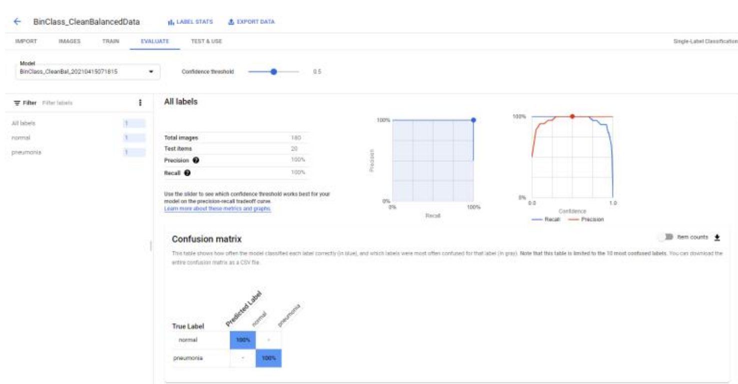
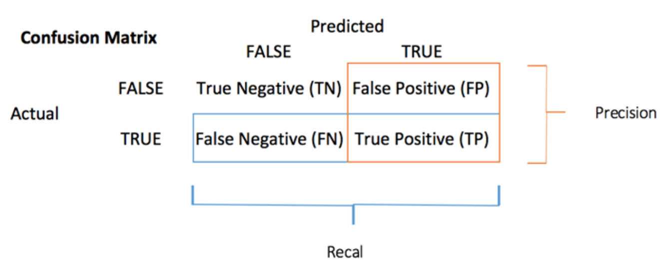
AutoML Modeling Report

MICHAEL KWONG HIN SANG

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	From Evaluate Tab in GCP AutoML Vision: Total images 180 Test Items 20 Google AutoML automatically splits the data into training and testing sets. 80% of data is used for training, 10% for validation and 10% for testing.
Confusion Matrix What do each of the cells in the confusion matrix describe? What values did you observe (include a screenshot)? What is the true positive rate for the “pneumonia” class? What is the false positive rate for the “normal” class?	 The confusion matrix is a table with four cells. It shows how the model classified the normal v/s pneumonia labels after it was trained on the data set. It describes the performance of the classification model or classifier. The first column shows how many normal samples were predicted as normal or pneumonia. So, 100% of the normal samples were predicted as normal (in blue in first column). None of the normal samples were predicted as pneumonia. The <u>true positive rate for pneumonia class</u> means that the model projected correctly the pneumonia images as pneumonia class. The confusion matrix shows the true positive rate is 100%. The second column shows how many pneumonia samples were predicted as normal or pneumonia. So, none of the pneumonia samples were predicted as normal. 100% of the pneumonia samples were predicted

	as pneumonia (in blue in second column). The false positive rate for normal class means that the model projected that there were normal images classified as pneumonia but that is false. The confusion matrix shows the false positive rate is 0%. The confusion matrix here shows that the precision of the model is very high when all normal images were classified as normal class, and all pneumonia images were classified as pneumonia class.
Precision and Recall What does precision measure? What does recall measure? What precision and recall did the model achieve (report the values for a score threshold of 0.5)?	Precision measures how accurately the model predicted that the normal images were normal class label and the pneumonia images were pneumonia class label. Recall measures the true positive rate or sensitivity. Of all the positive examples, it gives what proportion the model predicted accurately. For a score threshold of 0.5, the precision and recall the model achieved were both at 100%.  

<table>
<tr>
<td></td>
<td>

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

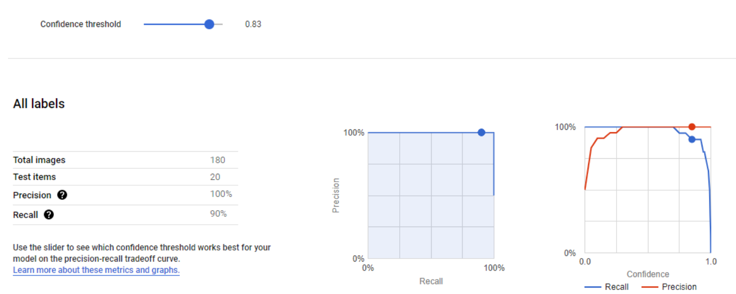
</td>
</tr>
<tr>
<td>

Score Threshold
When you increase the threshold what happens to precision? What happens to recall? Why?

</td>
<td>

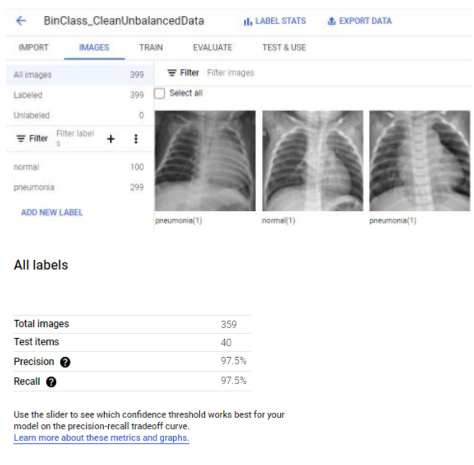
When I increased the threshold, precision stays the same at 100% but recall reduces as shown below.

This is because a higher threshold reduces the recall which is the sensitivity of the positive classification the model gets right. If confidence is low, it means more classification will be done over that low barrier of confidence but images will have more risk to be misclassified. If confidence is higher, it means less images will be classified since that is higher barrier of confidence but images classified will have less risk to be misclassified.



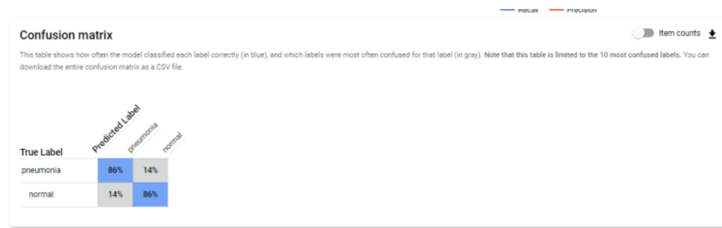
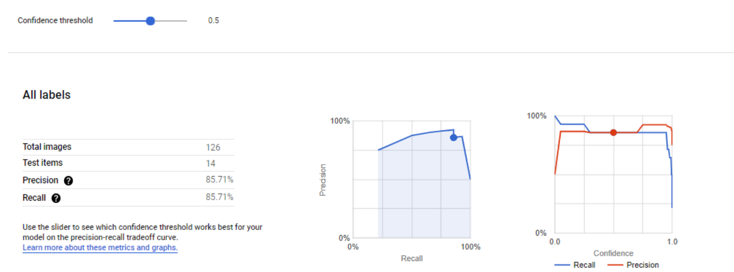
</td>
</tr>
</table>

Binary Classifier with Clean/Unbalanced Data

<table>
<tr>
<td>Train/Test Split
How much data was used for training? How much data was used for testing?</td>
<td>Data used for training 359
Data used for testing 40

</td>
</tr>
<tr>
<td>Confusion Matrix
How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.</td>
<td>

After introducing unbalanced data in this model, the confusion matrix now shows that there are misclassified images. 3% of the pneumonia images were misclassified as normal class. These results are potentially due to the unbalanced data skewing the data to an incorrect class.</td>
</tr>
<tr>
<td>Precision and Recall
How have the model's precision and recall been affected by the unbalanced data (report the values for a score threshold of 0.5)?</td>
<td>With clean/unbalanced data and score threshold of 0.5, the model's precision and recall are lower than with clean/balanced data.</td>
</tr>
<tr>
<td>Unbalanced Classes
From what you have observed, how do unbalanced classed affect a machine learning model?</td>
<td>When training a machine learning model, unbalanced data skew model towards an outcome.</td>
</tr>
</table>

Binary Classifier with Dirty/Balanced Data

Confusion Matrix How has the confusion matrix been affected by the dirty data? Include a screenshot of the new confusion matrix.	 After using dirty data, there are more pneumonia images misclassified as normal class (top right of matrix) and more normal images misclassified as pneumonia class (bottom left of matrix).
Precision and Recall How have the model's precision and recall been affected by the dirty data (report the values for a score threshold of 0.5)? Of the binary classifiers, which has the highest precision? Which has the highest recall?	 With the dirty data, precision and recall were reduced to 85.71% for a score threshold of 0.5. Of the binary classifiers (Clean/Balanced, Clean/Unbalanced, Dirty/Balanced), the highest precision is with Clean/Balanced with Precision of 100% and Recall of 100%.
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Dirty data negatively affect the success metrics used when evaluating the performance of a model: precision, recall and confusion matrix.

3-Class Model

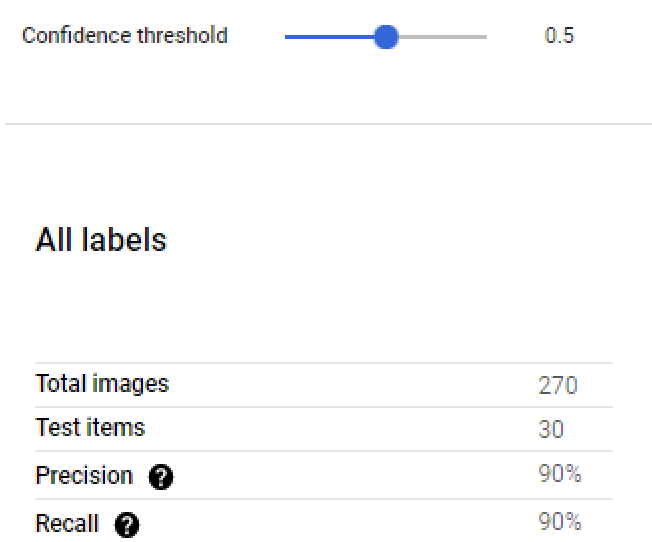
<table>
<tr>
<td>Confusion Matrix
Summarize the 3-class confusion matrix. Which classes is the model most likely to confuse? Which class(es) is the model most likely to get right? Why might you do to try to remedy the model’s “confusion”? Include a screenshot of the new confusion matrix.</td>
<td>


The 3-class model imported 300 images: 100 normal images, 100 bacterial pneumonia images and 100 viral pneumonia images. There are 3 classes: normal, bacterial pneumonia and viral pneumonia.

The class that the model is most likely to confuse is the Viral pneumonia. From confusion matrix, 70% viral pneumonia images classified as viral pneumonia class.

The classes that the model is most likely to get right are the normal and bacterial pneumonia classes. From confusion matrix, they are at 100% prediction.

To remedy the model’s confusion, I would increase the images to train in the dataset.
</td>
</tr>
<tr>
<td>Precision and Recall
What are the model’s precision and recall? How are these values calculated (report the values for a score threshold of 0.5)?</td>
<td>
The model’s precision and recall are as shown below:



</td>
</tr>
</table>

Confidence threshold 0.5

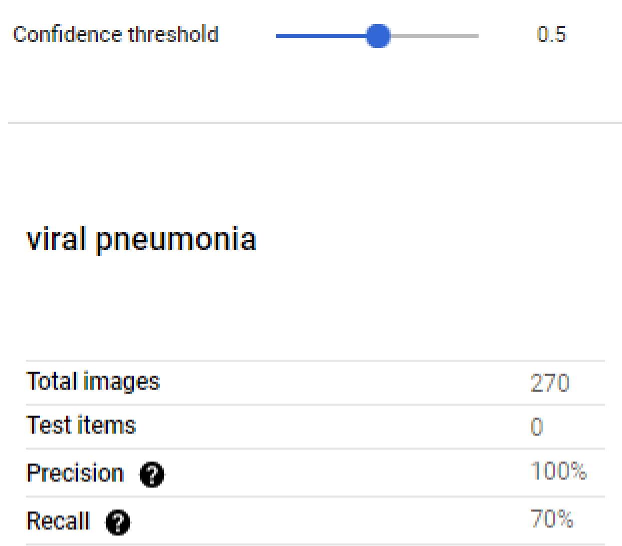
bacterial pneumonia

Total images	270
Test items	0
Precision	83.33%
Recall	100%

Confidence threshold 0.5

normal

Total images	270
Test items	0
Precision	90.91%
Recall	100%

<table>
<tr>
<td></td>
<td>


Confidence threshold 0.5

viral pneumonia

Total images	270
Test items	0
Precision	100%
Recall	70%

In an imbalanced classification problem with more than two classes, the 3-class model precision is calculated as the sum of true positives across all classes divided by the sum of true positives and false positives across all classes.

- Precision = Sum c in C TruePositives_c / Sum c in C (TruePositives_c + FalsePositives_c)

In an imbalanced classification problem with more than two classes, the 3-class model recall is calculated as the sum of true positives across all classes divided by the sum of true positives and false negatives across all classes.

- Recall = Sum c in C TruePositives_c / Sum c in C (TruePositives_c + FalseNegatives_c)
</td>
</tr>
<tr>
<td>F1 Score
What is this model’s F1 score?</td>
<td>F1 score is used in Model Evaluation.

F1 is the mean of precision and recall values.</td>
</tr>
</table>

	$recall = \frac{true\ positives}{true\ positives\ +\ false\ negatives}$ $precision = \frac{true\ positives}{true\ positives + false\ positives}$ $F_1 = 2 * \frac{precision * recall}{precision + recall}$ F1 = 2 * (0.90 * 0.90) / (0.90 + 0.90) F1 = 2 x (0.81 / 1.80) F1 = 0.90