****

**CAPSTONE PROJECT**

**Detection and Localization of Pneumonia**

**in Chest Radiographs**

**Team Members (in alphabetical order):**

Shubham Gajbhiye

Sajag Goel

Ramya Nagarajan

**Table of Contents**

1. Abstract
2. Introduction
3. Problem Statement and Dataset
4. Exploratory Data Analysis

4.1 Basic analysis of csv files

4.2 Distribution based on Classes and Target

4.3 Analysis of DICOM data

4.4 Distribution of Lung Opacities

4.5 Data Pre-processing

5. Proposed Algorithms for Object Detection

5.1 Model Selection

5.2 Mask RCNN

5.2.1 Model Overview

5.2.2 Execution

5.2.3 Results

5.2.4 Predictions

5.3 MobileNet-UNet

5.3.1 Model Overview

5.3.2 Execution

5.3.3 Results

5.3.4 Predictions

5.4 CheXNet-UNet

5.4.1 Model Overview

5.4.2 Execution

5.4.3 Results

5.4.4 Predictions

6. Comparative Results

7. Implications

8. Limitations and Improvements

9. Learning's

10. References

**1. Abstract**

The goal of this project was to build a suitable model for automated detection and localization of pneumonia on chest X-ray images. The training dataset comprised of 26684 chest radiographs in DICOM format which was part of the Kaggle Pneumonia Detection Challenge1. EDA and pre-processing involved extracting patient information from the DICOM files and consolidating relevant data into a JSON format for model building and analysis. Three object detection models; Mask RCNN, MobileNet-UNet and CheXNet-UNet have been explored. With a recall of 0.845, the CheXNet-UNet architecture seems to work best for the given dataset.

**2. Introduction**

Pneumonia is an infection of the lungs caused by pathogens like bacteria, viruses and fungi present in the air we breathe. Under normal circumstances, the lungs are filled with air, however the infection causes inflammation in the air sacs which get filled with fluid or pus causing difficulty in breathing. Infants under the age of two and the older population with impaired immune system are the most vulnerable. In a report published by the World Health Organization2, nearly 150 million people are infected by pneumonia each year out of which 11-20 million are severe enough to require hospitalization. In addition, it is also a leading cause of death in children under the age of 5 with an estimated 1.4 million deaths per year.

Chest X-rays are the most common and affordable diagnostic tools for pneumonia detection. However, interpretation of X-rays can be quite challenging as it involves evaluating a 2D image depicting complex 3D organs. It is quite likely that one might fail to diagnose an early onset of pneumonia or other conditions of the lungs. Therefore it would be beneficial to have tools that can help experts quickly and accurately diagnose the disease and get the patient started on their treatment in a timely manner.

Artificial Intelligence has made a steady progress in the field of radiology where CNN based deep learning algorithms have become the standard choice for medical image classification3. In this project we have evaluated and compared the results from three CNN based object detection models. Section 3 provides a brief description of the data, section 4 summaries the results from EDA, section 5 explores the performance of various models, section 6 provides a comparative summary, limitations and possible improvements have been discussed in section 7 and the key learning's from the project are present in section 8 of the report.

**3. Problem Statement and Dataset**

The emphasis of this project is on identifying the presence of pneumonia and localizing the inflammation on chest X-rays.

The dataset comprises of 26684 train and 3000 test images in the DICOM format. The features of the train images are presented in two csv files. In the first file, the image corresponding to a patient id is grouped based on one of the three classes: 'Lung Opacity', 'Normal', 'Not-Normal/No-Lung opacity' as shown below

A screenshot of a cell phone

Description automatically generated

The other involves grouping based on 'Target'. The 'Lung Opacity' class is labeled as '1' whereas the 'Normal' and 'Not-Normal/No-Lung opacity' classes are all grouped as Target '0'. In addition, bounding box coordinates (x, y, width and height) have been provided for Target '1'.

A picture containing photo, sitting, table

Description automatically generated

**4. Exploratory Data Analysis**

**4.1 Basic analysis of csv files**

Preliminary analysis of the two csv files indicates that there are 26684 unique patient ids in the training data. There are no missing values in the csv file where the patient ids are grouped with respect to the 3 Classes. However, nearly 68% of the data appears to be missing from the other file where the patient ids are grouped with respect to the target; 1 and 0. This suggests that only 32% of the train images correspond to target = 1 and have bounding box information. Displayed below are some of the X-ray images grouped by Classes.

A picture containing photo, monitor, different, screen

Description automatically generated

**4.2 Distribution based on Classes and Target**

As mentioned in Section 3, the images are grouped into 3 classes: 'Lung Opacity', 'Normal', 'Not-Normal/No-Lung opacity'. The distribution across 'Classes' appears to be nearly uniform as shown in the bar plot below.

The 3 Classes are mapped into 2 Target values: Target 0 is mapped to 'Normal' & 'No Lung Opacity/Not Normal' classes whereas; Target 1 is mapped to the 'Lung Opacity' class. Nearly, 68% corresponds to Target 0 and the remaining 32% corresponds to Target 1.

Class DistributionA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

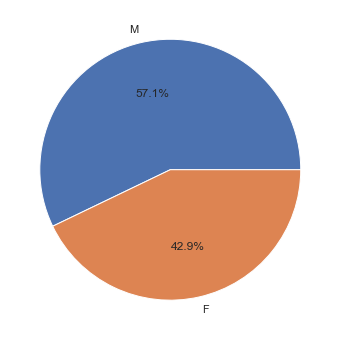
Description automatically generated Target Distribution

**4.3 Analysis of DICOM Data**

The X-ray images are presented in the standard DICOM format used in medical imaging. Relevant patient information (gender and age) and image details (resolution and pixel intensities) have been extracted for subsequent analysis. All images have a uniform size of 1024x1024. The image files are read using the pydicom package and displayed by calling the pixel intensity values.

Analysis of the Dicom data indicates that the average age of the population is 46.5 years. In addition, pneumonia is more prevalent in the male population with a distribution of 57% male and 43% female; however the distribution remains consistent across the 3 classes as shown below.

A screenshot of a cell phone

Description automatically generated

The chest X-rays have been recorded in 2 view positions: PA (posterior-anterior) and AP (anterior posterior) with a distribution of 58% (PA) and 42% (AP). In general, for chest radiographs PA images are known to have a higher overall quality and hence preferred over AP view position.

**4.4 Distribution of Lung Opacities**

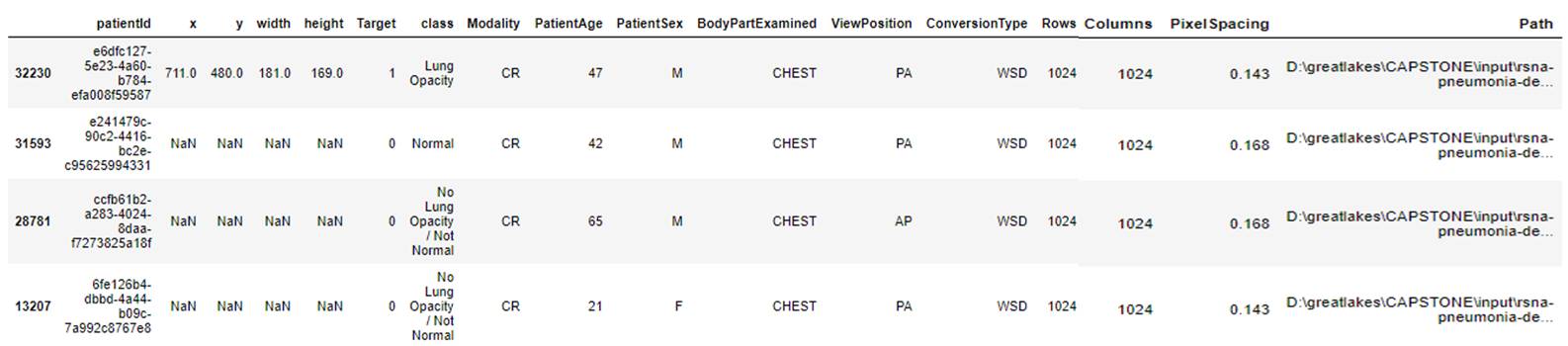
In order to analyze the localization regions of lung opacities, nearly 2000 images corresponding to Target = 1 i.e. the 'Lung Opacity' class were sampled and the centers of the bounding box coordinates were plotted as displayed below. The distribution indicates that for people infected with pneumonia, the inflammation is localized more in the center of both the lungs.

A picture containing table, large, water

Description automatically generated

**4.5 Data pre-processing**

As a final step, the data from the two csv files, Dicom images and the image path were consolidated into a JSON format to be used later for building the models. The JSON file features are shown below:



**5. Proposed Algorithms for Object Detection**

**5.1 Model Selection**

For the given dataset, the training data has 6012 images (from a total of 22684) that have been classified as 'pneumonia' and labeled, '1'. The rest are 'non-pneumonia' images which have been labeled as 0. For labels=1, the location of lung opacities have been provided via bounding box coordinates. Based on our EDA, for pneumonia containing images, the number of bounding boxes range from 1 up to a maximum of 3, indicating that for a single image there could be multiple locations associated with lung opacities. The task therefore is to not only identifying the presence (or absence) of pneumonia but also to localize multiple opacities (objects) in images; hence it would be appropriate to implement suitable Object Classification algorithms.

Accordingly, we have implemented three approaches:

1. Mask RCNN

2. MobileNet-UNet

3. CheXNet-Unet

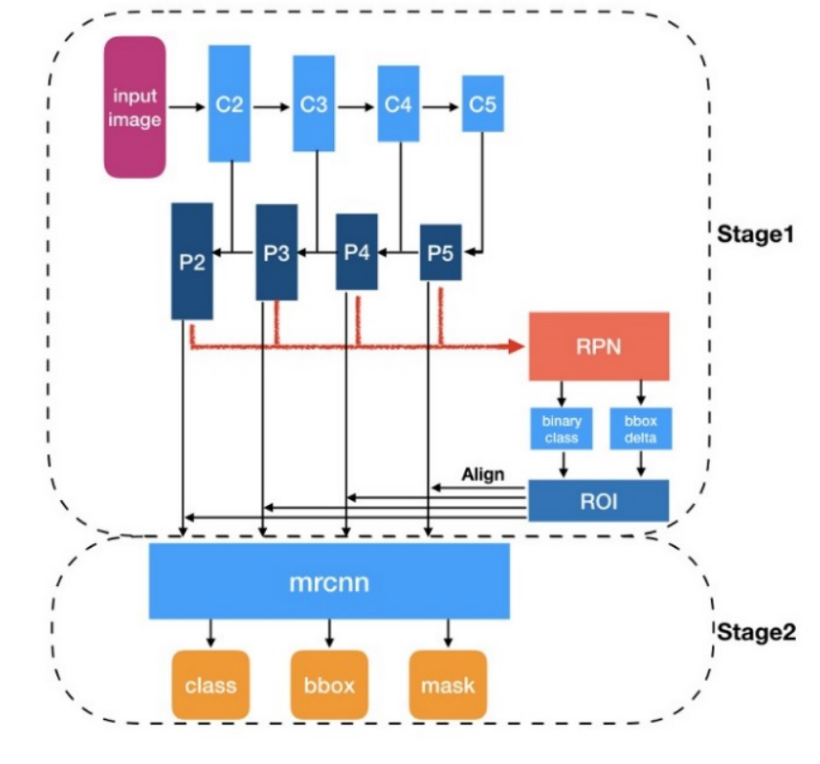
**5.2 Mask RCNN**

5.2.1 Model Overview

Mask RCNN is a deep neural network aimed to solve instance segmentation problem in machine learning or computer vision. In other words, it can separate different objects in a image or a video. You give it a image, 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.

Backbone is a FPN style deep neural network. It consists of a bottom-up pathway, a top-bottom pathway and lateral connections. Bottom-up pathway - ResNet50, which extracts features from raw images. Top-bottom pathway generates feature pyramid map which is similar in size to bottom-up pathway. Lateral connections are convolution and adding operations between two corresponding levels of the two pathways. FPN outperforms other single ConvNets mainly because it maintains strong semantically features at various resolution scales.

Though Mask RCNN Is evolved from Faster RCNN, it is better because in the first stage it uses ROI Align instead of ROI Pooling thus reducing loss and misalignment as well as there is a second stage of semantic segmentation



5.2.2 Execution

The train dataset with a total of 30,227 images was split into train/validation/test set in a 60/20/20 ratio. In this model we did not compile the data like other models where we reduced the training data to unique rows of 26,694 images so that we can feed more data. Each row represents coordinates of only 1 bounding box in case target is 1. Model was trained on the machine having 4 GB of Nvidia 960M graphic card with 9th Generation i7 processor and 16GB ddr4 ram

The original image resolution was 1024x1024. As a first step, was to understand what image size can be processed on the given GPU and it was identified that images of size 256x256 are able to process through different learning rate variations and number of epochs. Also when we tries running the size lower than 256x256 it started resulting in higher val loss that’s why those sizes were ignored to be utilized. At a time 2 images were processed

A close up of a map

Description automatically generated

Image on the left shows the different learning rates across different epochs and the image on the right shows the values of val loss and train loss across different epochs. As can be seen from the image best val loss value is 1.133 across the epoch 32

5.2.3 Results

While doing the inference of the images, we considered 1 image to be executed at a time and instead of considering epoch 35 to be the starting point, we considered epoch 32 (the best) to be the start point of the inferencing.

Every image predicted from this model generated a prediction accuracy called confidence and we can define the minimum accuracy/confidence to predict the masks correctly on the image. The team considered all the different values of this confidence from 0.950 to 0.995 with a step-up value of 0.005 and following are the results of predictions across the given test data of 6046 images (20% of the original train data)

A screenshot of a computer

Description automatically generated

As we need to predict pneumonia correctly and not lead to removal of any image predicted falsely, we need to lower the number of False Negatives and at the same take care of the fact that false positives are not high as well. This leads us to use the min confidence value of 0.97 which has highest F Score of 0.643

5.2.4. Predictions

The predicted masks at a minimum confidence value of 0.97 are shown below for some of the test images corresponding to label =1 (pneumonia) and label = 0 (non-pneumonia). From left to right, the first image shows the ground truth (masks) superimposed on the X-ray image while the predicted mask is displayed with the bounding box on the right.

1. Patient having pneumonia predicted correctly

A picture containing shirt

Description automatically generated

1. Patient not having pneumonia predicted correctly

A picture containing photo, mirror, wearing, refrigerator

Description automatically generated

1. Patient not having pneumonia predicted incorrectly

A picture containing photo, shirt, person, refrigerator

Description automatically generated

**5.3 MobileNet-UNet**

5.3.1 Model Overview

The model framework is comprised of a MobileNet based network without the final fully connected layer for feature extraction and a UNet decoding methodology for image segmentation. This encoder-decoder system was chosen due to (i) the simplicity and lower computational requirements associated with MobileNets and (ii) the wide applicability of UNet architecture in medical imaging.

Down-sampling is achieved by the MobileNet encoder through a series of depth wise separable convolutions which requires only one-eighth of the computational resources when compared to standard convolutions4. The encoded output is fed to the UNet decoder which up-samples the features through transposed convolutions followed by concatenation with the corresponding feature maps in the encoder.

5.3.2 Execution

The train dataset with a total of 26694 images was split into train/validation/test set in a 60/20/20 ratio. The original image resolution was 1024x1024. As a first step, the model performance was recorded across different input image sizes (224, 192, 160, and 128). Each input size was also tuned to its maximum batch size. As shown below, as the input size decreases so does the validation loss, in addition there is a corresponding increase in the dice score. Thus, an input image size of 128x128 was chosen for further tuning.

|  |  |  |  |
| --- | --- | --- | --- |
| **Input Size** | **Max Batch Size** | **Val\_loss** | **Val\_dice\_score** |
| 224x224 | 8 | 3.488 | 0.3064 |
| 192x192 | 12 | 1.618 | 0.4628 |
| 160x160 | 16 | 0.9901 | 0.5053 |
| 128x128 | 24 | 0.8469 | 0.5315 |

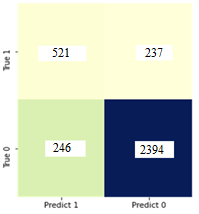
The next step involved monitoring the model performance for chosen input size and for different values (1.0, 0.75. 0.50, 0.25) of the width multiplier, alpha. The trend is shown below, where the lowest validation loss and the highest dice score was obtained for alpha=1.0

The final model was trained on an input image size of 128x128, with alpha=1.0, learning rate of 0.001 which was reduced by a factor of 1 at regular intervals. The validation loss was evaluated by compiling the model on the best saved weights. The model parameters and the learning curve are shown below.

|  |  |
| --- | --- |
| **Input size** | **128x128** |
| Batch size | 24 |
| Epochs | 30 |
| val\_loss | 0.812 |
| dice\_score | 0.558 |

5.3.4 Results

The test set with nearly 3400 images was evaluated by varying the prediction threshold from 0.5 to 0.95 and recording the corresponding performance metrics like Accuracy, Precision, Recall, F-score and Specificity. Since the project outcome is sensitive to False Negatives (i.e. the ability to correctly predict pneumonia containing images), we chose Recall as the deciding metric. Optimal results were obtained at a prediction threshold of 0.5 with a Recall of 0.742. The corresponding confusion matrix and the performance metrics are given below.



|  |  |
| --- | --- |
| **True Positives** | **521** |
| True Negatives | 2394 |
| False Positives | 246 |
| False Negatives | 237 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Images Sampled** | **Prediction**  **Threshold** | **Accuracy** | **Precision** | **Recall** | **F-1 score** | **Specificity** |
| 3398 | 0.5 | 0.857 | 0.679 | 0.687 | 0.683 | 0.906 |

5.3.5 Predictions

The predicted masks at a minimum threshold of 0.5 are shown below for some of the test images corresponding to label =1 (pneumonia) and label = 0 (non-pneumonia).

For label=1, from left to right, the first image shows the ground truth (masks) superimposed on the X-ray image while the predicted mask is displayed with the bounding box on the extreme right. Similarly, for label=0, the model does not predict any bounding boxes as expected.

Label = 1

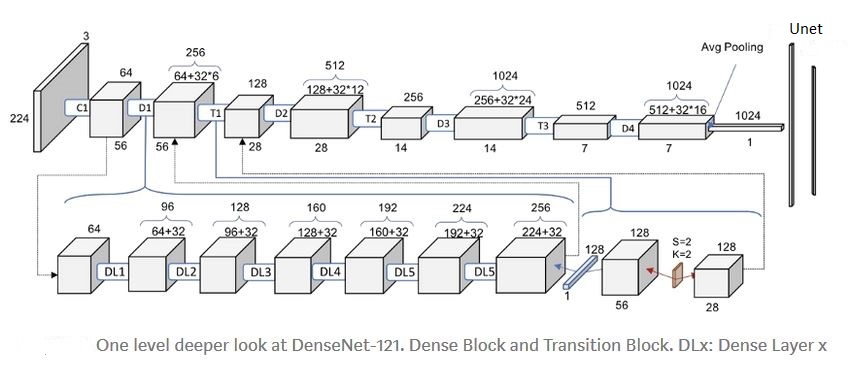
 Label =1

Label = 0

**5.4 CheXNet-UNet**

5.2.1 Model Overview

The framework for model used comprised of a CheXNet (DenseNet) with a flavor of Unet Toppings. CheXNet act as a feature extraction and Unet act as decoding methodology for image segmentation. This encoder-decoder system was chosen due to (i) The major advantage is that in this architecture every particular layer is receiving output from all of its previous layers which make the architecture denser in nature and loss is info can be recover from the next dense layer (ref below fig) (ii) CheXnet is capable of classifying 16 different class of diseases in X-ray images (iii) the wide applicability of UNet architecture in medical imaging.

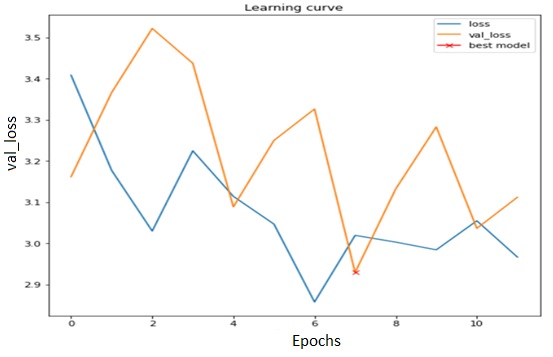


Down-sampling is achieved by the DenseNet encoder and encoded output is fed to the UNet decoder which up-samples the features through transposed convolutions followed by concatenation with the corresponding feature maps in the encoder.

5.2.2 Execution

The train dataset with a total of 26694 images was split into train/validation/test set in a 70/20/10 ratio. The original image resolution was 1024x1024 which is converted into 224x224 so that given image is in model input format.

|  |  |  |  |
| --- | --- | --- | --- |
| **Input Size** | **Max Batch Size** | **Val\_loss** | **Val\_dice\_score** |
| 224x224 | 16 | 2.9 | 0.4 |

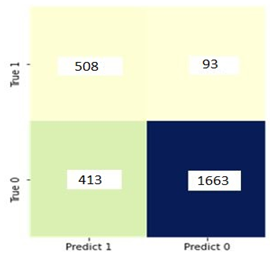


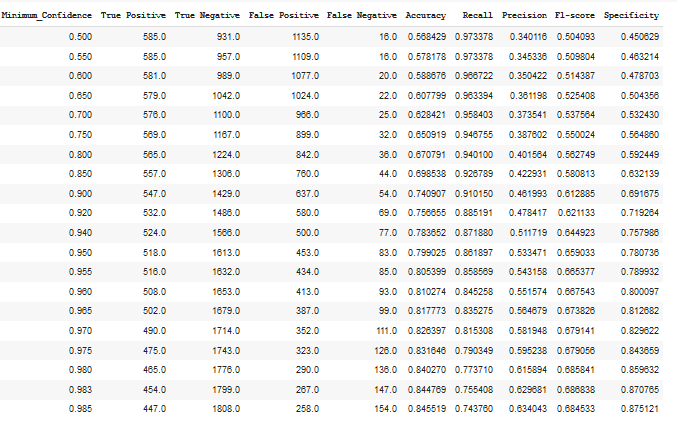
The learning curve show that at epochs 7 model has less val\_loss as compare to loss at any other epochs. So we selected and saved weights at epochs 7 and tried to predict result on test images.

5.2.3 Results

The test set with 2667 nearly images was evaluated by varying the prediction threshold from 0.5 to 0.95 and recording the corresponding performance metrics like Accuracy, Precision, Recall, F-score and Specificity. Since the project outcome is sensitive to False Negatives (i.e. the ability to correctly predict pneumonia containing images), we chose Recall as the deciding metric. Optimal results were obtained at a prediction threshold of 0.96 with a Recall of 0.845. The corresponding confusion matrix and the performance metrics are given below.

|  |  |
| --- | --- |
| **True Positives** | **508** |
| True Negatives | 1663 |
| False Positives | 413 |
| False Negatives | 93 |



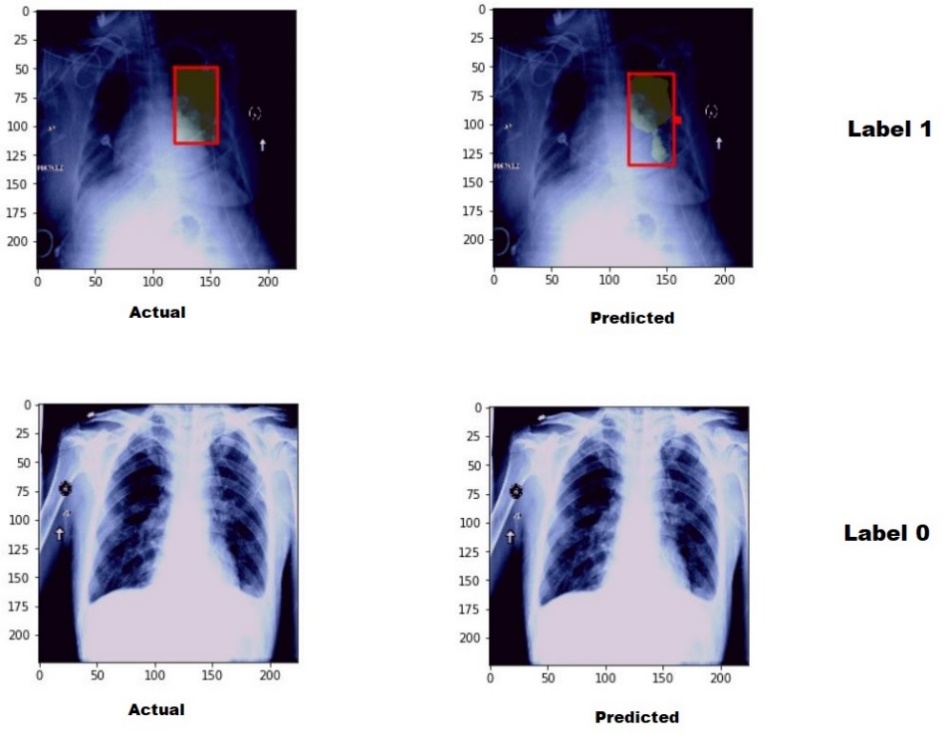


In above fig optimal results were obtained at a prediction threshold of 0.96 with a Recall of 0.845 because True Positive and False Negative deviation is not much as compare to previous records.

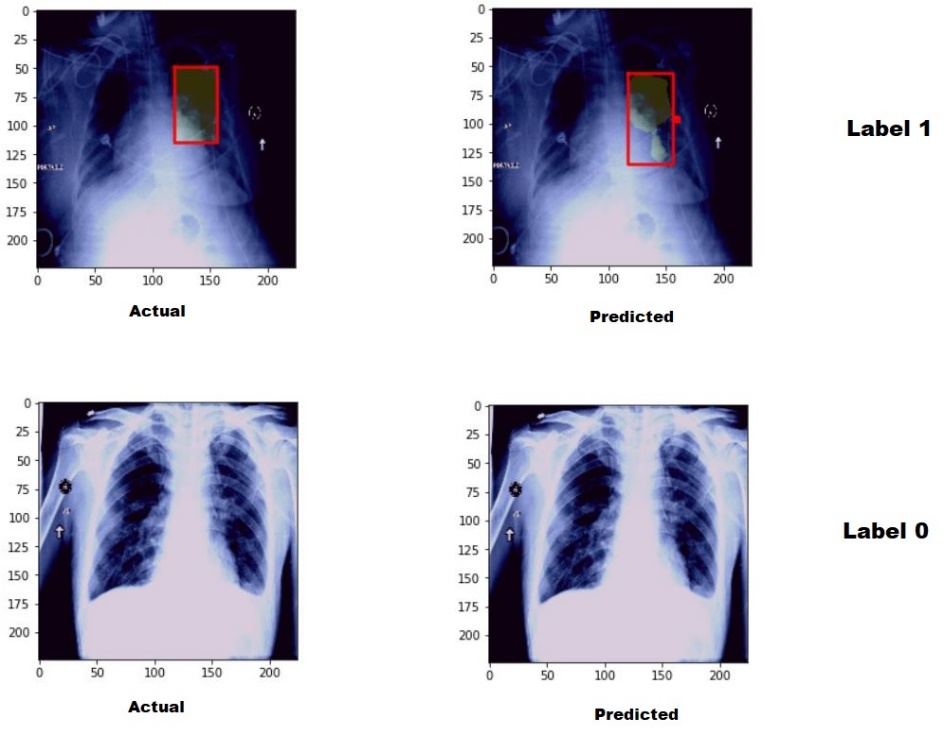
5.2.4. Predictions

The predicted masks at a minimum confidence value of 0.96 are shown below for some of the test images corresponding to label =1 (pneumonia) and label = 0 (non-pneumonia). From left to right, the first image shows the ground truth (masks) superimposed on the X-ray image while the predicted mask is displayed with the bounding box on the right.

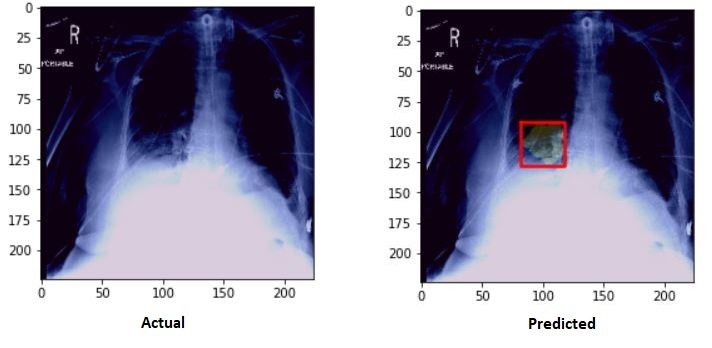
1. Patient having pneumonia predicted correctly



1. Patient not having pneumonia predicted correctly



1. Patient not having pneumonia predicted un-correctly



**6. Comparative Results**

The performances of the three models have been compared based on the evaluation metrics as displayed in the following table

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Val\_loss** | **Accuracy** | **Precision** | **Recall** | **F-1 score** | **Specificity** |
| Mask RCNN | 1.133 | 0.751 | 0.564 | 0.747 | 0.643 | 0.753 |
| Mobilenet-Unet | 0.8662 | 0.845 | 0.621 | 0.742 | 0.676 | 0.873 |
| CheXNet-Unet | 2.9 | 0.810 | 0.551 | 0.845 | 0.667 | 0.800 |

For this project, it is crucial to not incorrectly label pneumonia containing images as normal as this would lead to unfavorable patient outcomes. Therefore, the number of False Negatives or the magnitude of Type II error needs to be minimized. Recall is the ratio of (True Positives) /(True Positives + False Negatives), and a suitable metric to evaluate the 3 models. Based on this criterion, the CheXNet-Unet model with a Recall of 0.845 outperforms the other two.

**7. Implications**

The implications of the project are four fold:

* For a given chest X-ray the model can predict if the patient is suffering from pneumonia
* In addition, the location of the inflammation can be identified with an optimal degree of confidence based on the bounding box coordinates
* Analysis of chest X-rays can be a time consuming process, these models can help speed up the interpretation process
* CNN based deep learning models have the ability to recognize complex patterns in images as a result they can assist physicians with the correct assessment of diseases. The sensitivity of these models can also help in identifying an early onset of disease which can work in favor of the patient.

**8. Limitations and Improvements**

The limitations and possible improvements are as follows:

* Nearly 60% of the dataset corresponds to Target=0, therefore there will be an inherent bias towards accurately predicting images associated with label = 0 i.e. the 'non-pneumonia' images.
* Since the aim of the project is to localize pneumonia patches in Target=1, the above data imbalance might not yield satisfactory results.
* Target 0 has images associated with 2 Classes: 'Normal' and 'No Lung Opacity/Not Normal'. It is very likely that ‘Not-Normal’ class contributes to the False Positives
* A more balanced dataset would improve predictions
* Bounding boxes coordinates for the 'Not-normal' class would help differentiate the ‘pneumonia patches’ from other abnormalities thereby reducing the count of False Positives in the predictions

**9. Learning's**

* A broad understanding of the applicability of AI in image analysis, medical images in particular
* Use of pydicom package to extract the elements of a DICOM image file
* The importance of custom loss and metric functions for model evaluation
* Collaboration between team members for different tasks

**10. References**

1. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>
2. <https://www.who.int/bulletin/volumes/82/12/rudan1204abstract>
3. Kallianos, K., Mongan, J., Antani, S., Henry, T., Taylor, A., Abuya, J., Kohli, M. How far have we come? Artificial intelligence for chest radiograph interpretation. Clin. Radiol. 2019, 74, 338–345
4. Howard, Andrew & Zhu, Menglong & Chen, Bo & Kalenichenko, Dmitry & Wang, Weijun & Weyand, Tobias & Andreetto, Marco & Adam, Hartwig. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.
5. Mask RCNN Model used from github - <https://github.com/matterport/Mask_RCNN>
6. T. I. Mohammad, A. A. Md, T. M. Ahmed, and A. Khalid, “Abnormality detection and localization in chest x-rays using deep convolutional neural networks,” 2017, <http://arxiv.org/abs/1705.09850>.
7. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning , https://arxiv.org/abs/1711.05225
8. D. K. Kermany and M. Goldbaum, *Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification*, Mendeley Data, London, UK, 2018.