Training & Evaluating Neural Network model using Google Cloud AutoML

Overview

In this project, we will train four different models using four variants of the pneumonia dataset. The dataset contains children's' chest x-ray images, and that they are classified into two classes, "normal" and "pneumonia". We want to observe how properties of the data impact the models' results.

Dataset

The dataset used in this project is Kaggle Chest X-Ray Images (Pneumonia) Dataset. It has 2 categories (Pneumonia/Normal).

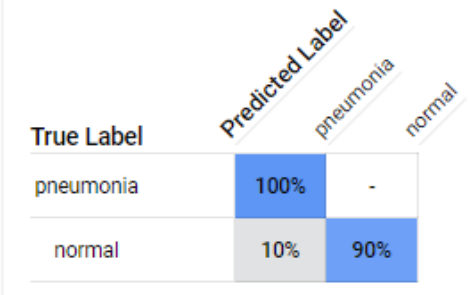
Model 1: Binary Classifier with Clean/Balanced Data

We'll start by training a model simply using 100 images from the "normal" class and 100 images from the "pneumonia" class

Train/Test Split

Training: 100 images labelled as normal and 100 images labelled as pneumonia
Testing \ validation: 10 images were used for testing and 10 images for validation

Confusion Matrix



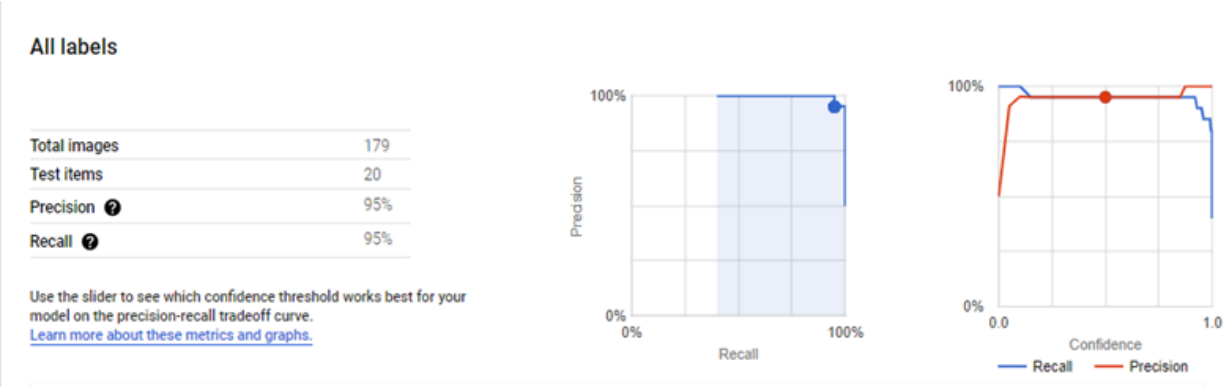
Each cell describes the number of times the model classified an image in each category shown in percentage.

In the first row, the matrix shows that 100% of pneumonia images was correctly classified as "pneumonia. this section includes the correctly predicted actual positive so that TPR is 100%.

On the same row a none of pneumonia images was predicted as normal is shown in the

second cell (no FN for pneumonia class).
In the second row, 10% of normal images was classified as “pneumonia”. which tells that this number of normal images were misclassified and that FPR is 10%
in the next cell, 90% of normal images were classified correctly as “Normal”
In the pneumonia actual, TPR for this class is 90%

Precision and Recall



The **precision** is the ratio of true positives to predicted positives. It measures the number of images the model predicted correctly in its class compared to all predictions.

Recall is the ratio of true positives predictions to actual positives.
The Model achieved a precision and recall of 95%

Score Threshold

When the threshold is increased above 0.8 the precision value will keep increasing but the recall value will decrease.
Because precision takes in consideration the number of images predicted as positive and when we increase the model confidence the precision value will increase. While, for recall value is affected by the presence false negative prediction which decreases its value.

Model 2: Binary Classifier with Clean/Unbalanced Data

Here, we used 100 images from the “normal” class, and 200 more "pneumonia" class images

Train/Test Split

Training: 100 images labelled as normal
300 images labelled as pneumonia

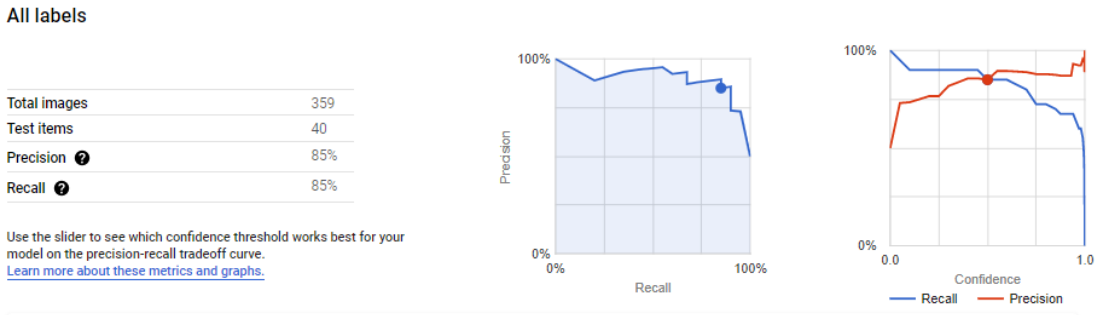
Testing \ Validation: 10 images for testing \ 10 validation

Confusion Matrix

The matrix shows that the model has predicted 83% of pneumonia images and 90% of normal images correctly in their respective classes.

Also, 10% of normal images was classified as “pneumonia” while 17% of pneumonia images were classified as “Normal”.

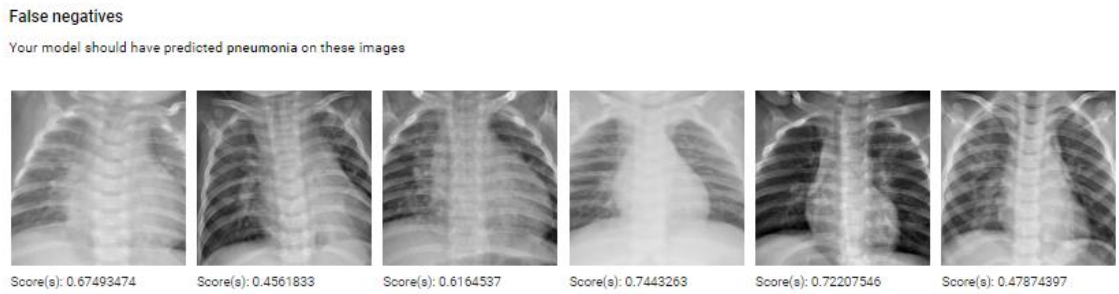
Precision and Recall



The model achieved 85% in both metrics (85% precision, 85 % recall) at 0.5 confidence threshold.

The effect of unbalanced classes on machine learning model

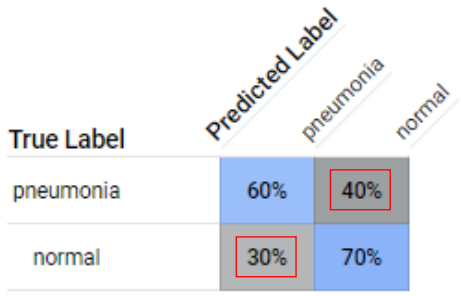
Unbalanced classes increased the number of false positive on normal class 5 images were incorrectly predicted as normal.
Increased the number of False negatives on pneumonia class the model predicted 5 images with pneumonia as normal.



Model 3: Binary Classifier with Dirty/Balanced Data

For this this model, we are using 100 images from the “normal” class and 100 images from the “pneumonia” class
Dirty Data: 30 images of each class is switched (mislabeled)

Confusion Matrix

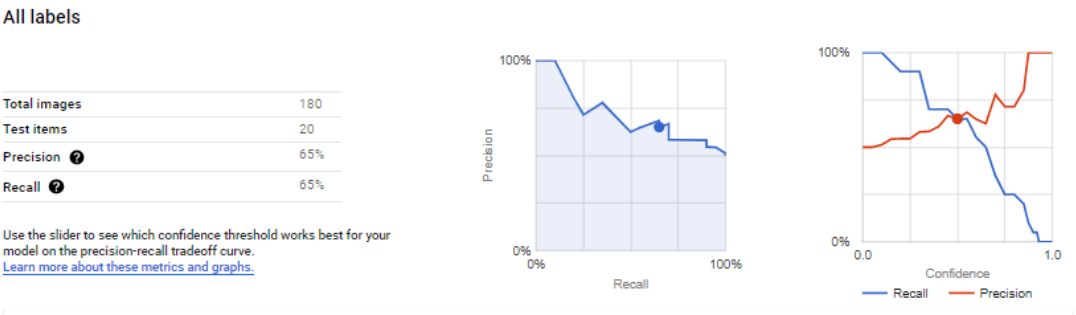


Dirty data has affected the accuracy of the model. The number of False Negatives has increased in each class.

40% of pneumonia images were classified as normal

30% of normal images classified as pneumonia

Precision and Recall

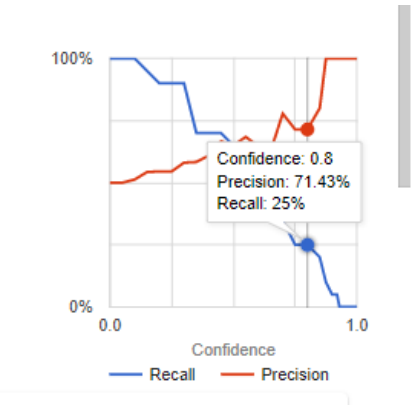


The model precision and recall have dropped to 65% at 0.5 threshold compared to the model metric when clean data was used for training.

By comparing all 3 classifiers, Clean Balanced classifier has performed well among all three with the highest precision and recall value of 95%
Which indicates the importance of training the model with balanced and clean dataset.

The effect of unbalanced classes on machine learning model

Training the model with dirty data has affected the accuracy of the model it can be seen when we increase our confidence threshold to 0.8 the recall drops to 25% which indicates that the Ratio of true positives to actual positives is low. The model is misclassifying actual positives and predicting them as negative.



Model 4: 3-Class Model

In this model, "pneumonia" images actually have two different classes: "bacterial" pneumonia and "viral" pneumonia. We will use 100 "normal" images, 100 "bacterial pneumonia" images, and 100 "viral pneumonia" images (for a total of 3 classes).

Confusion Matrix

True Label \ Predicted Label	normal	viral pneumonia	bacterial pneumonia
normal	100%	-	-
viral pneumonia	10%	80%	10%
bacterial pneumonia	-	10%	90%

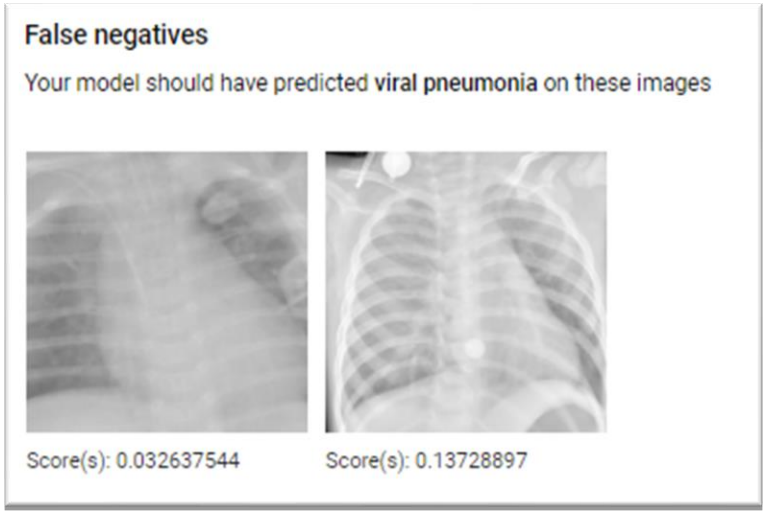
if we look at the values in the diagonal shape, we see that they mark the highest numbers in the matrix which indicates that the model has correctly predicted most of the data in its correct class.

The model predicted all images of normal class correctly.

Confusion appears in bacterial pneumonia class in predicting 10% of the class images as viral pneumonia

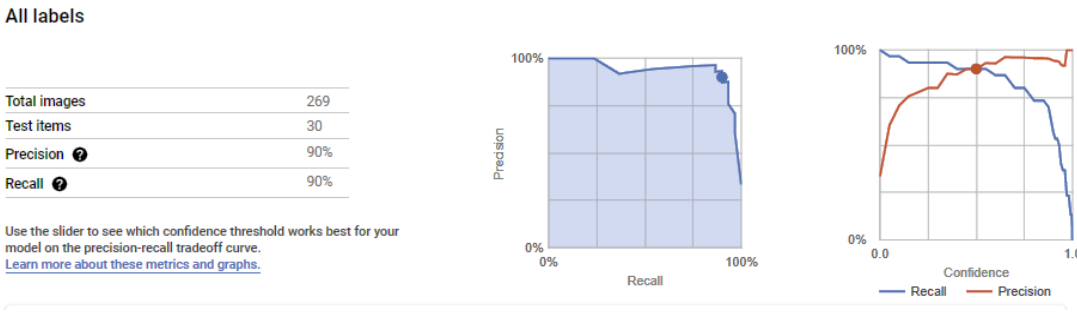
Another area of confusion is when the model has classified 10% of the viral pneumonia as normal and 10% as bacteria pneumonia

A sample of the false negative is shown below



To lessen the confusion, I might train the model on more images of viral pneumonia and select images with better quality.

Precision and Recall



At 0.5 score of thresholds
Precision: 90%
Recall: 90%

To calculate the model precision, we need 1st to calculate the precision to each class

P Normal = TP / TP + FP = 100 / 100 + 10 = 100 / 110 = 0.909 = 0.91
P Viral = TP / TP + FP = 80 / 80 + 10 = 0.888 = 0.89
P Bacterial = TP / TP + FP = 90 / 90 + 10 = 0.9
P Model = (P Normal + P Viral + P Bacterial) / 3 = (0.91 + 0.89 + 0.9) /3 = 0.9

We will calculate the Recall in the same manner

R Normal = TP / TP + FN = 100 / 100 + 0 = 1
R Viral = TP / TP + FN = 80 / 80 + 10 + 10 = 80 / 100 = 0.8
R Bacterial = TP / TP + FN = 90 / 90 + 10 = 90 / 100 = 0.9
R Model = (R Normal + R Viral + R Bacterial) /3= (1 + 0.8 + 0.9) /3 = 0.9

F1 Score

F1 = (2 * precision * recall) / (precision + recall)

F1 = (2 * 0.9 * 0.9) / (0.9 + 0.9)

F1 = 1.62 / 1.8 = 0.9

End of report..