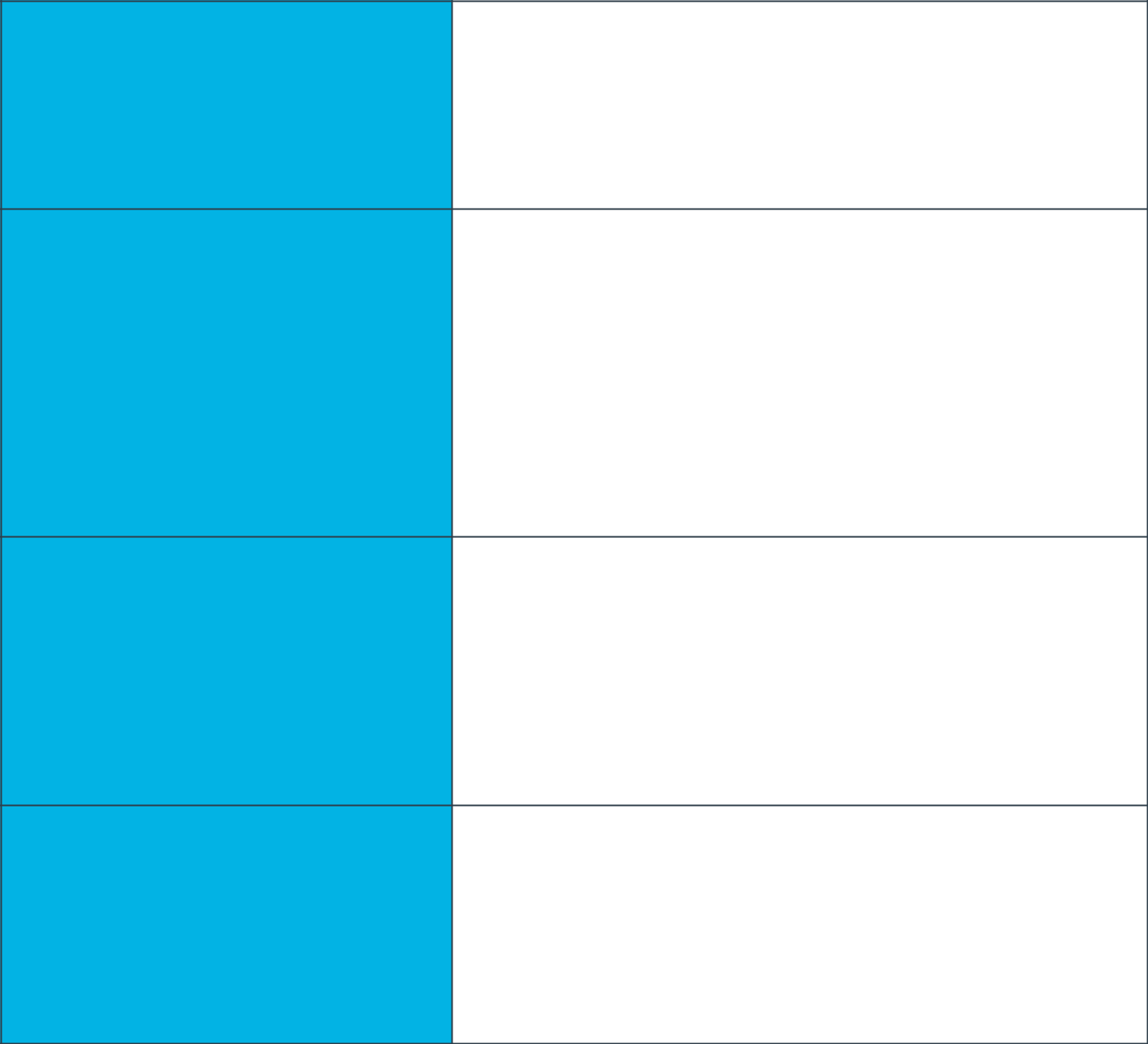
AutoML Modeling Report

*Nagesh Waghmare*



**Binary Classifier with Clean/Balanced Data**



**Train/Test Split**

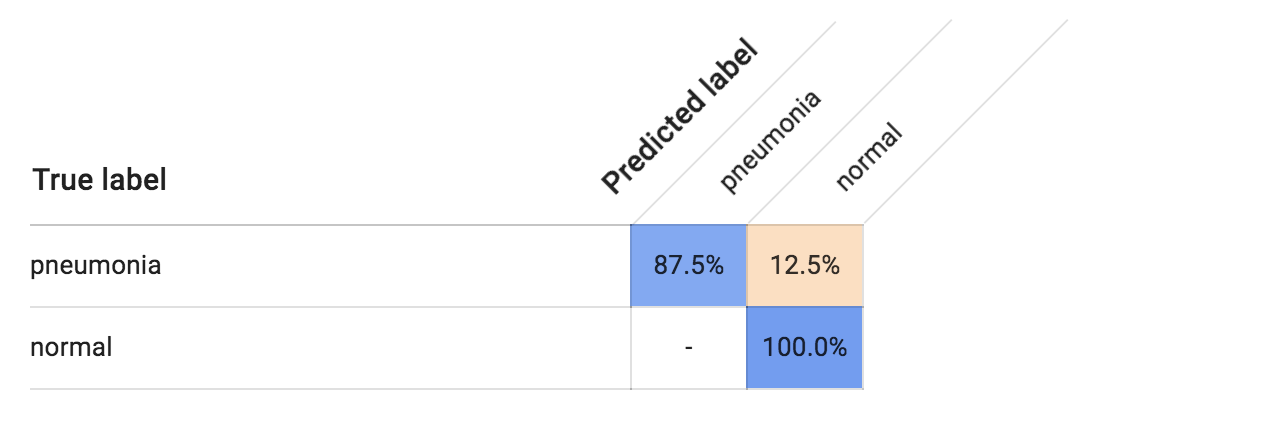
There are 300 images of pneumonia and 299 images of normal in the dataset. Out of them, 476 images were used in training, 67 in validation and 56 in test.

How much data was used for

training? How much data was used

for testing?

Confusion matrix describes the performance of classification model on test data. This table shows how often the model classified each label correctly, and which labels were most often confused for that label. The four cells represent TP, TN, FP, FN. The TP for pneumonia class is 87.5% and FP rate for normal class is 12.5%



**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?

**Precision & Recall**

Precision tells us what portion of positive identifications are actually correct. A high precision model produces fewer false positives. On the other hand, recall tells us what portion of actual positives was identified correctly. A high recall model produces fewer false negatives. The model achieved a precision of 92.9% and recall of 92.9%.

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)?

**Score Threshold**

When you increase the score threshold, the precision goes up and recall seems to decrease. This is because when you increase the score threshold you want to be more confident when you make a prediction. Hence by increasing the score threshold, your will classify fewer images but it will have lower risk of misclassifying the images.

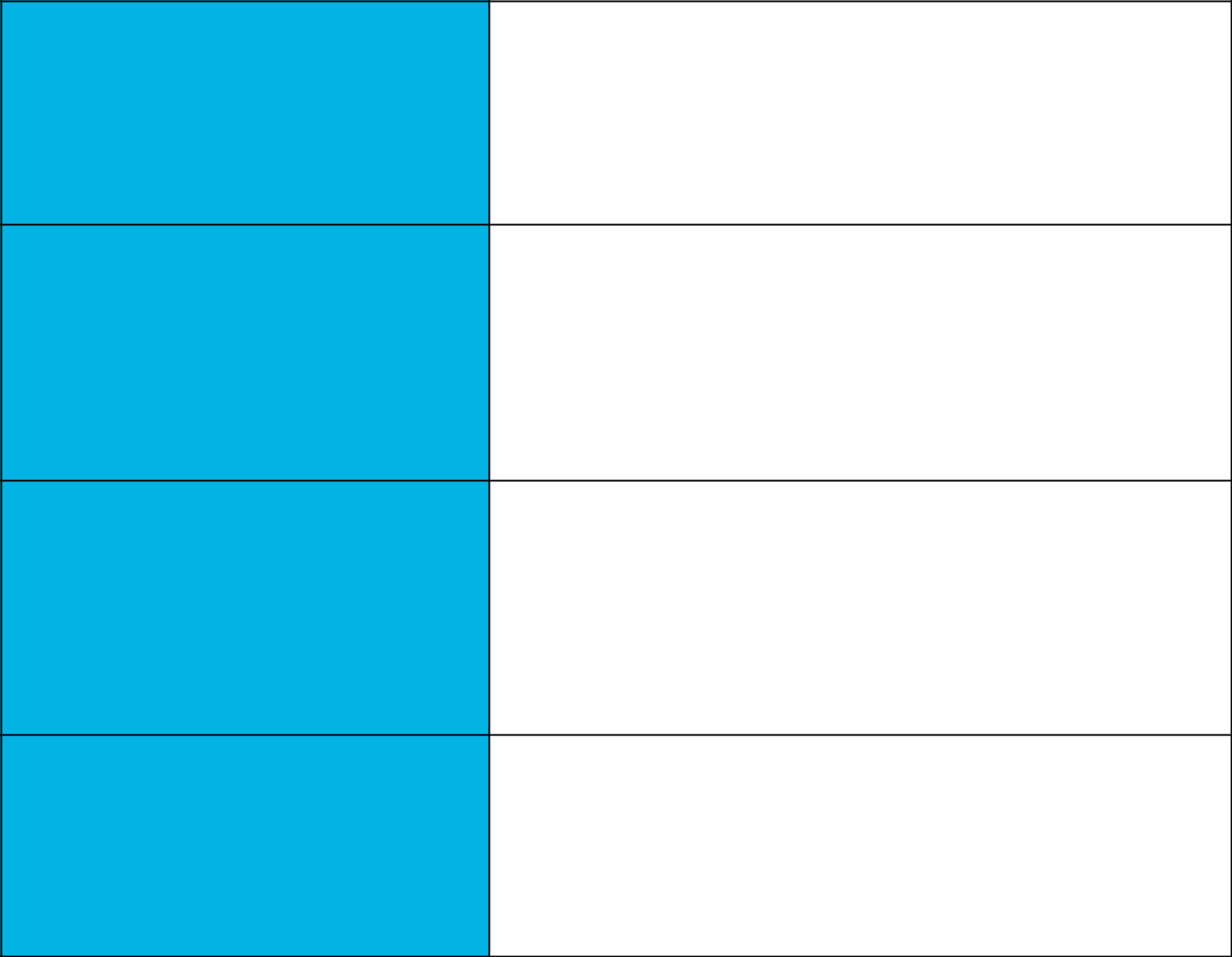
When you increase the score

threshold, what happens to

precision? What happens to recall?

Why?

**Binary Classifier with Clean/Unbalanced Data**



**Train/Test Split**

A total of 399 images were present in dataset. Out of them 304 were used in training, 42 in validation and 53 in test.

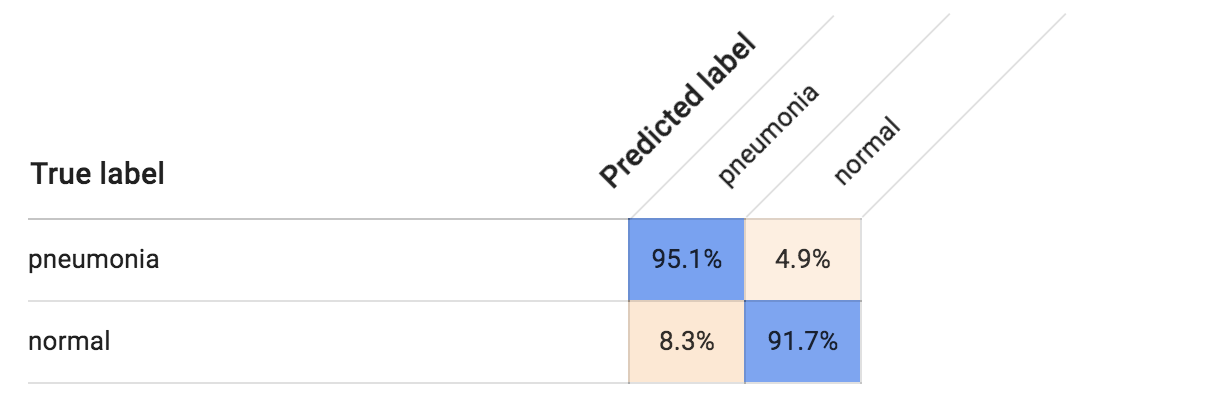
How much data was used for

training? How much data was used

for testing?

**Confusion Matrix**

The TP % went up for pneumonia. Earlier all the normal images were classified correctly. But now there seems to be a slight dip in % to 91.7%.



How has the confusion matrix

been affected by the unbalanced

data? Include a screenshot of the

new confusion matrix.

**Precision & Recall**

How have the model’s precision

Both precision and recall went up to 94.3%

and recall been affected by the

unbalanced data? (Report the

values for a score threshold of 0.5.)

Unbalanced data introduces bias. Model will have a bias towards predicting the label that has more data in training. As you can see from the results, we did not change the no of images for pneumonia but decreasing the no of normal images lead to increase in no of TP for pneumonia. As a test, I took 10 images, 5 each from normal and pneumonia and saw how well model was performing. On pneumonia images, the model did really well as it got all 5 images right with very high confidence. Whereas on normal images it classified 3 out of 5 images correctly. But the confidence on correctly classified images was not high. So it achieved an accuracy of 80%.

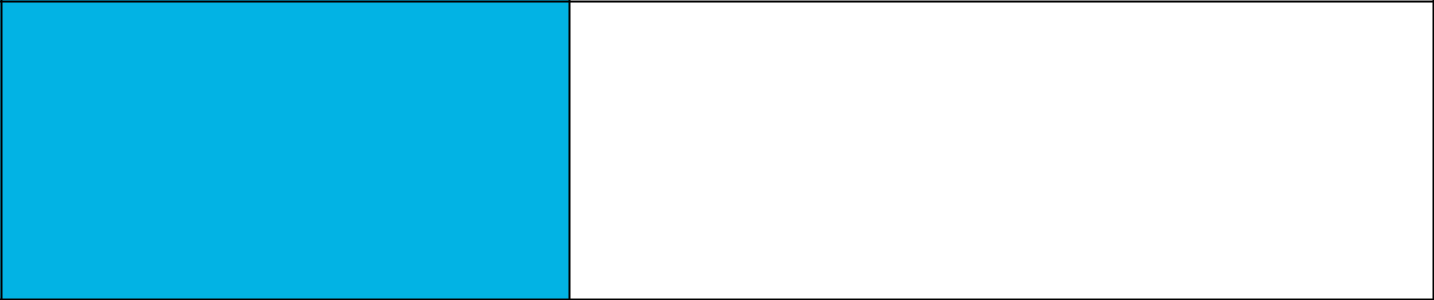
**Unbalanced Classes**

From what you’ve observed, how

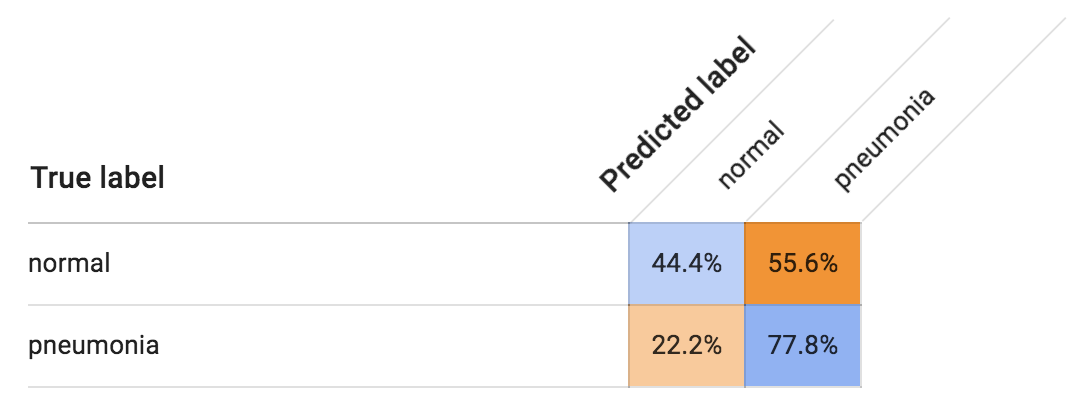
do unbalanced classes affect a

machine learning model?

**Binary Classifier with Dirty/Balanced Data**



As you see from the confusion matrix, the model total confuses in predicting and performs really bad.



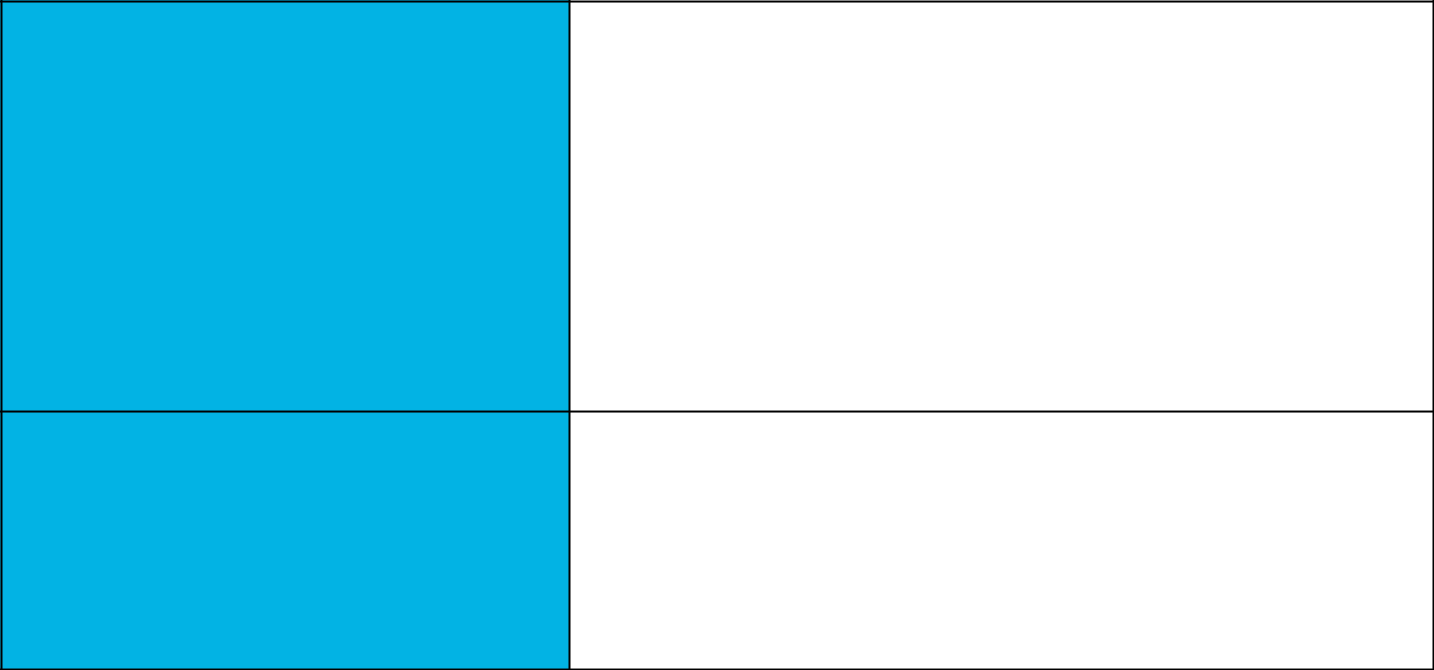
**Confusion Matrix**

How has the confusion matrix

been affected by the dirty data?

Include a screenshot of the new

confusion matrix.

**Precision & Recall**

How have the model’s precision

The precision and recall values went to a very low value of 61.1%. The highest precision and recall values were observed with clean and unbalanced data.

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**

Machine learning model clears struggles to find patterns among classes as the data is mixed up. Model sees same patterns in both labels and hence performs poorly.

From what you’ve observed, how

do dirty data affect a machine

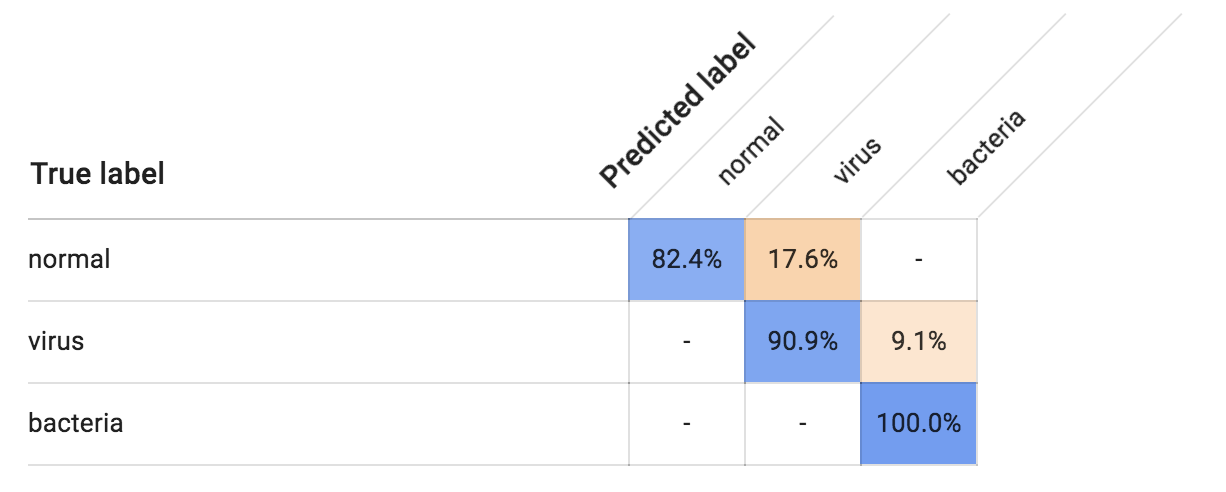
learning model?

**3-Class Model**



Model is most likely to confuse with normal class. It also confuses a little bit about virus label. Model will not confuse about bacteria class as can be seen from confusion matrix.

We can add more images to each class as there only 100 images for each class now.



**Confusion Matrix**

Summarize the 3-class confusion

matrix. What classes are the model

most likely to confuse? What

class(es) is the model most likely to

get right? What might you do to try

to remedy the model’s “confusion”?

Include a screenshot of the new

confusion matrix.

Precision and recall are calculated foreach class individually and by taking an average of them you get precision and recall values for whole model. Precision for normal, virus and bacteria classes are 82.4%, 90.9%, 100%. Similarly for recall divide the value of TP by sum of the column. Hence precision and recall values for whole model are 91.7%(((82.4+90.9+100)/298)\*100.0) and 89.2% respectively.

**Precision & Recall**

What are the model’s precision and

recall? How are these values

calculated? (Report the values for a

score threshold of 0.5.)

**F1 Score**

Model’s F1 score is 0.903.

What is this model’s F1 score?