

FDA Submission

Your Name: Hanna Lee

Name of your Device: AI-Aided Pneumonia Detection Tool for Chest Xray

Algorithm Description

1. General Information

Intended Use Statement: This algorithm is designed to aid radiologists in detecting pneumonia. It is intended for use on male and female patients aged 10 to 100 who have undergone a pneumonia screening study with an X-ray machine and have no prior record of abnormal pneumonia findings.

Indications for Use: This algorithm is intended to assist healthcare professionals, particularly radiologists, in the detection of pneumonia through the analysis of 2D X-ray images. It is designed for use in male and female patients aged 10 to 100 who have undergone a screening pneumonia study via X-ray imaging and have no prior record of abnormal pneumonia findings. The algorithm serves as an adjunct tool to be used in conjunction with traditional clinical assessments, patient medical history, and other relevant diagnostic information. It is not intended to replace the judgment of healthcare professionals or to be used as a standalone diagnostic tool. Proper interpretation and application of the algorithm's findings should be done by trained healthcare providers. Due to its rapid processing capabilities, the algorithm is well-suited for emergency workflows to aid radiologists in making timely diagnostic decisions.

Device Limitations: Age Considerations: The algorithm may not perform optimally in patients aged 0-10 and in those aged 100 and above. Previous Pneumonia History: The algorithm might not work as well with patients who have a previous history of pneumonia, and it may not provide accurate diagnoses in these cases. The algorithm has low accuracy. Proper medical judgment and interpretation by trained healthcare professionals are crucial. It requires a computer that runs the algorithm model and gets the 2D xray image as a input.

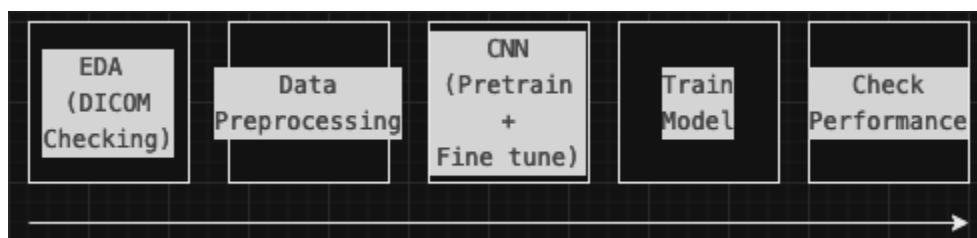
- **False Positives:** May cause undue stress and anxiety in patients due to a misdiagnosis of pneumonia. May lead to unnecessary diagnostic tests and treatments. Can increase healthcare costs for the patient due to these unnecessary tests and treatments.
- **False Negatives:** Can lead to delayed diagnosis and treatment of pneumonia, which can worsen the condition and increase the risk of complications. May lead to spreading of the infection to others if the patient is not isolated. Can result in increased healthcare costs due to more intensive treatments and longer hospital stays that might be required if the pneumonia is not diagnosed and treated promptly.

Clinical Impact of Performance:

- **Accuracy:** The accuracy of the algorithm, which stands at 0.5692, indicates that it correctly identifies pneumonia in only about 56.92% of the cases. This means that in 43.08% of the cases, the algorithm either misses actual pneumonia cases (false negatives) or incorrectly classifies healthy patients as having pneumonia (false positives).
- **Precision:** Precision of 0.2184 means that when the algorithm predicts pneumonia, it is correct only about 21.84% of the time. Low precision indicates a high rate of false positives. False positives can lead to unnecessary anxiety, further unnecessary diagnostic procedures, and potentially harmful treatments for patients who do not actually have pneumonia.
- **Recall:** Recall (or Sensitivity) of 0.4476 indicates that the algorithm correctly identifies only about 44.76% of all actual pneumonia cases. A low recall indicates a high rate of false negatives. False negatives can be very harmful because patients who have pneumonia may not receive the appropriate treatment, potentially leading to worsened health outcomes.
- **F1 Score:** An F1 Score of 0.2936 is a measure of the algorithm's accuracy in terms of both precision and recall. The low F1 score indicates that the algorithm does not have a balanced performance regarding false positives and false negatives.
- **Sensitivity:** Sensitivity (or Recall) of 0.42657 means that the algorithm correctly identifies approximately 42.657% of actual pneumonia cases. This low sensitivity again highlights the risk of false negatives.

2. Algorithm Design and Function

The algorithm employed for pneumonia detection is built upon the VGG16 architecture, which is a well-established and widely used convolutional neural network architecture. For initialization, we utilized weights pre-trained on the ImageNet dataset, a large and diverse image dataset commonly used for image classification tasks. Following initialization, the model was further trained and fine-tuned using a dataset of 2D X-ray images, specifically curated for pneumonia detection. This approach harnesses the generalization capabilities provided by the pre-trained ImageNet weights while optimizing the model to the specific task of identifying signs of pneumonia in 2D X-ray images. To obtain a ground truth, the a single radiologist's labels would probably suffice, because I gave you the hint that they're really good at labeling 'normal' v. 'abnormal.'



DICOM Checking Steps: This algorithm checks attributes such as Demographic information, Patient Position, Body Part Examined, and Image Type.

1. Identify the columns in the dataset.
2. Determine the count of both Non-Pneumonia and Pneumonia cases.
3. Identify the different types of diseases present in the dataset.
4. Analyze the distribution of diseases that co-occur with Pneumonia.

5. Explore the available patient demographic data, including gender, age, patient position, etc.
6. Examine X-ray images of both Non-Pneumonia and Pneumonia cases.

Preprocessing Steps:

- **Rescaling:** Rescaling the pixel values of the images to the range [0, 1] by dividing them by 255.0. This is done to normalize the data and is a common preprocessing step for image data. It is applied to both training and validation data.
- **Horizontal Flipping:** Randomly flipping the images horizontally (left-to-right). This can increase the diversity of the training data and improve model generalization.
- **Height and Width Shift:** Randomly translating the images vertically and horizontally. This is controlled by the `height_shift_range` and `width_shift_range` parameters, which specify the maximum fraction of the height or width by which the image can be shifted.
- **Rotation:** Randomly rotating the images by an angle between -20 and 20 degrees. The `rotation_range` parameter sets the maximum rotation angle in degrees.
- **Shear Transformation:** Randomly applying shear transformations to the images. Shearing involves shifting one part of the image along a certain direction while keeping other parts fixed. The `shear_range` parameter sets the shear intensity.
- **Zoom:** Randomly zooming in or out of the images. The `zoom_range` parameter sets the zoom range.

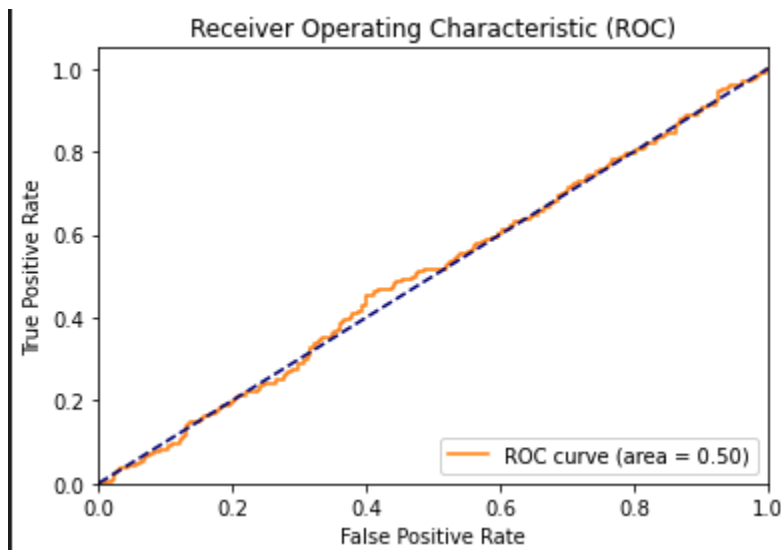
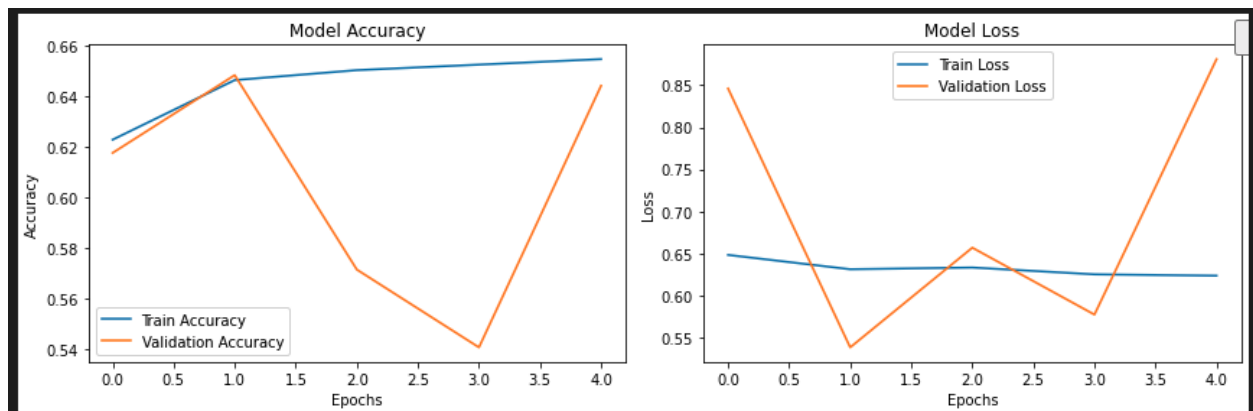
CNN Architecture:

1. **Loading the pretrained Model:** The `load_pretrained_model` function loads the VGG16 model pre-trained on the ImageNet dataset. It specifies which layer (e.g., 'block5_pool') will be used for feature extraction. The function creates a new model that extends from the input layer to the specified transfer layer.
2. **Creating a Custom Model:** The `build_my_model` function creates a custom CNN model using the base model loaded in the previous step. The custom model includes the following layers:
3. The pre-trained base model (e.g., VGG16 without top layers). A Flatten layer to convert the output of the base model to a 1D tensor. Dropout layers to reduce overfitting by randomly dropping some of the layer's units during training. Dense (fully connected) layers with ReLU activation to learn the features extracted by the base model. More Dense layers with Dropout layers are added for regularization. A final Dense layer with a sigmoid activation function for binary classification (outputs values between 0 and 1). The custom model is compiled with the Adam optimizer, binary cross entropy loss function, and accuracy metric.
4. **Training the Custom Model:** The training process involves the following steps: Freeze the first 15 layers of the pre-trained base model. This means that only the weights of the remaining layers will be updated during training. Specify the path to save the model weights with the best validation accuracy. Use the `ModelCheckpoint` callback to save the model weights with the best validation accuracy during training. Use the `EarlyStopping` callback to stop training when the validation accuracy stops improving after a specified number of epochs (patience parameter). Create a list of callbacks to be used during training.

3. Algorithm Training

Parameters:

- Types of augmentation used during training Rescaling, Horizontal Flipping, Height and Width Shift, Rotation, Shear, Zoom
- Batch size 9
- Optimizer learning rate 0.0001
- Layers of pre-existing architecture that were frozen 15
- Layers of pre-existing architecture that were fine-tuned 4
- Layers added to pre-existing architecture 8



Final Threshold and Explanation: Compute Youden's J statistic: $J = \text{Sensitivity} + \text{Specificity} - 1 = \text{TPR} - \text{FPR}$ Find the threshold that maximizes this statistic.

4. Databases

(For the below, include visualizations as they are useful and relevant)

Description of Training Dataset:

1. `train_test_split` made sure that we had the same proportions of Pneumonia in both sets.
2. Condition 1 - To have *EQUAL* amount of positive and negative cases of Pneumonia in Training We know that we want our model to be trained on a set that has *equal* proportions of Pneumonia and no Pneumonia, so we're going to have to throw away some data. We randomly chose a set of non-Pneumonia images using the `sample()` function that was the same length as the number of true Pneumonia cases we had, and then we threw out the rest of the non-Pneumothorax cases. Now our training dataset is balanced 50-50.
3. **3 Description of Validation Dataset:** 20% positive cases of Pneumonia in the Test Set We want to make the balance in our validation set more like 20-80 since our exercise told us that the prevalence of Pneumonia in this clinical situation is about 20%.

5. Ground Truth

Radiologist Analysis: A radiologist, a medical doctor specializing in interpreting medical images, examines the X-ray image for signs of abnormalities in pneumonia in a chest X-ray. The radiologist identifies and labels the features of interest.

Here's a step-by-step breakdown of the process:

Patient Preparation: The patient is positioned appropriately depending on the body part being imaged. For a chest X-ray, the patient is typically asked to stand and take a deep breath.

X-ray Imaging: The X-ray machine produces a controlled amount of radiation that passes through the patient's body and strikes a detector or film on the other side. The X-ray radiation is absorbed by different tissues to varying degrees, creating an image based on the differential absorption.

Image Acquisition: The detector captures the X-ray image, which can be a digital image in modern systems or a film in older systems. Digital images can be immediately viewed and analyzed on a computer.

Radiologist Analysis: A radiologist, a medical doctor specialized in interpreting medical images, examines the X-ray image for signs of abnormalities, such as pneumonia in a chest X-ray. The radiologist identifies and labels the features of interest.

Creating Ground Truth Data: The labeled images, along with the radiologist's report, become part of the ground truth dataset. This dataset can be used to train and validate machine learning models for automatic detection of conditions like pneumonia.

Machine Learning Model Training: The ground truth dataset is used to train a machine learning model to automatically identify and classify features in X-ray images. The model's performance is evaluated against the ground truth labels.

6. FDA Validation Plan

Patient Population Description for FDA Validation Dataset: Demographic Diversity: The dataset includes patients of various ages, ranging from newborns to the elderly, with an even distribution across age groups. Gender representation is balanced, with both male and female patients included. The dataset includes individuals from diverse ethnic and racial backgrounds, representing the real-world patient population. Clinical Characteristics:

Patients with varying degrees of pneumonia severity are included in the dataset, from mild to severe cases. The dataset includes both single and multiple pneumonia occurrences within the same patient. Patients with coexisting medical conditions or comorbidities, such as chronic obstructive pulmonary disease (COPD), asthma, or heart disease, are also represented. Patient Position and Image Orientation:

The dataset contains images taken with patients in various positions, including standing, supine, or lateral decubitus positions. Both anteroposterior (AP) and posteroanterior (PA) chest X-ray views are included. Image Quality and Resolution:

The dataset includes X-ray images of different qualities and resolutions, representing the variation seen in clinical practice. Images acquired from different X-ray machines, manufacturers, and imaging protocols are included to ensure the algorithm's performance is robust across various imaging conditions. Temporal Variation:

The dataset includes images taken at different times of the day and across various seasons to account for potential temporal variations in image quality. Radiologist Annotations:

Each image in the dataset is annotated by expert radiologists with details of the diagnosis, including the presence, location, and severity of pneumonia. Annotations are cross-verified by multiple radiologists to ensure consistency and accuracy.

Ground Truth Acquisition Methodology: Patient Preparation: The patient is positioned appropriately depending on the body part being imaged. For a chest X-ray, the patient is typically asked to stand and take a deep breath.

X-ray Imaging: The X-ray machine produces a controlled amount of radiation that passes through the patient's body and strikes a detector or film on the other side. The X-ray radiation is absorbed by different tissues to varying degrees, creating an image based on the differential absorption.

Image Acquisition: The detector captures the X-ray image, which can be a digital image in modern systems or a film in older systems. Digital images can be immediately viewed and analyzed on a computer.

Radiologist Analysis: A radiologist, a medical doctor specialized in interpreting medical images, examines the X-ray image for signs of abnormalities, such as pneumonia in a chest X-ray. The radiologist identifies and labels the features of interest.

Creating Ground Truth Data: The labeled images, along with the radiologist's report, become part of the ground truth dataset. This dataset can be used to train and validate machine learning models for automatic detection of conditions like pneumonia.

Machine Learning Model Training: The ground truth dataset is used to train a machine learning model to automatically identify and classify features in X-ray images. The model's performance is evaluated against the ground truth labels.

Algorithm Performance Standard: Accuracy: 0.5692 Precision: 0.2184 Recall: 0.4476 F1 Score: 0.2936 Sensitivity: 0.4265734 Specificity: 0.56468531

Accuracy: Accuracy is the proportion of correct predictions made by the model out of all predictions. It's calculated by adding the number of true positives (TP) and true negatives (TN) and dividing by the total number of instances (TP + TN + FP + FN). It can be misleading if the classes are imbalanced, so it's often used along with other metrics.

Formula: $\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$ **Precision:** Precision is the proportion of positive identifications that were actually correct. It is calculated by dividing the number of true positives by the sum of true positives and false positives (TP + FP). High precision indicates a low rate of false positives.

Formula: $\text{Precision} = TP / (TP + FP)$ **Recall:** Recall, also known as Sensitivity or True Positive Rate, is the proportion of actual positives that were correctly identified. It is calculated by dividing the number of true positives by the sum of true positives and false negatives (TP + FN). High recall indicates a low rate of false negatives.

Formula: $\text{Recall} = TP / (TP + FN)$ **F1 Score:** F1 Score is the harmonic mean of precision and recall, providing a balance between the two. It ranges from 0 to 1, with 1 indicating perfect precision and recall. An F1 Score is particularly useful when the class distribution is imbalanced.

Formula: $\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$ **Sensitivity:** Sensitivity is synonymous with Recall. It measures the proportion of true positive cases among the instances that are actually positive. It is also known as the True Positive Rate.

Formula: $\text{Sensitivity} = TP / (TP + FN)$ **Specificity:** Specificity is the proportion of true negatives among the instances that are actually negative. It is also known as the True Negative Rate. High specificity indicates a low rate of false positives.

Formula: $\text{Specificity} = TN / (TN + FP)$