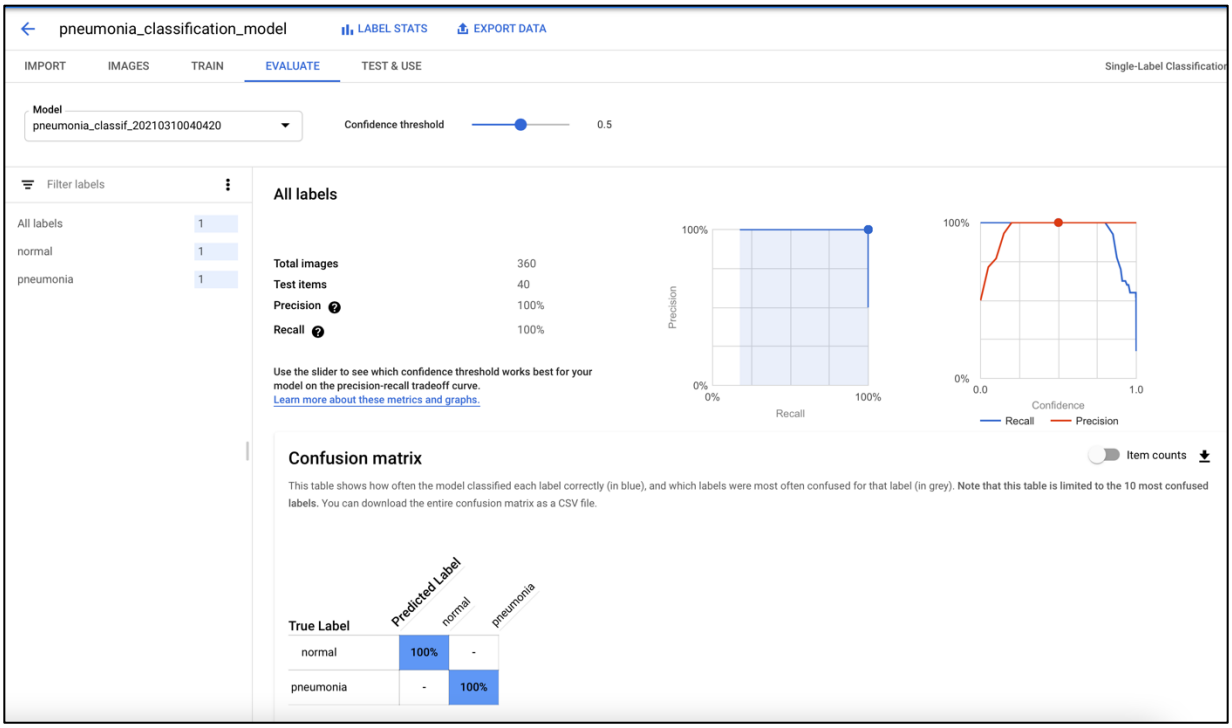
AutoML Modeling Report

Manal Almehmadi

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	400 data images used for training: - 200 images for normal cases - 200 images for pneumonia cases 40 data images for testing. AutoML automatically uses 80% of your images for training, 10% for validating, 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?	Each of the cells in the confusion matrix describes how frequently each label has been classified by the model; and also shows how correctly the desired categories (normal and pneumonia) was identified. Calculating a confusion matrix can give you a better idea of what your classification model is getting right and what types of errors it is making. True positive rate for the "pneumonia" class: "pneumonia" was predicted positive and it's true i.e I predicted that the patient actually had "pneumonia" and it's true. (in my model is 100%) False positive rate for the "normal" class: "normal" case was positively predicted and it's false i.e I predicted that the patient is "normal" and does NOT have pneumonia, but turns out prediction was false (in my model is 0%) The results from the confusion matrix show that model did great job with no errors, all normal and pneumonia images were classified correctly. The errors are at 0%.



Precision and Recall

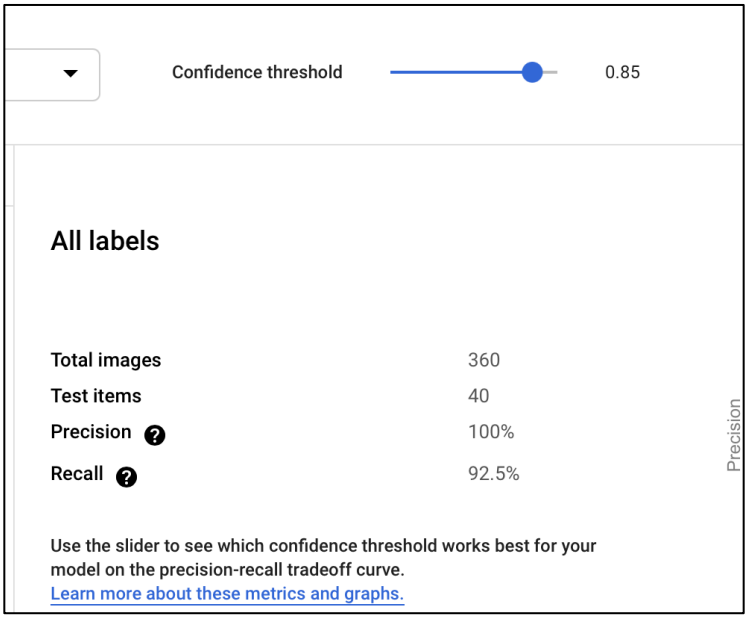
What does precision measure? What does recall measure? What precision and recall did the model achieve (report the values for a score threshold of 0.5)?

Precision measure the percentage of correct predications against total number of predictions.
Model precision answers the question of when the model makes a prediction, how likely is that prediction to be correct?

Recall measures the percentage of correctly identified instances total possible instances.
Model recall answers the question of how good is a model at identifying actual occurrences of objects in the data. This will give us an understanding of whether or not the model can recognize these objects

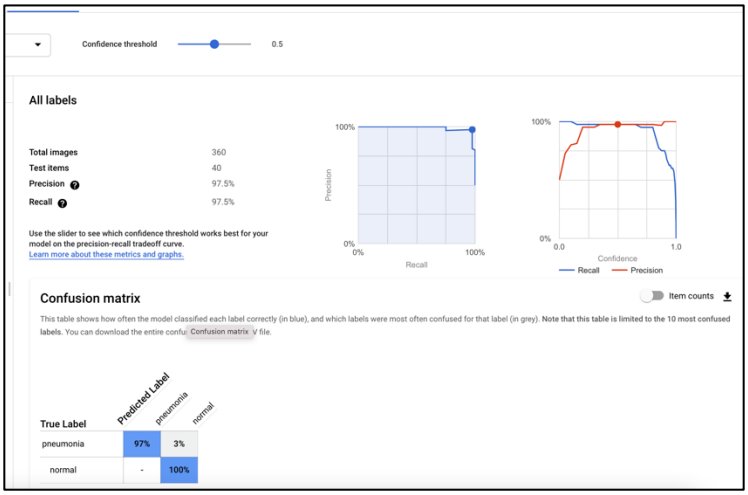
For a score threshold of 0.5, the model achieved a precision of 100.0% and recall of 100.0% (see image).

Total images	360
Test items	40
Precision	100%
Recall	100%

Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	When score threshold is increased to 0.85, it is seen that recall is lower and precision stay the same (100%).  The score threshold refers to the level of confidence the model must have to assign a category to a test item. If your score threshold is low, your model will classify more images, but runs the risk of misclassifying a few images in the process. If your score threshold is high, your model will classify fewer images, but it will have a lower risk of misclassifying images.

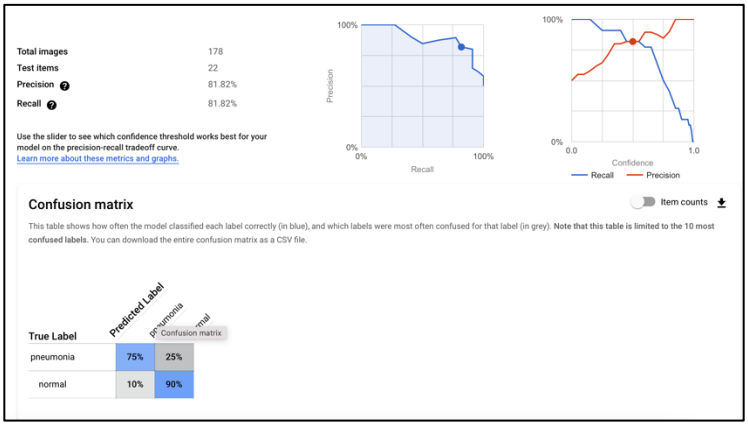
Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	400 data images used for training: - 100 images for normal cases - 300 images for pneumonia cases 40 data images for testing.

Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	The model classified normal cases correctly during testing 100% of the time; and pneumonia cases was classified correctly 97% times. Effects of unbalanced data in the confusion matrix shows the model detects all normal cases correctly with no error but some pneumonia cases the model classified them incorrectly as normal cases with percentage 3%. so, it's clear that the model still does good job with tiny error in label pneumonia cases. 
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)?	Unbalanced data caused the model to achieve the same precision and recall of 97.5% each. The model will produce fewer false positive and fewer false negative.
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	Imbalanced data is problem with classification models where the classes are not represented equally. An unbalanced dataset will bias the prediction model towards a particular class.

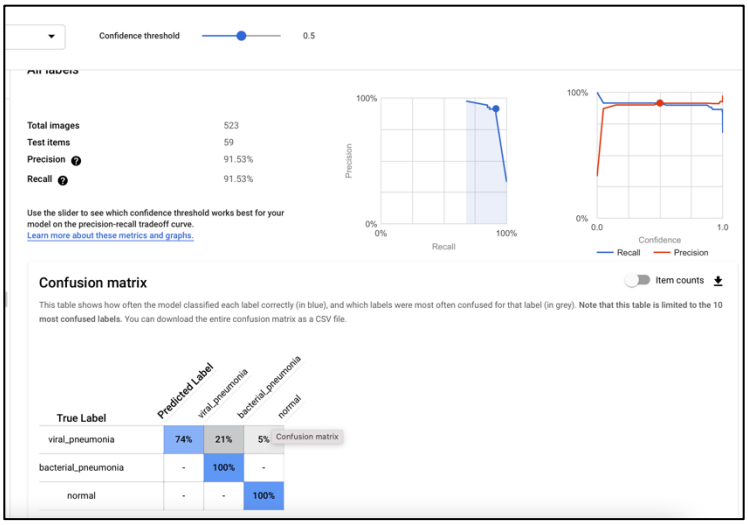
Binary Classifier with Dirty/Balanced Data

Confusion Matrix How has the confusion matrix been affected by the dirty data? Include	The dirty data affects performance of model as shown in confusion matrix. Pneumonia class are more commonly mislabeled, most

a screenshot of the new confusion matrix.	often confused with normal class. It's labeled correctly 79% of time and confused with normal class 25% of time. Also, normal class is mislabeled 10% 
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?	Dirty data caused the model to achieve the same precision and recall of 81.82% each. The binary classifier with clean/balanced data has the highest precision (97.5%) and highest recall (97.5%).
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Dirty data significantly impacts the model performance. The model struggles to find patterns among classes as the data is mixed up.

3-Class Model

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.	The viral pneumonia is the only class that the model confused classify it. For balanced data with 194 images in each class (normal, bacterial pneumonia and viral pneumonia), the confusion matrix shows that the model highly classified normal and bacterial pneumonia labels 100% of time. Viral pneumonia label is classified correctly just 74% of time and most often confused label it as bacterial pneumonia 21% of time and rarely classified it as normal cases with 5%.



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)?

The value of the precision and recall in the multiclass classification is 91.53%;
Precision and recall for the whole model are calculated foreach class individually and then taken an average of them.

Recall = TP / (TP+FN)
R (Viral) = 14/ (14+4+1) = 0.73684
R (bac) = 1
R (normal) = 1
R = (0.73684+1+1) / 3 = 0.912281 ~ 91.23%

Precision = TP / (TP/FP)
P (Viral) = 14/ (14) = 1
P (bac) = 20 / (4 + 20) = 0.833
P (normal) = 20 / (20 + 1) = 0.95
P = (1+0.833+0.95) / 3 = 0.927 ~ 93%

True Label	Predicted Label: viral_pneumonia	bacterial_pneumonia	normal
viral_pneumonia	14	4	1
bacterial_pneumonia	-	20	-
normal	-	-	20

F1 Score What is this model's F1 score?	F1 = 2*(precision * recall) / (precision + recall) = 2*(91.53% * 91.53%) / (91.53% + 91.53%) = 2*0.8377 / 1.8306 = 0.915 ~ 91.5%