

FINAL REPORT FOR AIML CAPSTONE PROJECT PNEUMONIA DETECTION

Group Members:

1. **Vinay Mahamunkar**
2. **Keerthana Narisetty**
3. **Vidyashree Biligere Narayanaswamy**
4. **Lns prakash goteti**
5. **Kaushik M**

Table of Contents

Summary of problem statement

Overview of final process

Step By Step walkthrough the solution

Model Evaluation

Comparison to Benchmark

Visualizations

Implications

Limitations

Closing Reflections

Summary of problem statement

A pneumonia detection model is a deep learning model that is trained to identify and diagnose pneumonia in medical images, such as chest X-rays or CT scans. The goal of such a model is to assist radiologists and other medical professionals in quickly and accurately detecting cases of pneumonia, which can be a serious and potentially life-threatening condition. The model is typically built using a convolutional neural network (CNN) architecture and trained on a dataset of labelled chest X-ray images. Once trained, the model can analyse new images and make predictions about whether the image contains evidence of pneumonia. The model's performance is evaluated using metrics such as accuracy, precision, recall, and F1-score, and it should be validated on a large dataset and reviewed by radiologists before being used in a clinical setting.

In the present work the dataset contains X-ray images of patients' lungs. Here we develop models for both segmentation techniques to detect lung opacities on chest X-ray film images, i.e to locate the position of inflammation in the DICOM image. The present work aims at developing CNN architectures to detect lung opacities on chest X-ray film images and subsequently building classifiers pertaining to this disease. The subsequent part of the work aims at applying transfer learning, fine tuning the parameters and improving the model accuracy architectures for building classifiers pertaining to this disease.

1. Introduction:

Respiratory infections are found to be prominent cases for hospitalization, for example, in Iraq they represent 60% mean consultations corresponding to 45% of the average population hospitalized [1]. On the other hand, 22-42% of adult pneumonia patients require hospitalization and 5-10% of them require ICU [2].

Considering Among adults suffering from pneumonia, it is estimated that between 22 and 42% require hospitalization and between 5 and 10% need an intensive care unit, and the lethality varies between 5 and 50% depending on the severity of the condition, which is higher in the elderly and immunosuppressed patients [2]. According to UNICEF data [3], pneumonia is increasingly becoming a life threatening disease among children compared to other diseases among them claiming the lives of over 700,000 children under five every year, or around 2,000 every day. South Asia is prominent as 2,500 cases per 100,000 children are noticed by West and Central Africa (1,620 cases per 100,000 children).

Why do we need to solve it?

Pneumonia accounts for over 15% of all deaths of children under 5 years old internationally. In 2015, 920,000 children under the age of 5 died from the disease. Pneumonia accounts for over 500,000 visits to emergency departments in the United States. There were 50,000 deaths in 2015 keeping the ailment on the list of top 10 causes of death in the United States.

Why can't doctors themselves handle the issue?

Well, reading the X-ray/CXR is a complicated thing because of following reasons:

1. Lung cancer, fluid overload (pulmonary edema), bleeding, volume loss (atelectasis or collapse), or post-radiation or surgical changes appears as increased opacity on CXR.
2. Outside of the lungs, fluid in the pleural space (pleural effusion) also appears as increased opacity on CXR.
3. Positioning of the patient and depth of inspiration can alter the appearance of the CXR
4. Clinicians are faced with reading high volumes of images every shift.

To improve the efficiency and reach of diagnostic services, an automated Pneumonia detection system is very much necessary.

In the present work, our objective is to build a pneumonia detection system i.e., to locate the position of inflammation in the DICOM image. In other words, to build an algorithm that needs to automatically locate lung opacities on chest radiographs. We have developed CNN model to detect lung opacities on chest X-ray film images. The subsequent part of the work aims at applying transfer model learning, fine tuning the parameters and improving the model accuracy architectures for building classifiers pertaining to this disease. The dataset contains X-ray images of patients' lungs.

The data description for the present problem is provided below:

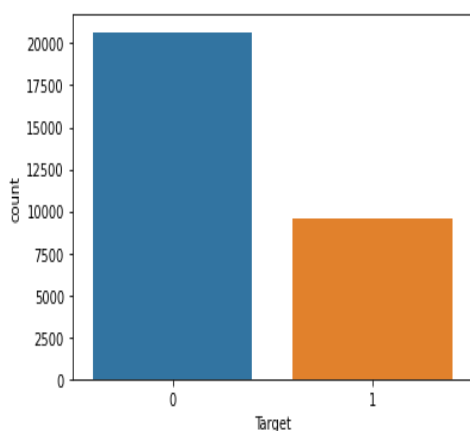
- o In the dataset, some of the features are labeled “Not Normal No Lung Opacity”. This extra third class indicates that while pneumonia was determined not to be present, there was nonetheless some type of abnormality on the image and oftentimes this finding may mimic the appearance of true pneumonia. Dicom original images: - Medical images are stored in a special format called DICOM files (*.dcm). They contain a combination of header metadata as well as underlying raw image arrays for pixel data.
- o Dataset has been attached along with this project. Please use the same for this capstone project.
- o Original link to the dataset: <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data>

Overview of final process

First step was to import the data from csv files then we merged the patient ids to its classes and annotations. We took metadata from dicom images performed EDA on the dataset for better understanding of data.

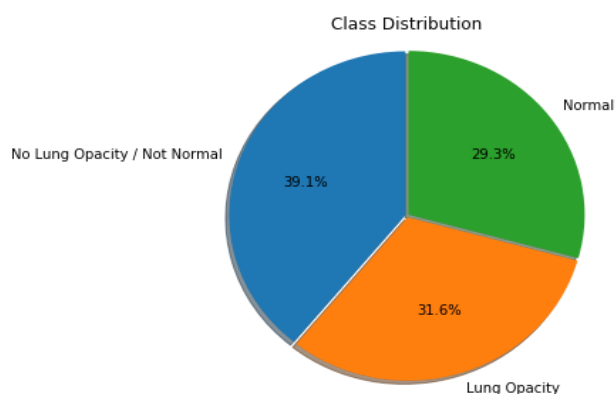
Data findings from the data:

The csv file 'stage_2_train_labels.csv' contains the patient id , the bounding box coordinates (x, y, width and height) along with the Target column consisting of values 0 and 1. There are 30,227 records in this file. The other file is 'stage_2_detailed_class_info.csv' which contains the patient id and the classes: **No Lung Opacity/Not Normal, Normal, Lung Opacity**.



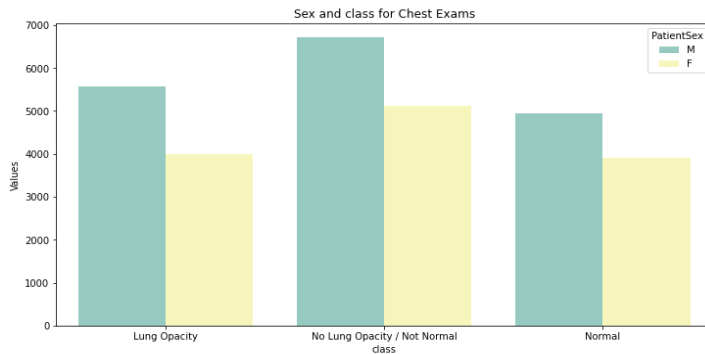
one can notice that the instances with class 0 are more than double that of class 1 instances which can be considered as class imbalance problem and to be handled as part of classification task during subsequent phases of this work. One can notice the relative numbers from the adjacent count plot

% of population across opacity groups:

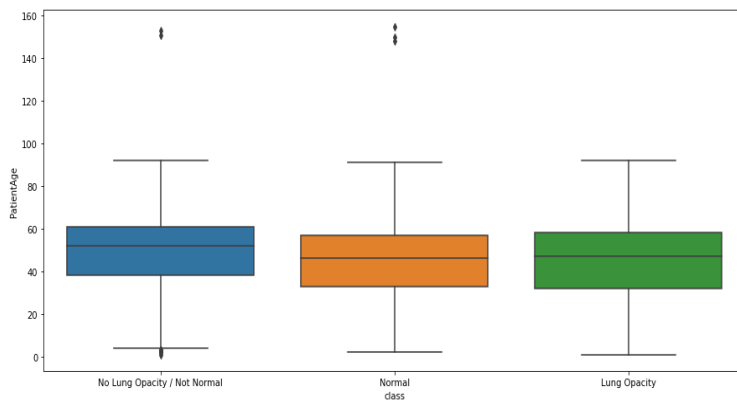


This pie chart suggests data is divided 30% each for all classes

We have performed detailed analysis on the gender v/s classes. It can also be seen that the number of male patients are more than that of female across the label categories as depicted from the count plot below:



Outlier analysis is performed with respect to the patient age attribute and the opacity. The associated box plots are furnished in the Figure 7 below:



Number of patients in age category

<=75	13318
<=50	12157
<=26	3972
<=100	780

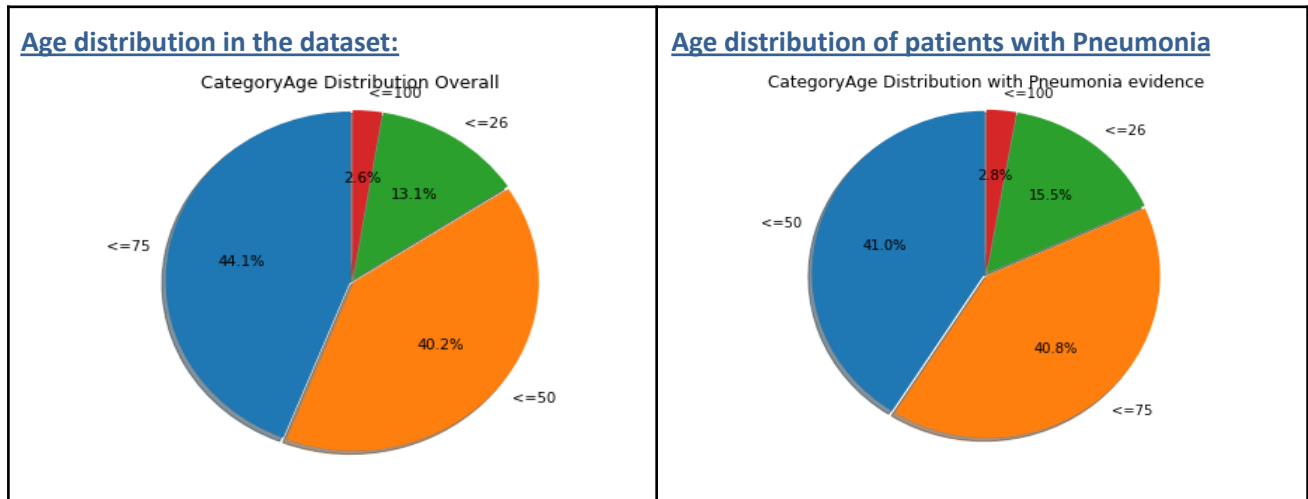
It is also observed that:

- The mean age is 46 years , whereas the minimum age is 1 year and the max age is 155 which seems to be an outlier.
- 50% of the patients are of around 49 age , the std deviation is 16 which suggest that age is not normally distributed.
- There are 8851 normal cases, people with lung opacity are 9555 and No Lung Opacity / Not Normal are 11821.
- Patients with evidence of Pneumonia are associated with Lung Opacity class and target = 1.
- Patients with no definitive evidence of Pneumonia are either of Normal or No Lung Opacity / Not Normal class and target = 0.

Based on these features following visualizations are generated:

It is observed that

- Numbers of patients are highest for age group 51-75 for overall case
- But the number of patients is highest for the age group 27-50 for Target=1 case.



Metadata in the DCIM image:

A single DCIM image was taken from the dataset and the metadata has been displayed. It includes the patient details such as name, age, modality, gender, view position; details about the image itself, date and time among other parameters. Additional parameters (age, gender, view position, pixel spacing) have been appended to the amalgamated data frame for analysis.

Data frame attribute analysis:

Gender: The dataset consists of a higher percentage of male patients compared to females.

Patient age: The histogram plot indicates that there is a greater representation of patients in the 40- 60 age range.

ViewPosition: It indicates if the x-ray is taken from posterior or anterior position.

Preprocessing of image data

In the preprocessing phase we resized the images for faster computation and added 3 channels to dicom images.

Original size: 1024X1024

Resized size: 128X128

We created a basic CNN model for image classification taking target as target variable after that fine tuned the model for better accuracy and also user transfer learning on the model . Transfer learning model are as follows:

VGGNet16 and VGGNet19:

VGGNet-16 is a convolutional neural network (CNN) architecture developed by Visual Geometry Group (VGG) at Oxford University. It is composed of 16 layers of convolutional layers, as well as fully connected layers, and is designed to classify images into 1000 object categories.

Our model accuracy for **VGGNet16 is 75%** and **VGGNet19 is 76%**

Inception Network :

Inception network is a convolutional neural network (CNN) architecture used for solving image recognition and detection problems.

Our model accuracy for **Inception net is 78%**

RestNet :

ResNet is a convolutional neural network (CNN) architecture developed by Microsoft Research [1], which is characterized by its very deep layers reducing the amount of parameters required, making it a popular choice for many computer vision tasks.

Our model accuracy for **RestNet is 82%**

DenseNet :

DenseNet is a convolutional neural network (CNN) architecture developed by researchers at the University of California, Berkeley. It is characterized by its densely connected layers, which are connected to one another in a feed-forward fashion.

We have achieved the accuracy of **84% through DenseNet**

Object detection model

After transfer learning we create an object detection model for detecting swelling in lungs which indicates pneumonia. We created a network of the model and custom functions to calculate loss and accuracy. We used jaccard loss also known as iou loss and mean iou to calculate IOU of the model. We use a data loader function to load images and annotations in the model.

After the model is fitted through the train data we evaluate using validation data and test it on test images.

Step By Step walkthrough the solution

1. **Importing data:** First step is to import from the csv files. File stage_2 labels.csv contains patient id, annotations for the bounding box of patients who have pneumonia, target referring if patient has pneumonia or not. File detailed_classinfo file contains class of pneumonia patients with patient id of patients.
2. **Mapping images to its classes and annotations:** To map images to its given classes and annotations we firstly merge the data frames of labels and classes after that we created a function to take metadata out of dicom images and attach to its patient. We added a column to the dataframe for the image path of patients.
3. **Preprocessing and Visualisation of different classes:** We performed EDA on data to understand the data better and use that information in model building. We preprocessed the images before taking them into the model building phase. The original size of images was 1024X1024 which can slow down model building process so we resized to 128X128 for faster model building process also added 3 channels to dicom image which is grey scale originally. Converted images into a numpy array and took the target as the target column.
4. **Displaying images with bounding box:** We wrote a function to show bounding boxes on images. Function takes sample data from the user and checks if the target is one or zero based on that it draws a bounding box on the infected area of the lungs using annotations from dataframe.
5. **Design, train and test basic CNN models for classification:** After pre-processing data we create a basic CNN model for classification of images to improve accuracy we use densenet123 .
6. **Fine tune the trained basic CNN models for classification:** To fine tune basic CNN model we added more convolution layers fitted data through more epochs and added learning rate to optimizer to get better accuracy, precision and recall.
7. **Apply Transfer Learning model for classification:** we used densenet, vggnet16, vggnet19, resnet, inceptionnet models to train model on our data to find how per trained models affect the accuracy.
8. **Design, train and test RCNN & its hybrids based object detection models to impose the bounding box or mask over the area of interest:** To create object detection model we need to firstly pre-process the data for data we create dictionary called pneumonia locations which contains patientid and annotations of bounding boxes and a list of files names of images after that we create a data generator which takes image data and filenames create a mask for image using bounding box annotations and return this data after this we create a network for the bounding box prediction model compile it with jaccard loss function, iou function for accuracy of the model and load the data with data loader function fit this data in model.
9. **Pickle the model for future prediction:** after the model is generated and fitted with weights we save the model using pickle module for future prediction.

Model Evaluation

Image classification model:

There are several techniques for evaluating image classification models. The most typical approach for evaluating classification performance is using metrics such as precision, recall, f-measure, and accuracy. These metrics are computed from a confusion matrix, which compares the predicted class with the actual class.

Precision — Out of all the examples that are predicted as positive, how many are really positive?

Recall — Out of all the positive examples, how many are predicted as positive?

These are important measures in the medical field along with specificity and sensitivity. high precision of class 1 is important because class 1 indicates the presence of pneumonia. If precision is low we will miss the cases of pneumonia which can have fatal repercussions. The same can be said about recall of the model.hence having high precision and recall of model used in the medical field is important.

Here are precision,recall,f1 score and accuracy percentage of all the models.

Model	Precision%	Recall%	F1 score%	Accuracy%
Basic CNN model	62	54	52	69
Fine tuned Basic CNN model	56	60	58	72
DenseNet with Validation data	72	60	66	80
DenseNet	73	79	76	84
VggNet16	57	81	67	75
VggNet19	64	54	59	76
ResNet	79	56	66	82
InceptionNet	64	72	68	78

colour	rank
	highest acc
	second highest acc
	third highest acc

Densenet has the highest accuracy followed by the resnet model.

Object detection model :

When evaluating an object detection model, there are several metrics that are commonly used to assess its performance. These metrics are typically used to measure the model's ability to accurately locate and classify objects within an image.

We are using intersection over union metric with the jaccard loss function.

Intersection over Union (IoU) is a metric used to evaluate the performance of object detection models. It measures the overlap between the predicted bounding box and the ground-truth bounding box. The IoU value ranges between 0 and 1, where 1 represents a perfect match between the predicted and ground-truth bounding boxes, and 0 represents no overlap.

The IoU is calculated by taking the ratio of the area of the intersection of the predicted and ground-truth bounding boxes to the area of the union of the two bounding boxes:

$$\text{IoU} = (\text{Intersection Area}) / (\text{Union Area})$$

A common practice is to consider a detection as a true positive when IoU is greater than a certain threshold, typically 0.5.

mean Iou for the validation data of the model is 73%, validation accuracy of model is 96.04%.

Comparison to Benchmark

For benchmarking we used Cornell University Machine Learning Group Paper - CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning.

This research paper explains how they used a chest x-ray dataset to train a deep learning model for detection of 14 diseases which included pneumonia. The paper was released in 2017 and for comparison they asked radiologists to look through x-rays and identify the disease; the cheXnet model outperformed radiologists in disease detection. Here are the results.

Pathology	Wang et al. (2017)	Yao et al. (2017)	CheXNet (ours)
Pneumonia	0.633	0.713	0.7680

When compared to our CNN classification model after fine tuning of parameters we have accuracy of 72% which is close to benchmark.

Densenet,resnet,inceptionet are outperforming the benchmark model.

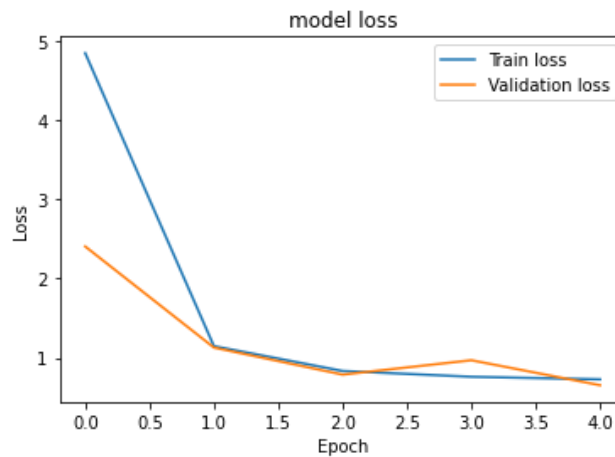
link to the research paper : <https://arxiv.org/abs/1711.05225>

Visualizations

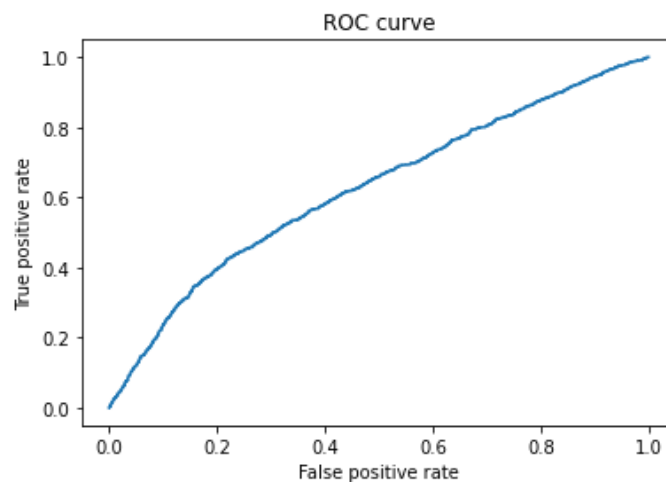
Here are some visualizations used to evaluate the model

Classification Report of basic cnn model:

	precision	recall	f1-score	support
0	0.70	0.94	0.80	3082
1	0.54	0.14	0.23	1452
accuracy			0.69	4534
macro avg	0.62	0.54	0.52	4534
weighted avg	0.65	0.69	0.62	4534



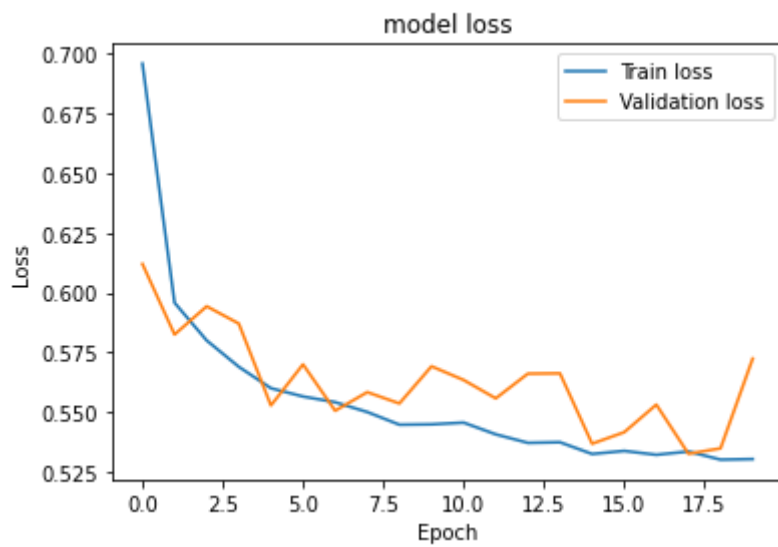
comparison of loss of train and validation data of basic cnn model.



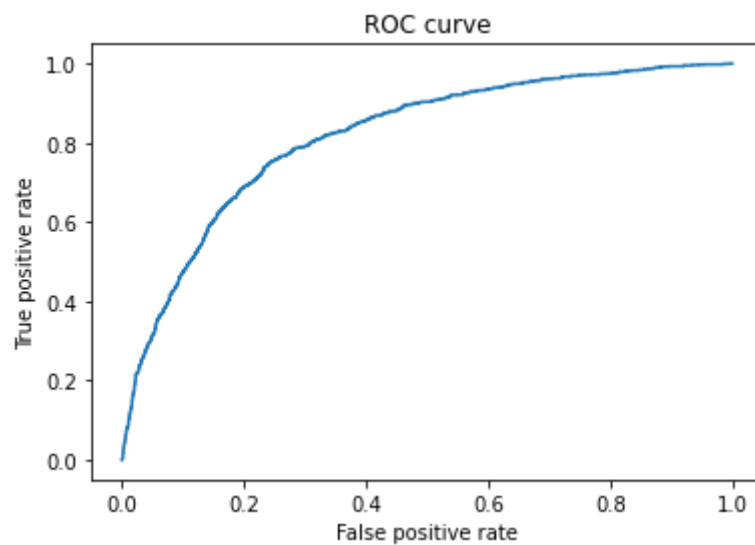
Roc curve of basic cnn model

classification report of fine tuned cnn model:

	precision	recall	f1-score	support
0	0.81	0.77	0.79	3082
1	0.56	0.60	0.58	1452
accuracy			0.72	4534
macro avg	0.68	0.69	0.68	4534
weighted avg	0.73	0.72	0.72	4534



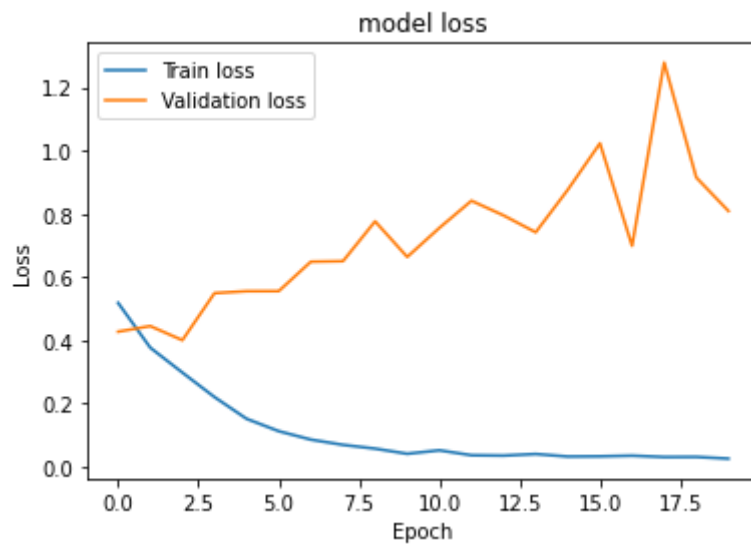
comparison of loss of train and validation data of fine tuned cnn model.



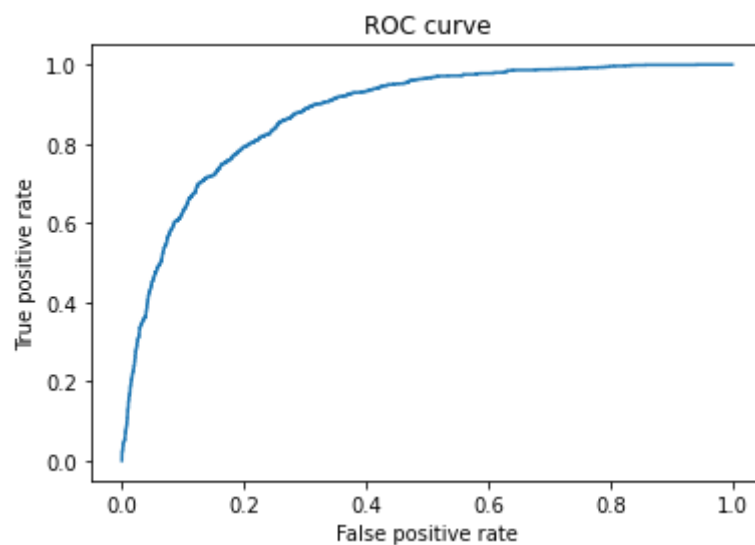
Roc curve of fine tuned cnn model

Classification Report of Densenet:

	precision	recall	f1-score	support
0	0.90	0.87	0.88	4162
1	0.73	0.79	0.76	1884
accuracy			0.84	6046
macro avg	0.81	0.83	0.82	6046
weighted avg	0.85	0.84	0.84	6046



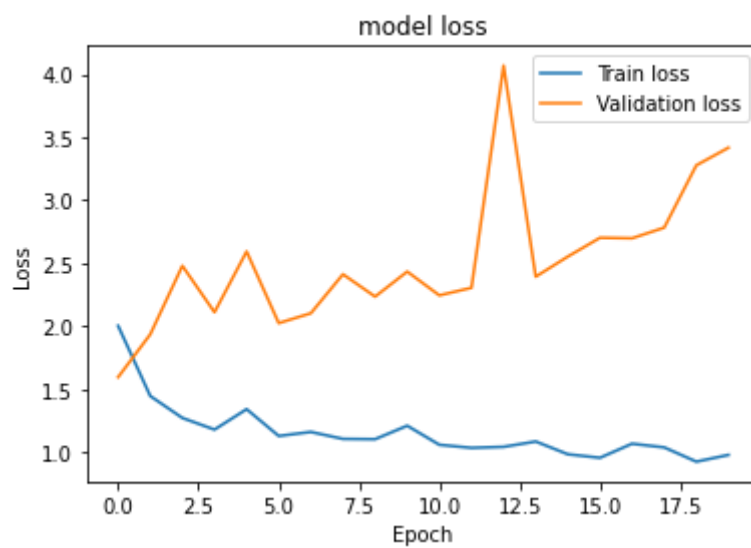
comparison of loss of train and validation data of dense net model



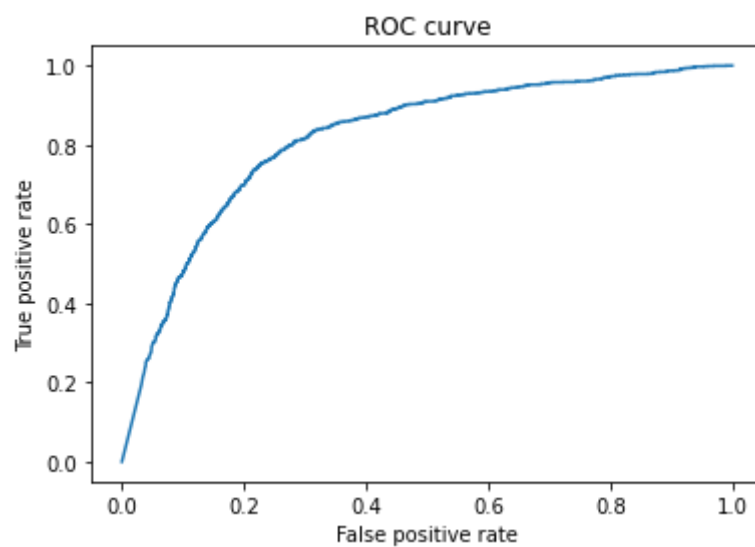
Roc curve densenet model

Classification Report of Vgg16:

	precision	recall	f1-score	support
0	0.90	0.72	0.80	4162
1	0.57	0.81	0.67	1884
accuracy			0.75	6046
macro avg	0.73	0.77	0.73	6046
weighted avg	0.79	0.75	0.76	6046



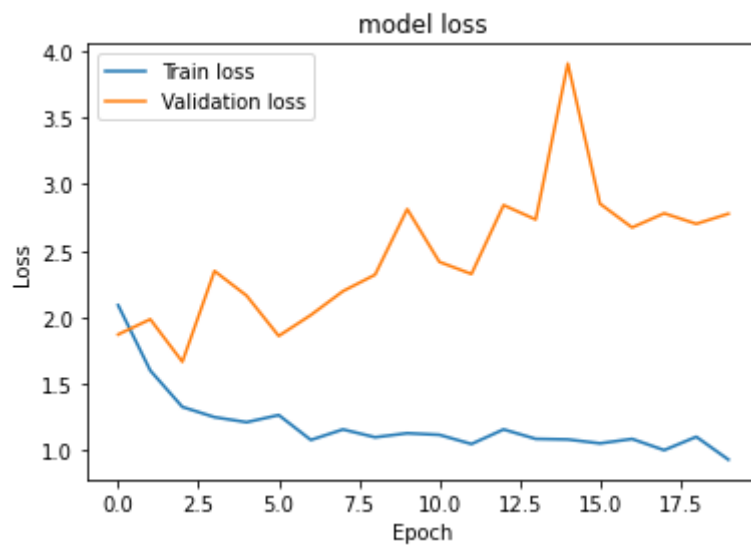
comparison of loss of train and validation data of Vgg16 model



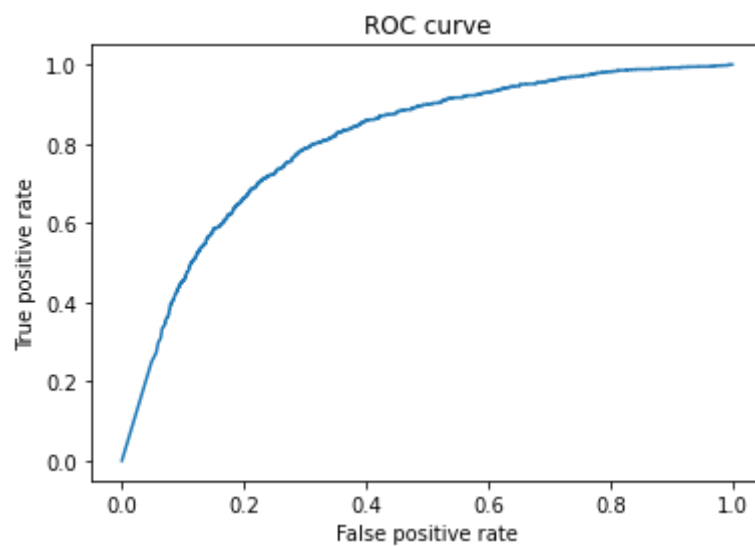
Roc curve Vgg16 model

Classification Report of Vgg19:

	precision	recall	f1-score	support
0	0.81	0.86	0.83	4162
1	0.64	0.54	0.59	1884
accuracy			0.76	6046
macro avg	0.72	0.70	0.71	6046
weighted avg	0.75	0.76	0.76	6046



comparison of loss of train and validation data of Vgg19 model



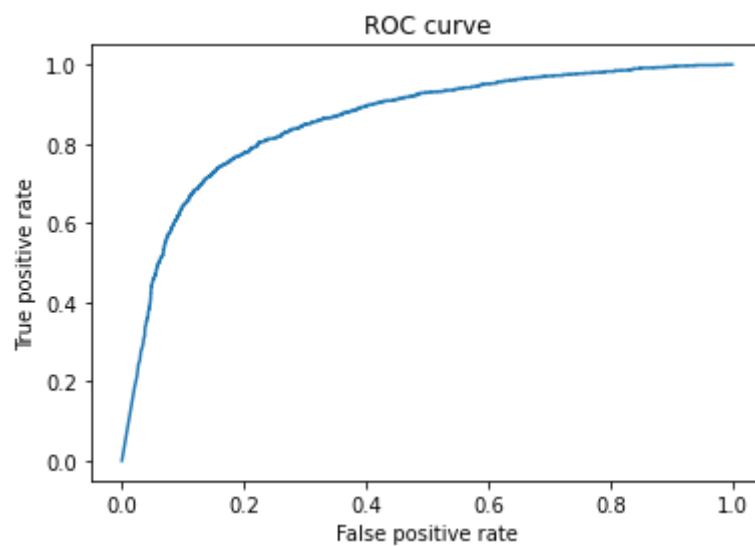
Roc curve of Vgg19 model

Classification Report of Resnet:

	precision	recall	f1-score	support
0	0.82	0.93	0.88	4162
1	0.79	0.56	0.66	1884
accuracy			0.82	6046
macro avg	0.81	0.75	0.77	6046
weighted avg	0.81	0.82	0.81	6046



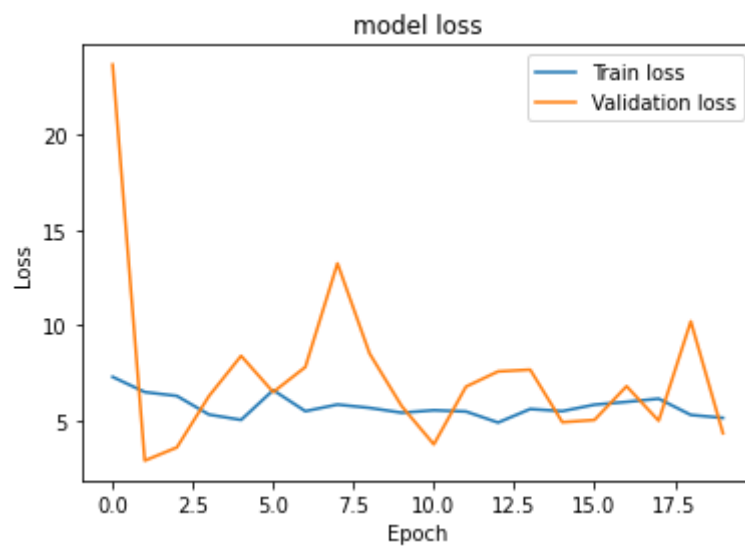
comparison of loss of train and validation data of resnet model



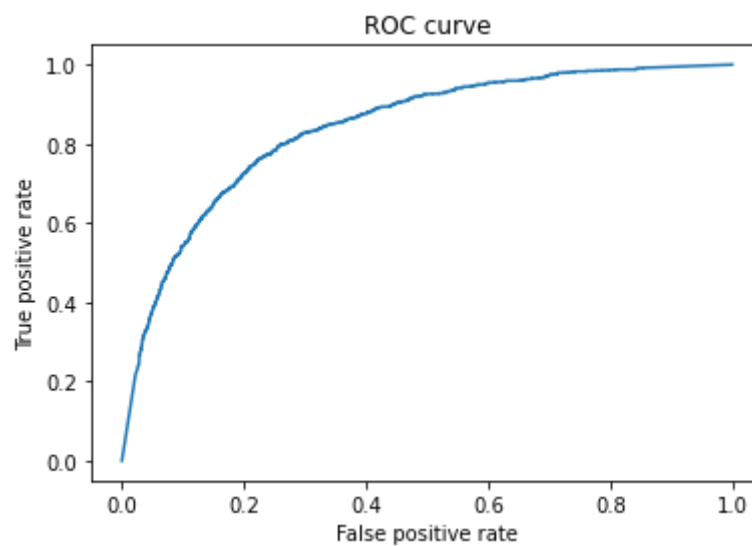
Roc curve of resnet model

Classification Report of inceptionnet:

	precision	recall	f1-score	support
0	0.87	0.81	0.84	4162
1	0.64	0.72	0.68	1884
accuracy			0.78	6046
macro avg	0.75	0.77	0.76	6046
weighted avg	0.79	0.78	0.79	6046

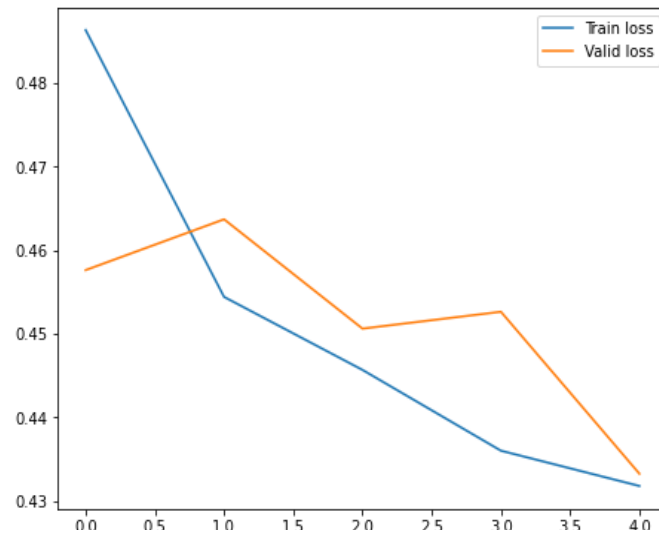


comparison of loss of train and validation data of inceptionnet model

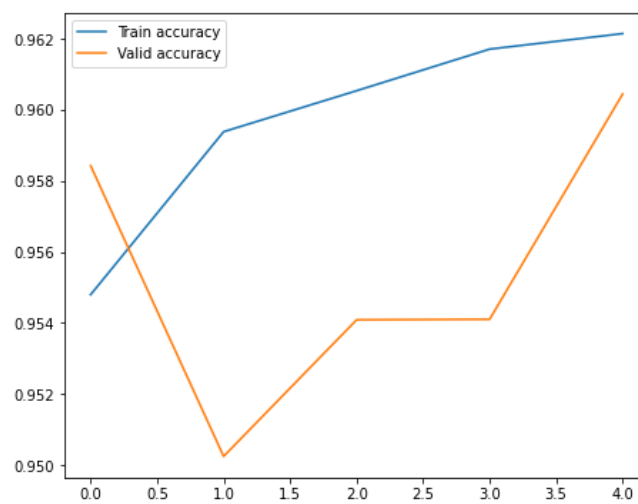


Roc curve of inceptionnet model

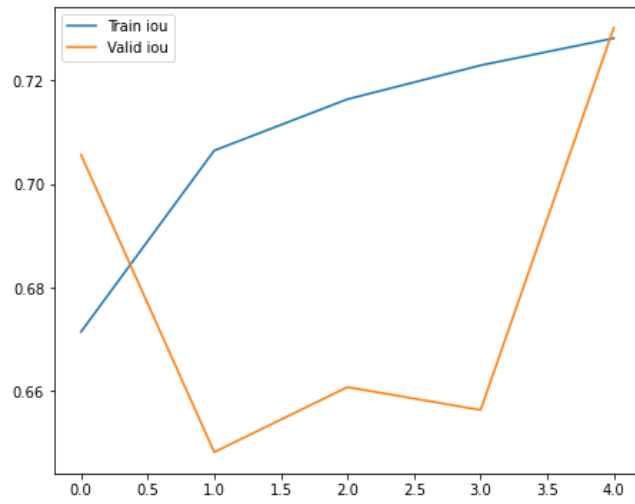
Object detection model visualizations:



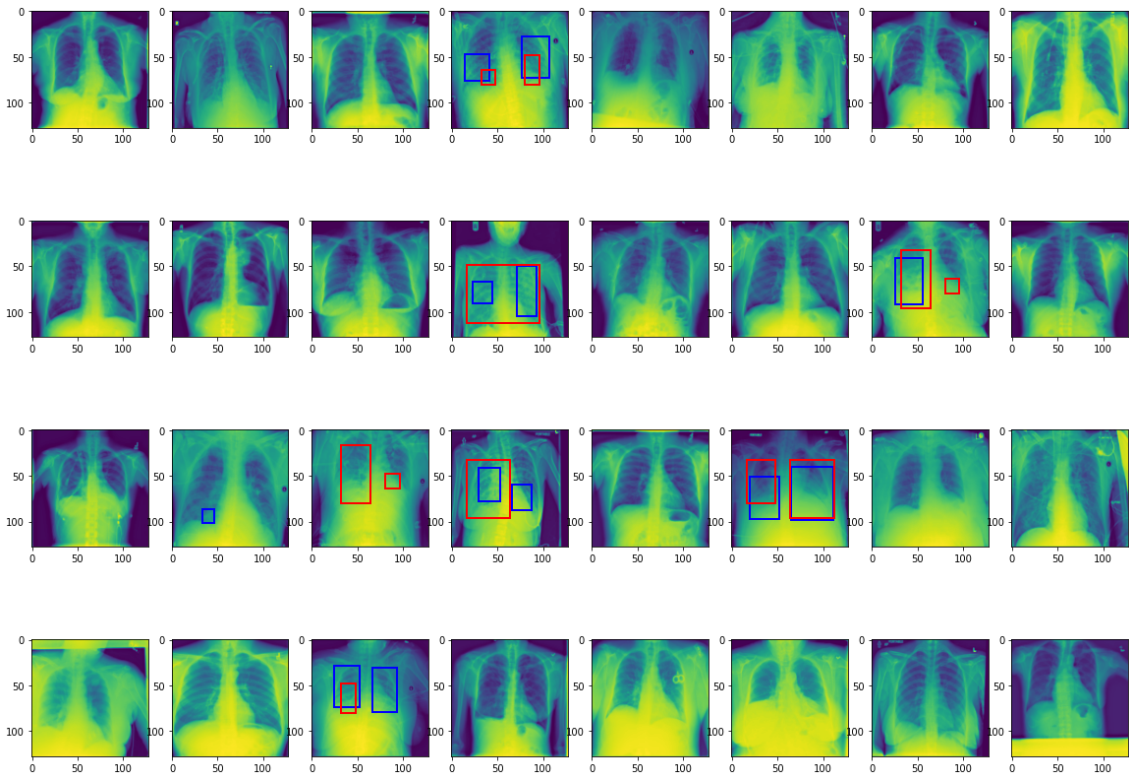
comparison of train loss and validation loss



comparison of train accuracy and validation accuracy



comparison of train IOU and validation IOU



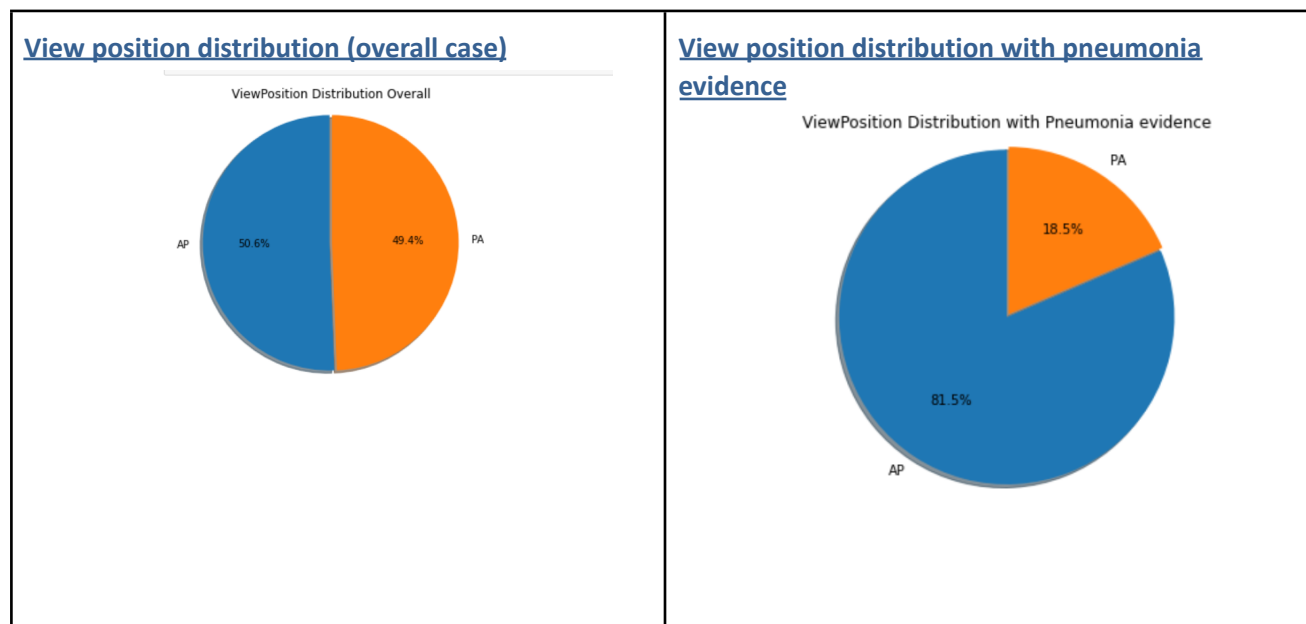
comparison of round truth and predicted values of bounding boxes.

Implications

Due to the moderate accuracy rate and high miss rate, the model is recommended to be used as an assistant at the screening stage to predict whether a patient is suffering from Pneumonia but not recommended to use to state the patient not to be having Pneumonia.

From Data findings we found out that X Rays taken from backside of the body has more chance of detection of pneumonia, here are visualizations to support the statement:

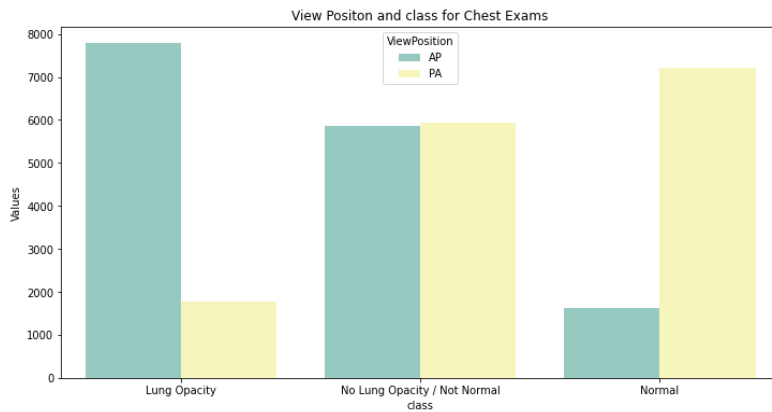
- **Posterior/Anterior (PA):** X-ray is taken from the back part of the chest. So it hits the posterior part before the anterior part. Patient needs to stand against the X-ray machine.
- **Anterior/Posterior (AP):** X-ray is taken from the front part of the chest. So it hits the anterior part before the posterior part. This is taken when Patient cannot stand against an X-ray machine but heart size is exaggerated.



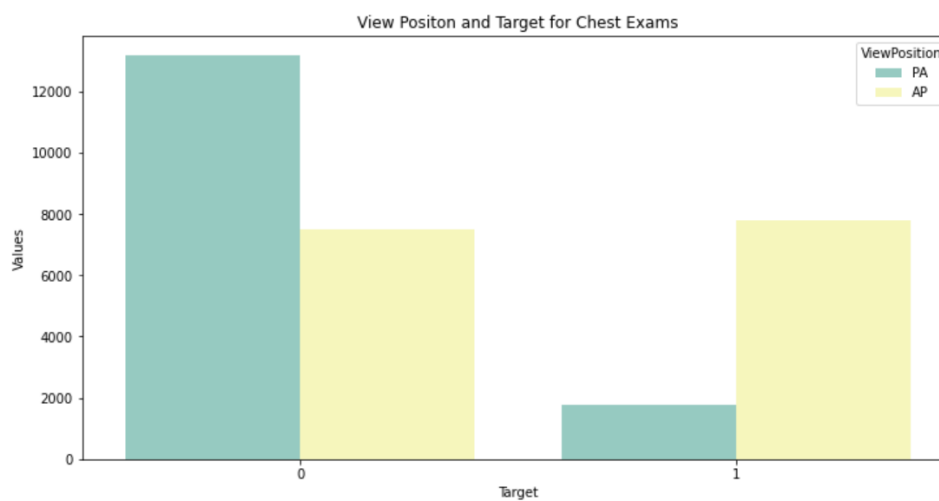
Observation:

1. ViewPosition = PA and ViewPosition = AP are almost equally distributed in the overall train data
2. When Target=1, View position = AP dominates.

View positions and various states of opacity



View positions and Target



The solution has the following advantages. Our solution will help in quick decision making (in urgent cases), saves cost, etc.

High accessibility: The model can be deployed on cloud and with a web application, therefore it will be easier for the end users to access the application on any computer with a browser installed.

Higher availability: Diseases, like pneumonia, require early diagnostics for doctors to detect it. False-negative and false-positive results can both be rather destructive. Computer vision excludes the possibility of human error to some degree and serves as an assistant for radiologists. And it's available 24/7.

Speed, urgency: Solution allows doctors to focus more on patients rather than on examining X-ray images. Speed might be crucial for urgent situations.

Cost savings: It results in cost rationalization in terms of lesser staff for examining X-ray images.

Limitations

- During EDA it is observed that Gender, Age and View Position attributes are having positive correlation with the targets. But, in this model these features are not utilized explicitly.
- Due to the number of false negatives. We recommend this model to be used at screening stage to predict whether a patient is suffering from Pneumonia but not recommended to use to state the patient not to be having Pneumonia.
- Overall accuracy is moderate.
- As an additional feature making a mobile app or a web application would have been more useful to the doctors.
- This model could be used as an assistant at the screening stage, not as an accurate to replace humans.

Closing Reflections

During the lifecycle for the project there were many learnings. Such as,

- Dealing with large data sets and running multiple GPUs.
- Using multiple algorithms for the same problem and evaluating the models

Improvements that could be done are,

- Improving the model to make decisions accurately on false negatives and increase the overall precision.
- Due to resource constraints we were able to use a sample of data. In the next step we can use better computational sources and use whole data for better accuracy and precision and accuracy.
- A mobile app for the doctors where they can see the reports for the concerned patients will speed up the process of diagnosing the affected person and can be life saving.
- Features to upload as bulk radiographs of patients and see the reports. an application which can be installed on doctors computer and doctor and upload bulk x-rays and model can suggest if a patient has pneumonia or not with some percent of confidence.
- Integration of cloud deployed model to their own applications running on their computers

Acknowledgments

We have explored the additional information provided at Kaggle web site, available at:

<https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/overview/acknowledgements> during the present work.

We have used research paper of Cornell university :CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning as benchmark for research

<https://arxiv.org/abs/1711.05225>