FDA Submission

Your Name: Tushar Ranjan

Name of your Device: VGG16-Based Pneumonia Detection Algorithm

Algorithm Description

1. General Information

Intended Use Statement: This algorithm is intended to assist radiologists with the diagnosis of pneumonia. It assigns a probability to a chest X-ray, and if that probability is greater than 0.3644, it is considered "positive". This algorithm is NOT intended to provide a definitive diagnosis. It is only intended to assist radiologists in the task of diagnosing pneumonia.
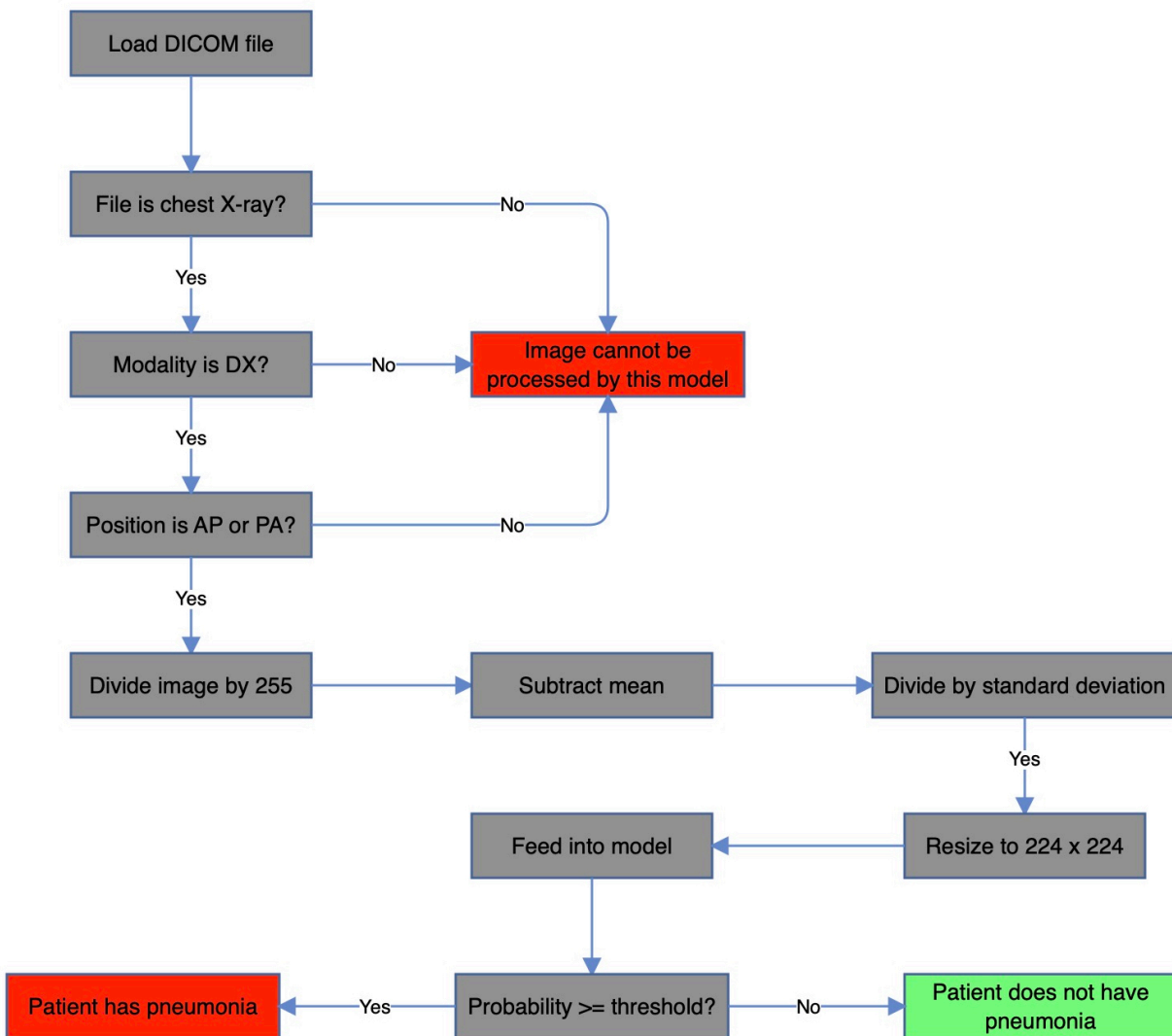
Indications for Use: This algorithm can be used on male or female patients between the ages of 1 and 100. The input X-rays could have been taken in either the "AP" or "PA" position. Comorbidities that indicate for use are edema, infiltration, atelectasis, consolidation, and effusion.

Device Limitations: The algorithm must be run on a system with a dedicated GPU and TensorFlow with at least version 2.1.

Clinical Impact of Performance:

- True Positive: Medical staff can quickly administer necessary care and treatment, potentially preventing complications.
- True Negative: Medical staff would not have to waste time and would be able to focus time and energy on the actual conditions affecting the patient.
- False Positive: Medical staff would end up wasting time treating the patient for a disease they don't actually have.
- False Negative: Pneumonia would go untreated, potentially leading to complications which would require more care and treatment. This is the worst possible outcome.
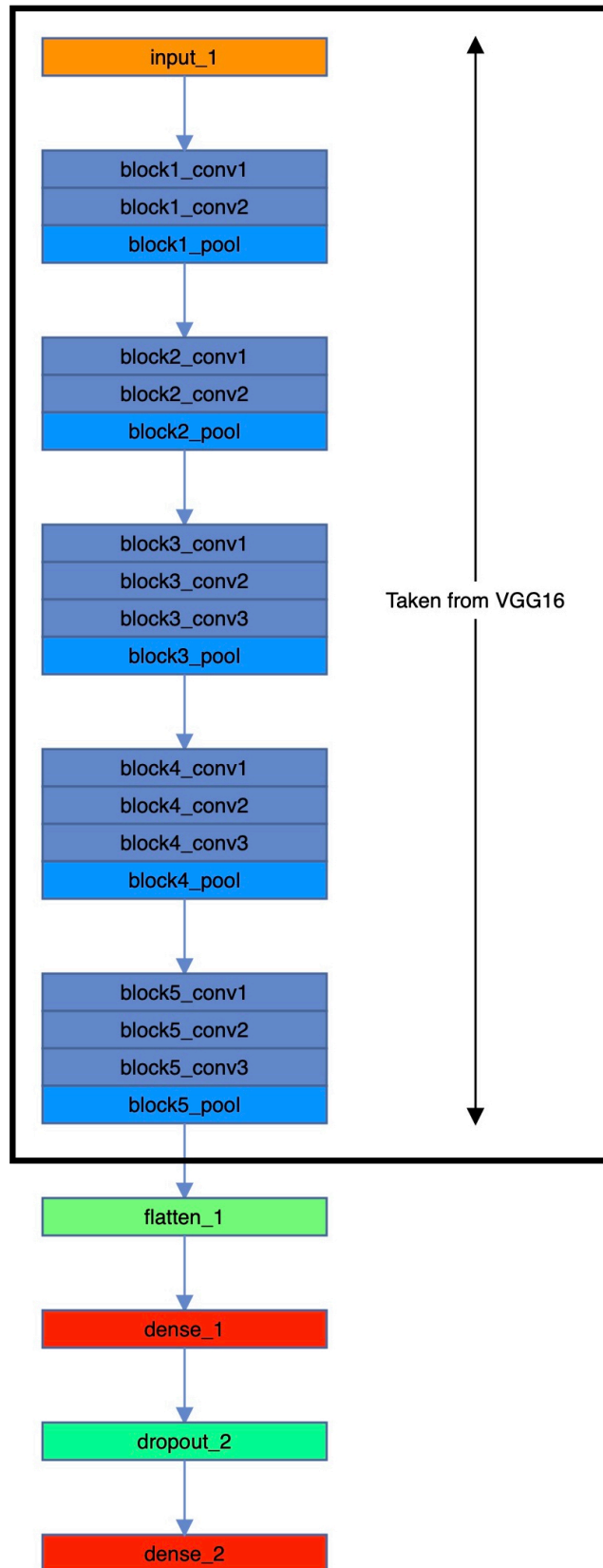
2. Algorithm Design and Function

Load DICOM file → **File is chest X-ray?**
- No → **Image cannot be processed by this model**
- Yes → **Modality is DX?**
  - No → **Image cannot be processed by this model**
  - Yes → **Position is AP or PA?**
    - No → **Image cannot be processed by this model**
    - Yes → **Divide image by 255** → **Subtract mean** → **Divide by standard deviation** → Yes → **Resize to 224 x 224** → **Feed into model** → **Probability >= threshold?**
      - Yes → **Patient has pneumonia**
      - No → **Patient does not have pneumonia**

DICOM Checking Steps:

- Only chest X-rays are accepted
- The modality must be "DX"
- The position must be "AP" or "PA"

Preprocessing Steps:

- Divide the image by 255
- Subtract the mean from each image
- Divide the image by its standard deviation
- Resize the image to 224x224

CNN Architecture:

input_1

block1_conv1
block1_conv2
block1_pool

block2_conv1
block2_conv2
block2_pool

block3_conv1
block3_conv2
block3_conv3
block3_pool

block4_conv1
block4_conv2
block4_conv3
block4_pool

block5_conv1
block5_conv2
block5_conv3
block5_pool

Taken from VGG16

flatten_1

dense_1

dropout_2

dense_2

| Layer Name	Layer Type	Layer Shape	Number of Parameters
input_1	InputLayer	224 x 224 x 3	0
block1_conv1	Conv2D	224 x 224 x 64	1792
block1_conv2	Conv2D	224 x 224 x 64	36928
block1_pool	MaxPooling2D	112 x 112 x 64	0
block2_conv1	Conv2D	112 x 112 x 128	73856
block2_conv2	Conv2D	112 x 112 x 128	147584
block2_pool	MaxPooling2D	56 x 56 x 128	0
block3_conv1	Conv2D	56 x 56 x 256	295168
block3_conv2	Conv2D	56 x 56 x 256	590080
block3_conv3	Conv2D	56 x 56 x 256	590080
block3_pool	MaxPooling2D	28 x 28 x 256	0
block4_conv1	Conv2D	28 x 28 x 512	1180160
block4_conv2	Conv2D	28 x 28 x 512	2359808
block4_conv3	Conv2D	28 x 28 x 512	2359808
block4_pool	MaxPooling2D	14 x 14 x 512	0
block5_conv1	Conv2D	14 x 14 x 512	2359808
block5_conv2	Conv2D	14 x 14 x 512	2359808
block5_conv3	Conv2D	14 x 14 x 512	2359808
block5_pool	MaxPooling2D	7 x 7 x 512	0
flatten_1	Flatten	25088	0
dense_1	Dense	1024	25691136
dropout_1	Dropout	1024	0
dense_2	Dense	1	1025

3. Algorithm Training

Parameters:

- Types of augmentation used during training:
  - Height shift range: 0.1
  - Width shift range: 0.1
  - Rotation range: 15°
  - Shear range: 0.15
  - Zoom range: 0.2
- Batch size:
  - Training: 100
  - Validation: 625
- Optimizer learning rate: 10^{-4}
- Layers of pre-existing architecture that were frozen
  - input_1
  - block1_conv1
  - block1_conv2
  - block1_pool
  - block2_conv1
  - block2_conv2
  - block2_pool
  - block3_conv1
  - block3_conv2
  - block3_conv3
  - block3_pool
  - block4_conv1
  - block4_conv2
  - block4_conv3
  - block4_pool
  - block5_conv1
- Layers of pre-existing architecture that were fine-tuned
  - block5_conv2
  - block5_conv3
  - block5_pool
- Layers added to pre-existing architecture
  - flatten_1
  - dense_1
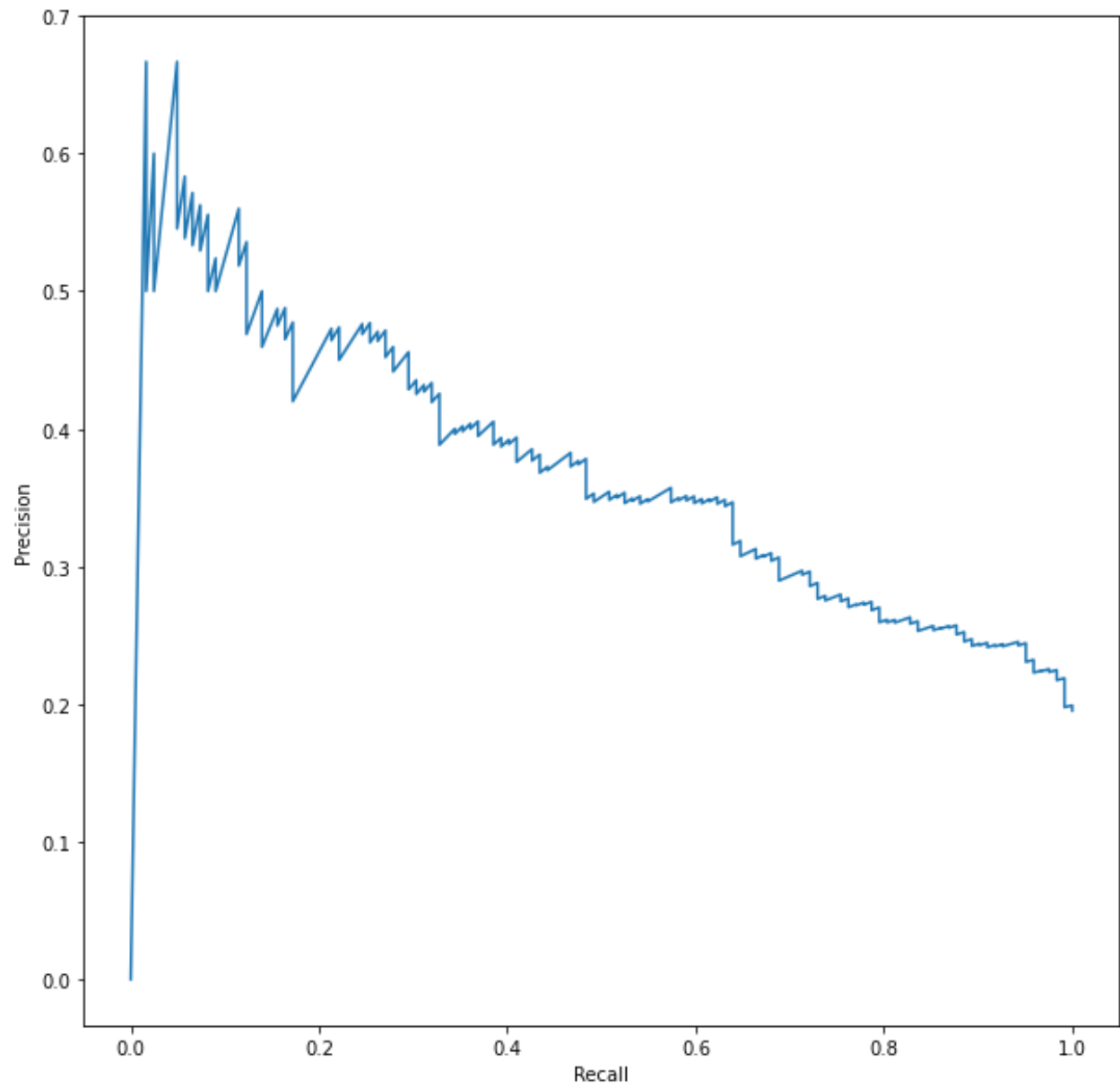 - dropout_1
  - dense_2

Training Results:

- ROC Curve
  - Area under curve: 0.714
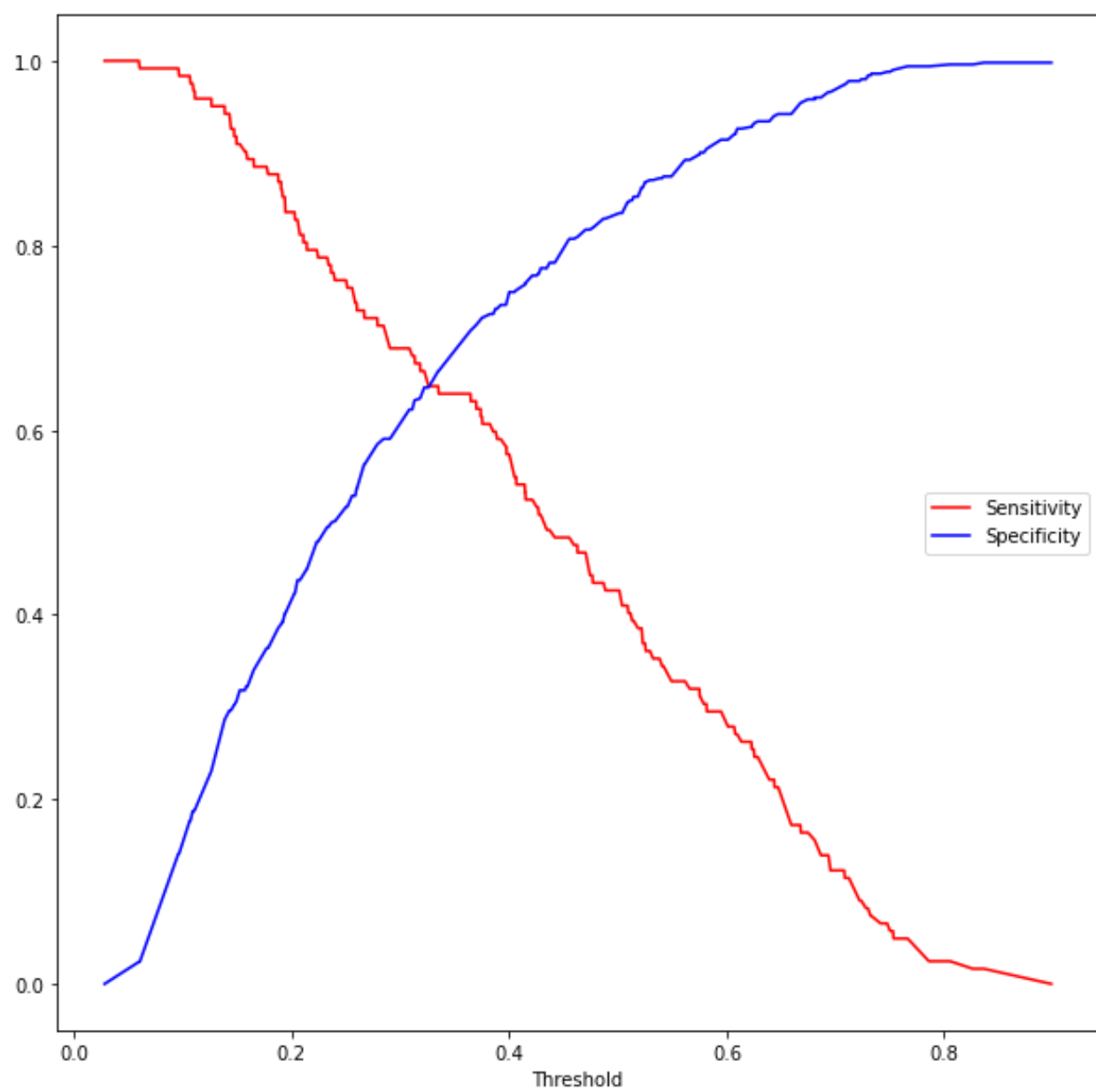 - Threshold closest to (0, 1): 0.3644
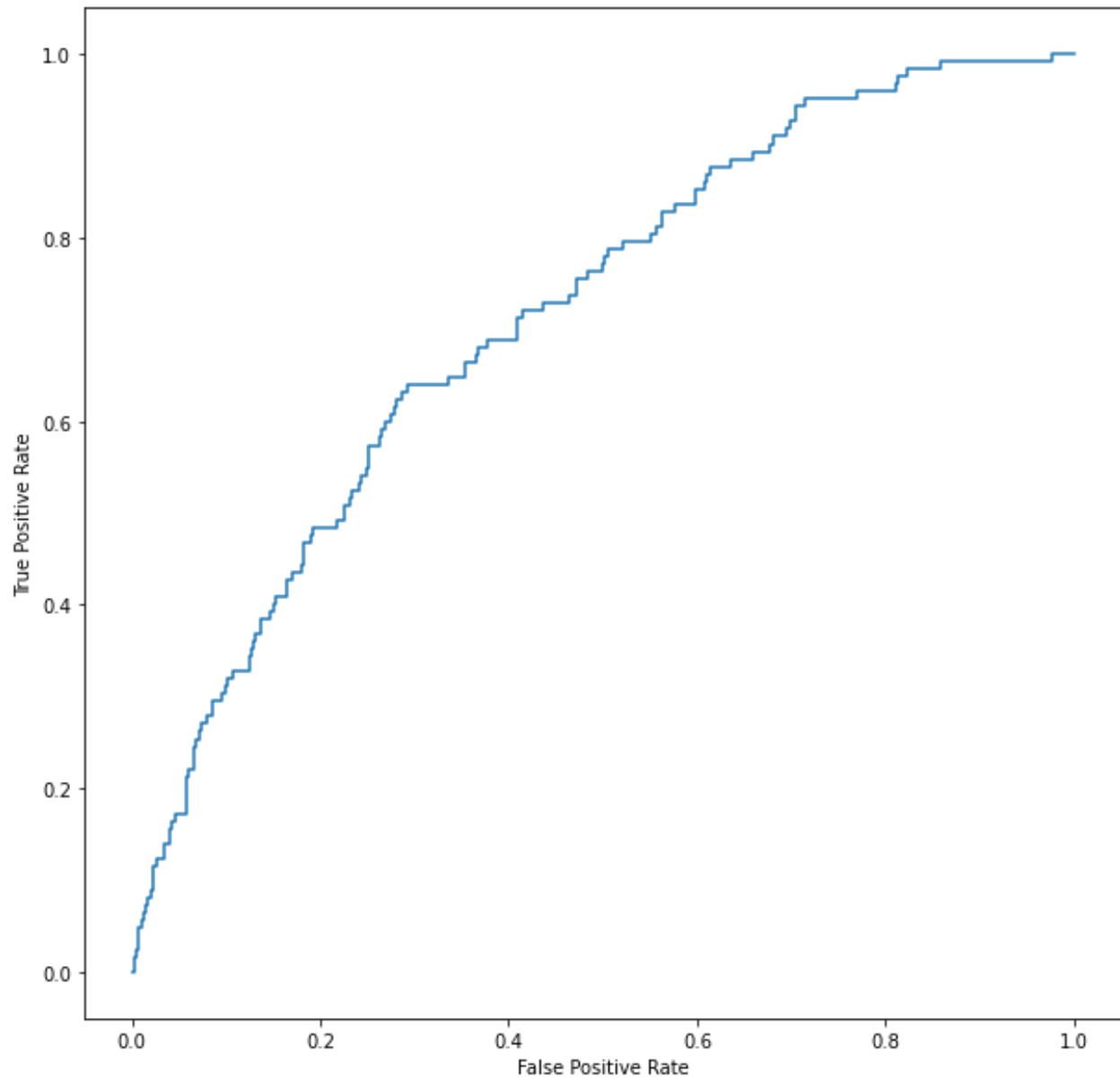
- Precision/Recall/F1
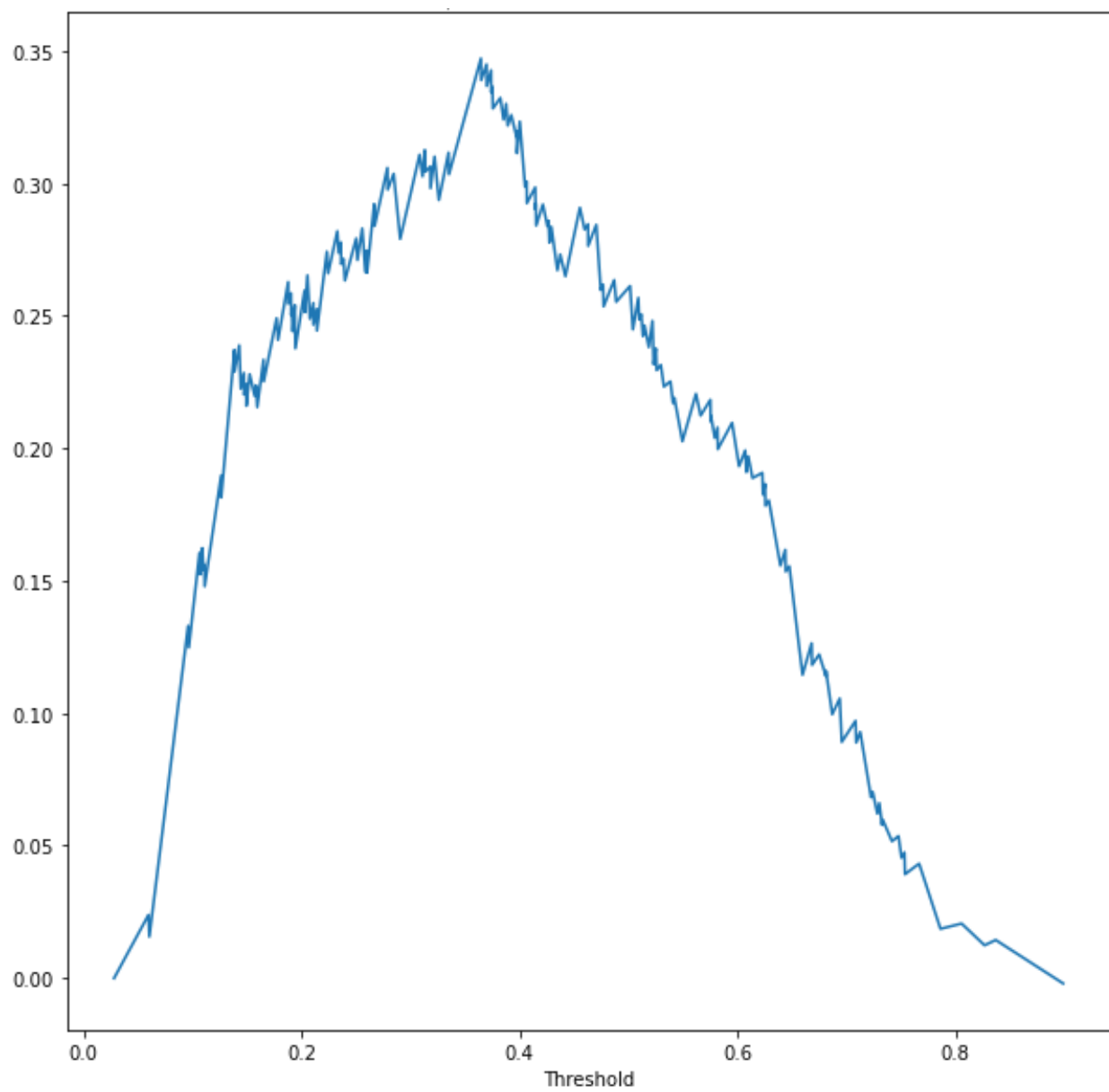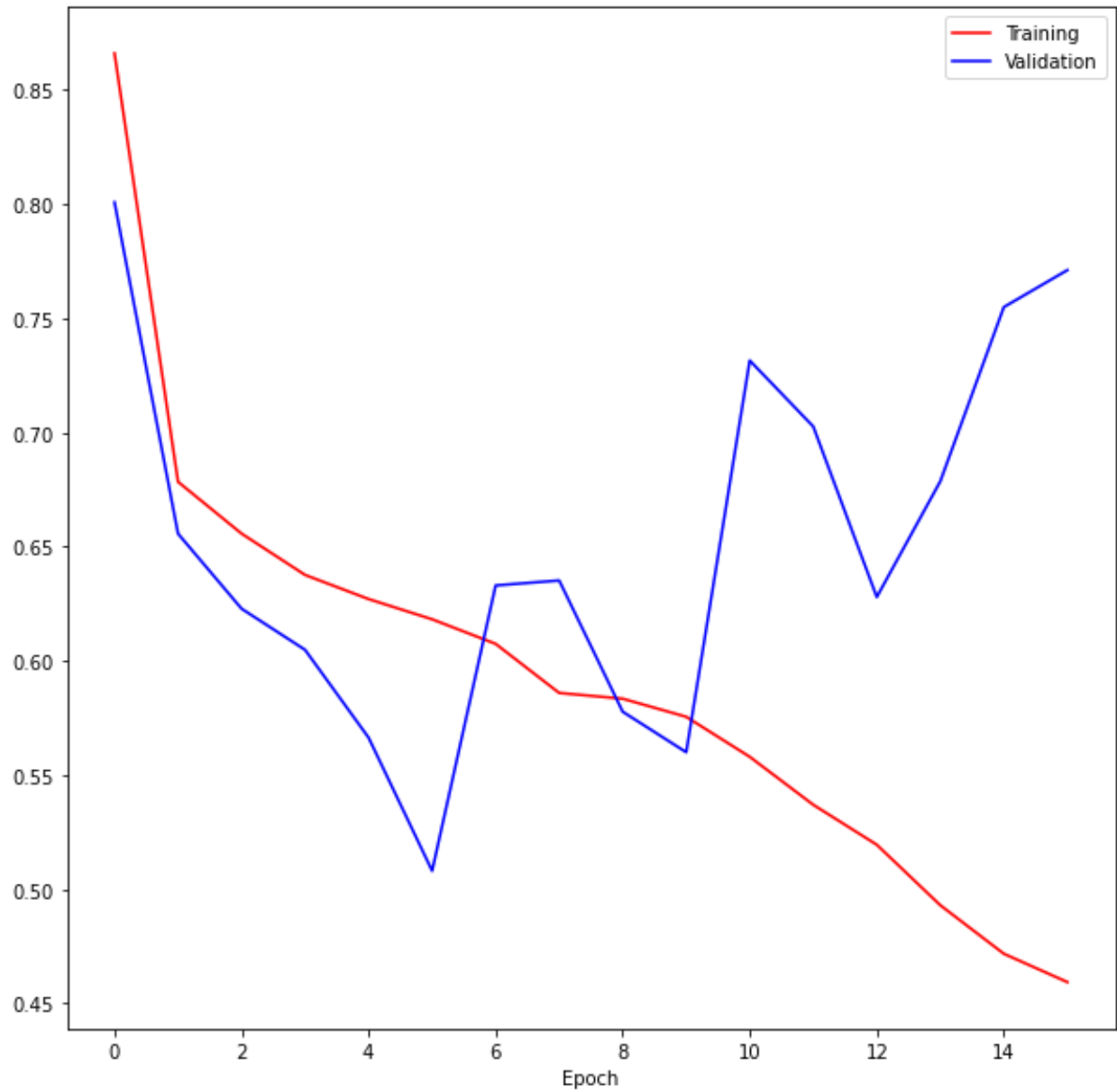 - Threshold with highest F1: 0.3644
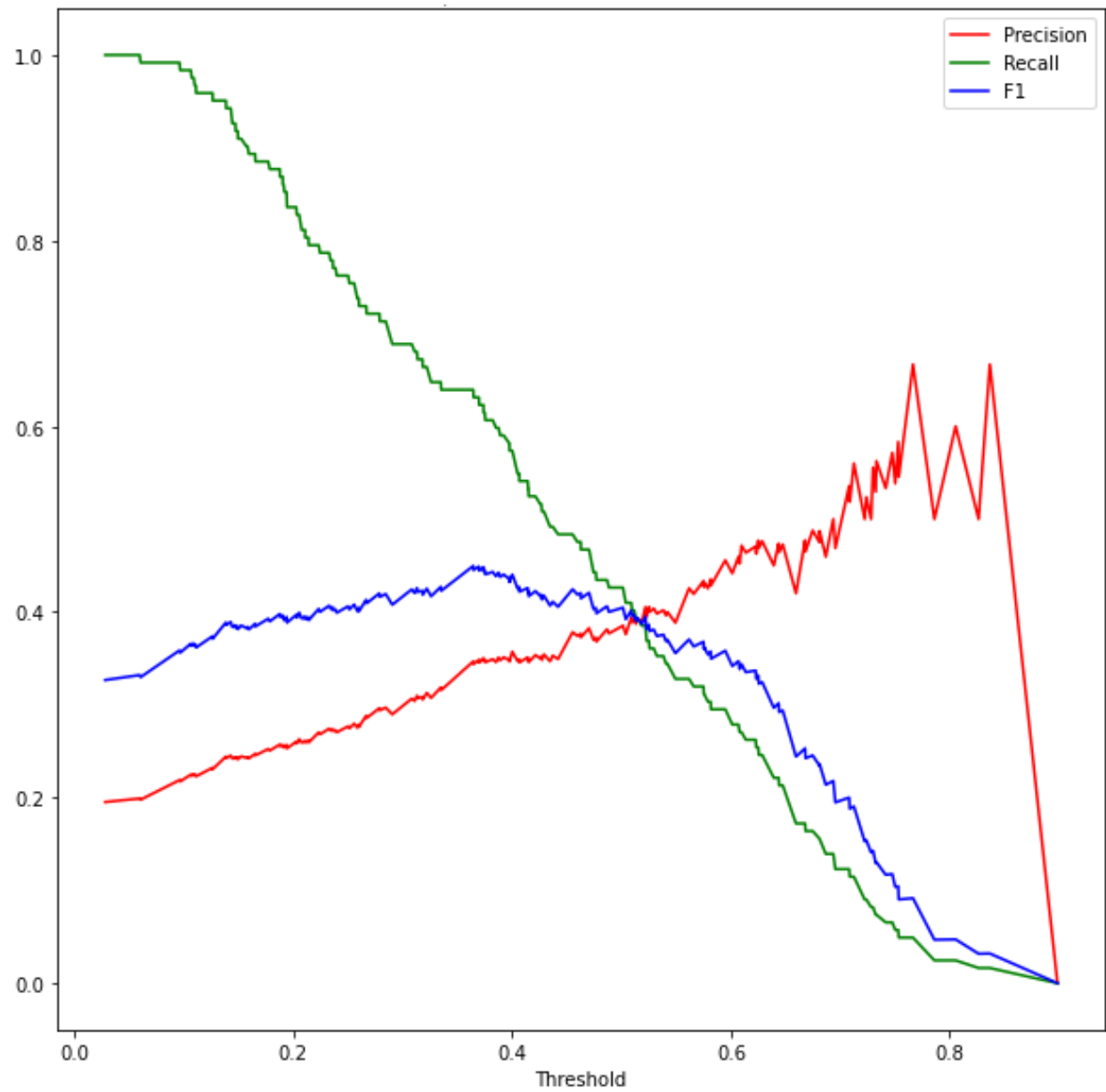
- Precision vs. Recall curve

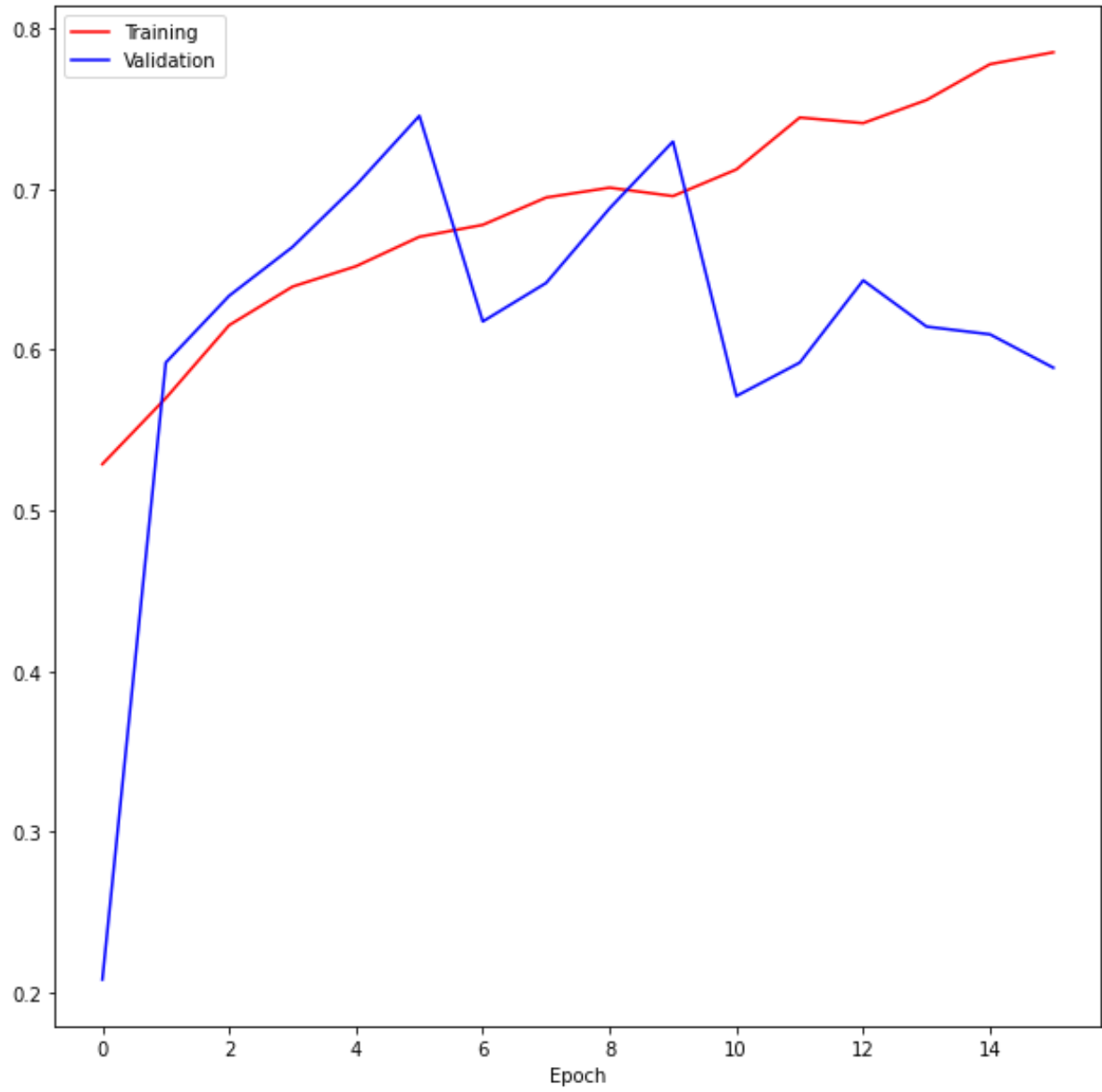- Sensitivity/Specificity vs. Threshold curve

- Youden's J Statistic vs. Threshold curve
  - Threshold with maximum J statistic: 0.3644

- Loss Curve

- Accuracy curve

Final Threshold and Explanation: The chosen threshold for inference was 0.3644, as this was the threshold which resulted in the maximum F1 statistic, the maximum value of Youden's J statistic, and the point on the ROC curve closest to (0, 1). Specifically, the threshold of 0.3644 yielded the following statistics when used on 625 validation samples:

| Statistic	Value
True Positives (TP)	78
False Positives (FP)	152
True Negatives (TN)	351
False Negatives (FN)	44
Sensitivity	0.639344262295082
Specificity	0.697813121272366
Youden's J Statistic	0.337157383567448
Precision	0.339130434782609
Recall	0.639344262295082
F1	0.443181818181818

4. Databases

The dataset consisted of 112,120 chest X-rays. However, there were only 30,805 unique patient ID's, which means that each patient had an average of 3.64 X-rays taken. Only 1,431 X-rays (approximately 1.28%) were labeled as positive for pneumonia.

The age distributions for the entire dataset and for the subset of X-rays labeled positive for pneumonia were roughly the same, as shown in the plots and tables below.
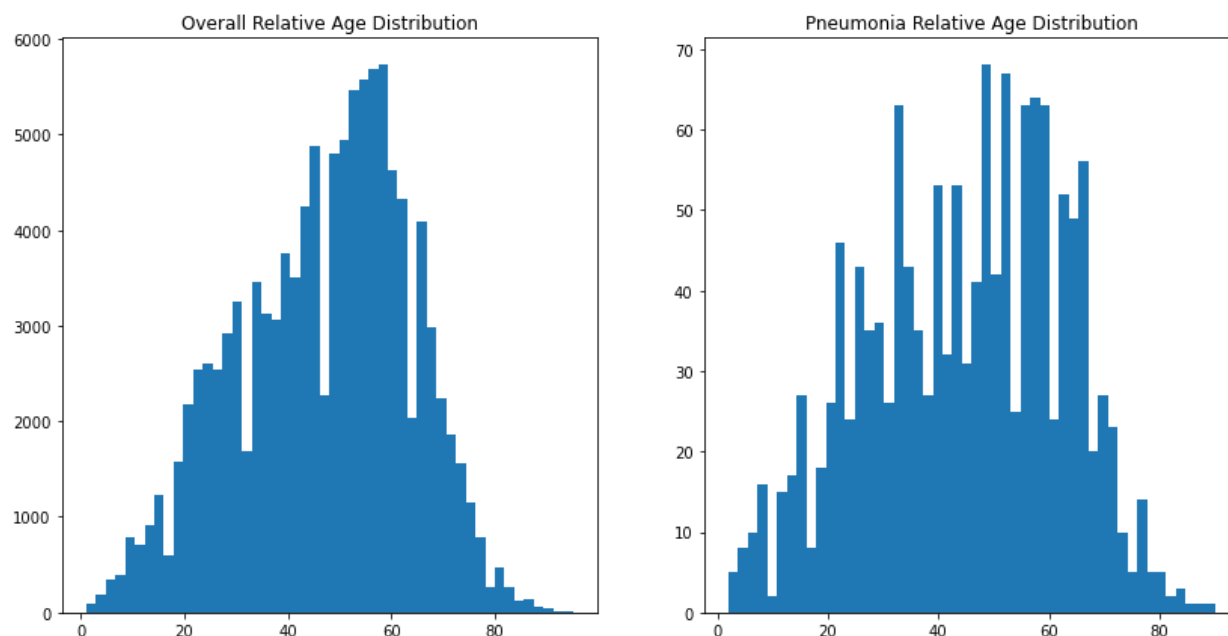
	Overall	Pneumonia-positive
Mean Age	46.87	44.67
Median Age	49	46
Standard Deviation	16.60	17.64
50% of X-rays in range	35 - 59	31 - 59
75% of X-rays in rage	25 - 65	23 - 65

The gender ratio of the overall dataset was 56.49% male, whereas the gender ratio for all pneumonia-positive X-rays was 58.56% male. For the entire dataset, 60% of the X-rays were taken in the PA position, whereas only 44% of the X-rays that were labeled positive for pneumonia were taken in the PA position. X-rays with higher follow-up numbers also appeared to be relatively more common among X-rays that were marked as positive for pneumonia, as the two histograms below show.
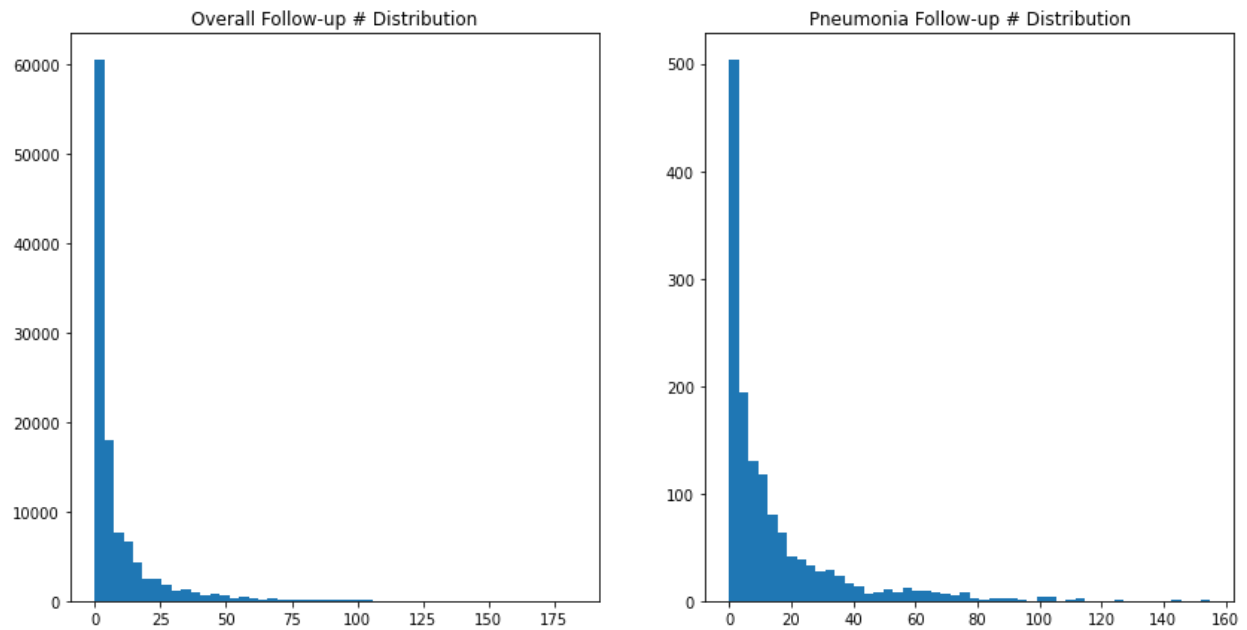
The correlation coefficients between maleness and pneumonia positivity, age and pneumonia positivity, X-ray position and pneumonia positivity, and follow-up number and pneumonia positivity, along with their p-values, are listed in the table below.

	Correlation coefficient	P-value
Age	-0.015	4.4×10^{-7}
Maleness	0.005	0.112
PA X-ray position	-0.037	1.5×10^{-35}
Follow-up number	0.04	1.7×10^{-47}

Edema, infiltration, atelectasis, consolidation, effusion, fibrosis, and pneumothorax were also found to be correlated with pneumonia. Those correlations and their p-values are shown below in descending order.

Condition	Correlation coefficient	P-value
Edema	0.174	0
Infiltration	0.073	2.3×10^{-132}
Atelectasis	0.030	1.3×10^{-23}
Consolidation	0.025	2.9×10^{-17}
Effusion	0.024	3.8×10^{-16}
Fibrosis	-0.007	0.02
Pneumothorax	-0.010	0.001

The dataset - which consisted of 112,120 X-rays, including 1,431 X-rays labeled positive for pneumonia - was originally split into a training subset and a validation subset. The training subset consisted of 80% of the data from the original dataset, including 80% of the X-rays labeled positive for pneumonia. The remaining data all went into the validation subset.

Description of Training Dataset: The training dataset was built from the training subset by combining all the pneumonia-positive X-rays in the training subset and a random sampling of the same size from the pneumonia-negative X-rays in the training subset. This resulted in a training dataset of 2,290 X-rays, half of which were labeled positive for pneumonia.

Description of Validation Dataset: The validation dataset was built from the validation subset by combining all the pneumonia-positive X-rays in the validation subset and a random sampling of four times the same size from the pneumonia-negative X-rays in the validation subset. This resulted in a training dataset of 1,430 X-rays, 20% of which were labeled positive for pneumonia.

The process by which the two datasets were created is illustrated below.
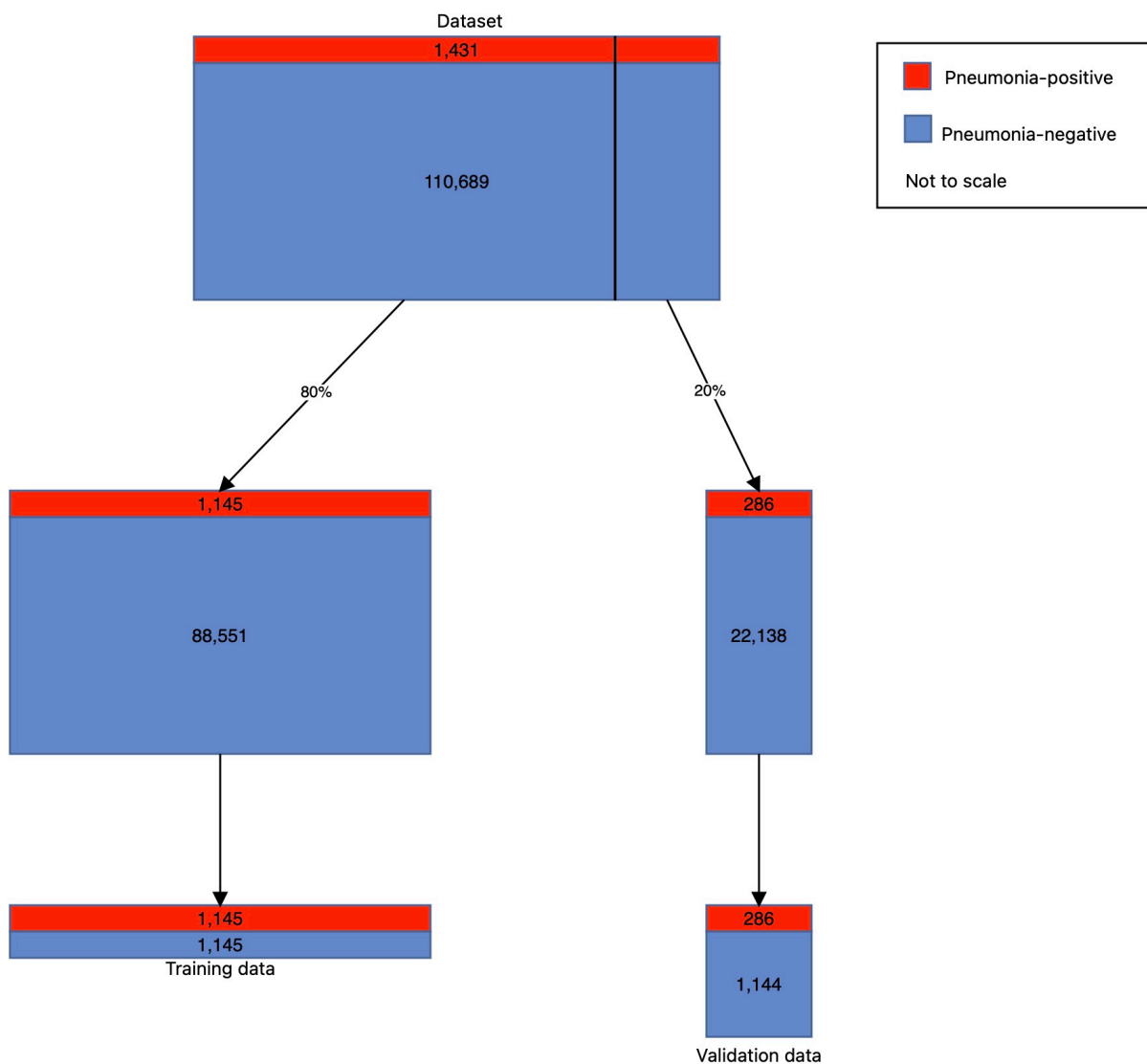
5. Ground Truth

The original dataset[1] of chest X-rays was made available by the National Institutes of Health. It consists of 112,120 X-rays taken from 30,805 unique patients. Each X-ray is labeled as negative or positive for 14 possible diseases: atelectasis, consolidation, infiltration, pneumothorax, edema, emphysema, fibrosis, effusion, pneumonia, pleural thickening, cardiomegaly, nodule mass, or hernia. A "No finding" label is given to those X-rays that are negative for all 14 diseases.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset: The FDA validation dataset should consist only of chest X-rays taken of people between 1 and 100 years of age. The modality for all X-rays should be DX. The position for all X-rays should be either PA or AP.

Ground Truth Acquisition Methodology: The labels for the X-rays were generated by text-mining the associated radiological reports using natural language processing (NLP). It can be assumed that the labels are at least 90% accurate.

Algorithm Performance Standard: According to Rajpurkar, et al.[2], the CheXNet deep learning model was able to achieve an F1 score of 0.435, with a 95% confidence interval between 0.387 and 0.481, and an AUROC of 0.768 when trying to diagnose pneumonia by looking at chest X-rays. The model described in this paper was able to achieve an F1 score of 0.443 by choosing a threshold of 0.3644, as well as an AUROC of 0.714.

---

[1] https://www.kaggle.com/datasets/nih-chest-xrays/data

[2] https://arxiv.org/pdf/1711.05225.pdf