

Predicting Covid 19 Positive Cases From Lung X Ray Images

Model Use in a Medical Context:

All data would have to be reviewed by a data science team and a medical expert panel before approval.

Metric to optimize is recall, because we want to maximize the number of model identified Covid-19 positive X-Ray images that are actually positive.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

Model Metrics Comparison Table:

Category	Model	Precision	Recall	F1	Accuracy
Covid-19	VGG16	0.93	1.00	0.96	0.97
Normal	VGG16	1.00	0.92	0.96	0.97
Viral Pneumonia	VGG16	1.00	1.00	1.00	0.97
Covid-19	ResNet50	0.92	0.92	0.92	0.85
Normal	ResNet50	1.00	0.69	0.82	0.85
Viral Pneumonia	ResNet50	0.72	0.93	0.81	0.85
Covid-19	DenseNet121	1.00	1.00	1.00	0.93
Normal	DenseNet121	1.00	0.77	0.87	0.93
Viral Pneumonia	DenseNet121	0.82	1.00	0.90	0.93

The chosen model was (VGG16) because of the following reasons:

- Covid-19 recall of 1 on unseen test data so that all positive cases are correctly identified
- Viral Pneumonia Recall of 1 on unseen test data so that all positive cases are identified
- The tradeoff is a slightly lower recall of 0.92 for Normal patients. This was the highest recall for Normal patients between the tested models. The team would have to discuss cost implications of misdiagnosis, ethical implications, etc.
- The team should take into consideration there is a limited amount of Covid-19 images and model performance should improve as we continue to gather more data
- Correct diagnosis of an x ray image is very challenging and the model can classify the image in minutes to facilitate this part of the diagnostic process
- Predictions can be obtained within minutes
- Image processing is automated by the model
- The model possesses scalability and flexibility an important feature as the data grows
- Predictions can be served from the cloud to any device

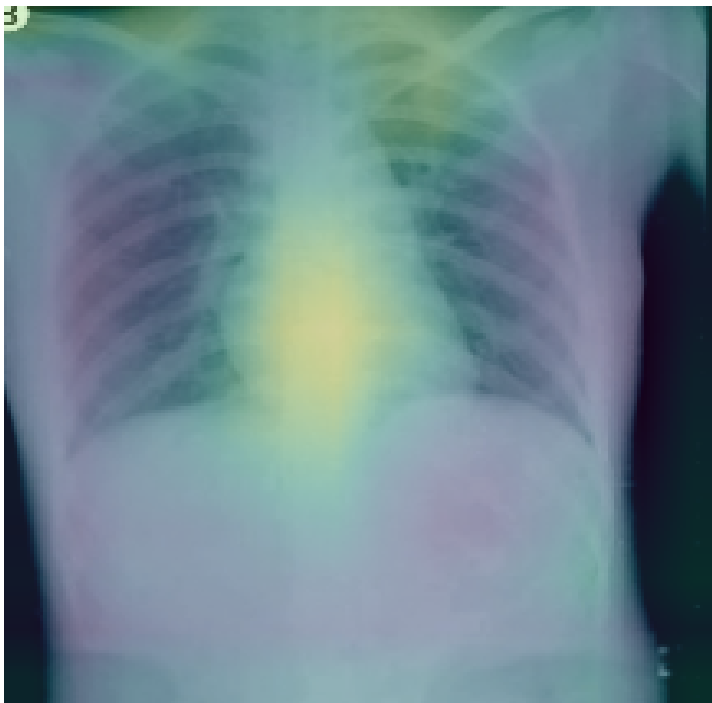
Interpretability and Further Diagnostic Assistance

Images displaying the regions where the model activations occurred can have the dual purpose of showing in a verifiable way what features the model was focusing on when identifying the correct

images and also further assist with the diagnosis regarding what areas were impacted or not impacted by the disease.

- Note areas in light yellow. These areas can be identified anatomically by the medical team. These are activation regions in the model for the Covid-19 Category

GradCAM



FURTHER STUDY

Image segmentation can be used to segregate regions of the body and through an interdisciplinary effort they can be tagged with the correct region names. This additional set of segmented images can be run side to side with activated images and superimposed to immediately see the name of the region that is being activated.