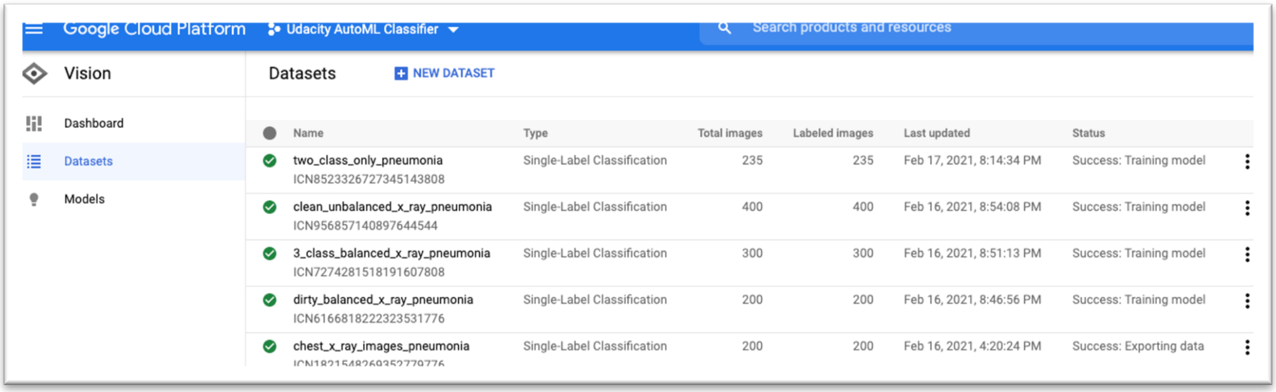
AutoML Modeling Report

Hussain Al-Balhareth

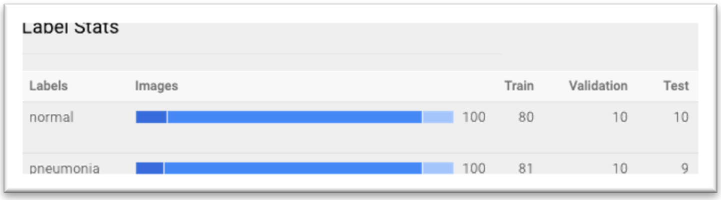


Name	Type	Total images	Labeled images	Last updated	Status
two_class_only_pneumonia ICN8523326727345143808	Single-Label Classification	235	235	Feb 17, 2021, 8:14:34 PM	Success: Training model
clean_unbalanced_x_ray_pneumonia ICN9568571408976445444	Single-Label Classification	400	400	Feb 16, 2021, 8:54:08 PM	Success: Training model
3_class_balanced_x_ray_pneumonia ICN7274281518191607808	Single-Label Classification	300	300	Feb 16, 2021, 8:51:13 PM	Success: Training model
dirty_balanced_x_ray_pneumonia ICN6166818222323531776	Single-Label Classification	200	200	Feb 16, 2021, 8:46:56 PM	Success: Training model
chest_x_ray_images_pneumonia ICN1821548269352779776	Single-Label Classification	200	200	Feb 16, 2021, 4:20:24 PM	Success: Exporting data

Figure: AutoML five trained models

Binary Classifier with Clean/Balanced Data

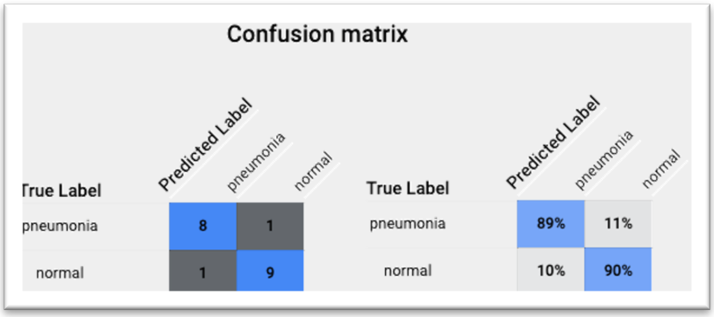
Train/Test Split
How much data was used for training? How much data was used for testing?



Label Stats

Labels	Images	Train	Validation	Test
normal	100	80	10	10
pneumonia	100	81	10	9

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?

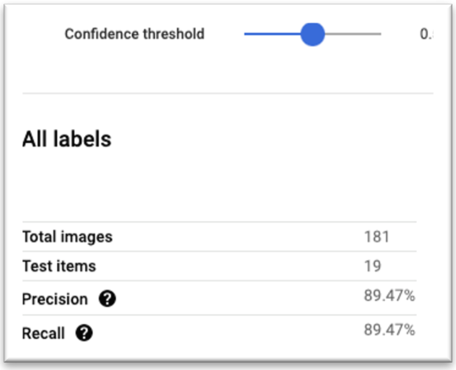
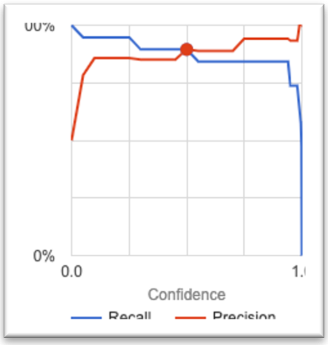


Confusion matrix

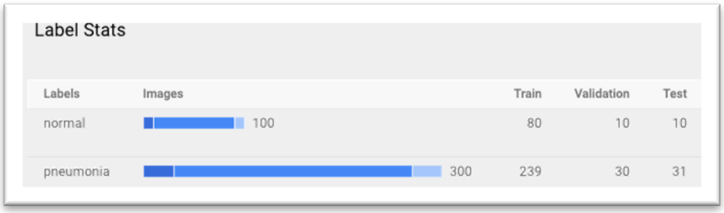
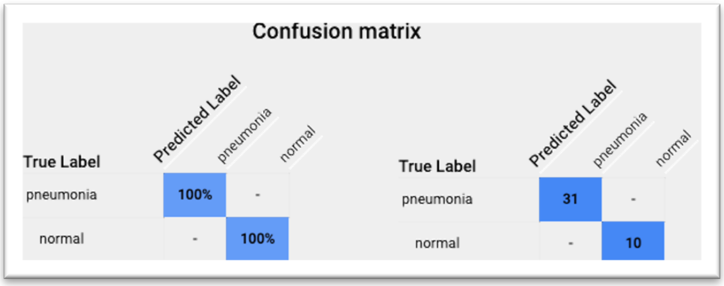
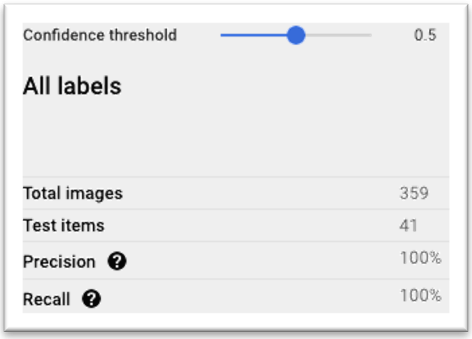
True Label \ Predicted Label	pneumonia	normal
pneumonia	8	1
normal	1	9

True Label \ Predicted Label	pneumonia	normal
pneumonia	89%	11%
normal	10%	90%

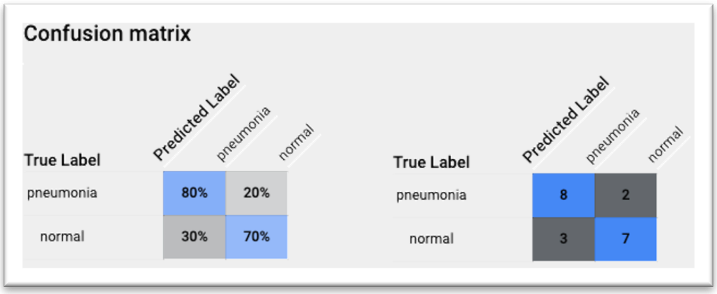
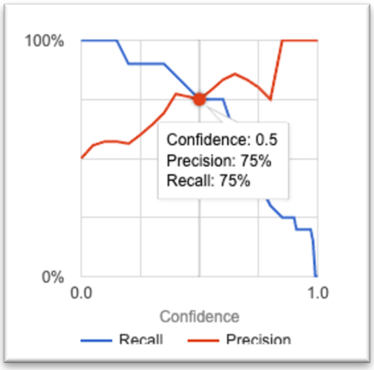
The blue (diagonal) values refer to true predictions (positive or negative) while off-diagonal (grey) refers to false predictions (positive or negative). For example, in

	the first row, a pneumonia (positive) case was predicted correctly (TP) 8 times (89% of the time) and was only once falsely predicted as normal (negative) or 11% of the time. With regard to the normal (negative) case, it was predicted correctly (TN) 8 times (90% of the time) and was only once falsely predicted (FP) as pneumonia 10% of the time.
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 true predictions over total predications. While recall measures true predictions over total ground truth.  Confidence threshold 0. All labels Total images 181 Test items 19 Precision 89.47% Recall 89.47%
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	 Precision score is proportional to confidence and vice versa with recall case. For example, when the model predicts an image as normal with 60% confidence and 40% confidence, then in this case the outcome prediction would be normal and so on.

Binary Classifier with Clean/Unbalanced Data

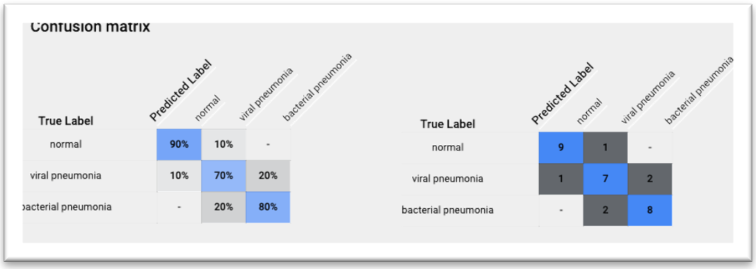
Train/Test Split How much data was used for training? How much data was used for testing?	 Label Stats Labels · Images · Train · Validation · Test normal · 100 · 80 · 10 · 10 pneumonia · 300 · 239 · 30 · 31
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	 Confusion matrix True Label · Predicted Label: pneumonia · normal pneumonia · 100% · - normal · - · 100% True Label · Predicted Label: pneumonia · normal pneumonia · 31 · - normal · - · 10 Yes, it improved to 100%.
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)?	 Confidence threshold 0.5 All labels Total images 359 Test items 41 Precision 100% Recall 100% It (unexpectedly) increased to 100%.
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	It weirdly gets higher scores probably due to higher exposure to training. Normally, such unbalance could create a bias or tendency in the model towards one prediction over another.

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.	 Confusion matrix True Label — Predicted Label: pneumonia, normal pneumonia: 80%, 20% normal: 30%, 70% True Label — Predicted Label: pneumonia, normal pneumonia: 8, 2 normal: 3, 7 It gets worse.
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?	 It became very sensitive to confidence level with overall decrease in the performance by 15% to 75%.
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	It impacts the results negatively.

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.

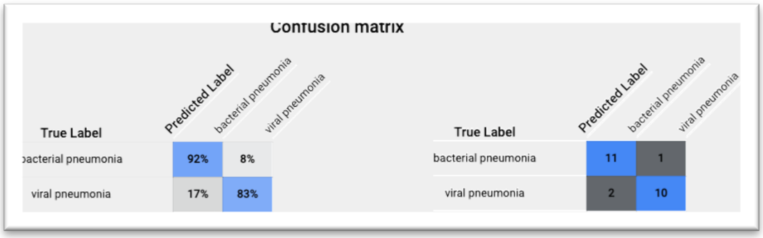


Confusion matrix

True Label	Predicted Label normal	viral pneumonia	bacterial pneumonia
normal	90%	10%	-
viral pneumonia	10%	70%	20%
bacterial pneumonia	-	20%	80%

True Label	Predicted Label normal	viral pneumonia	bacterial pneumonia
normal	9	1	-
viral pneumonia	1	7	2
bacterial pneumonia	-	2	8

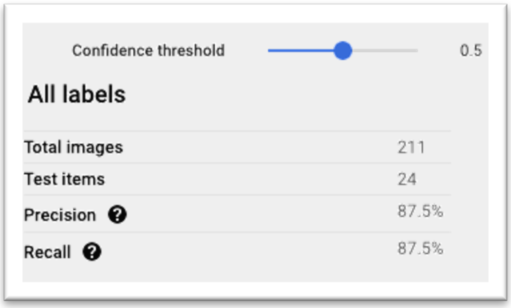
Viral pneumonia is most likely to confuse with least true prediction score (70%). While normal is most likely to get right with the highest prediction score (90%). To remedy that, I tried another model with a two-class pneumonia dataset (viral or bacterial) and got the following result:



Confusion matrix

True Label	Predicted Label bacterial pneumonia	viral pneumonia
bacterial pneumonia	92%	8%
viral pneumonia	17%	83%

True Label	Predicted Label bacterial pneumonia	viral pneumonia
bacterial pneumonia	11	1
viral pneumonia	2	10



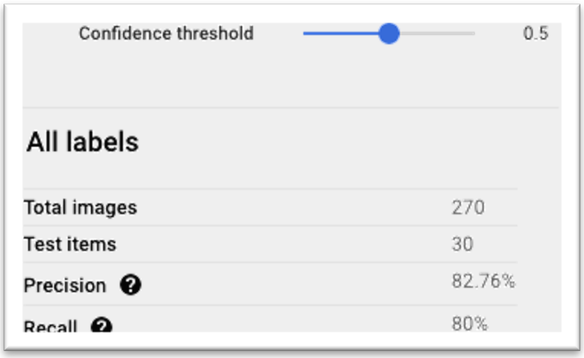
Confidence threshold 0.5

All labels

Total images	211
Test items	24
Precision	87.5%
Recall	87.5%

So, the overall precision and recall scores were both enhanced, e.g. increased by 7.5%.

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)?



Confidence threshold 0.5

All labels

Total images	270
Test items	30
Precision	82.76%
Recall	80%

$$precision = \frac{\frac{9}{10} + \frac{7}{10} + \frac{8}{10}}{3} = 0.8$$

	$recall = \frac{\frac{9}{10} + \frac{7}{10} + \frac{8}{10}}{3} = 0.8$
F1 Score What is this model’s F1 score?	$F1 = 2 \times \frac{82.76\% * 80\%}{82.76\% + 80\%} = 81.35\%$