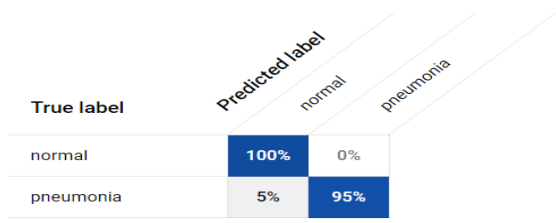


AutoML Modeling Report

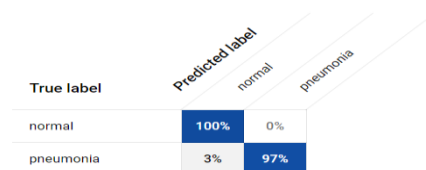


Taif Alharbi

Binary Classifier with Clean/Balanced Data

Train/Test Split How much data was used for training? How much data was used for testing?	The total number of images used in this case is 200, where 140 were used for training, 20 for validation, and 40 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?	<p>The confusion matrix evaluates the classification model on test data, indicating correct label classification and common misclassifications. The four cells represent TP, TN, FP, and FN. The pneumonia class has a 95% true positive rate, while the normal class has a 5% false positive rate.</p>  <table><tr><th rowspan="2">True label</th><th colspan="2">Predicted label</th></tr><tr><th>normal</th><th>pneumonia</th></tr><tr><th>normal</th><td>100%</td><td>0%</td></tr><tr><th>pneumonia</th><td>5%</td><td>95%</td></tr></table>	True label	Predicted label		normal	pneumonia	normal	100%	0%	pneumonia	5%	95%
True label	Predicted label											
	normal	pneumonia										
normal	100%	0%										
pneumonia	5%	95%										
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 represents the proportion of correctly identified positive cases, indicating the ability to minimize false positives. Recall, on the other hand, represents the proportion of actual positive cases that were correctly identified, indicating the ability to minimize false negatives. In this case, the model obtained a precision100% and recall of 95%, indicating high accuracy in recall minimizing false negatives.											
Score Threshold When you increase the threshold what happens to precision? What happens to recall? Why?	Raising the score threshold increases precision but decreases recall. This occurs because a higher score threshold requires greater confidence for making predictions. Consequently, fewer images are classified, but the risk of misclassification decreases.											

Binary Classifier with Clean/Unbalanced Data

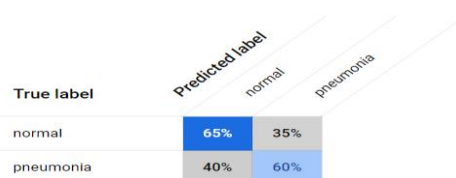
Train/Test Split How much data was used for training? How much data was used for testing?	The total number of images used in this case is 400, where 280 were used for training, 40 for validation, and 80 for testing									
Confusion Matrix How has the confusion matrix been affected by the unbalanced data? Include a screenshot of the new confusion matrix.	<p>The addition of 200 more pneumonia images in the dataset resulted in a notable change in the confusion matrix. Specifically, the true positive (TP) value increased from 95% to 97%, indicating a higher number of correct positive identifications. However, the true negative (TN) value decreased, suggesting a reduced tendency to correctly identify normal images.</p>  <table><tr><th>True label \ Predicted label</th><th>normal</th><th>pneumonia</th></tr><tr><th>normal</th><td>100%</td><td>0%</td></tr><tr><th>pneumonia</th><td>3%</td><td>97%</td></tr></table>	True label \ Predicted label	normal	pneumonia	normal	100%	0%	pneumonia	3%	97%
True label \ Predicted label	normal	pneumonia								
normal	100%	0%								
pneumonia	3%	97%								
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 value of both precision and recall are good both went up.									
Unbalanced Classes From what you have observed, how do unbalanced classes affect a machine learning model?	The results indicate that an imbalanced dataset introduces bias in the machine learning model. This bias refers to the model's inclination to classify more input objects, such as images, into the classes that have a larger representation in the training data. In this case, the model tends to classify more images as 'pneumonia' due to the larger number of training images available for that class.									

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 introduction of a balanced but mislabeled "dirty data" dataset had a significant and negative impact on the confusion matrix. The ML model's performance deteriorated across all aspects of the prediction process, as evident from the provided snapshot. This poor performance can be attributed to the presence of mislabeled and incorrect data, which resulted in the ML model making incorrect decisions.



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 precision and recall values, in this case, went to an extremely low value of 62.5%. The highest precision and recall values were observed with clean and unbalanced data.

Dirty Data

From what you have observed, how does dirty data affect a machine learning model?

When mislabeled dirty data is introduced, it creates conflicting information for the trained ML model to handle. As a consequence, the model makes multiple misclassifications due to the erroneous nature of the data, leading to its overall poor performance.

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 exhibits a higher likelihood of confusion with the normal class, and it also has some confusion with the virus label. However, there is less confusion observed with the bacteria class, as evident from the confusion matrix. To address this issue, it would be beneficial to augment the dataset by adding more images to each class, especially considering the currently limited count of 100 images per class.

True label	Predicted label		
	bacterial_pneumonia	viral_pneumonia	normal
bacterial_pneumonia	80%	15%	5%
viral_pneumonia	0%	95%	5%
normal	5%	15%	80%

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)?

In the given 3-class confusion matrix scenario, with a threshold score of 0.5, the precision and recall values are calculated to be 89.5% and 85% respectively. In the case of a 3×3 confusion matrix, precision, and recall values are initially calculated for each class individually. These individual values are then averaged to determine the overall precision and recall values.

F1 Score

What is this model's F1 score?

The model's F1 score is 0.8719.