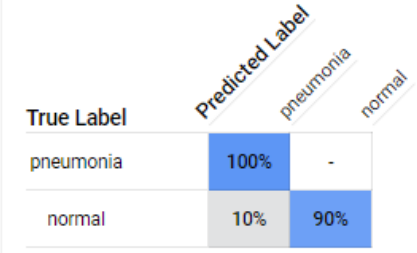
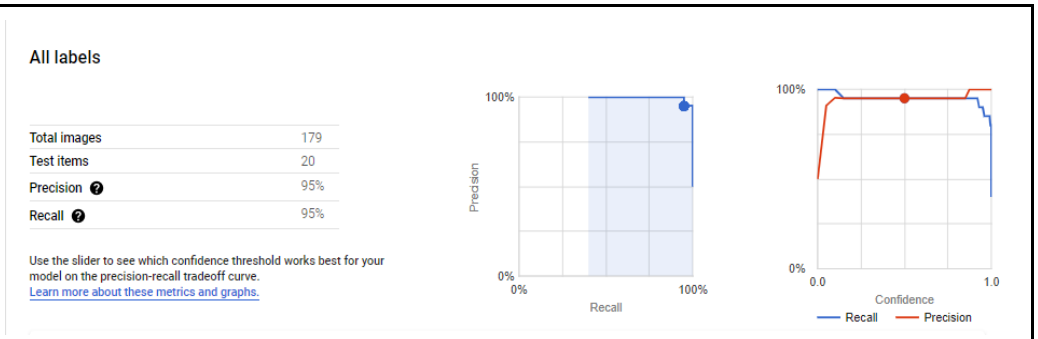
AutoML Modeling Report

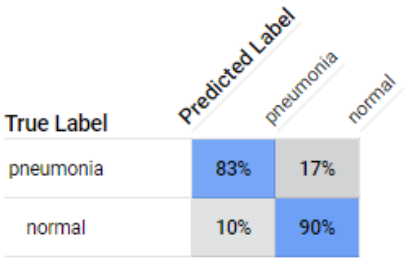
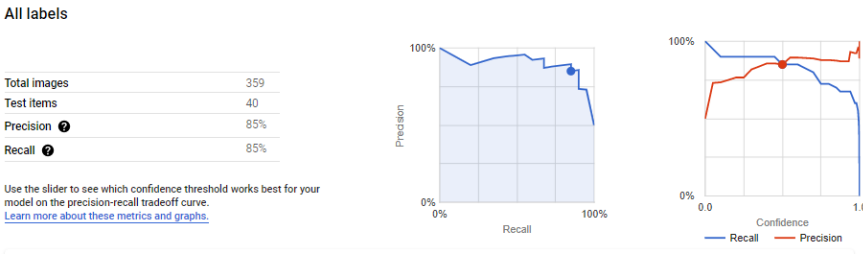
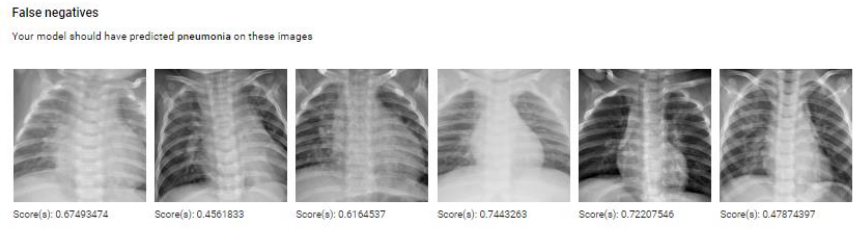
Ghadeer Alodaib

Binary Classifier with Clean/Balanced Data

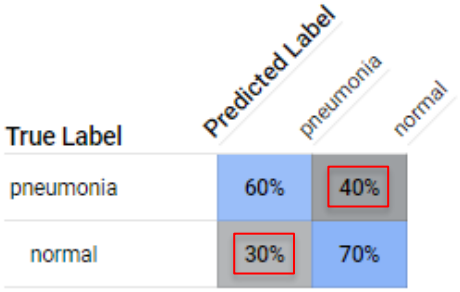
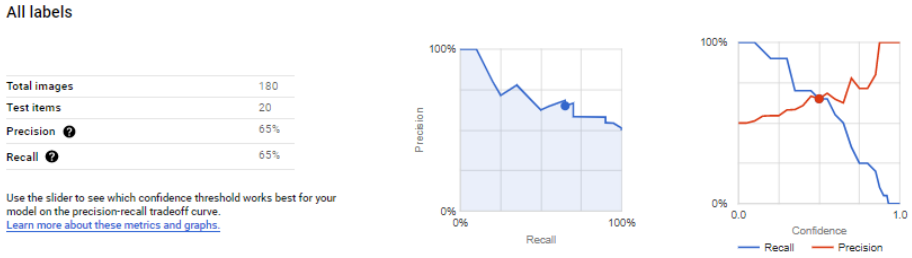
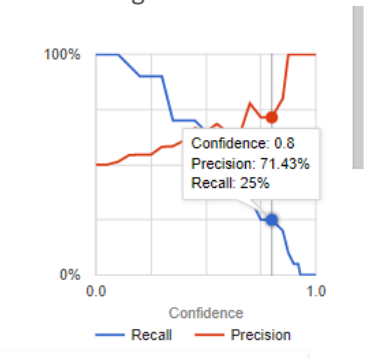
Train/Test Split How much data was used for training? How much data was used for testing?	Training: 100 images labeled as normal and 100 images labeled as pneumonia Testing \ validation: 10 images were used for testing and 10 images for validation
Confusion Matrix What do each of the cells in the confusion 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 cell describes the number of times the model classified an image in each category shown in percentage **In the first row**, the matrix shows that 100% of pneumonia images was correctly classified as "pneumonia. this section includes the correctly predicted actual positive so that TPR is 100%. On the same row a none of pneumonia images was predicted as normal is shown in the second cell (no FN for pneumonia class) **In the second row,** 10% of normal images was classified as "pneumonia". which tells that this number of normal images were misclassified and that FPR is 10% in the next cell, 90% of normal images were classified correctly as "Normal" In the pneumonia actual, TPR for this class is 90% 
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)?	The **precision** is the ratio of true positives to predicted positives. It measures the number of images the model predicted correctly in its class compared to all predictions **Recall** is the ratio of true positives predictions to actual positives. The Model achieved a precision and recall of **95%**

	
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	When the threshold is increased above 0.8 the precision value will keep increasing but the recall value will decrease. Because precision takes in consideration the number of images predicted as positive and when we increase the model confidence the precision value will increase. While, for recall value is affected by the presence false negative prediction which decreases its value.

Binary Classifier with Clean/Unbalanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	**Training**: 100 images labelled as normal 300 images labelled as pneumonia **Testing \ Validation:** 10 images for testing \ 10 validation
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	The matrix shows that the model has predicted **83%** of pneumonia images and **90%** of normal images correctly in their respective classes. Also, **10%** of normal images was classified as "pneumonia" while **17%** of pneumonia images were classified 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 model achieved 85% in both metrics (85% precision, 85 % recall) at 0.5 confidence threshold. 
Unbalanced Classes From what you have observed, how do unbalanced classed affect a machine learning model?	Increased the number of false positive on normal class 5 images were incorrectly predicted as normal Increased the number of False negatives on pneumonia class the model predicted 5 images with pneumonia as normal. 

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.	Dirty data has affected the accuracy of the model The number of False Negatives has increased in each class. 40% of pneumonia images were classified as normal 30% of normal images classified as 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?	The model precision and recall have dropped to 65% at 0.5 threshold compared to the model metric when clean data was used for training. **By comparing all 3 classifiers, Clean Balanced** classifier has performed well among all three with the highest precision and recall value of **95%** Which indicates the importance of training the model with balanced and clean dataset. 
Dirty Data From what you have observed, how does dirty data affect a machine learning model?	Training the model with dirty data has affected the accuracy of the model it can be seen when we increase our confidence threshold to 0.8 the recall drops to 25% which indicates that the Ratio of true positives to actual positives is low. The model is misclassifying actual positives and predicting them as negative. 

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.

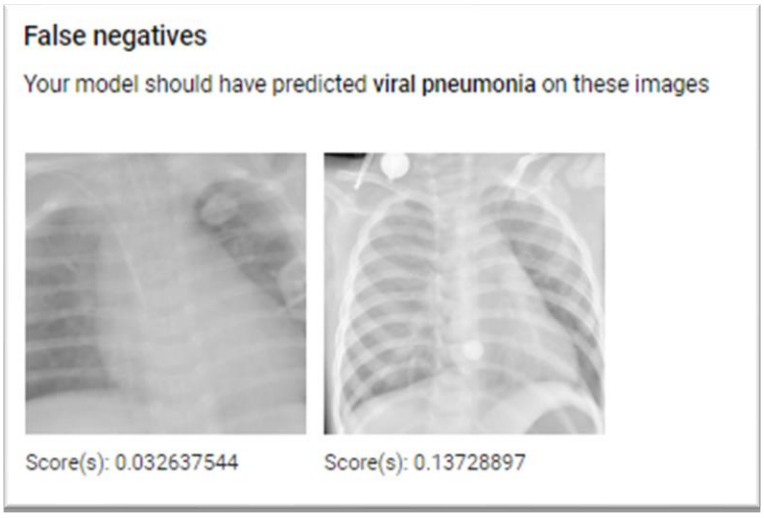
if we look at the values in the diagonal shape, we see that they mark the highest numbers in the matrix which indicates that the model has correctly predicted most of the data in its correct class.
The model predicted all images of normal class correctly.

Confusion appears in bacterial pneumonia class in predicting 10% of the class images as viral pneumonia

True Label \ Predicted Label	normal	viral pneumonia	bacterial pneumonia
normal	100%	-	-
viral pneumonia	10%	80%	10%
bacterial pneumonia	-	10%	90%

Another area of confusion is when the model has classified 10% of the viral pneumonia as normal and 10% as bacteria pneumonia

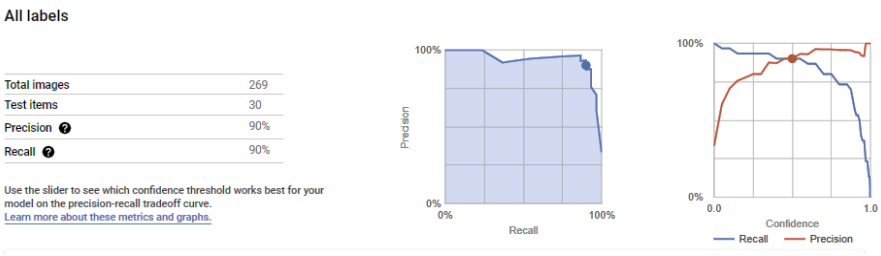
A sample of the false negative is shown below



To lessen the confusion, I might train the model on more images of viral pneumonia and select images with better quality.

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 0.5 score of thresholds
Precision: 90%
Recall: 90%



To calculate the model precision, we need 1st to calculate the **precision** to each class

P Normal = TP / TP + FP = 100 / 100 + 10 = 100 / 110 = 0.909 = 0.91

P Viral = TP / TP + FP = 80 / 80 + 10 = 0.888 = 0.89

P Bacterial = TP / TP + FP = 90 / 90 + 10 = 0.9

P Model = (P Normal + P Viral + P Bacterial) / 3 = (0.91 + 0.89 + 0.9) /3 = **0.9**

We will calculate the **Recall** in the same manner

R Normal = TP / TP + FN = 100 / 100 + 0 = 1

R Viral = TP / TP + FN = 80 / 80 + 10 + 10 = 80 / 100 = 0.8

R Bacterial = TP / TP + FN = 90 / 90 + 10 = 90 / 100 = 0.9

R Model = (R Normal + R Viral + R Bacterial) /3= (1 + 0.8 + 0.9) /3 = **0.9**

F1 Score
What is this model's F1 score?

F1 = (2 * precision * recall) / (precision + recall)

F1 = (2 * 0.9 * 0.9) / (0.9 + 0.9)

F1 = 1.62 / 1.8 = **0.9**

Thank you

Ghadeer Alodaib