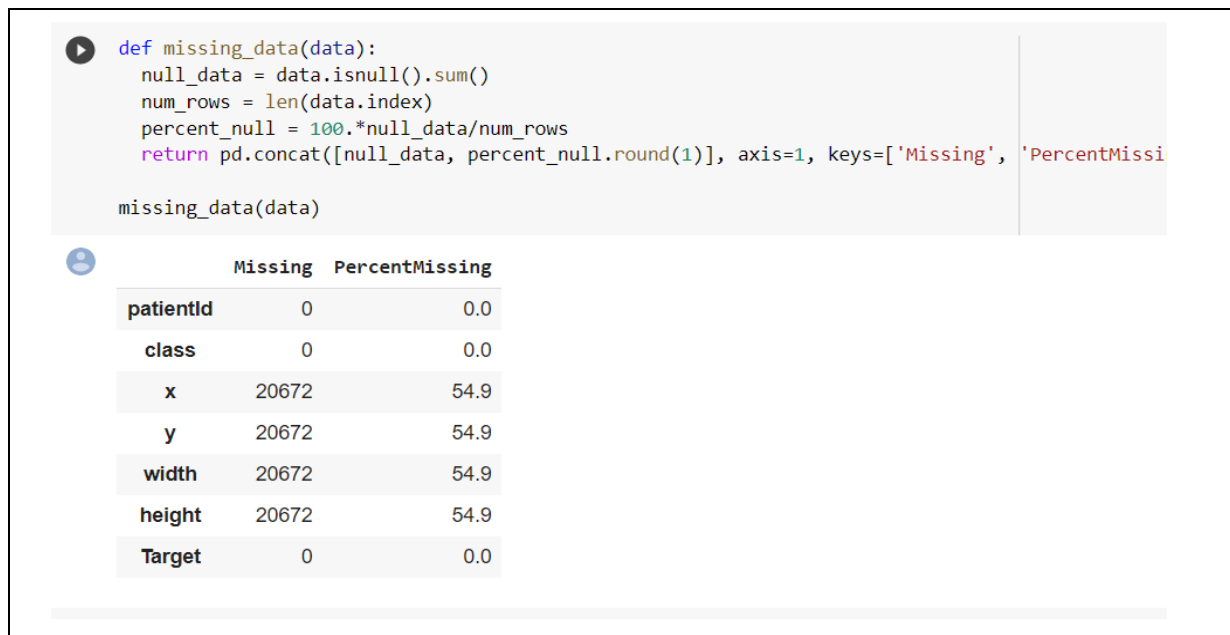
In the class_info.csv file, there are two columns patientId and class column that describe the three conditions of lungs (see below fig.).

	patientId	class
0	0004cfab-14fd-4e49-80ba-63a80b6bddd6	No Lung Opacity / Not Normal
1	00313ee0-9eaa-42f4-b0ab-c148ed3241cd	No Lung Opacity / Not Normal
2	00322d4d-1c29-4943-afc9-b6754be640eb	No Lung Opacity / Not Normal
3	003d8fa0-6bf1-40ed-b54c-ac657f8495c5	Normal
4	00436515-870c-4b36-a041-de91049b9ab4	Lung Opacity

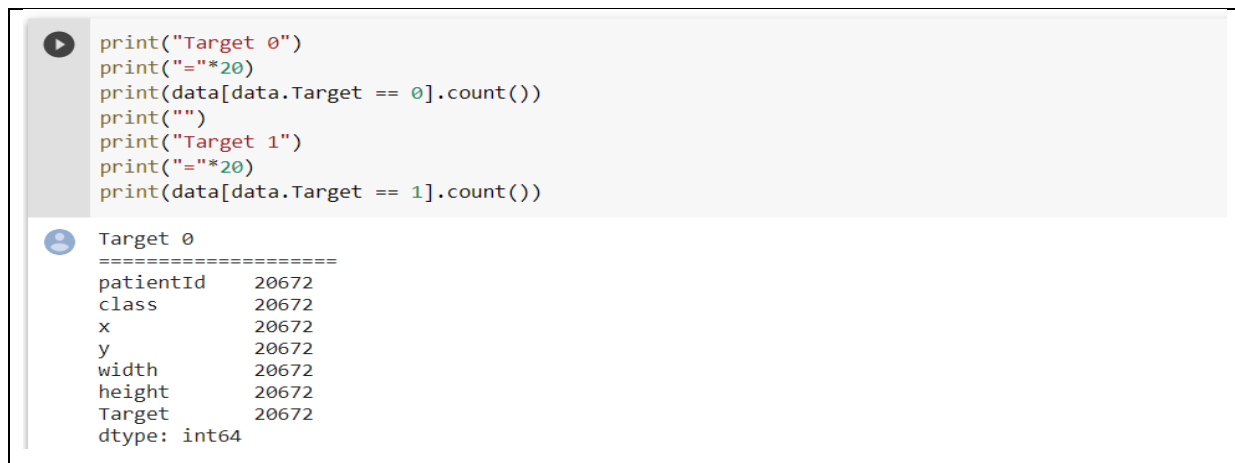
The frequency of patients in each class and their respective percentages are shown in the below figure. 23.5 percent of the patients are Normal, and the remaining are Not Normal and Lung opacity.

[]	data['class'].value_counts()
	Lung Opacity 16957
	No Lung Opacity / Not Normal 11821
	Normal 8851
	Name: class, dtype: int64
	data['class'].value_counts()*(100.0)/len(data.index)
	Lung Opacity 45.063648
	No Lung Opacity / Not Normal 31.414600
	Normal 23.521752
	Name: class, dtype: float64

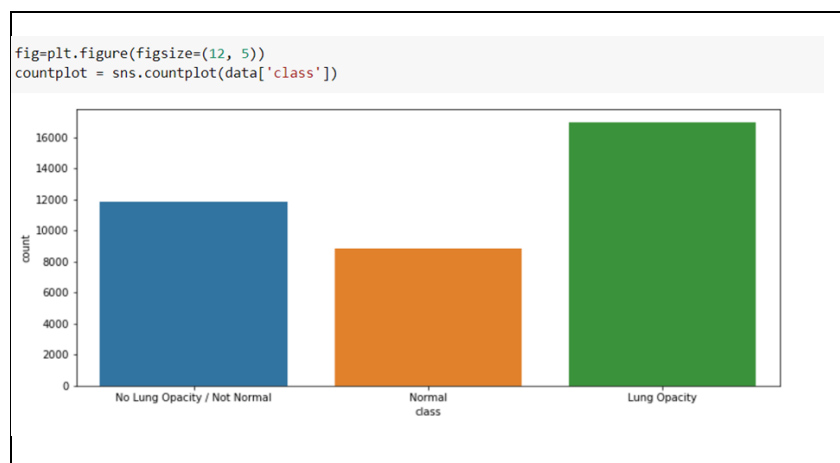
The bounding box dataset has missing values in x, y, height, and width column. A total of 20672 missing values are present in each column.



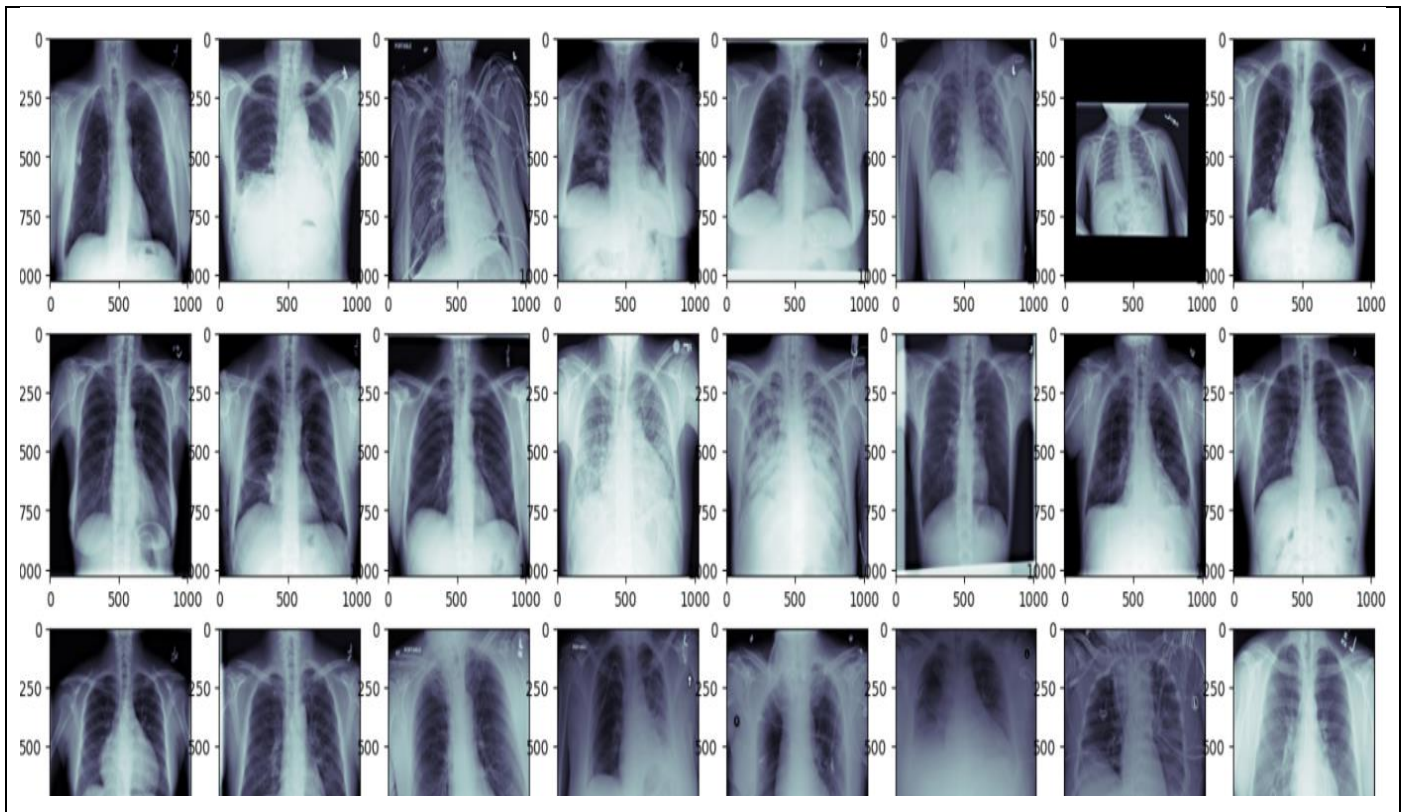
The below figure shows that the number of missing values is equal the number of No Lung Opacity (target=0). This indicates that the missing values are not random missing or errors, they are the images of lungs of normal person (or person without Pneumonia).



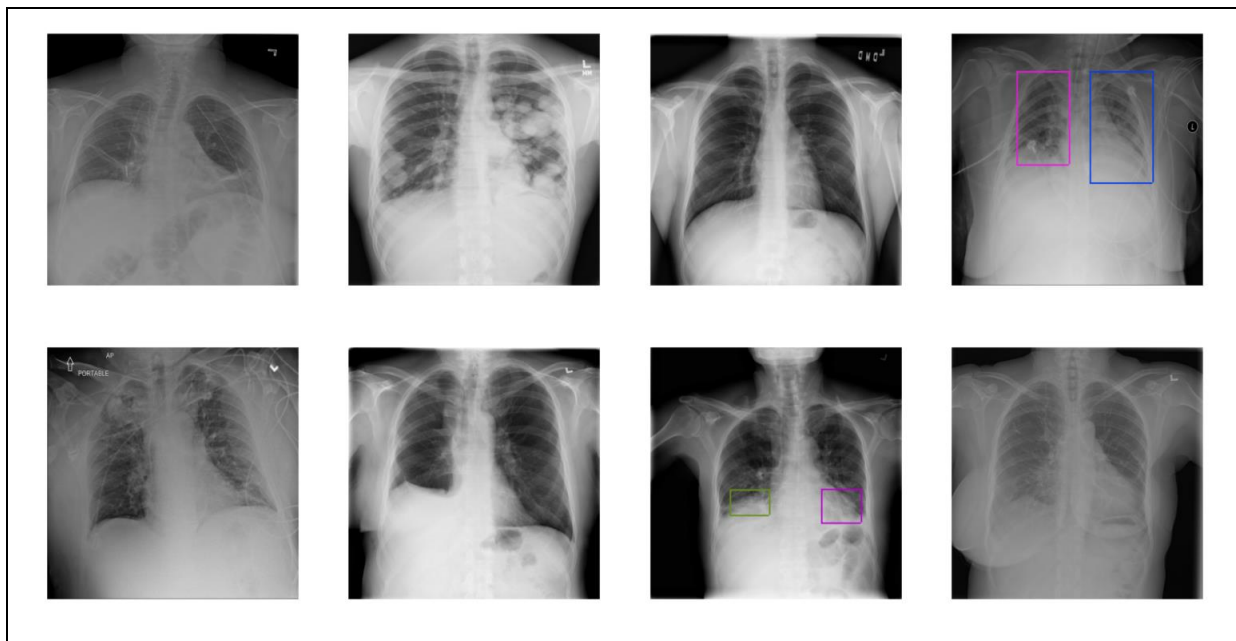
So, the remaining 16957 are the positive means Lung Opacity case. The below figure shows the count plot of the three classes.



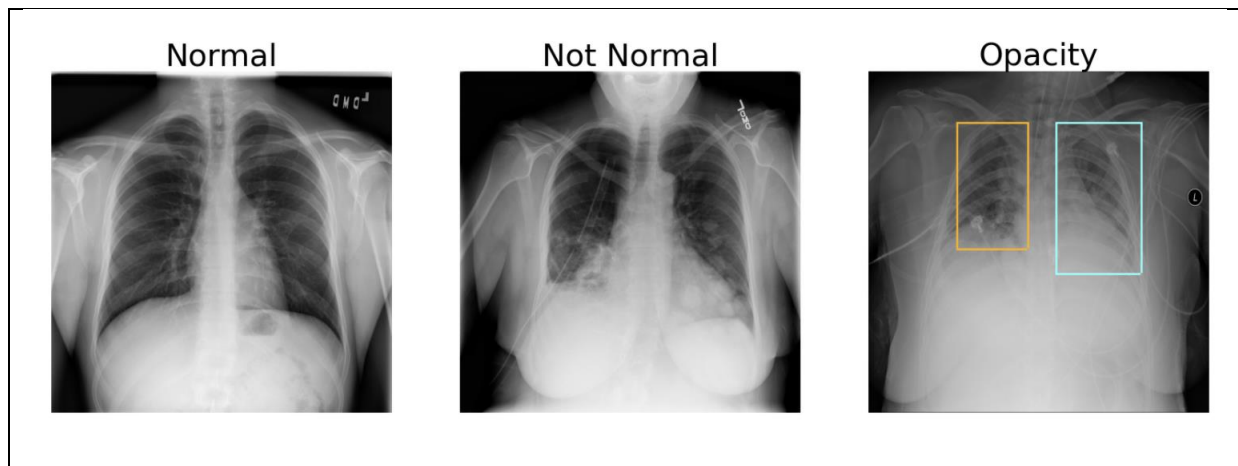
There are 26684 training images and 3000 test images. Visualizations of the few samples are shown below.



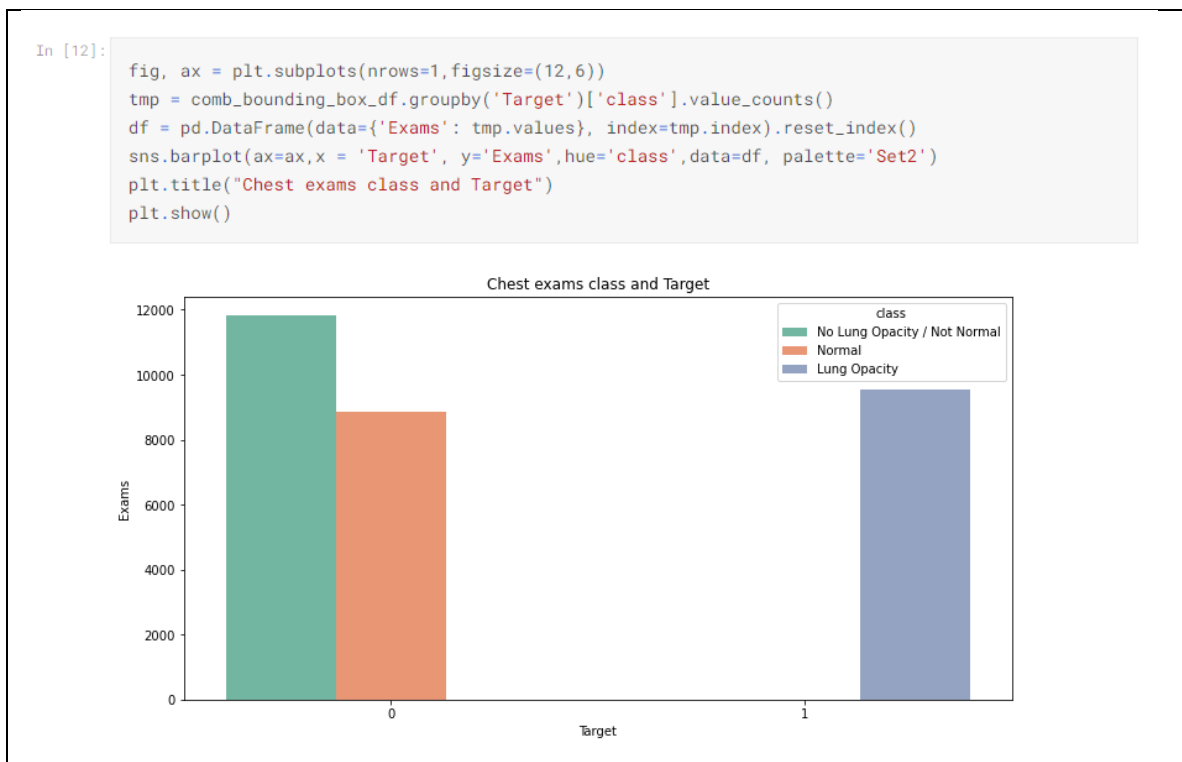
The images, along with the bounding box, is presented below. The figure indicates that some images have more than one bounding box, whereas some do not even have one.



As we can determine from the above visualizations, the task in hand is a regression problem. There is a need for building a feasible model that can regress the bounding box in the images. Moreover, there is an extra class that is Not normal/ No lung opacity. This class shows there is an anomaly in the lungs, which can be easily misread as pneumonia. So, there is a need to examine that class a little more briefly. A visualization showing all three classes together is shown below.



A bar plot of the target classes is shown below.



The information regarding the patients is available in the metadata of the Dicom images. Visualizations of that data may give a better understanding of the pneumonia disease itself. The below picture shows the data frame of the extracted dicom data.

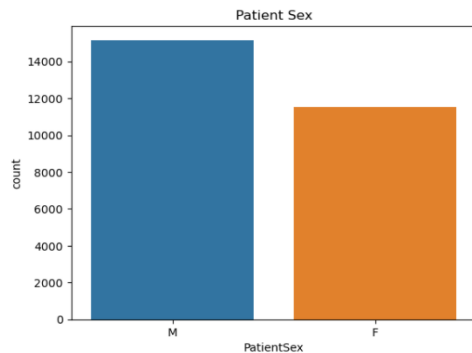
```
[53]: image_meta_df['class'] = detail_class_df['class']
image_meta_df['patientId'] = detail_class_df['patientId']
image_meta_df.head()
```

Out[53]:

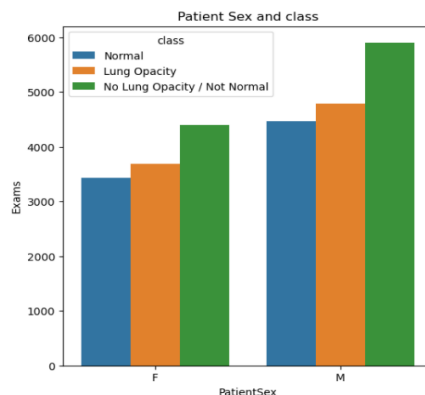
	PatientAge	BodyPartExamined	ViewPosition	PatientSex	path	class	patientId
0	8	CHEST	PA	M	../input/rsna-pneumonia-detection-challenge/st...	No Lung Opacity / Not Normal	0004cfab-14fd-4e49-80ba-63a80b6bdd6
1	16	CHEST	AP	M	../input/rsna-pneumonia-detection-challenge/st...	No Lung Opacity / Not Normal	00313ee0-9eaa-42f4-b0ab-c148ed3241cd
2	62	CHEST	PA	M	../input/rsna-pneumonia-detection-challenge/st...	No Lung Opacity / Not Normal	00322d4d-1c29-4943-afc9-b6754be640eb
3	65	CHEST	AP	M	../input/rsna-pneumonia-detection-challenge/st...	Normal	003d8fa0-6bf1-40ed-b54c-ac657f8495c5
4	26	CHEST	AP	F	../input/rsna-pneumonia-detection-challenge/st...	Lung Opacity	00436515-870c-4b36-a041-de91049b9ab4

Gender is one of the variables present in the data which can be explored. We can infer that the dataset has more male examples from the below images than the female examples. In this case, there are more men with pneumonia, around 4800 compared to around 3300 women with pneumonia. The abnormality in the lungs is very high in men compared to women. There is a massive difference in the Not normal/ No lung opacity class between males and females.

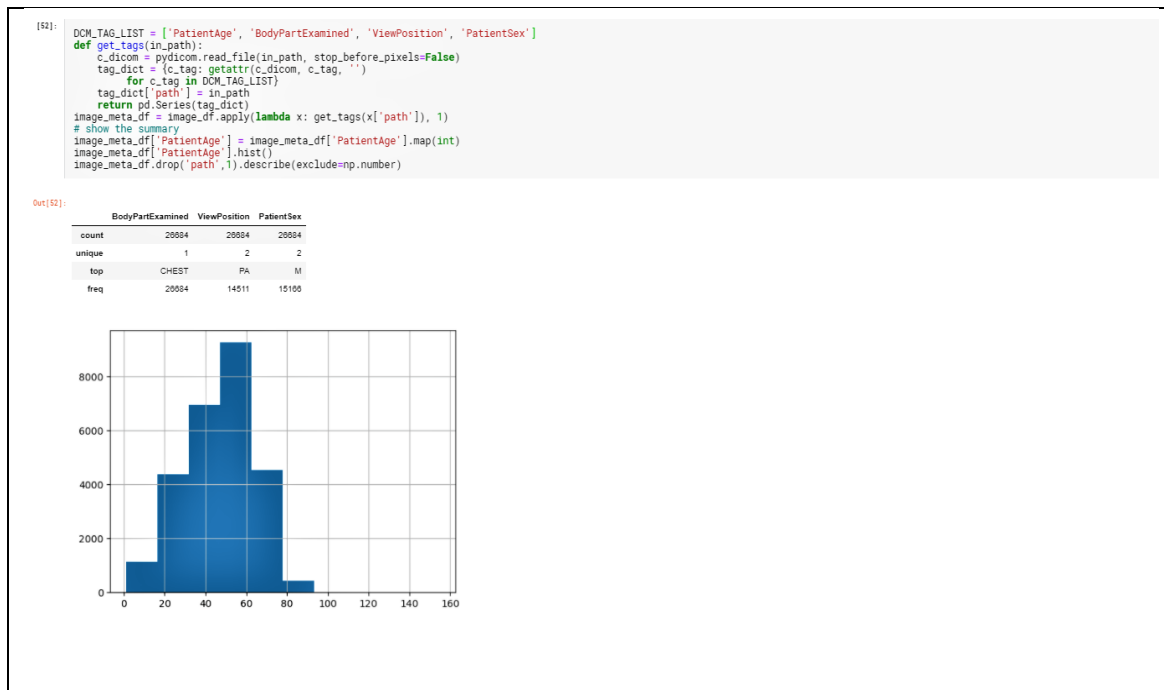
```
[56]: sns.countplot(image_meta_df['PatientSex'])
plt.title("Patient Sex")
plt.show()
```



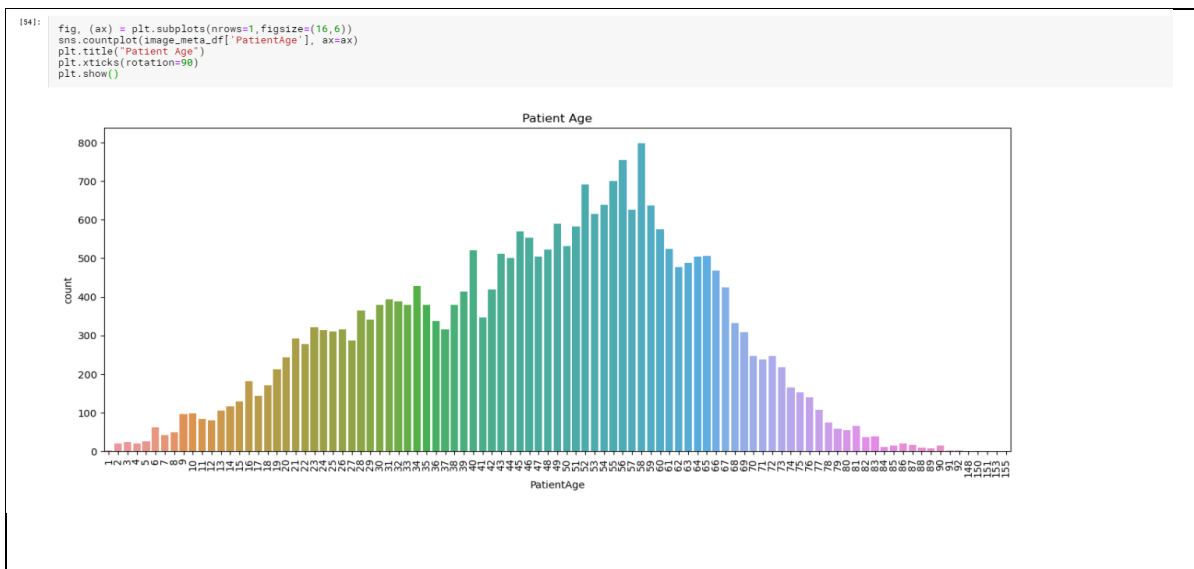
```
[55]: tmp = image_meta_df.groupby(['class', 'PatientSex'])['patientId'].count()
df1 = pd.DataFrame(data=tmp.values, index=tmp.index).reset_index()
tmp = df1.groupby(['Exams', 'class', 'PatientSex']).count()
df3 = pd.DataFrame(data=tmp.values, index=tmp.index).reset_index()
fig, (ax) = plt.subplots(nrows=1, figsize=(6,6))
sns.barplot(ax=ax, x = 'PatientSex', y='Exams', hue='class', data=df3)
plt.title("Patient Sex and class")
plt.show()
```



Similarly, we can explore the other variables such as body parts examined and age.

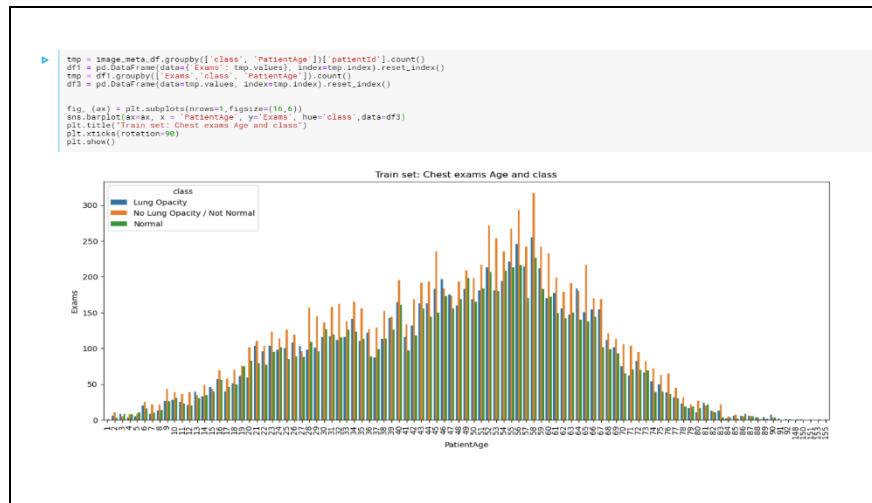


The above picture gives the histogram of the age distribution. We can see that most of the patients are above 40 years and below 65 years. The detailed plots are used to examine further.



From the above visualization, we can see that patients between 50 and 60 are more affected, with the dominant age being 58. The figure clearly shows that pneumonia is more prominent at an older age. Moreover, we can also see that the provided dataset consists of people aged between 40 and 65. We can explore a bit further by comparing Age with the class. The visualization that summarizes that relationship is plotted in below picture.

From the below plot, we can infer that Abnormalities in the lungs are more prominent in the people age between 45 and 65. So of no surprise, the lung opacities are maximum in that age group.



To summarize the Exploratory data analysis. We can say we are dealing with class imbalance, as most images do not have the bounding boxes. We can say that there is a need for the bounding box regressor model. The metadata provided can give insight to the disease as a whole. The visualizations of the original data and dicom metadata are extremely useful for the stakeholders and model evaluation.

MODEL BUILDING

Based on the findings from exploratory data analysis and problem statement, it is evident that the model should be a bounding box regressor that can identify the lung opacities, in turn predicting pneumonia. The model should have the ability to localize and identify the opacities. So, based on that, we came up with the following models.

CNN MODEL WITH RESNET BLOCKS:

This model consists of a series of residual blocks in the middle with downsampling then followed by output block, which leads to upsampling.

Approach

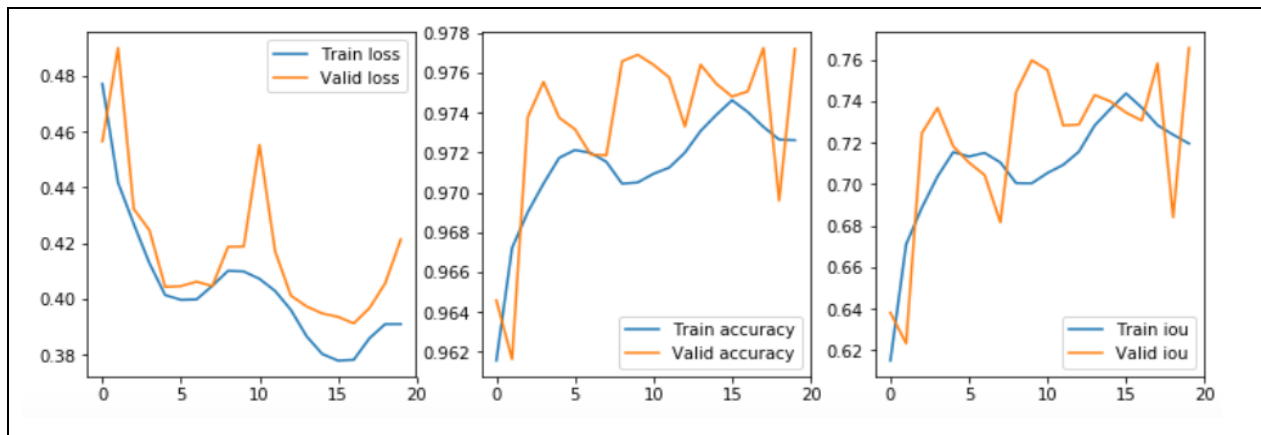
Firstly, a convolutional neural network is used to segment the image, using the bounding boxes directly as a mask. The connected components are used to separate multiple areas of predicted pneumonia. Finally, a bounding box is simply regressed around every connected component.

Network

The network consists of several residual blocks with convolutions and downsampling blocks with max pooling. At the end of the network, a single upsampling layer converts the output to the same shape as the input. As the input to the network is 256 by 256 (instead of the original 1024 by 1024) and the network downsamples several times without any meaningful upsampling (the final upsampling is to match in 256 by 256 mask), the final prediction is very crude. If the network down samples four times, the final bounding boxes can only change with at least 16 pixels.

Results

This model is initially trained for 20 epochs. This gives the mean IOU of 0.76 on the validation set, and model is a good fit. Accuracy increases slightly as epochs increased, and loss gradually decreases. Results for the above network on the initial network are as shown in the below graphs.



U-NET WITH MOBILENET BACKBONE

This model is predominantly used for the image segmentation problems. In this case, the UNet model is modified for pneumonia detection problem using the bounding boxes' coordinates to construct the masks.

Approach

The mobilenet model is loaded from the Keras library with imagenet competition weights and the top of the network is frozen. The image is given as an input with the corresponding bounding box, and feature maps are produced and downsampled gradually to fit the bounding box and then returned to the original shape of the input image. This process allows the network to propagate context information to higher resolution layers.

Network

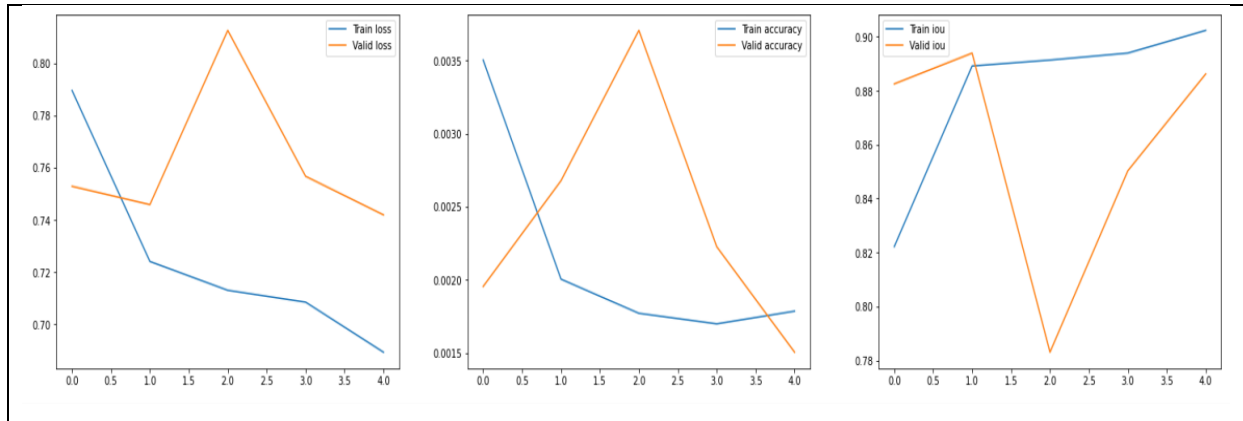
This network consists of u-shaped architecture, as the name suggests. During the downsampling, features are generated gradually. Whereas on upsampling, the features are duplicated to that of the original image with mask segmentation.

This network takes an image input of 224*224 pixels with three channels red, green, and blue. It has around 3 million parameters. A standard convolution network consisting of repeated use of convolutions, each followed by a rectified linear unit (ReLU) and a max-pooling process, is the contracting part. The spatial information is decreased during the contraction, while feature information is increased. Via a series of up-convolutions and concatenations with high-resolution characteristics from the contracting path, the expansive path incorporates the function and spatial details.

Results

During the initial run of 5 epochs, the u-net model showed promising results with a mean IOU of 0.88. While loss goes up and gradually decreases, the model's accuracy seems to take a downfall with the training. The results

for the initial run are shown in the graphs below. This model's drawback is its training time is very long, nearly 50mins per epoch on GPU.



DENSENET 121

We have an input image in a typical Convolutional Neural Network, which is then passed through the network to get a projected output mark in a way where the forward pass is pretty straightforward. In a DenseNet, each layer is connected to every other layer, hence the name Densely Connected Convolutional Network.

Approach

$L(L+1)/2$ direct connections exist for L layers. The feature maps of all the previous layers are used as inputs for each layer, and their own feature maps are used as inputs for each subsequent layer. The concatenation of feature maps from previous layers is the input of a layer within DenseNet.

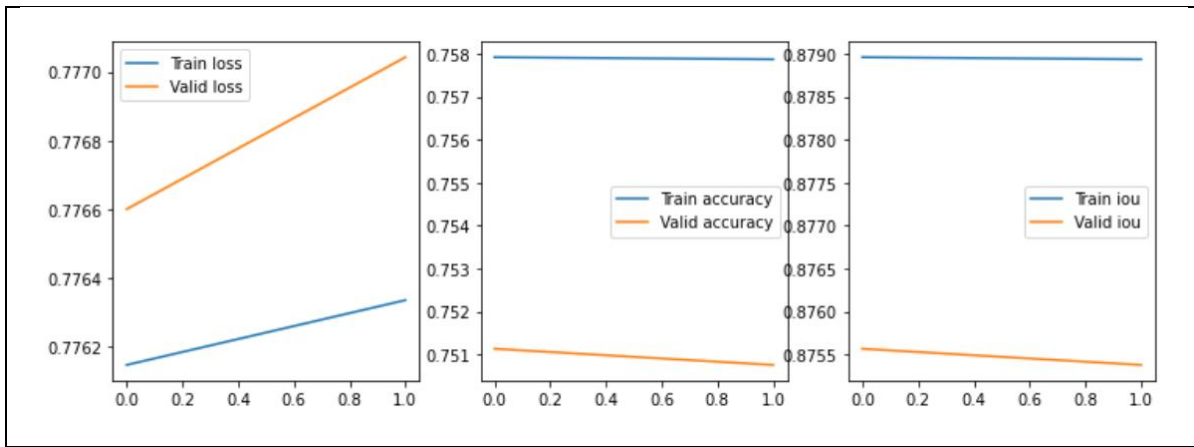
Network

There are four Dense Blocks with varying layer numbers in each architecture. For instance, in the four dense blocks, DenseNet-121 has [6,12,24,16] layers. We can see that a 7x7 stage 2 Conv Layer followed by a 3x3 stride-2 MaxPooling layer consists of the first part of the DenseNet architecture. A Classification Layer that accepts the feature maps of all network layers to perform the classification follows the fourth dense block.

Downsampling and upsampling are added to the dense layers to learn the mask implementation of the given input images of 224*224 with three channels (RGB). This network has 6million parameters.

Results

During the initial run of the model, it gave better results than U-net. Here accuracy and loss are better accompanied by 0.88 mean IOU. Training time takes longer than other models but less than u-net, around 37mins for each epoch. The results from the initial training run are shown in the graph below.



Mask RCNN

The Mask R-CNN is the Faster R-CNN extension, which adds an output model for each detected object to predict a mask. Mask R-CNN, especially compared to a simple or even state-of-the-art deep convolutional neural network model, is a sophisticated model to implement.

Approach

Matterport's Mask R-CNN is adapted for the problem at hand. This model generates bounding boxes and segmentation masks for each instance of an object in the image. It's based on Feature Pyramid Network (FPN) and a ResNet101 backbone.

Network

The first step is to install the library. Installation involves cloning the GitHub repository and running the installation script on the workstation. Coco dataset pretrained weights are loaded to use for transfer learning.

Mask RCNN architecture consists of two stages,

Stage 1: The first stage consists of two networks, a backbone network (ResNet, VGG, Inception, etc.) and a network of regional proposals. To offer a set of region proposals, these networks run once per picture. Proposals for a region are regions in the function map that contain the object.

Stage 2: In the second stage, the network predicts bounding boxes and object class for each of the proposed region obtained in stage1. Each proposed region can be of different size, whereas fully connected layers in the networks always require a fixed size vector to make predictions. These proposed regions' size is fixed by using either RoI pool (which is very similar to MaxPooling) or the RoIAlign method. The RoIAlign layer's output is then fed into Mask head, which consists of two convolution layers. It generates a mask for each RoI, thus segmenting an image in a pixel-to-pixel manner.

Results:

After the initial run of 5 epochs of training, training loss gradually reduces along with the `rpn_class_loss`, `rpn_bbox_loss`, `mrcnn_class_loss`, `mrcnn_bbox_loss`, and `mrcnn_mask_loss`. This model trains a little faster than the above models.

POTENTIAL CHANCES TO IMPROVE THE MODEL

There are some parts of the model, data, and hyperparameters in the models can be used to improve the performance so far,

1. The depth of the models can be increased along with the optimizers and learning rate.
2. A subset of the data can be trained using the above models for a more significant number of epochs to achieve convergence, which may give better results.