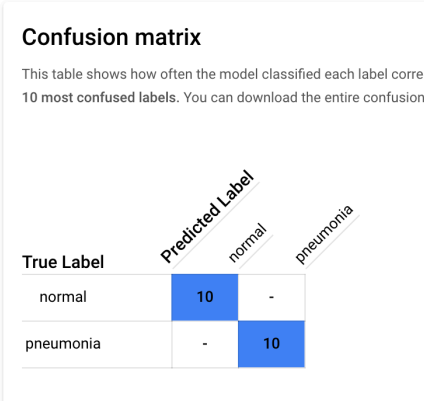
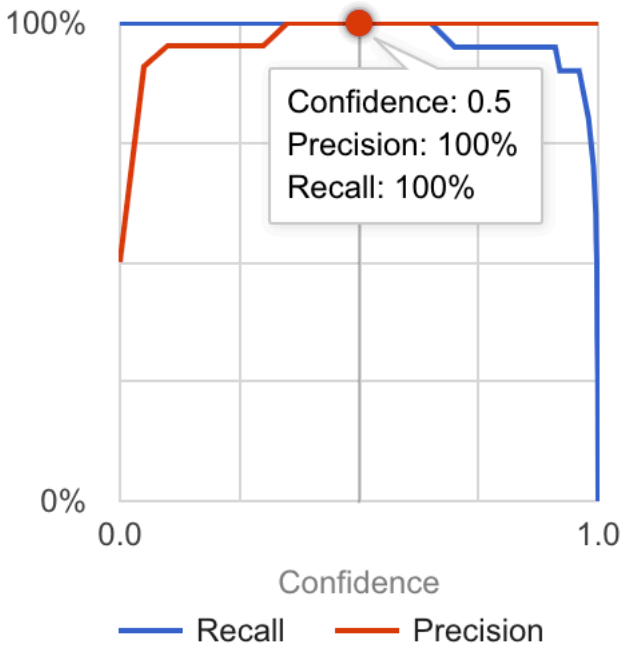
AutoML Modeling Report

SARA EL-ATEIF

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	**For Training:** **Normal :** 80 images **Pneumonia :** 80 images **For Validation:** **Normal :** 10 images **Pneumonia:** 10 images **For Testing:** **Normal :** 10 images **Pneumonia:** 10 images This means 80% of the data was used for training and 20% was used for testing (validation + test set).
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 intersection between the predicted label normal and the true label normal is called TP or **True Positive**. - The intersection between the predicted label pneumonia and the true label normal is called FP or **False Positive**.

	- The intersection between the predicted label normal and the true label pneumonia is called FN or **False Negative**. - The intersection between the predicted label pneumonia and the true label pneumonia is called TN or **True Negative**. From this confusion matrix we can see that our model has scored a 100% accuracy which means he got everything right over the reported 10 examples at least. **TP (pneumonia)** = 10 **FP (normal)** = 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:** is the fraction of relevant instances among the retrieved instances. Precision = TP/(TP+FP) **Recall:** is the fraction of retrieved relevant instances among all relevant instances. Recall = TP/(TP+FN) For a threshold of 0.5 the model achieved a 100% in both precision and recall. 

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

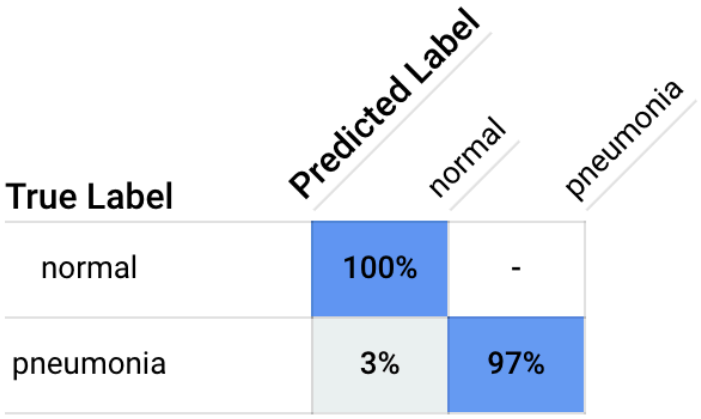
After increasing the threshold to a value of 0.85 we can see that the precision value didn't change but the value recall has decreased.
According to AutoML vision documentation : "If your score threshold is high, your model will classify fewer images, but it will have a lower risk of misclassifying images". This means that our model is getting less relevant as the FN values are getting higher. This behavior is something we would highly avoid in the medical field as it could be dangerous for the patient especially in critical conditions.

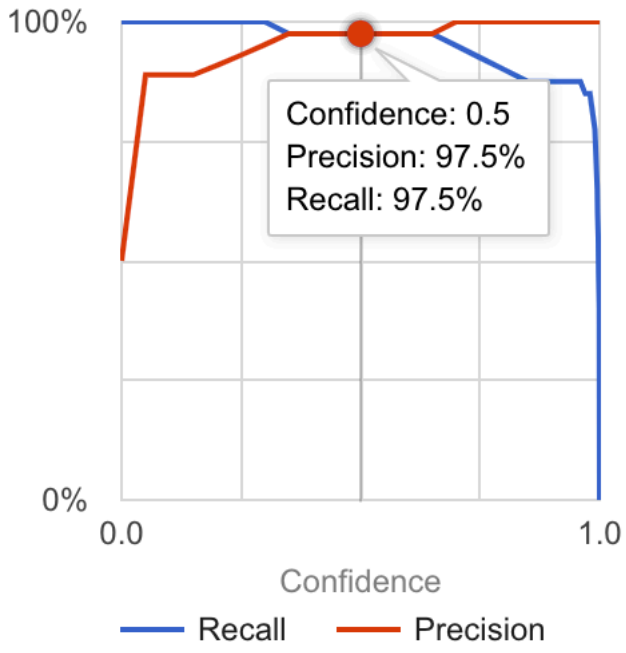
Binary Classifier with Clean/Unbalanced Data

Train/Test Split
How much data was used for training? How much data was used for testing?

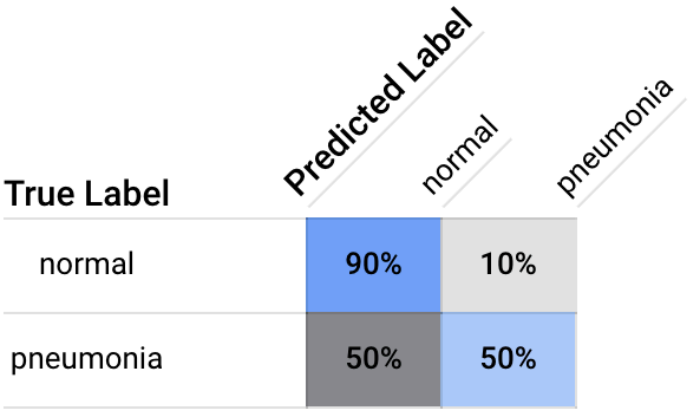
For Training:
Normal : 80 images
Pneumonia : 239 images

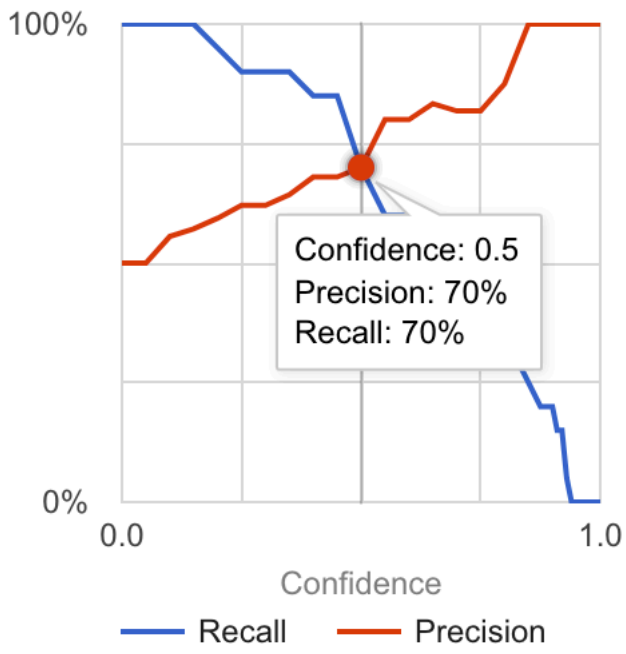
For Validation:

	Normal : 10 images **Pneumonia:** 30 images **For Testing:** **Normal :** 10 images **Pneumonia:** 30 images This means 80% of the data was used for training and 20% was used for testing (validation + test set).
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	 The perfect 100% score we had before has gone down and now we have: **TN** = 97% **FN** = 3%
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)?	Both the precision and recall values have decreased to a value of 97.5% at 0.5 threshold.

	
Unbalanced Classes From what you have observed, how do unbalanced classes affect a machine learning model?	The unbalanced classes affects the overall performance of the model. As in this case, both the precision, recall and the accuracy have decreased by 2.5 from the previously perfect 100% score.

Binary Classifier with Dirty/Balanced Data

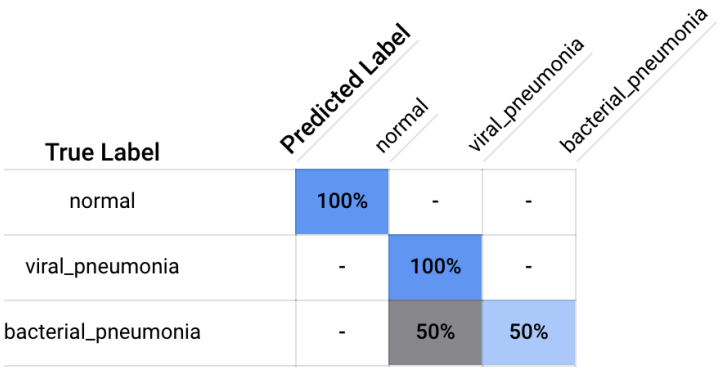
Confusion Matrix How has the confusion matrix been affected by the dirty data? Include a screenshot of the new confusion matrix.	 Before FP = 0, but now FP = 10%, what is more drastic is that FN has increased to 50% which is extremely bad for

	an ML model in the medical field. Additionally, our model has gotten worse at predicting when a patient has pneumonia.
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?	 At threshold = 0.5, the precision and recall have dramatically decreased to 70% as in 30% lower than the perfect score we first obtained which was 100%. The binary classifier that has the best precision and recall is the "Binary Classifier with Clean/Balanced Data", with precision = recall = 100%.
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Because of the dirty data the performance of the model has decreased by 30% from the 100% we achieved initially to become 70%. This means that if our data is dirty we encounter a high risk to suffer from bad performance and perhaps even find it difficult to improve it.

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 model is **most likely to confuse** between the viral_pneumonia and bacterial_pneumonia classes.

The model is **most likely to get right** the normal class.

To improve the performance of the model I might either tweak the model hyperparameters, train it some more or/and increase the data.

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)?

At threshold = 0.5 :
Precision = Recall = 83.33%

Precision = **[**Precision (normal) + Precision (viral_pneumonia) + Precision (bacterial_pneumonia)**] /3**

Recall = **[**Recall (normal) + Recall (viral_pneumonia) + Recall (bacterial_pneumonia)**]/3**

F1 = **[**F1 (normal) + F1 (viral_pneumonia) + F1 (bacterial_pneumonia)**]/3**

Precision(X) = TP(X) / (TP(X) + FP(X))

Recall(X) = TP(X) / (TP(X) + FN(X))

F1(X) = 2*Precision(X)*Recall(X)/(Precision(X)+Recall(X))

		Predicted %		
		Normal	**Viral**	**Bacterial**
Actual	**Normal**	Pnn=100	Pvn=0	Pbn=0
	Viral	Pnv=0	Pvv=100	Pbv=0
	Bacterial	Pnb=0	Pvb=50	Pbb=50

Normal :
TN = 100 + 0 + 50 + 50 = 200

	FP = 0 + 0 = 0 FN = 0 TP = 100 Precision = 100/100+0 = 1 Recall = 100/100+0 = 1 F1 = 2*1*1/(1+1) = 1 **Viral Pneumonia:** TN = 100 + 0 + 0 + 50 = 150 FP = 0 + 50 = 50 FN = 0 TP = 100 Precision = 100/(100+50) = 0.66 Recall = 100/100+0 = 1 F1 = 2*0.66*1/ (0.66 + 1) = 0.79 **Bacterial Pneumonia:** TN = 100 + 0 + 100 + 0 = 200 FP = 0 + 0 = 0 FN = 50 TP = 50 Precision = 50/50+0 = 1 Recall = 50/(50+50) = 0.5 F1 = 2*1*0.5/(1+0.5) = 0.66 **Precision =** (1 + 0.66 + 1)/3 = ~ **0.87** **Recall =** (1 + 1 + 0.5)/3 = ~ **0.83**
F1 Score What is this model's F1 score?	**F1 = [**F1 (normal) + F1 (viral_pneumonia) + F1 (bacterial_pneumonia)**]/3** **F1(X) =** 2*Precision(X)*Recall(X)/(Precision(X)+Recall(X)) **F1 score =** 2*0.87*0.83/(0.87+0.83) = ~ **0.85**