



Covid Lung Analysis

WHY?

- Due to the Covid pandemic, healthcare providers have been stretched thin making it difficult for patients to get the care they require.
- If there was a way to create a model that could quickly identify a patient with covid that wasn't invasive, it could quicken the process for patients to get the care they need



About the Data

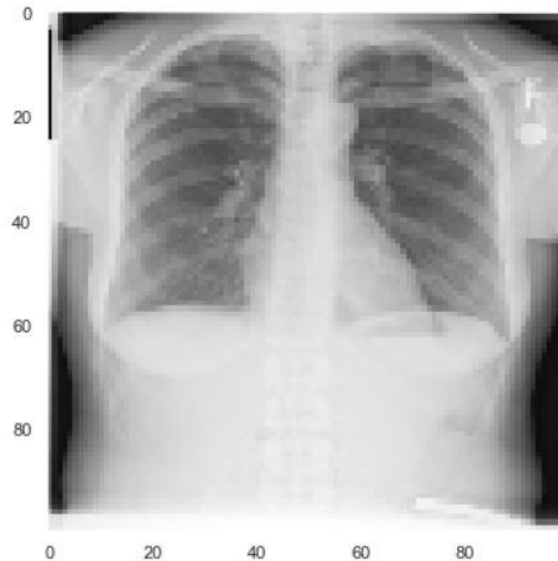
- Data originates from [Kaggle](#)
- The dataset contains lung x-ray images of 13,808 images in total.
 - 10,192 (74%) of the images are negative covid patients
 - 3,616 (26%) of the images are positive covid patients



Data wrangling

- Made the images uniform:
 - Rescaled images to (100,100)
 - Made the images grayscale

```
Image Dimension : (100, 100)  
Image Height    : 100  
Image Width     : 100  
Result          : Negative
```



Data Wrangling

- Manipulated the images using 3 different methods
 - No manipulation
 - Used a gaussian Blur filter
 - Used a Sharpen filter
- Tested them on a testing Neural Network

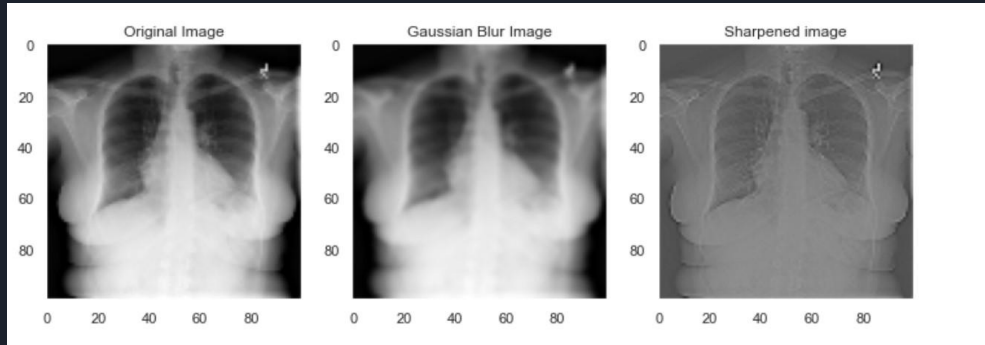
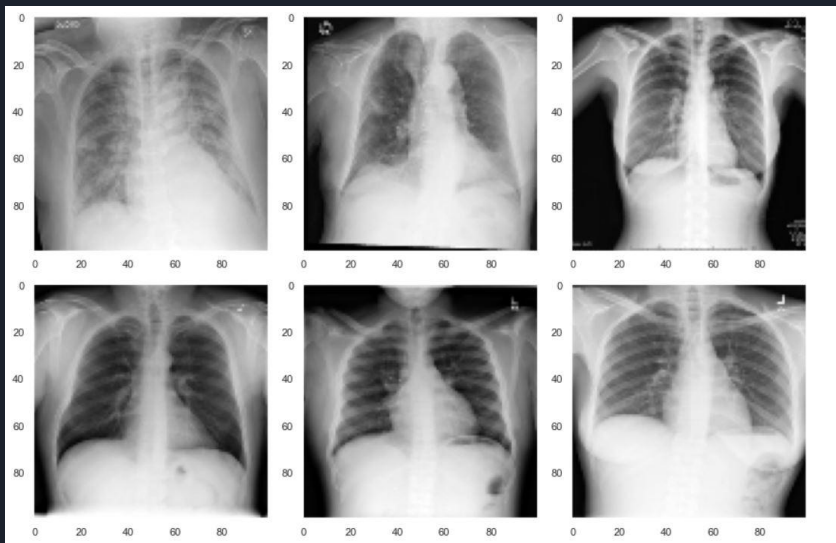


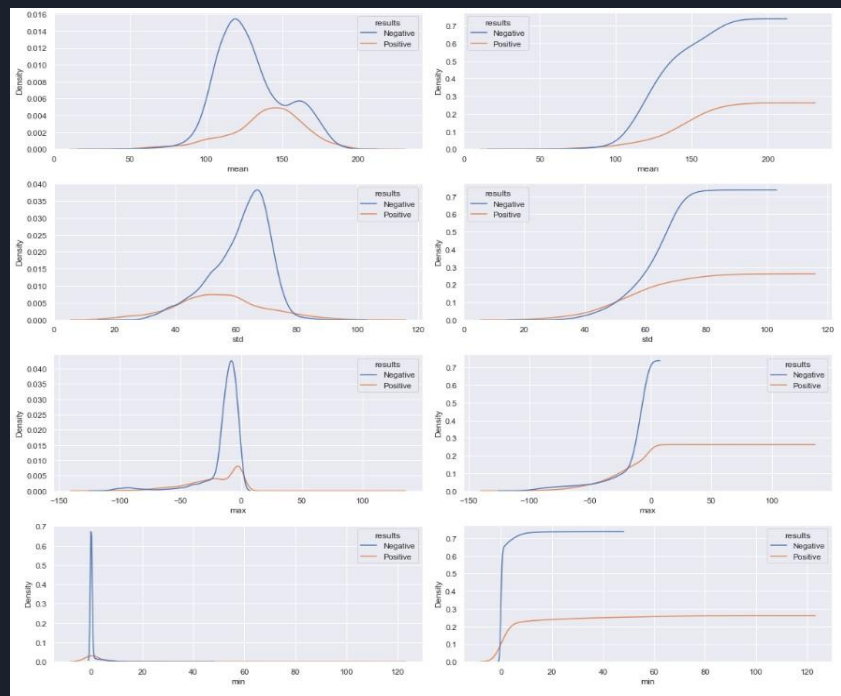
Image	Loss	Accuracy
Original	0.2962	0.86556
Gaussian Blur	0.2998	0.86218
Sharpened	0.2445	0.90418

EDA

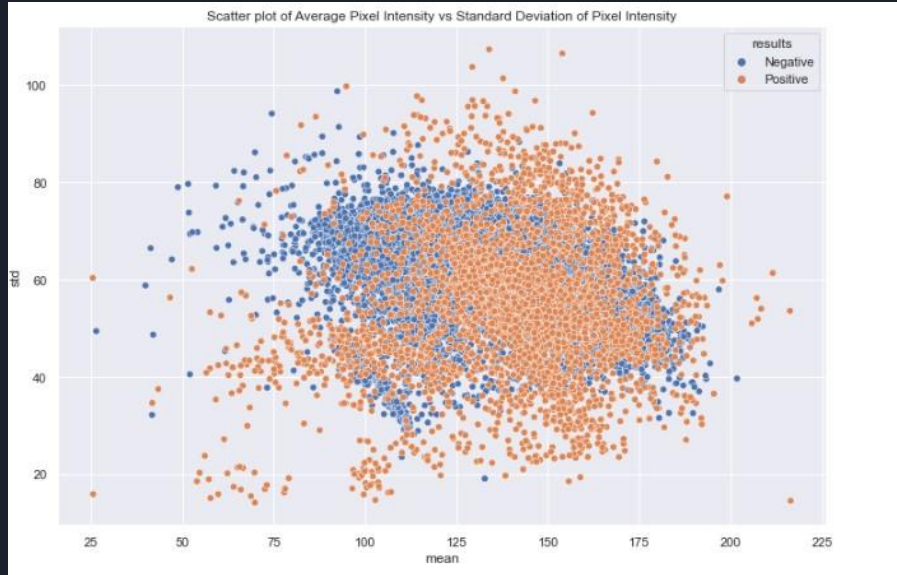
Slight differences between positive and negative Lung images



KDE plot confirms that there are potential subtle differences



EDA



- Scatterplot also shows subtle differences between positive and negative patients
- Positive patients have a slightly more concentrated cluster
- When positive patients stray from cluster, they stray further compared to negative patients



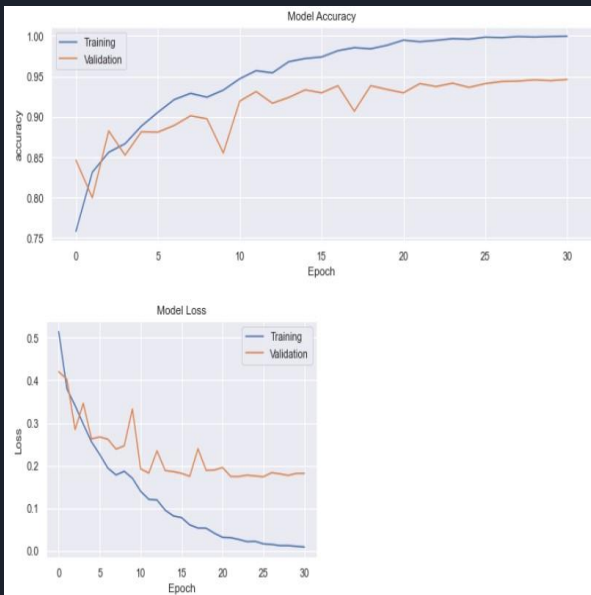
Modeling

- After deciding which image type to use, the next step was to test the hyperparameters
- Original Model
 - {units: 32, lr: 0.001, momentum: 0.9, units2: 128}
- Tuned Model
 - {units: 128, lr: 0.001, momentum: 0.6, units2: 128}

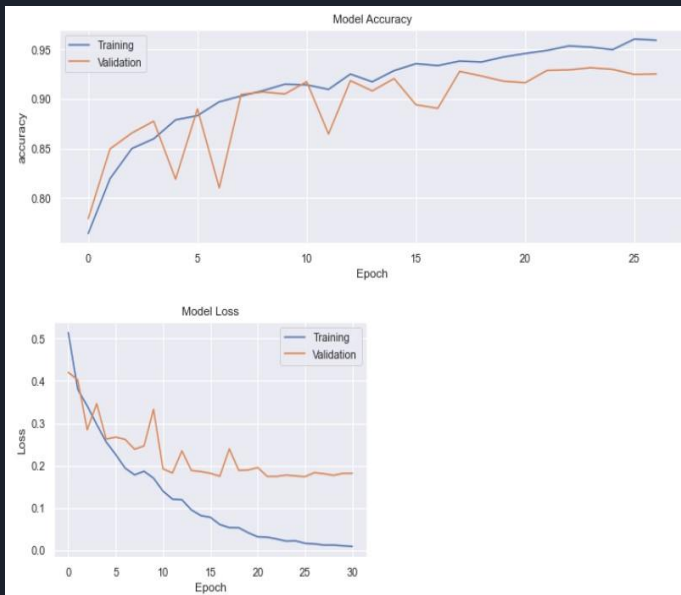
Model	Loss	Accuracy
Original	0.187	0.938
Tuned	0.181	0.932

Modeling

Original

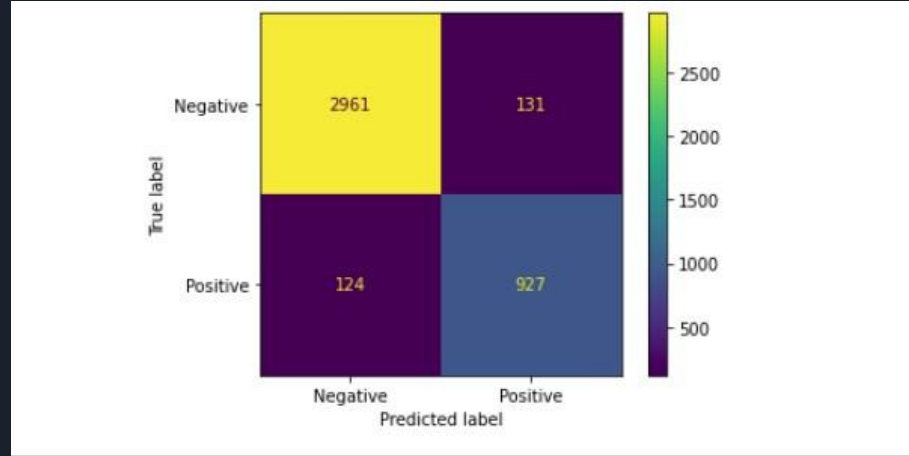


Tuned



Modeling

- Decided on the original model due to its simplicity
- Confusion Matrix shows high accuracy
- When making an incorrect prediction model tends to predict a false positive



Precision	Recall	Accuracy
0.8762	0.8820	0.9385



Conclusion

- On average there are subtle differences between a COVID positive patient and a negative patient.
- I was able to create a model with a 93.8% accuracy in predicting whether a patient is COVID positive using lung x-ray images



Future Improvements

- Spend more time in the tuning stage, adding more layers, using different layers, etc.
- Dealt with the unbalanced dataset in some way