Pneumonia Detection Challenge

Capstone Project

AIML.O.Jun'20B.Group9A

FINAL REPORT

PROJECT OVERVIEW

1 Problem statement

2 Data Pre-Processing

3 Exploratory Data Analysis

4 Model Building

5 Model Tuning and Performance

6 Industry Best Practices

PROBLEM STATEMENT

Pneumonia refers to an infection in one or both the lungs. It may be caused by microorganisms like bacteria, viruses, and fungi. It causes acute inflammation in the air sacs called alveoli present in the lungs, thereby filling these with fluid or pus, making it difficult to breathe.

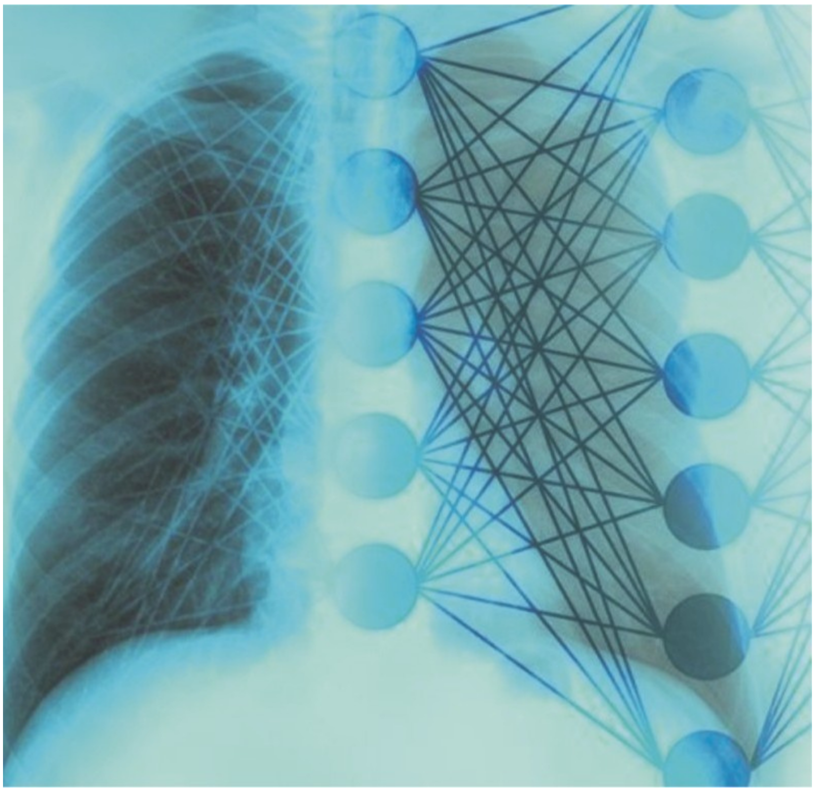
It is a major cause of deaths of children under 5 years of age. But with the onset of Covid, it is equally affecting and causing mortalities in people of all age groups.

Diagnosis of pneumonia is quite challenging and doctors have to rely on multiple parameters like clinical history, vital signs and laboratory exams. Furthermore, various other conditions in the lungs such as fluid overload (pulmonary edema), bleeding, volume loss (atelectasis or collapse), lung cancer, post-radiation or surgical changes, fluid in the pleural space (pleural effusion), etc. make the diagnosis more complicated.

However, medical imaging techniques like chest radiograph (CXR) or a CT Scan are treated as most reliable diagnostic tools for its detection. These techniques rest on the premise of differential absorption of radiation by various body parts.

PROBLEM STATEMENT

Tissues with sparse material, such as lungs which are full of air, do not absorb the X-rays and appear black in the image. Dense tissues such as bones absorb X-rays and appear white in the image. While tissue or fluid absorb moderate radiations and appear grey in the scan.
In a normal image the lungs are seen as black, but they have different projections on them - mainly the rib cage bones, main airways, blood vessels and the heart.
In case of pneumonia, a haziness (also referred to as consolidation) is present in the chest x-ray image.



PROBLEM STATEMENT

There can be another case which indicates that while pneumonia was determined not to be present, there was nonetheless some type of abnormality on the image and often times this finding may mimic the appearance of true pneumonia. This is very tricky to detect and needs expert medical practitioners to differentiate.
Therefore, computer-aided diagnosis systems are needed to guide clinicians. In the division of health care, artificial intelligence plays a major and vital role in order to perform better diagnosis with less error rate. In the present day's advancement in deep learning techniques that can be employed to predict the disease with higher accuracy.

In this project work, a comprehensive analysis is performed using CNN models such as mobile net, ResNet, Faster RCNN (Region Convolutional Neural network) and Mask RCNN. The model will be trained and tested with the 9555 images of the x-ray dataset and will aim to show that our model will achieve industry ready accuracy.

DATA PREPROCESSING

26684 DICOM format images with corresponding bounding boxes and patient ID's are taken in the form of a CSV file.

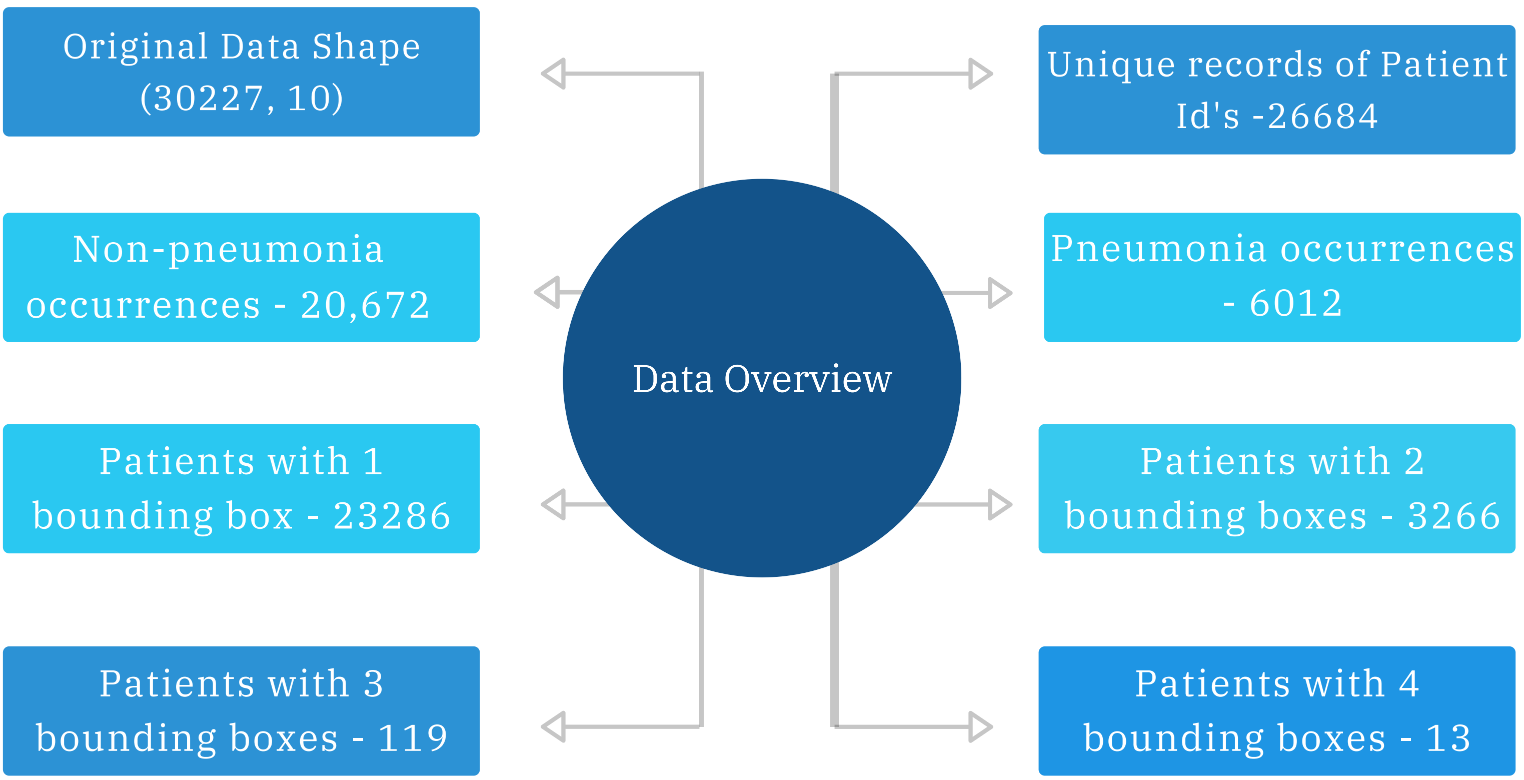
These images are classified into 3 classes - Normal, Lung opacity and No lung opacity.

All images are converted to DICOM format (Digital Imaging and Communications in Medicine) and is standard for handling storing, printing, and transmitting information in medical imaging

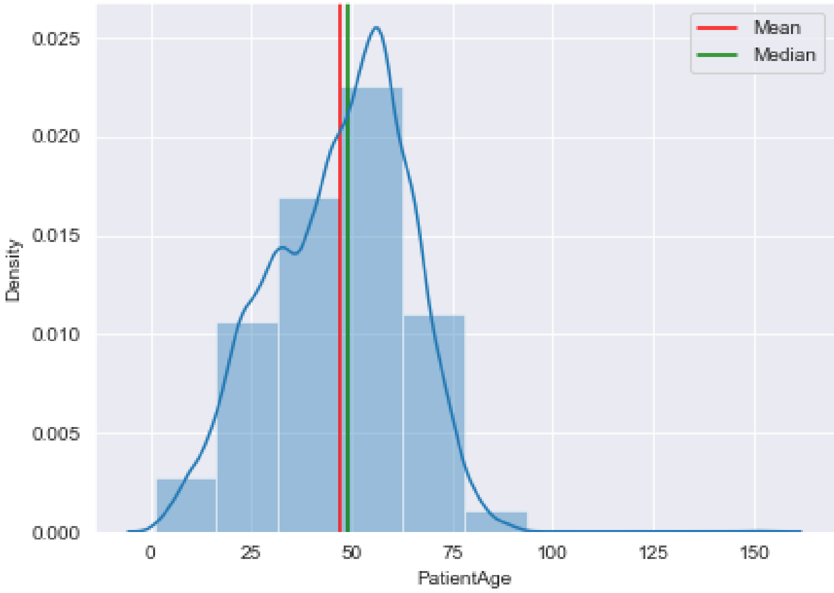
The DICOM images are resized into 256x256 image resolution and converted into an image array.

The bounding boxes corresponding to the patient id and will be utilized during training and testing process.

DATA OVERVIEW

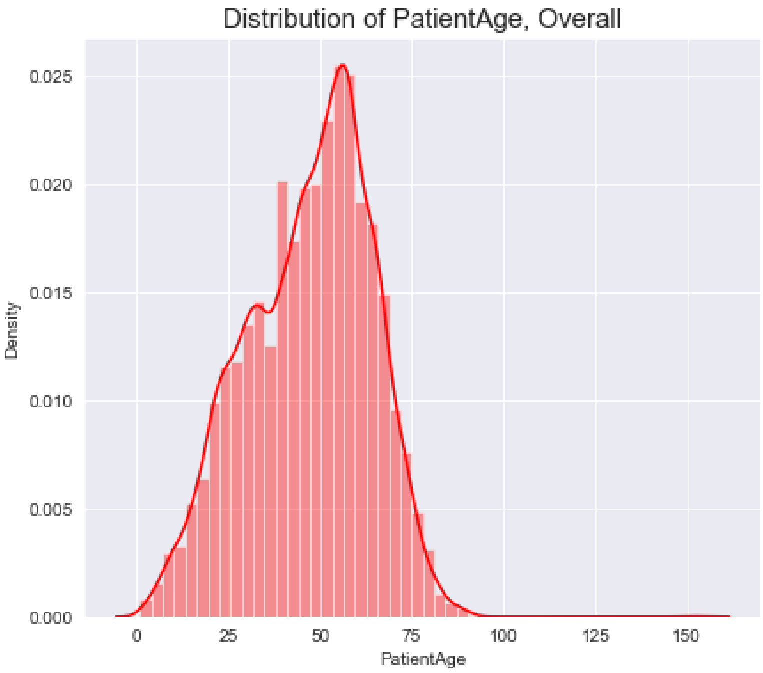


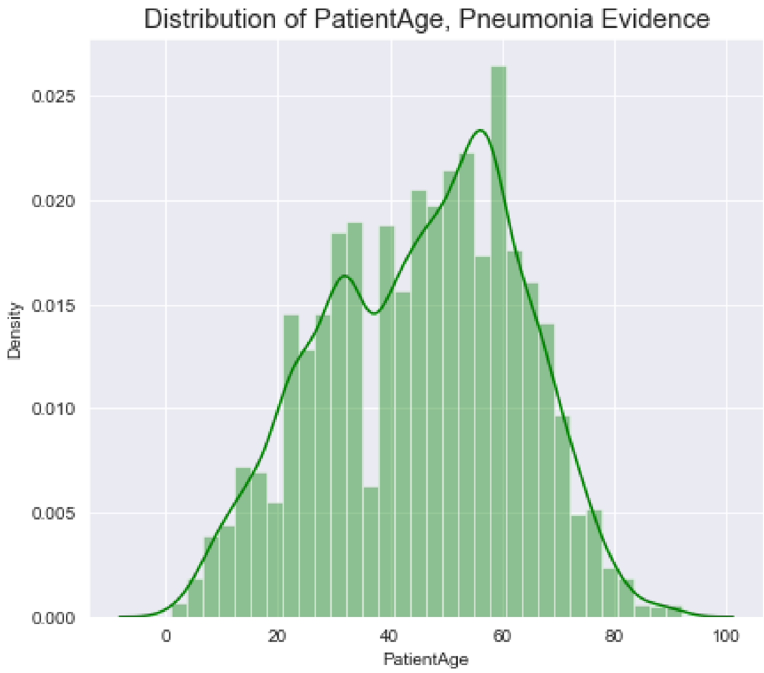
UNIVARIATE ANALYSIS



PATIENT AGE

The histogram is representative of the distribution of the patient age.
Mean age - 47 years
Median age - 49 years
Skewness measure is -0.23



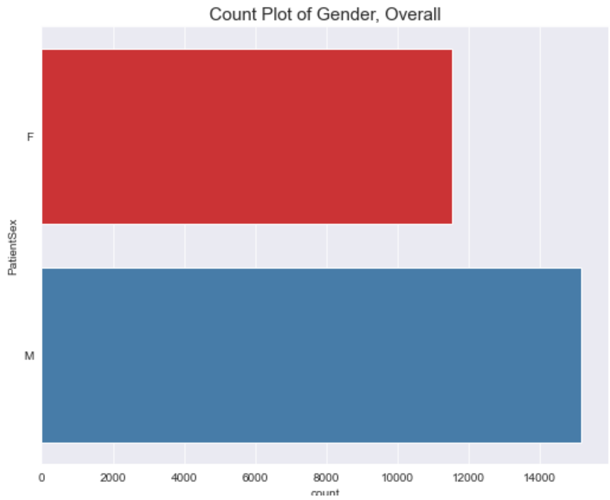


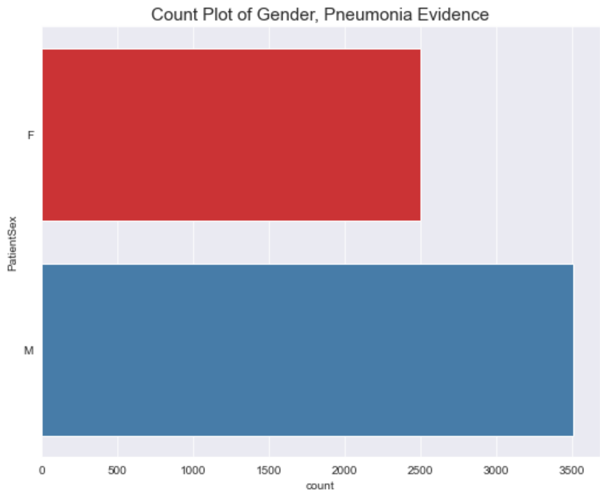
'PatientAge' is normally distributed with mean and median values almost coinciding for both kinds of patients.

25% each of the age values lie at less than 34 years and more than 59 years, while 50% of the age values lie between 34 and 59 years.

Maximum number of positive pneumonia cases were observed for patients in the age of 54 years.

UNIVARIATE ANALYSIS



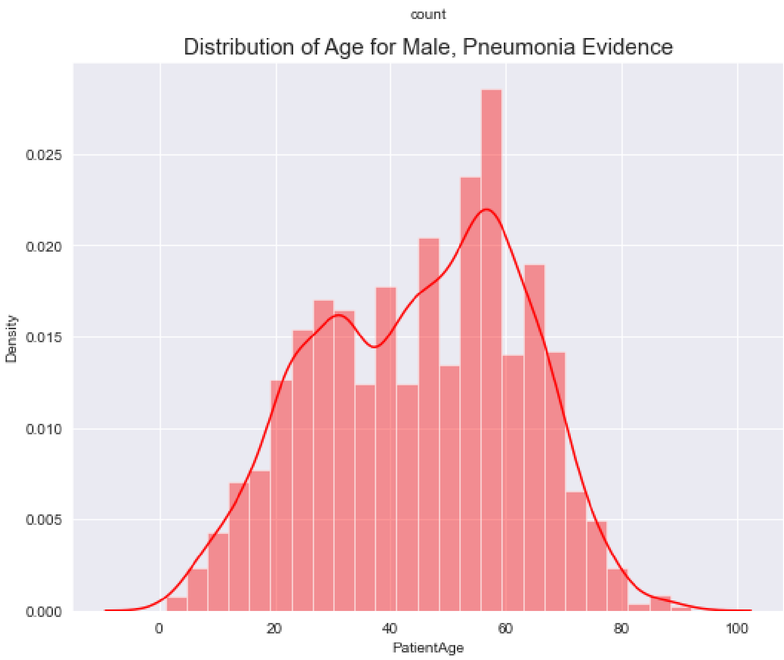


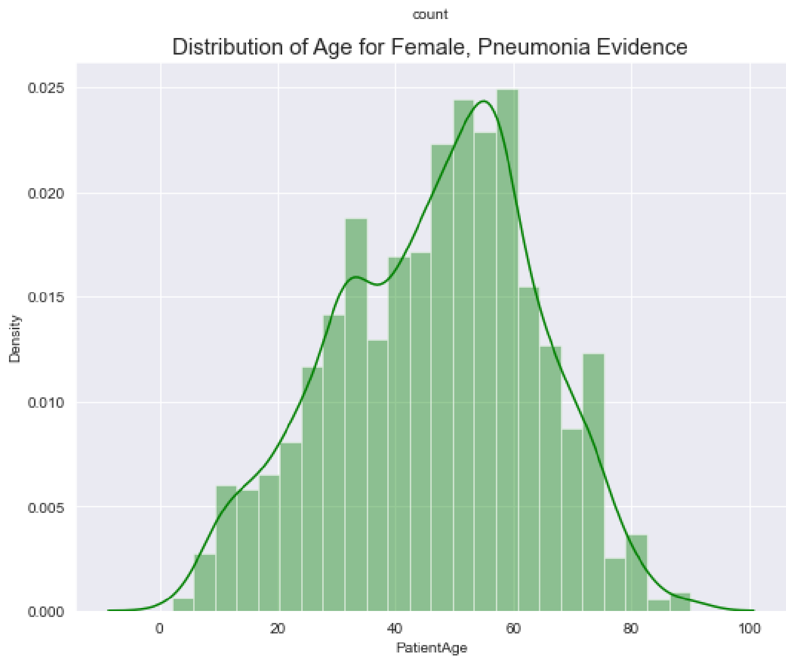
PATIENT SEX

The countplot indicates that overall 43% are females while male patients form the remaining 57%.

Also, for the pneumonic patients, females form 42%, while the remaining 58% are males.

This indicates a nearly balanced distribution of patients on the basis of gender.

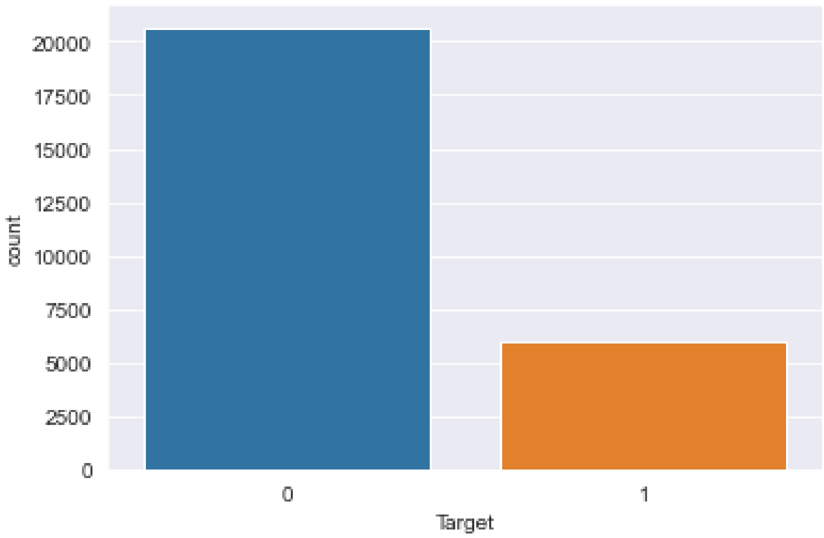




UNIVARIATE ANALYSIS

Distributions for both look similar.

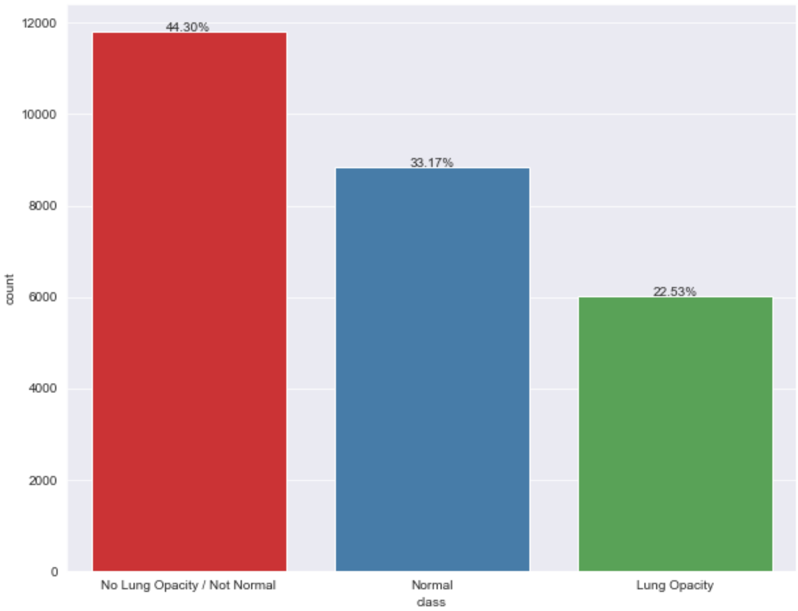
25% each of values lie at less than 34 and more than 59, while 50% of the values lie between 34 and 59.



TARGET COLUMN

Our dataset contains information about 20,672 (77.5%) patients with Target=0 (Normal), and 6,012(22.5%) patients with Target=1(Pneumonia).

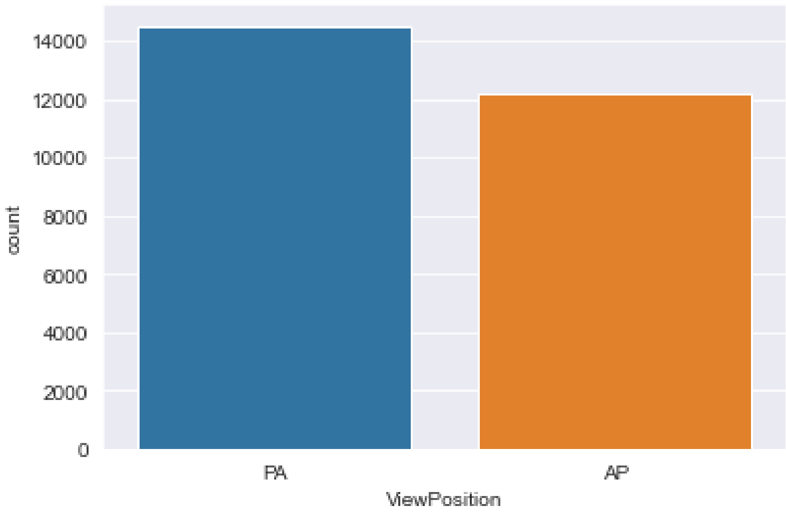
This shows the target imbalance in the dataset, with normal patients as many as more than 3 times the pneumonic patients.



CLASS COLUMN

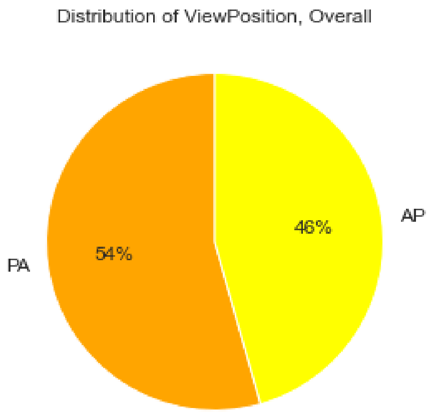
No Lung Opacity / Not Normal : 11821 or 44.29%
Normal : 8851 or 33.16%
Lung Opacity : 6012 or 22.53%

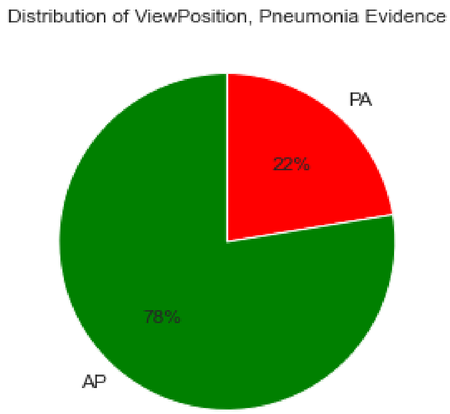
UNIVARIATE ANALYSIS



VIEW POSITION

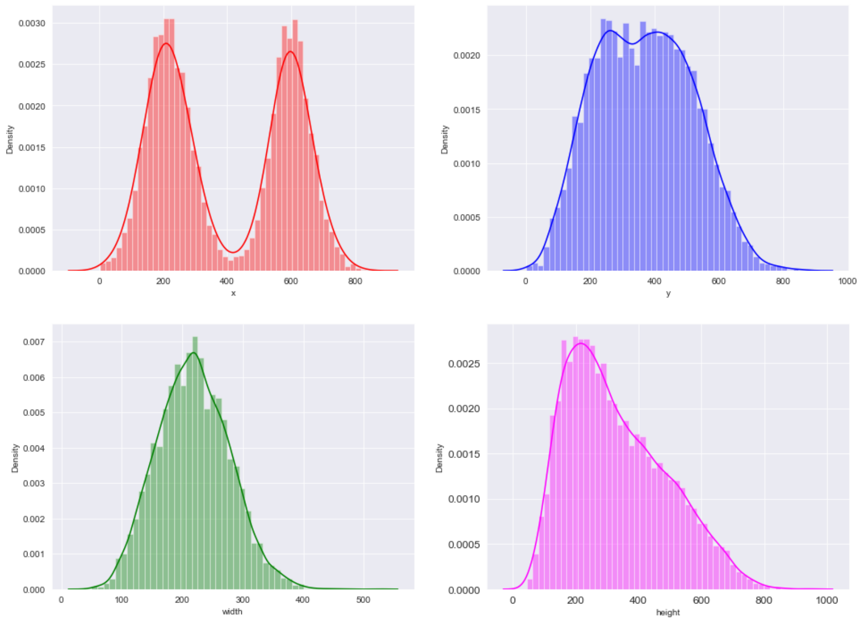
Posterior Anterior (PA) Scans - 14511
Anterior Posterior (AP) Scans - 12173





Almost equal distribution of PA and AP View Positions.
However for Pneumonia Evidence, AP Position gives significantly more positive results than PA Position.

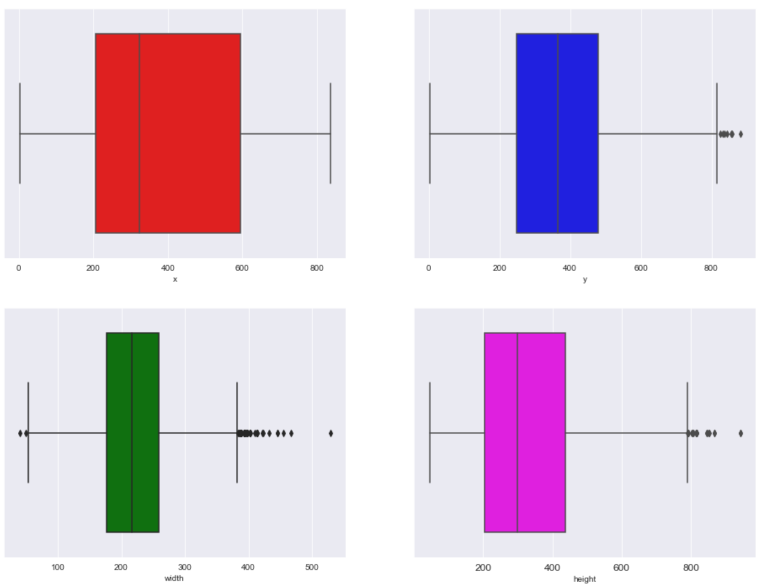
BOUNDING BOX ANALYSIS



BOUNDING BOX ANALYSIS

The distribution plots hint at near normal distributions of bounding box attributes.

The distribution plot of x shows 2 gaussians, hinting at 2 different clusters pertaining to each lung.



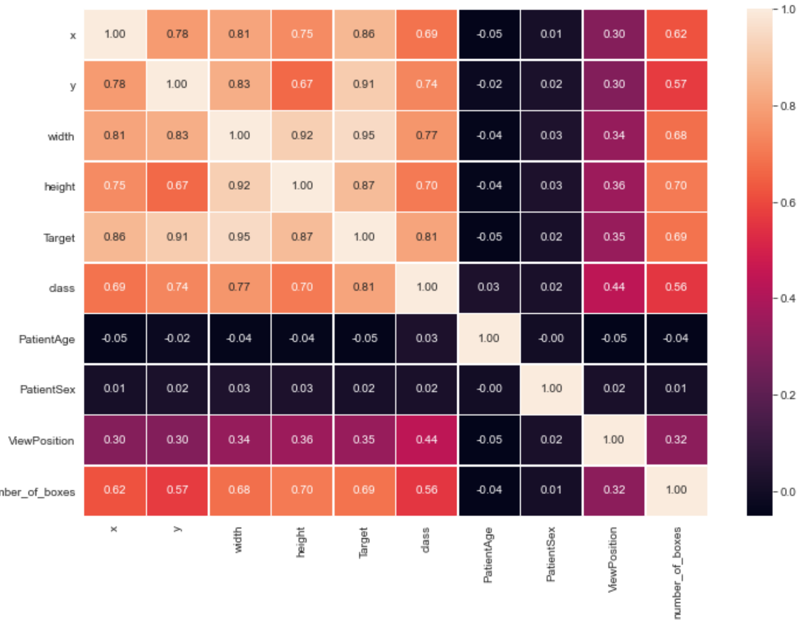
Boxplots of y, width and height show some outliers mostly on the upper side of the right whisker.

As x and y represent only the pixel positions of the bounding boxes, these are immaterial and can be ignored.

However, the width and height of the bounding boxes give us valuable insights about the patients. These attributes are directly proportional to the severity of infection for the patients.

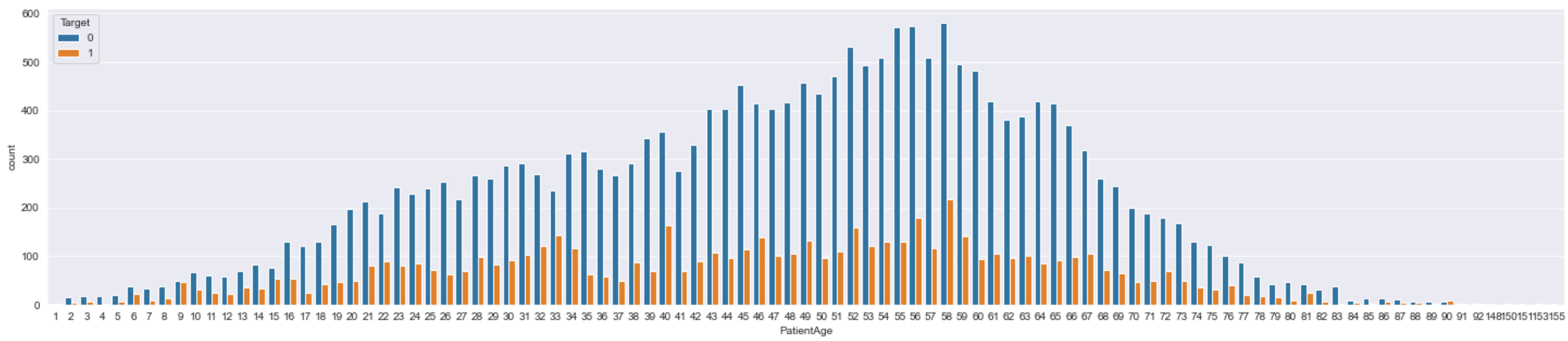
As we observe many outliers in the width and height columns, we can infer that there are many patients who are having more severe infections as compared to the others.

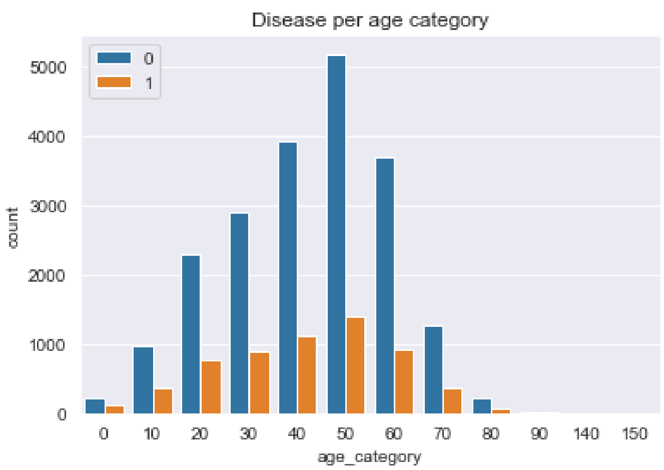
BIVARIATE AND MULTIVARIATE ANALYSIS



HEATMAP OF VARIABLES

Patient Age and Patient Sex do not show any significant correlation with any attribute.
'Target' and 'class' are highly correlated with 'x', 'y', 'width', and 'height'.
'x', 'y', 'width', and 'height' show high correlation with each other.
View Position show moderate correlation with 'Target' and 'Class'.

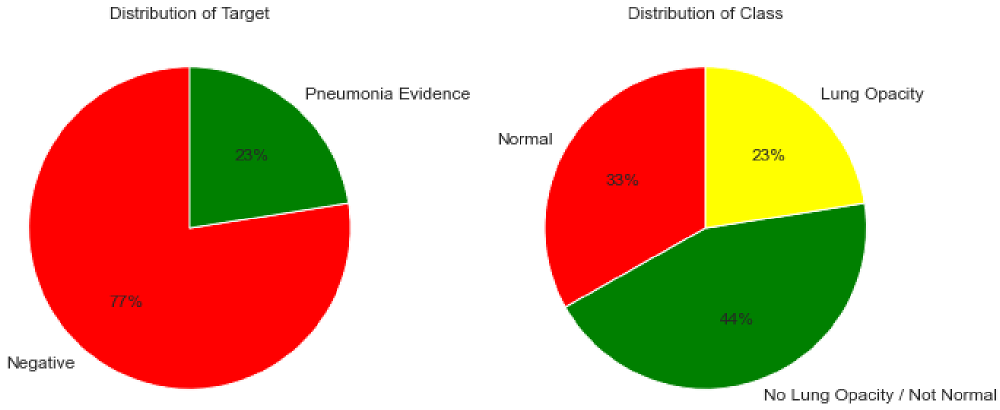




PATIENT AGE VS TARGET

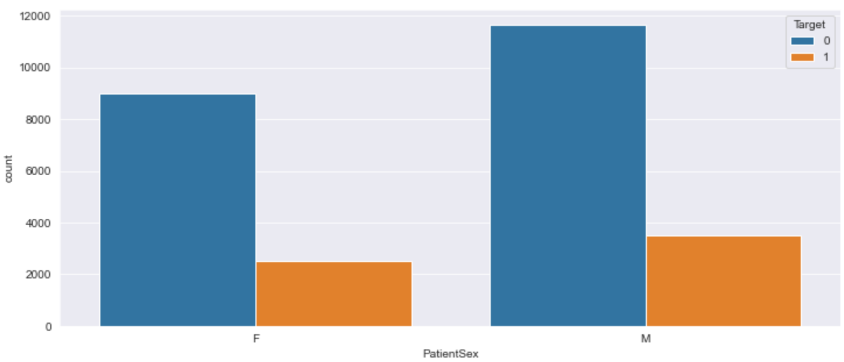
The instances of Normal patients are much higher than the Pneumonic patients in all age groups.

BIVARIATE AND MULTIVARIATE ANALYSIS



CLASS AGE VS TARGET

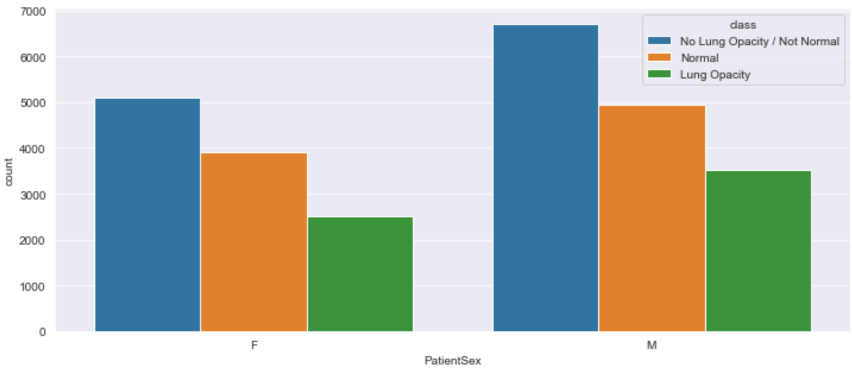
The 3 Classes are distributed along the dataset with approximately 1/3rd share of Normal, 44% for Not Normal / Not Opaque, and 22% for Opaque.
However, Target shows 77% share for Negative Patients and 23% share for Pneumonia Positive Patients.



PATIENT SEX VS TARGET

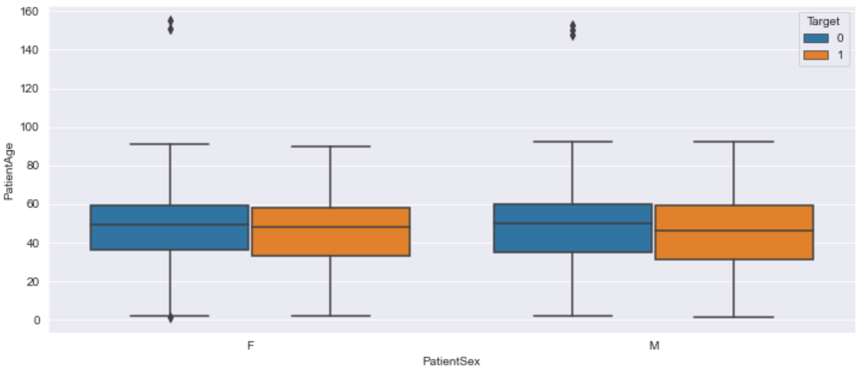
Instances with Target=0 (normal) are more than that of Target=1(Pneumonia) for both Males and Females.

BIVARIATE AND MULTIVARIATE ANALYSIS



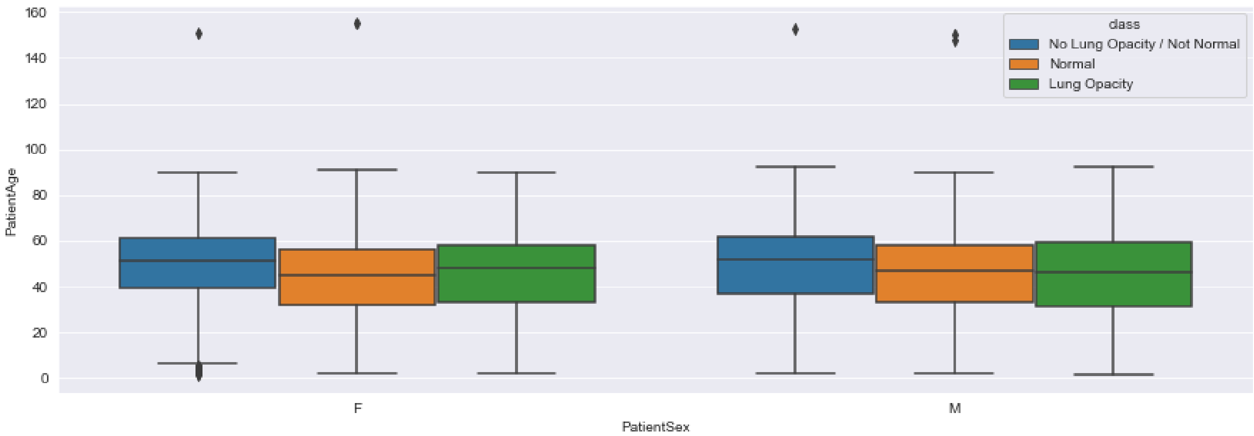
PATIENT SEX VS CLASS

The instances of No Lung Opacity / Not Normal are the highest, followed by Normal and Lung Opacity for both Males and Females.



PATIENT SEX VS AGE VS TARGET

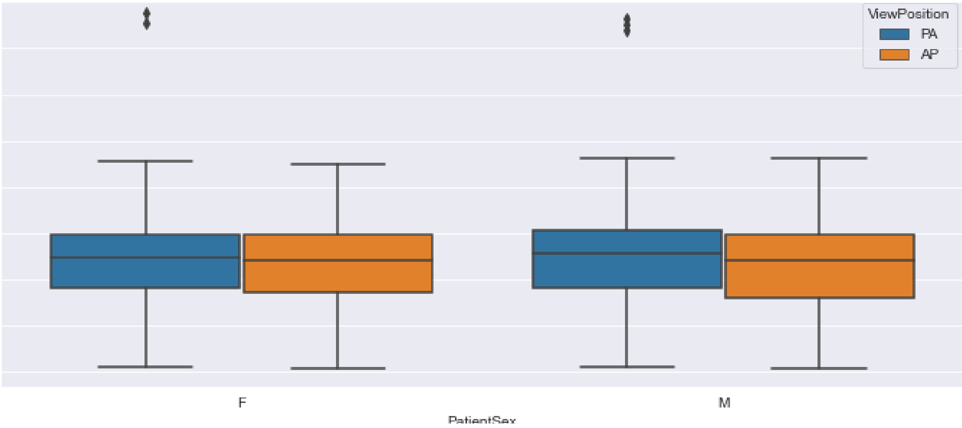
No significant relation between Target with Patient Age and Sex.



PATIENT SEX VS AGE VS CLASS

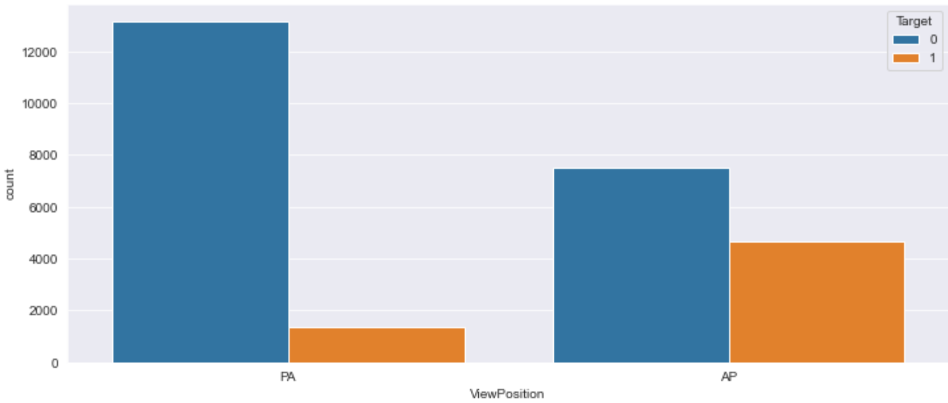
No significant relation between class with Patient Age and sex

BIVARIATE AND MULTIVARIATE ANALYSIS



PATIENT SEX VS AGE VS VIEW POSITION

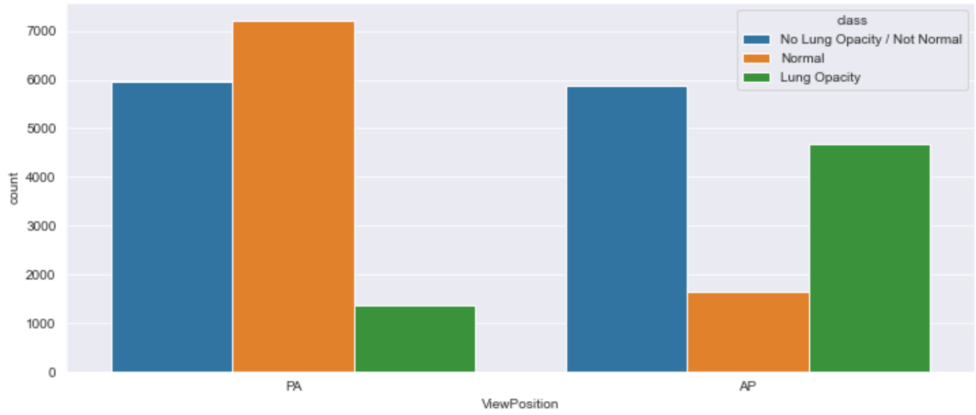
No significant relation between class with Patient Age and sex and view position



TARGET VS VIEW POSITION

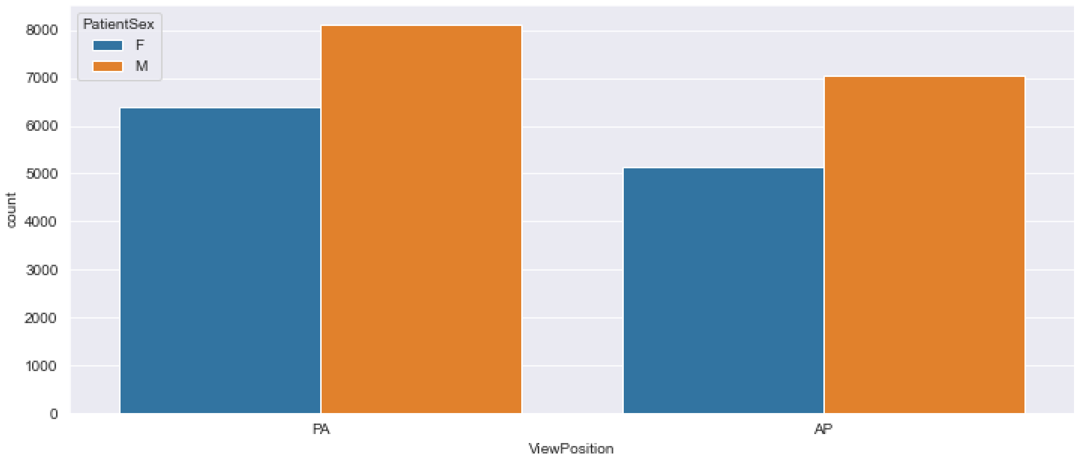
Scans taken in PA Position show more results as Negative. However, AP Position gives almost equal counts of '0' and '1' Targets.

BIVARIATE AND MULTIVARIATE ANALYSIS



CLASS VS VIEW POSITION

PA Position shows most cases as Normal or No Lung Opacity / Not Normal. However, AP Position shows more cases as No Lung Opacity / Not Normal. This indicates that as far as possible, X-Ray scans must be taken in AP Position for better results.



PATIENT SEX VS VIEW POSITION

The instances for both PA and AP Positions are show almost identical trend amongst Males and Females

NEW VARIABLE - SEVERITY SCORE

- Adding another dimension in the dataset which could help us perform the quantitative analysis on the X-Ray Image is the **severity score**. This can potentially indicate how severe the patient is infected, with the help of a parameter called the 'Severity Score'. This severity score shall be directly proportional to the intensity of lungs infection.

- The approach includes using bounding box parameters, namely 'height' and 'width' to calculate the area of each box and determining the number of bounding boxes for each patient.

- Basis domain expertise, a more severe patient shall have larger bounding box areas, meaning larger opaque areas. In addition, their numbers per X-ray image shall have a bearing on the severity.

- Instead of using each of these parameters independently, an approach to add the areas of all the bounding boxes for a unique patient id is used, to reach to a robust indicator of severity, which we shall call 'Severity Score'.

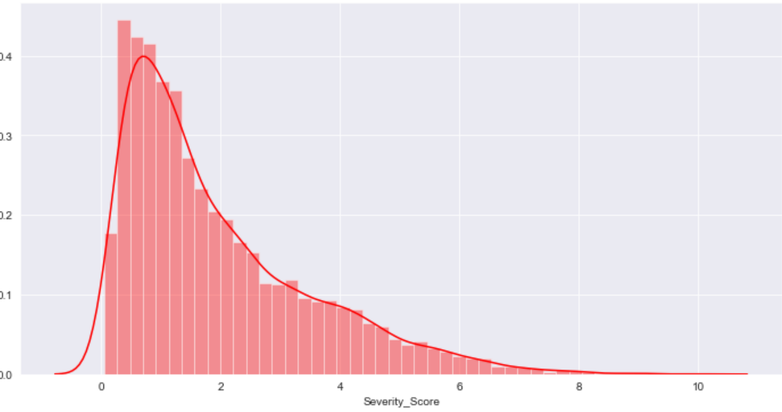
Calculating the areas of individual bounding boxes by multiplying their width with the corresponding height. A new data frame is created.

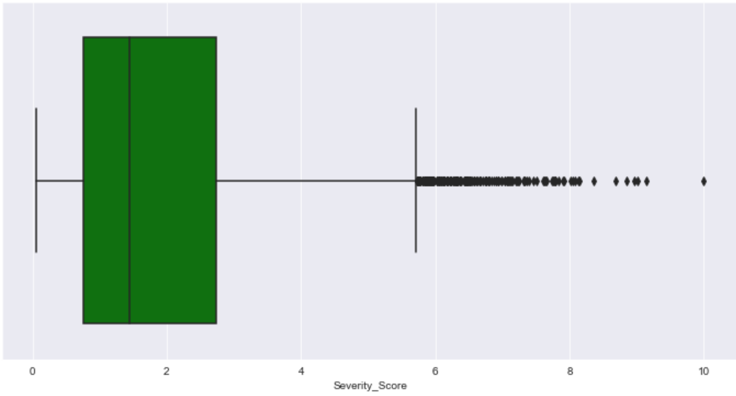
Adding the areas of all the bounding boxes corresponding to the individual patient IDs.

Calculating the Severity Score by dividing the Box Area with the maximum area so that we get the Severity Score on a scale of 0.0 to 10.0.

Merging PatientId, PatientAge, PatientSex, ViewPosition, Number_of_boxes a new dataframe is created for analysis.

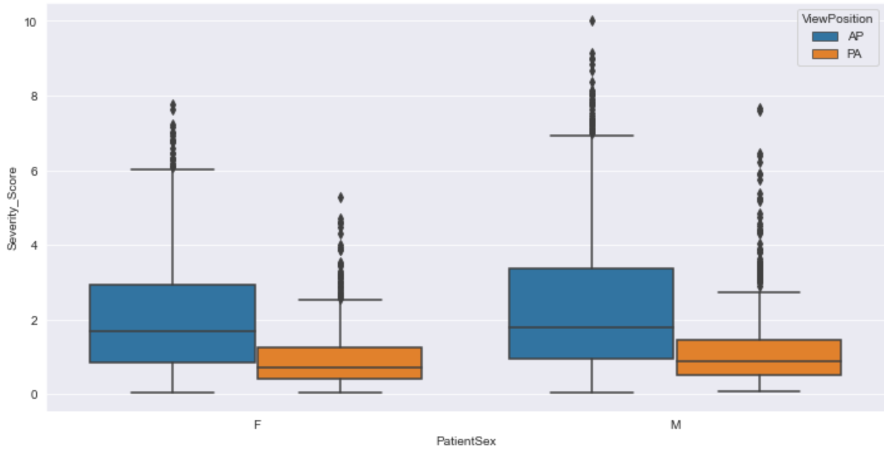
SEVERITY SCORE





SEVERITY SCORE

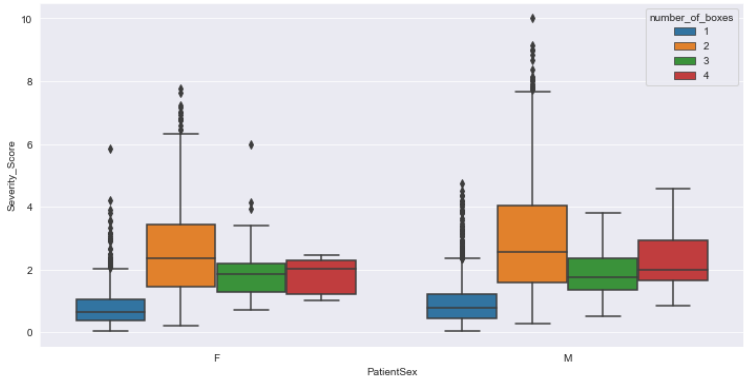
Severity Score ranges from 0.0 to 10.0
Here minimum value is 0.049 and maximum value is 10.0
75% of the total positive population have a severity score under 2.74
Only 25% population has severity score between 2.74 and 10.0 with a mean of 1.95
This indicates that the number of patients with high severity score is very less.



SEVERITY SCORE VS SEX VS VIEW POSITION

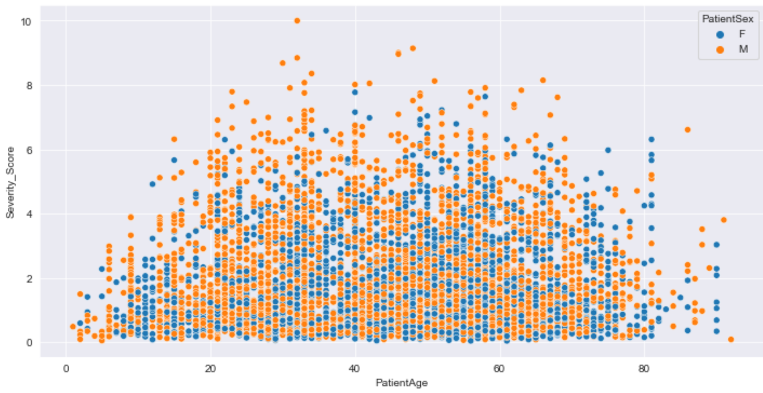
Males and females show almost similar distribution of Severity Score for lower scores.
However, high severity scores (>8) are visible only in male population.
AP position is more likely to give a larger share of high severity scores.
In fact, severity scores of more than 8 are only visible in X-rays taken with AP position.

SEVERITY SCORE



SEVERITY SCORE VS SEX VS BOUNDING BOXES

Males and females show almost similar distribution of Severity Score for lower scores.
However, high severity scores (>8) are visible only in male population.
Patients with just 1 bounding box in their scans tend to have the least Severity Score.
Contrary to assumption, high Severity Scores are visible only in patients having 2 bounding boxes and not in patients having 3 or 4 boxes.
In fact, severity scores of more than 4.0 are primarily visible in X-rays having 2 bounding boxes.



Males and females show almost similar distribution of Severity Score for lower scores across all ages.
However, moderate severity scores (6) are visible mostly in male population, and high severity scores (>8) are visible only in male population.
Lower age groups (age<20 years) and Higher age groups (age>70 years) have severity scores under 6.0
Age groups between 30 and 50 show maximum instances of high severity scores (>6).
This can be inferred that males between the age groups 30-50 are most likely to get severe illness and must be treated accordingly

Model Selection

MODEL SELECTION

The model evaluation process included the survey of VGG16, DenseNet, Inception Net Mobile Net, ResNet, Xception Net, Faster RCNN and Mask RCNN etc.

This Pneumonia Detection problem utilized the implementation of 5 models such as *Mobile Net, ResNet, Xception Net, Faster RCNN and Mask RCNN.*

The rationale behind selecting the models are because of their various advantages, lets say base models such as Mobile Net Rssnet and the Xception net have fewer parameters and it is simpler to train the model.

It is understood that ResNets model help in tackling the vanishing gradient problem using identity mapping. And networks with large number (even thousands) of layers can be trained easily without increasing the training error percentage.

MobileNets are light weight and have lower number of parameters , faster in performances with Small, low-latency, while Xception Net are linear stack of depth wise separable convolution layers with residual connections. “Xception” means “Extreme Inception”.

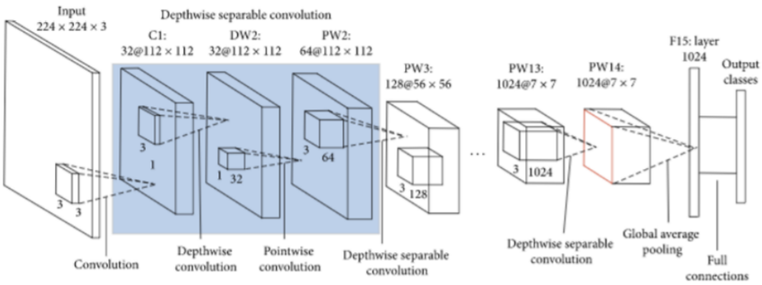
MODEL SELECTION

Advanced models such as the Faster RCNN and Mask RCNN will outperform when compared to the base models because these advanced models have better benefits.

Faster RCNN Model, consists of RPN generated region proposals. For all region proposals in the image, a fixed-length feature vector is extracted from each region using the ROI Pooling layer. The extracted feature vectors are then classified using the Fast R-CNN, the class scores of the detected objects in addition to their bounding-boxes are returned.

While in Mask RCNN, the network aimed to solve instance segmentation problem In other words, it can separate different objects in an image or a video it gives you the object bounding boxes, classes and masks. There are two stages of Mask RCNN. First, it generates proposals about the regions where there might be an object based on the input image. Second, it predicts the class of the object, refines the bounding box and generates a mask in pixel level of the object based on the first stage proposal. Both stages are connected to the backbone structure.

MOBILE NET



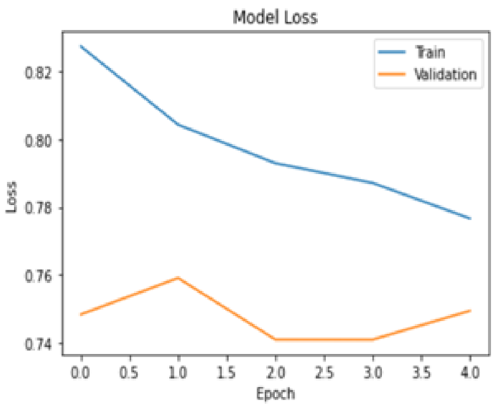
MobileNet-V2 Architecture

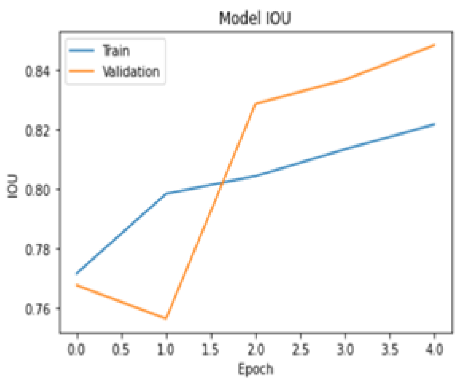
MobileNets primarily use depth wise separable convolutions in place of the standard convolutions used in earlier architectures to build lighter models.

Two new global hyperparameters(width multiplier and resolution multiplier) are used.

This model contains conv layers, activation functions with avg pooling layer in the last layer.

A fully connected layer with s with a softmax layer for classification and a regressor to predict the bounding box values are used.



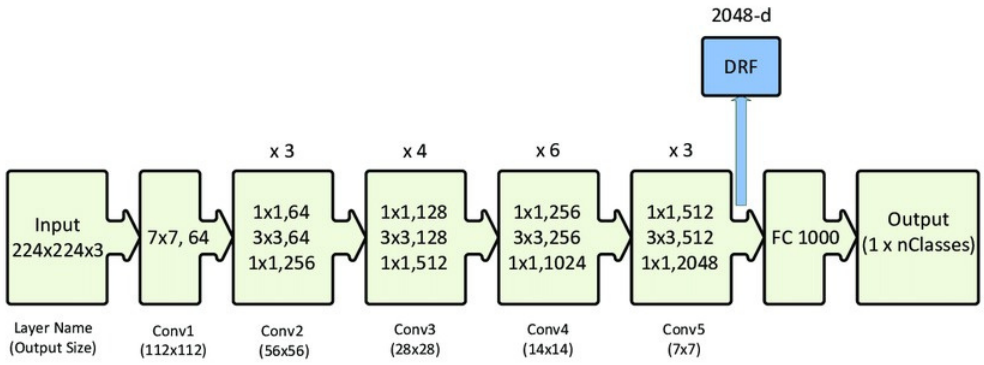


Model Accuracy - 87.89%

Train Loss - 0,7781

Validation Loss - 0.7494

RES NET

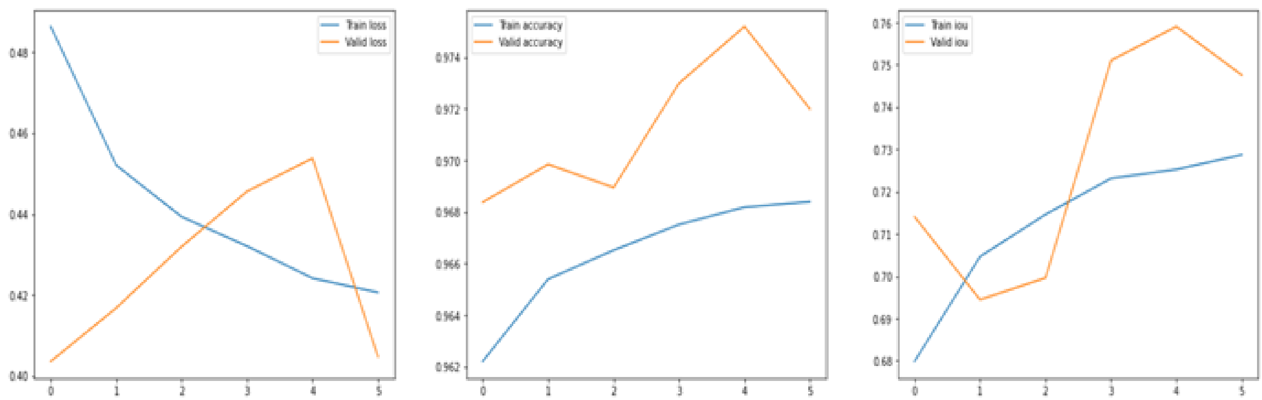


ResNet-50 architecture is shown with the residual units, the size of the filters and the outputs of each convolutional layer.

DRF extracted from the last convolutional layer of this network is also shown. Key: The notation k × k, n in the convolutional layer block denotes a filter of size k and n channels.

FC 1000 denotes the fully connected layer with 1000 neurons. The number on the top of the convolutional layer block represents the repetition of each unit. nClasses represents the number of output classes.

For this use-case Resnet as it has demonstrated very high accuracy, is fast on runtime and significantly reduces vanishing gradient problem is chosen. The same is achieved by using residual blocks with skip connections.



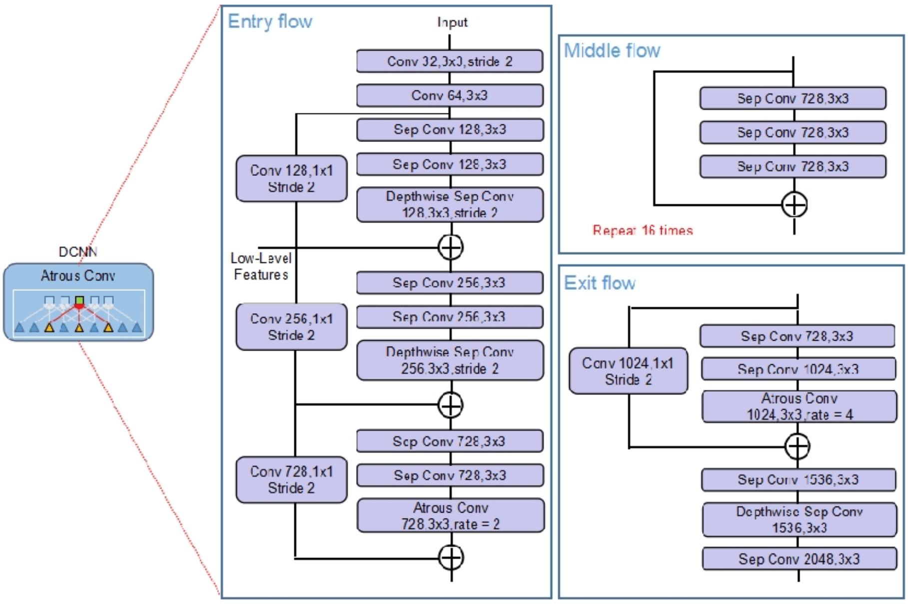
Metrics

Model Accuracy - 96%

Train Loss - 0.4207

Validation Loss - 0.4048

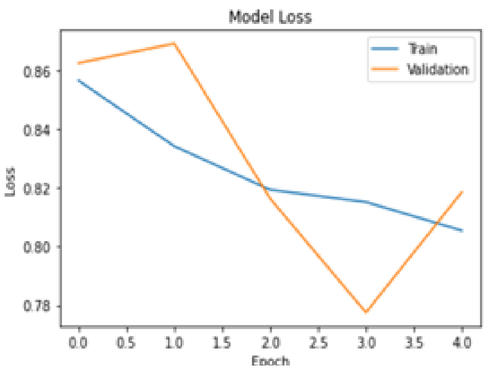
EXCEPTION NET

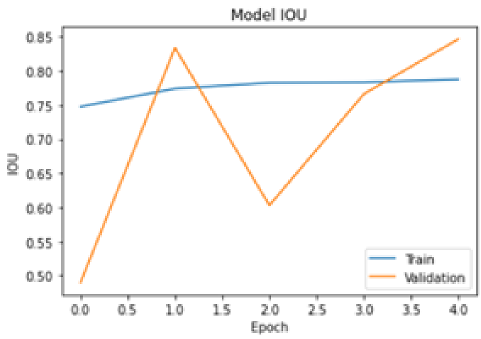


Xception is a deep convolutional neural network that introduced new inception layers. These inception layers are constructed from depthwise convolution layers, followed by a point-wise convolution layer. Xception achieved the third-best results on the ImageNet dataset

A concatenated neural network is designed by concatenating the extracted features of Xception and then connecting the concatenated features to a convolutional layer that is connected to the classifier.

This layer was added to extract a more valuable semantic feature out of the features of a spatial point between all channels, with each channel being a feature map.



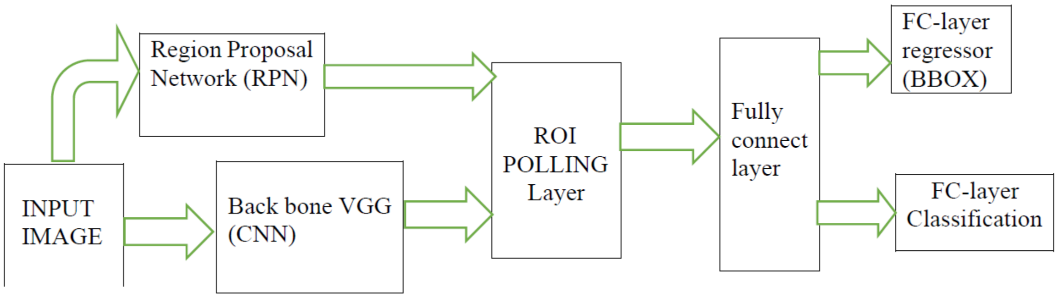


Accuracy - 87.94%

Train Loss - 0.8073

Validation Loss - 0.8186

FASTER R CNN

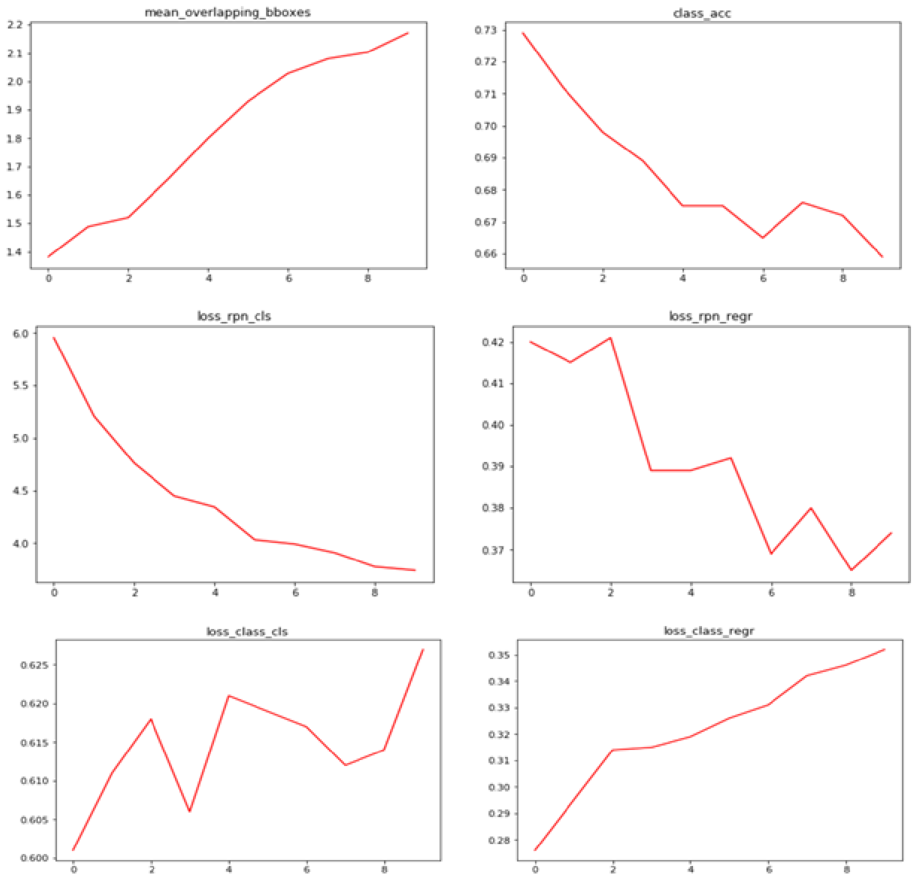


Faster R-CNN is one of the best methods for analyzing pneumonia by using all parameters efficiently. The figure above represents the architecture of faster R-CNN.

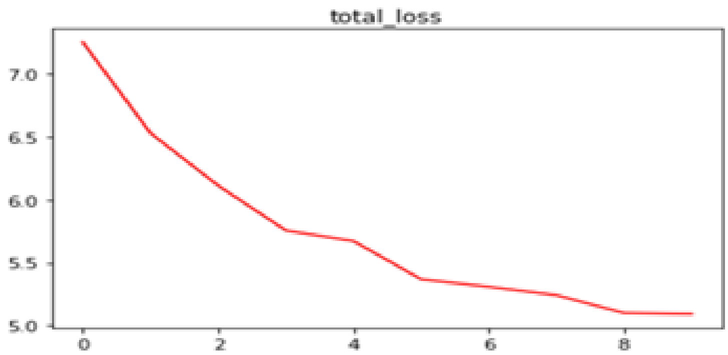
VGG-16 is used as a base model or backbone model and it consists of conv layers, activation function as Relu and max pooling layers.

These layers produce a feature by interpreting from the input x-ray image and RPN suggests proposals of the region of interest.

Pooling is applied to make all proposals a fixed size.



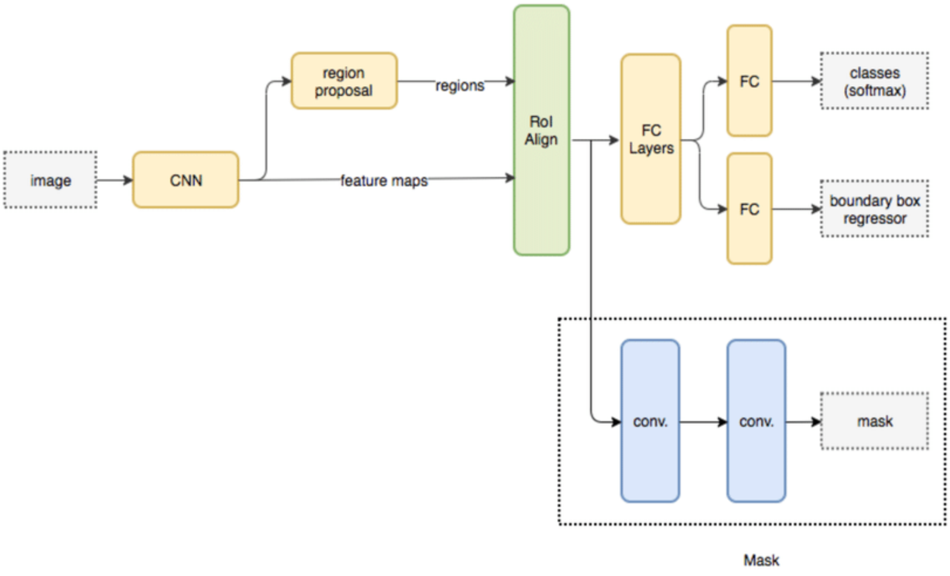
FASTER R CNN



RPN Class Loss - 3.73882

RPN Bounding Box Loss - 2.1710

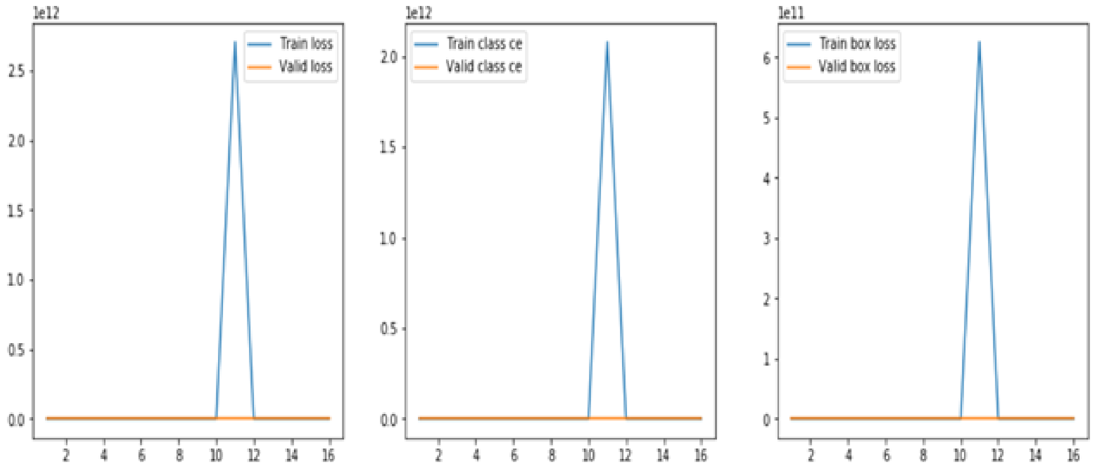
MASK R CNN



Mask-RCNN detector configuration parameters have been selected to reduce training time and that is capable of achieving state-of-the-art results on a range of object detection tasks.

Image Augmentation. Try finetuning some variables to custom values.

epoch, are limited to 5 and have set nominal values for the Detector Configuration to reduce run-time.



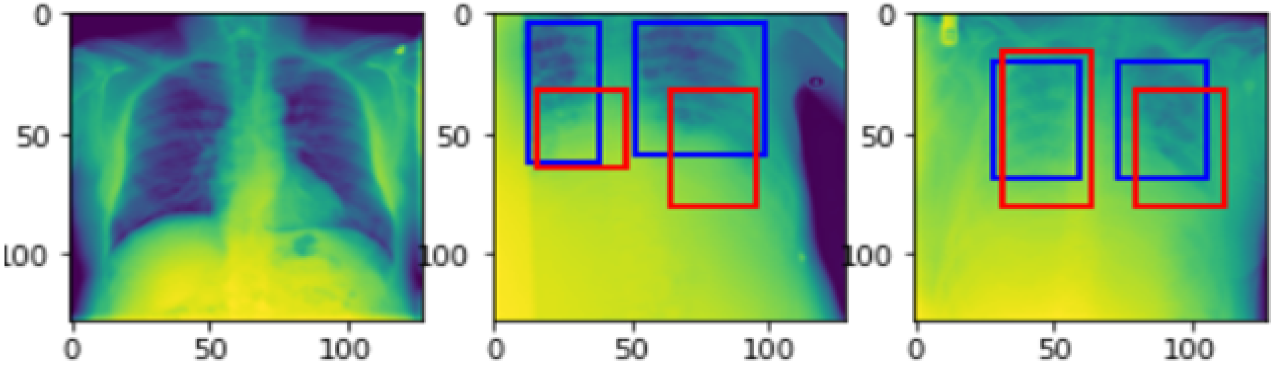
RPN Class Loss - .009287

RPN Bounding Box Loss - .336645

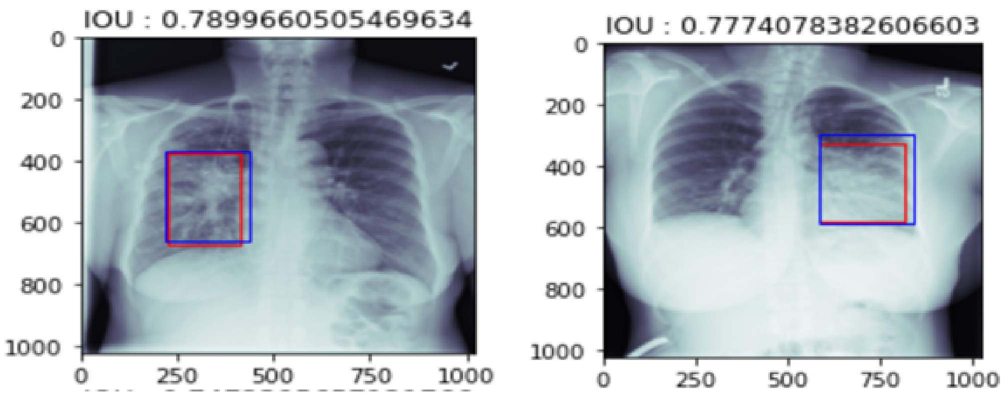
Model Output

MODEL OUTPUT

RES-NET

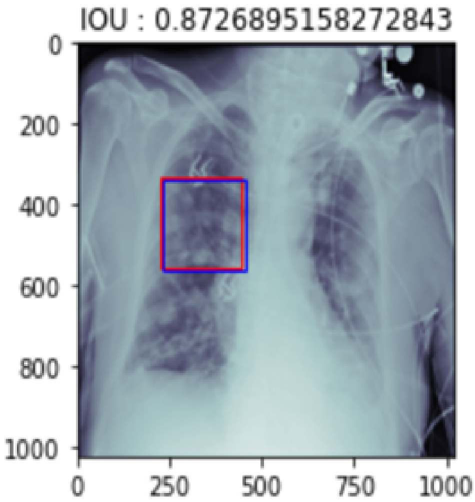


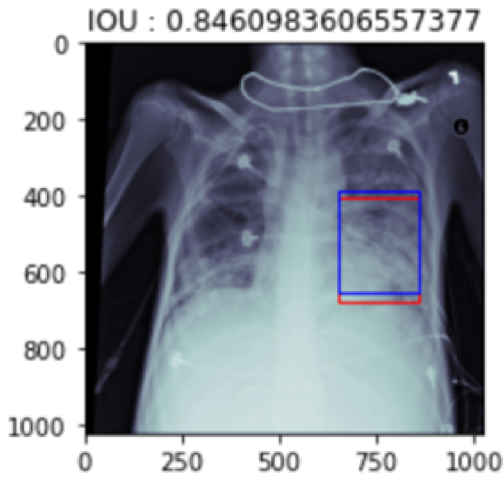
MOBILE NET



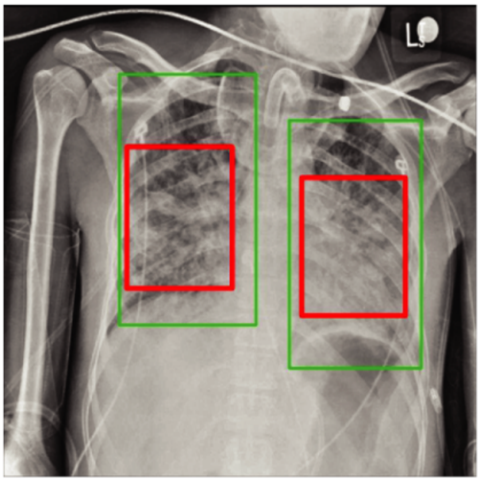
MODEL OUTPUT

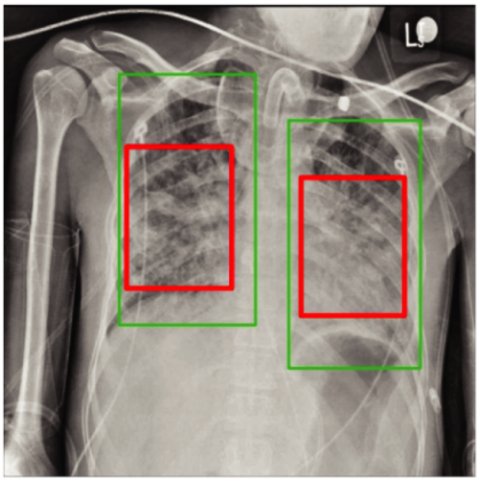
XCEPTION NET





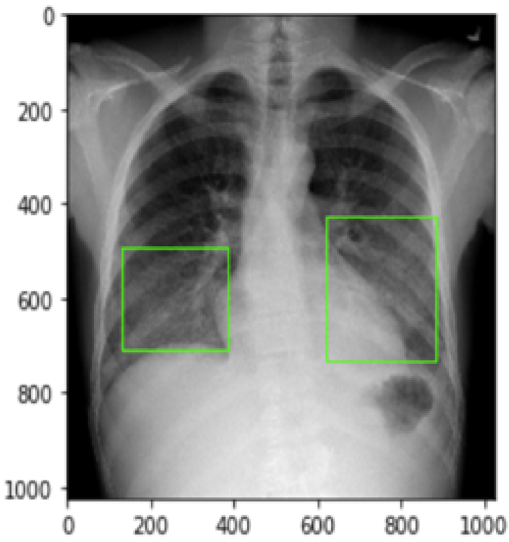
FASTER R CNN

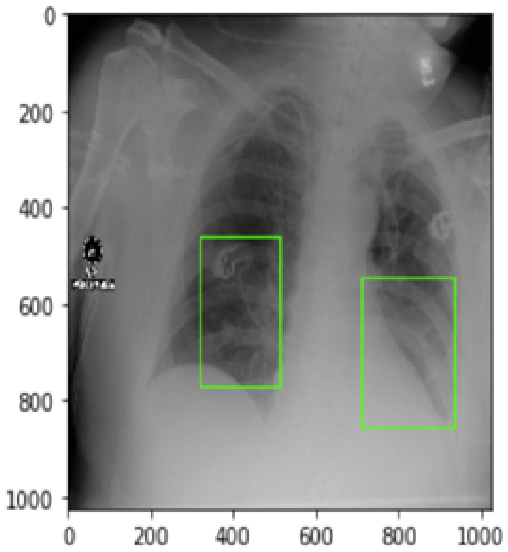




MODEL OUTPUT

MASK R CNN





Model Performance Comparison

MODEL COMPARISON

ADVANCED MODELS

MODELS	Train_loss	Val_loss	Mean_iou	Val Mean iou	Model Accuracy
ResNet	**0.4207**	**0.4048**	**0.7287**	**0.7476**	**96.84%**
MobileNet	**0,7781**	**0.7494**	**0.8193**	**0.8482**	**87.89%**
XceptionNet	**0.8073**	**0.8186**	**0.7856**	**0.8467**	**87.94%**

BASE MODELS

MODELS	**rpn_class_loss**	**rpn_bbox_loss**
Faster RCNN	**3.73882**	**2.1710**
Mask RCNN	**0.009287**	**0.336645**

Model Inferences

MODEL INFERENCES- BASE MODELS

- We find how the various attributes (obtained from both the files and images) are spread across the entire dataset.
- Input image Sizes is 224X224x1 (Mobile net, Xception Net) and 128x128x1 (Res net)
- Models are selected which are relevant to our work based on the data complexity. We have selected Res-net mobile net and the Xception net
- The 3 base model Resnet mobile net and Xception net model are trained for 5 epochs by considering Adam Optimizers and the corresponding model accuracy and the model losses have been captured.
- These models are performing well and it is able to detect and classify the dcm images, whether the dcm images consists of Pneumonia or non Pneumonia.
- Model performance can be improved by training to more number of epoch and also tuning the model hyperparameters.
- Post training, model weights are saved in checkout and finally we load model weights and test the model with test sets images and save the csv files which consist of anchor boxes values of the images which are Pneumonia patients records.

MODEL INFERENCES- ADVANCED MODELS

- In the Advanced models FRCNN and Mask RCNN Model are implemented with the limited data set which ranges from 1000-6012 images and also model is trained for 10 epochs with Adam optimizers.
- Input image Size is 256X256x3 png formate (Faster-RCNN) and 256x256x1 dicom images (Mask RCNN)
- The corresponding losses are obtained for each epoch such as Mean number of bounding boxes from RPN overlapping ground truth boxes, Loss RPN classifier, Loss RPN regression, Loss Detector classifier, Loss Detector regression, val_rpn_class_loss, val_rpn_bbox_loss, val_mrcnn_class_loss, val_mrcnn_bbox_loss, val_mrcnn_mask_loss, rpn_class_loss, rpn_bbox_loss, mrcnn_class_loss, mrcnn_bbox_loss, mrcnn_mask_loss.

- The challenges we faced are: Higher training time for each epoch. While, the model is trained for 10 epoch, model is performing well but we still can expect better model loss and better performance by training it for more number of epochs and also tuning hyper parameters.

Industry Benchmark and Future Prospects

INDUSTRY BENCHMARK AND FUTURE PROSPECTS

As industry is moving towards solving more complex problems using deep learning, the focus is shifting from conventional CNN architectures to more efficient models which can achieve both higher accuracy and better efficiency over existing CNNs, reducing parameter size and FLOPS by an order of magnitude. Generally, the models are made too wide, deep, or with a very high resolution. Increasing these characteristics helps the model initially but it quickly saturates and the model made just has more parameters and is therefore not efficient.

To overcome these shortcomings, the need is to shift to something more efficient.

The buzz word is 'EfficientNets' which the industry is shifting to as a backbone.

EfficientNet is a convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth/width/resolution using a simple yet highly effective compound coefficient.

INDUSTRY BENCHMARK AND FUTURE PROSPECTS

Unlike conventional practice that arbitrary scales these factors, the EfficientNet scaling method uniformly scales network width, depth, and resolution with a set of fixed scaling coefficients. The compound scaling method is justified by the intuition that if the input image is bigger, then the network needs more layers to increase the receptive field and more channels to capture more fine-grained patterns on the bigger image.

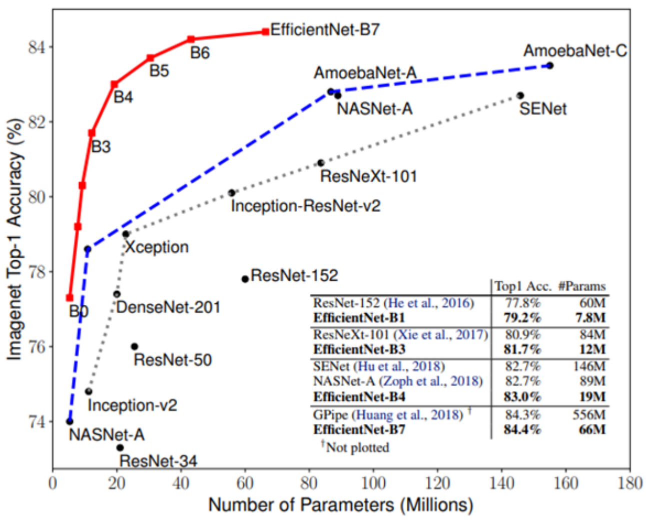
The concept of this new family of neural nets was introduced in May 2019 by two engineers from Google brain team named Mingxing Tan and Quoc V. Le published a paper called “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”.

The baseline model of this family is called B0 model and scaling this, the authors developed a full family of EfficientNets from B1 to B7 which achieved state of the art accuracy (84.4% top-1 accuracy) on ImageNet while being very efficient to its competitors as shown in the following table.

INDUSTRY BENCHMARK AND FUTURE PROSPECTS

Model	Top-1 Acc.	Top-5 Acc.	#Params	Ratio-to-EfficientNet	#FLOPS	Ratio-to-EfficientNet
EfficientNet-B0	**77.3%**	**93.5%**	**5.3M**	**1x**	**0.39B**	**1x**
ResNet-50 (He et al., 2016)	76.0%	93.0%	26M	4.9x	4.1B	11x
DenseNet-169 (Huang et al., 2017)	76.2%	93.2%	14M	2.6x	3.5B	8.9x
EfficientNet-B1	**79.2%**	**94.5%**	**7.8M**	**1x**	**0.70B**	**1x**
ResNet-152 (He et al., 2016)	77.8%	93.8%	60M	7.6x	11B	16x
DenseNet-264 (Huang et al., 2017)	77.9%	93.9%	34M	4.3x	6.0B	8.6x
Inception-v3 (Szegedy et al., 2016)	78.8%	94.4%	24M	3.0x	5.7B	8.1x
Xception (Chollet, 2017)	79.0%	94.5%	23M	3.0x	8.4B	12x
EfficientNet-B2	**80.3%**	**95.0%**	**9.2M**	**1x**	**1.0B**	**1x**
Inception-v4 (Szegedy et al., 2017)	80.0%	95.0%	48M	5.2x	13B	13x
Inception-resnet-v2 (Szegedy et al., 2017)	80.1%	95.1%	56M	6.1x	13B	13x
EfficientNet-B3	**81.7%**	**95.6%**	**12M**	**1x**	**1.8B**	**1x**
ResNeXt-101 (Xie et al., 2017)	80.9%	95.6%	84M	7.0x	32B	18x
PolyNet (Zhang et al., 2017)	81.3%	95.8%	92M	7.7x	35B	19x
EfficientNet-B4	**83.0%**	**96.3%**	**19M**	**1x**	**4.2B**	**1x**
SENet (Hu et al., 2018)	82.7%	96.2%	146M	7.7x	42B	10x
NASNet-A (Zoph et al., 2018)	82.7%	96.2%	89M	4.7x	24B	5.7x
AmoebaNet-A (Real et al., 2019)	82.8%	96.1%	87M	4.6x	23B	5.5x
PNASNet (Liu et al., 2018)	82.9%	96.2%	86M	4.5x	23B	6.0x
EfficientNet-B5	**83.7%**	**96.7%**	**30M**	**1x**	**9.9B**	**1x**
AmoebaNet-C (Cubuk et al., 2019)	83.5%	96.5%	155M	5.2x	41B	4.1x
EfficientNet-B6	**84.2%**	**96.8%**	**43M**	**1x**	**19B**	**1x**
EfficientNet-B7	**84.4%**	**97.1%**	**66M**	**1x**	**37B**	**1x**
GPipe (Huang et al., 2018)	84.3%	97.0%	557M	8.4x	-	-

It is clearly visible that EfficientNet Family shows exceptional accuracy with much lesser number of parameters as compared to popular model architectures like Inception, Xception, ResNet, and AmoebaNet. This can also be depicted graphically as follows:



INDUSTRY BENCHMARK AND FUTURE PROSPECTS

With considerably fewer numbers of parameters, the EfficientNet family of models are efficient and also provide better results.

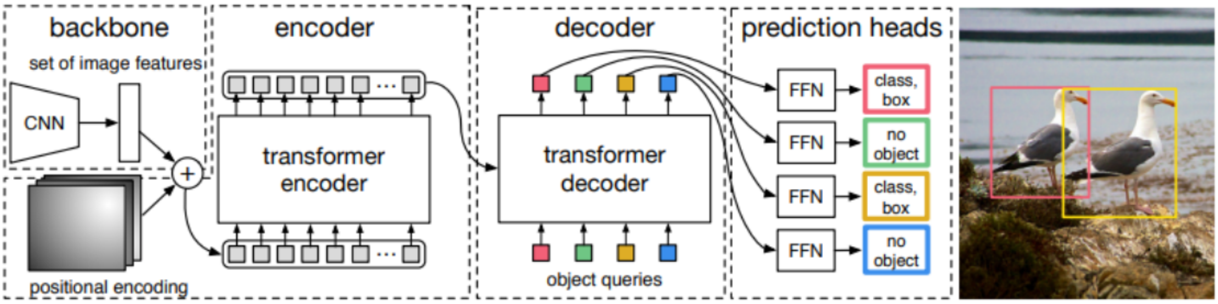
However, EfficientNet can take smaller images as input also but it will be overkill for a dataset like MNIST. EfficientNets are advisable to use for complex datasets.

As future work, we may build the Pneumonia Classification exercise on top of EfficientNet family and we expect considerable improvement in model performance while drastically reducing the computational expense.

In addition to exploring EfficientNet, the problem of object detection at scale is made more feasable through DETR.

The current deep learning algorithms perform object detection in a multi-step manner. They also suffer from the problem of near-duplicates, i.e., false positives. To simplify, the researchers at Facebook AI has come up with DETR, an innovative and efficient approach to solve the object detection problem.

INDUSTRY BENCHMARK AND FUTURE PROSPECTS



DETR takes a slightly different approach to object detection training by using transformers to replace complicated, hand-designed components that are commonplace in object detection pipelines.

The backbone – Features extracted from a Convolutional Neural Network and a positional encoding are passed

The transformer Encoder – A transformer is naturally a sequence processing unit and for the same reason, we the incoming tensors are flattened. It transforms the sequence into an equally long sequence of features.

The Transformer Decoder – takes in Object queries So its a decoder as a side input for conditioning information.

Prediction Feed-Forward Network (FFN) – The output for this is going through a classifier which outputs the class labels and bounding box output discussed earlier.

The current Pneunomia detection challenge can see the use of DETR for better performance and accuracy.

THE END