AutoML Modeling Report

Dixon Liang

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	A total of 269 images was used for the training. 30 images were used 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 contain a pair of the model classification and the true classification. In this model, every object was classified correctly in the model. The true positive rate for pneumonia class was 100% and the false positive rate for the normal class 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 measures the ratio of true positives to all the predicted positives. Recall measures the ratio of true positives to the actual positives. Both measures were 1 (100%) for the model at a score threshold of 0.5.
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	When the score threshold increases, the recall decreases as the confidence interval needs to be higher for the model to make classification. Thus, some of the true positives will be missed as there is not confidence to make a call. The precision will stay the same since the model would have correctly classified anything with a higher confidence anyways.

All labels

Total images	269
Test items	30
Precision ❓	100%
Recall ❓	100%

Use the slider to see which confidence threshold works best for your model on the precision-recall tradeoff curve.
Learn more about these metrics and graphs.

Confusion matrix

⬜ Item counts ⬇
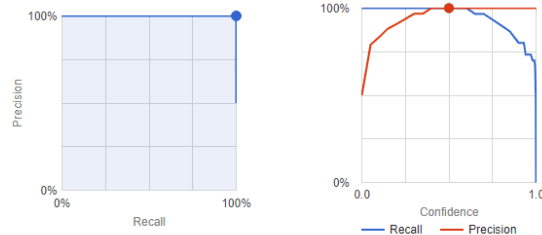
This table shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in gray). Note that this table is limited to the 10 most confused labels. You can download the entire confusion matrix as a CSV file.
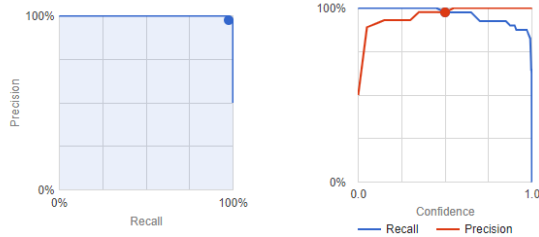
True Label \ Predicted Label	normal	pneumonia
normal	100%	-
pneumonia	-	100%

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	259 images were used for training. **1 was removed during the import due to a corrupted file.** 40 images were used for testing. 400 were imported to the model initially.
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	The confusion matrix has been affected in that 3% of the pneumonia cases have now been misclassified as "normal".
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)?	The precision and recall of the model have both fallen slightly (by 2.5%) each due to this misclassification.
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	Based on this exercise, an unbalanced data set will potentially bias the model and train it incorrectly in some scenarios.

All labels

Total images	359
Test items	40
Precision ❓	97.5%
Recall ❓	97.5%

Use the slider to see which confidence threshold works best for your model on the precision-recall tradeoff curve.
Learn more about these metrics and graphs.

Confusion matrix   ⬜ Item counts ⬇

This table shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in gray). **Note that this table is limited to the 10 most confused labels.** You can download the entire confusion matrix as a CSV file.

True Label	Predicted Label normal	pneumonia
normal	100%	-
pneumonia	3%	97%

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.	The confusion matrix looks a lot less clean than before. There are a lot more misclassifications in the form of false positives and true negatives. In fact it is hard to say that the model has done a good job at all, outside of classifying "normal" cases.
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?	Both the model's precision and recall have been adversely affected by the dirty data (both falling to about 65% from near 100%). Of the binary classifiers, the balanced/clean data has performed the best in both precision and recall.
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	From what I have observed, the dirty data is the most detrimental to model training.

All labels

Total images	180
Test items	20
Precision ❓	65%
Recall ❓	65%

Use the slider to see which confidence threshold works best for your model on the precision-recall tradeoff curve.
Learn more about these metrics and graphs.
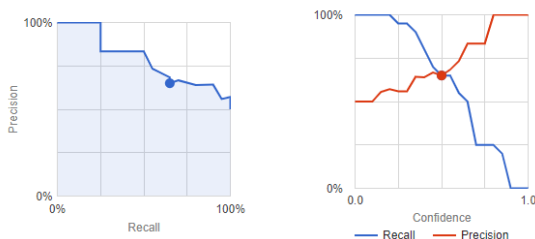
Confusion matrix

This table shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in gray). Note that this table is limited to the 10 most confused labels. You can download the entire confusion matrix as a CSV file.

Item counts

True Label \ Predicted Label	pneumonia	normal
pneumonia	50%	50%
normal	20%	80%

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.	In the 3-class confusion matrix, the model was able to classify all cases of viral and bacterial pneumonia correctly. However, in 10% of the normal cases, it mis-classified as bacterial pneumonia. Given that the model is mistaking some of the normal cases, I would want to take a look at what the possible similarities between the two – in particular, the normal cases that were misclassified which may have caused the model to classify as bacterial.
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 model's precision and recall are both 96.67% respectively. In this case, precision is calculated by combining all the true positives divided by all the true positives and false positives. Since we have three classes, it is the sum of the three true positives divided by all the positives for each class that were generated by the model (including false negatives). Recall is all the true positives divided by all the true positives and false negatives. Similarly, this will be the sum of all of the three true positives divided by the sum of all three of the true positives and all three of the false negatives for each class.
F1 Score What is this model's F1 score?	The model's F1 score is 2 * 0.9667 * 0.9667 / (0.9667 + 0.9667) = 0.9667.

All labels

Total images	269
Test items	30
Precision ❓	96.67%
Recall ❓	96.67%

Use the slider to see which confidence threshold works best for your model on the precision-recall tradeoff curve.
Learn more about these metrics and graphs.
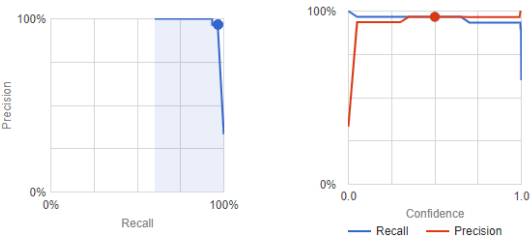
Confusion matrix

Item counts ⬇

This table shows how often the model classified each label correctly (in blue), and which labels were most often confused for that label (in gray). **Note that this table is limited to the 10 most confused labels.** You can download the entire confusion matrix as a CSV file.

True Label \ Predicted Label	viralpneumonia	normal	bacterialpneumonia
viralpneumonia	100%	-	-
normal	-	90%	10%
bacterialpneumonia	-	-	100%