UCS1063-Introduction to Machine Learning

Report

Sreedhar V – 185001161

Sudharshan R – 185001173

Shri Charan RS – 185001147

Yeshwanth SV – 185001204

COVID-19 DETECTION USING X RAY IMAGES:

Project Goal/Problem Statement:

- Covid-19 affects the bronchitis in the lungs and causes mild deformation in the structure.
- Our project here is to analyse the regularities in deformation and decide if a patient is affected by Covid-19 or not.
- Our goal here is to construct a model which creates and analyses the features of lungs with and without Corona virus infection without human intervention.

(i) X-ray image of Chest with Covid

(ii) X-ray image of Chest without Covid

Mechanism Involved:

COVID-19 is a respiratory disease, one that especially reaches into your respiratory tract, which includes your lungs.

COVID-19 can cause a range of breathing problems, from mild to critical. Older adults and people who have other health conditions like heart disease, cancer, and diabetes may have more serious symptoms.

Here's what the new coronavirus does to your lungs:

Think of your respiratory tract as an upside-down tree. The trunk is your trachea, or windpipe. It splits into smaller and smaller branches in your lungs. At the end of each branch are tiny air sacs called alveoli. This is where oxygen goes into your blood and carbon dioxide comes out.

The new coronavirus can infect the upper or lower part of your respiratory tract. It travels down your airways. The lining can become irritated and inflamed. In some cases, the infection can reach all the way down into your alveoli.

COVID-19 is a new condition, and scientists are learning more every day about what it can do to your lungs. They believe that the effects on your body are similar to those of two other coronavirus diseases, severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS).

Doctors can see signs of respiratory inflammation on a chest X-ray or CT scan. On a chest CT, they may see something they call "ground-glass opacity" because it looks like the frosted glass on a shower door.

Necessity of the project:

The Covid -19 pandemic had a huge impact on our lifestyle. Since covid can be detected 100% only through blood tests.

But blood tests take a lot of time around 5-6 hours. On a country like India where there is huge population density, it may be hard to cope with the number of arising cases and also the doubts of covid infection.

On account of this issue the projects seems to fill in a lot of Gray area.

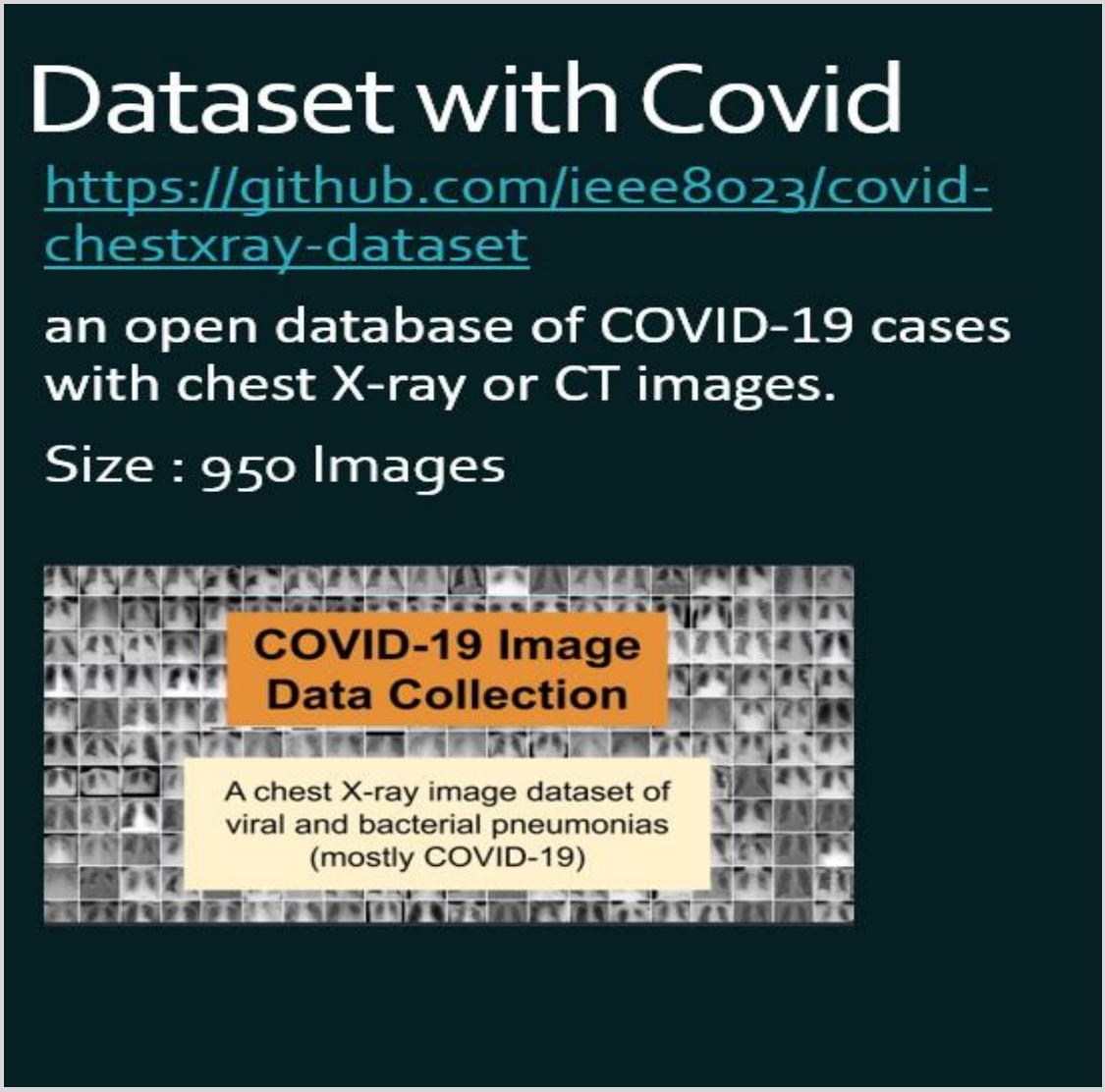
Feasibility:

The model we created had almost 97% – 98 % accuracy. On sticking to the lower end of 97% say in a group of 100 people affected with Covid and a person is not recognised to have covid and is let free without quarantine, then the person may affect the rest of the population.

On seeing it otherwise, say of 100 normal people if 3 were categorized as having covid and were quarantined with the other covid patients. Then the person is subjected to covid transmission even though it wasn't his/her fault.

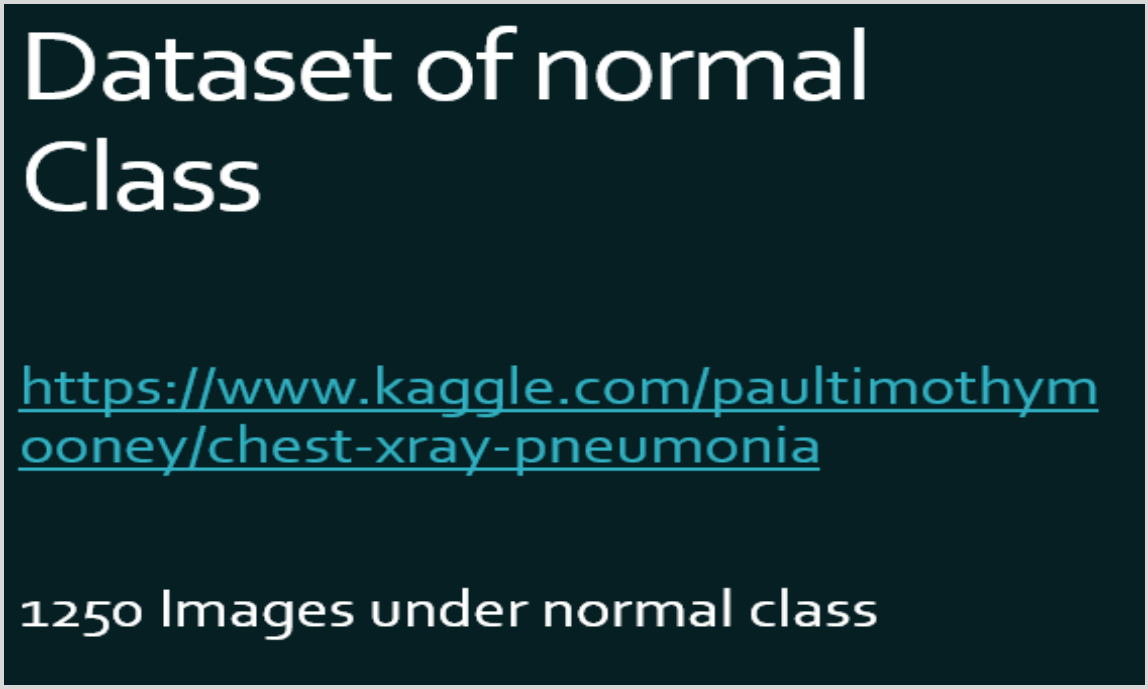
But where the project is used can be in commercial purposes. Like for calculation of covid count. Its use as a Covid detection is not supported as for the reasons stated above.

Datasets used:




```
data.head()
```

	patientid	offset	sex	age	finding	RT_PCR_positive	survival	intubated	intubation_present	went_icu	...	date	location	folder	filename	doi	
0	2	0.0	M	65.0	Pneumonia/Viral/COVID-19	Y	Y	N	N	N	...	January 22, 2020	Cho Ray Hospital, Ho Chi Minh City, Vietnam	images	auntminnie-a-2020_01_28_23_51_6665_2020_01_28_...	10.1056/nejmc2001272	https:/
1	2	3.0	M	65.0	Pneumonia/Viral/COVID-19	Y	Y	N	N	N	...	January 25, 2020	Cho Ray Hospital, Ho Chi Minh City, Vietnam	images	auntminnie-b-2020_01_28_23_51_6665_2020_01_28_...	10.1056/nejmc2001272	https:/
2	2	5.0	M	65.0	Pneumonia/Viral/COVID-19	Y	Y	N	N	N	...	January 27, 2020	Cho Ray Hospital, Ho Chi Minh City, Vietnam	images	auntminnie-c-2020_01_28_23_51_6665_2020_01_28_...	10.1056/nejmc2001272	https:/
3	2	6.0	M	65.0	Pneumonia/Viral/COVID-19	Y	Y	N	N	N	...	January 28, 2020	Cho Ray Hospital, Ho Chi Minh City, Vietnam	images	auntminnie-d-2020_01_28_23_51_6665_2020_01_28_...	10.1056/nejmc2001272	https:/
4	4	0.0	F	52.0	Pneumonia/Viral/COVID-19	Y	NaN	N	N	N	...	January 25, 2020	Changhua Christian Hospital, Changhua City, Ta...	images	nejmc2001573_f1a.jpeg	10.1056/NEJMc2001573	https:/



Dataset Analysis:

The covid lung x-ray has been taken from a github link maintained by a group of doctors. The dataset contains all kinds of alignment and CT scans. We filtered only the X-ray images in posterior-anterior alignment.

The healthy set of X-ray images of lungs has been taken from Kaggle's Pneumonia dataset, from which we filtered only the Normal lung X-Rays.

Model Building and Work-process:

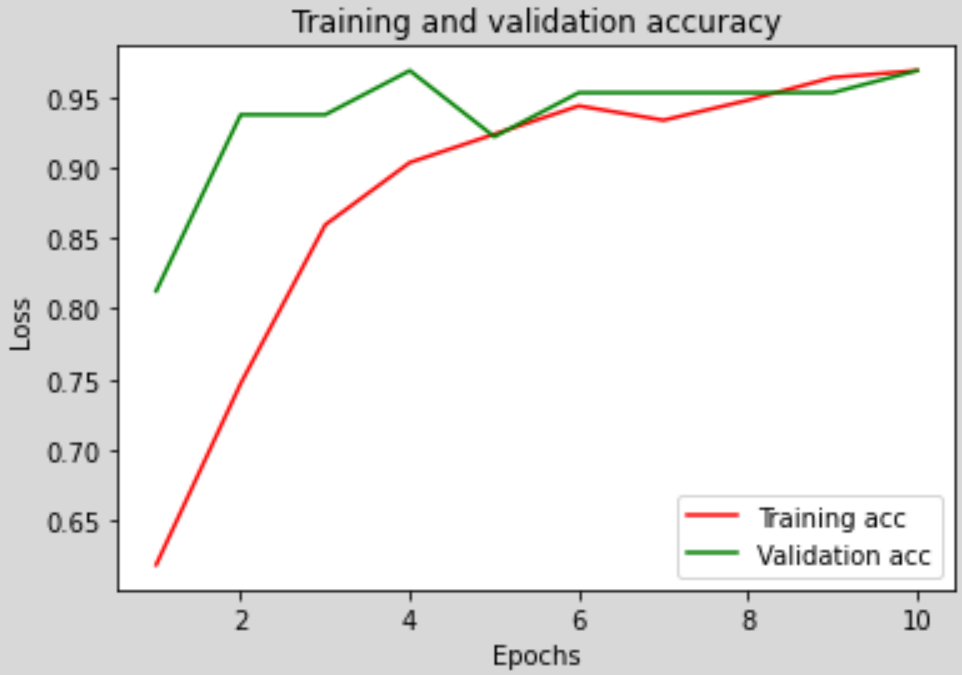
Our Goal is to Achieve a high accuracy model using CNN. On completion of this project, we got an accuracy of around 97% at the end of 10^{th} epoch having 8 steps per epoch. On regards of confusion matrix, we had False Positive and False Negative values pretty low around 1 and 6 respectively.

On processing the data, we found that there were 196 images of the batch of X-rays with Covid-19. So, in-order to get a proper model we chose 196 random images from the pneumonia data-set under NORMAL class.

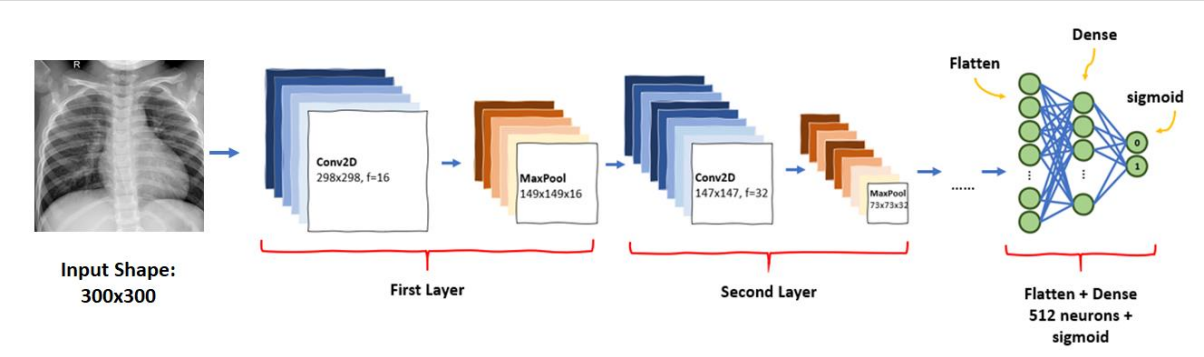
We divided the data into Train and test with 80% and 20% respectively. The features it recognised were around 6 lakhs on a self-developed model. On

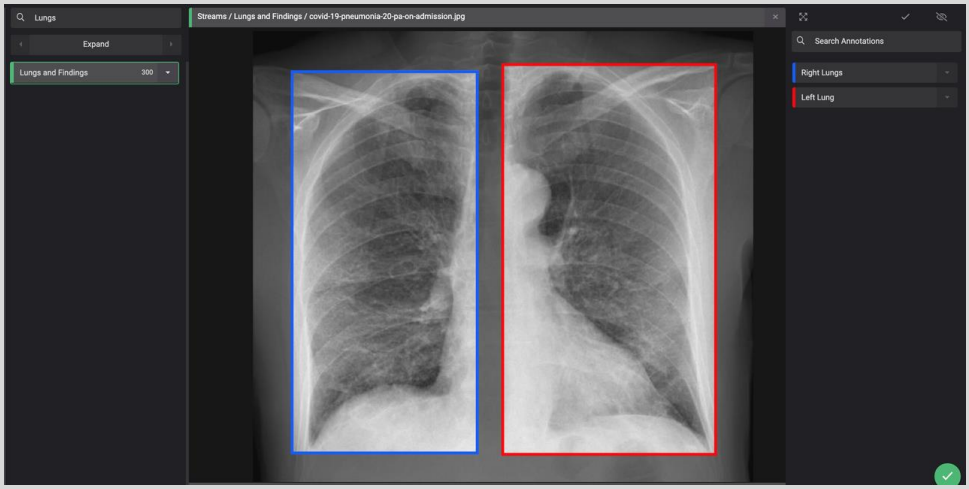
using it with a pre-trained model like VGG, it might overfit as it needed around 14 lakh features.

After we created the model and trained it for 10 epochs with 8 steps per epoch, we pretty much had a reasonable model with 97% accuracy. On the first epochs it had only around 55 % accuracy.



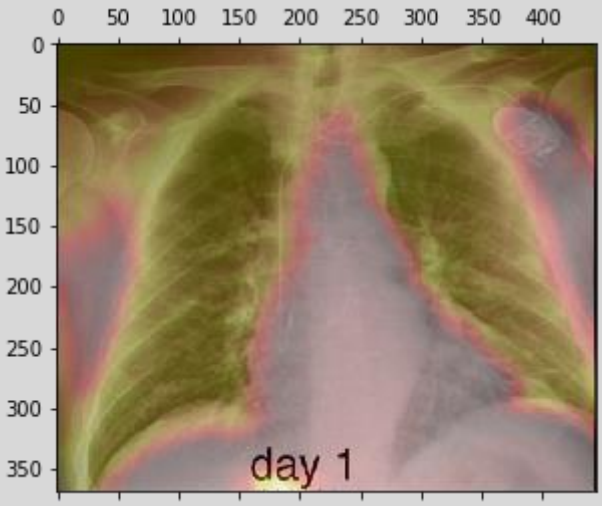
CNN:



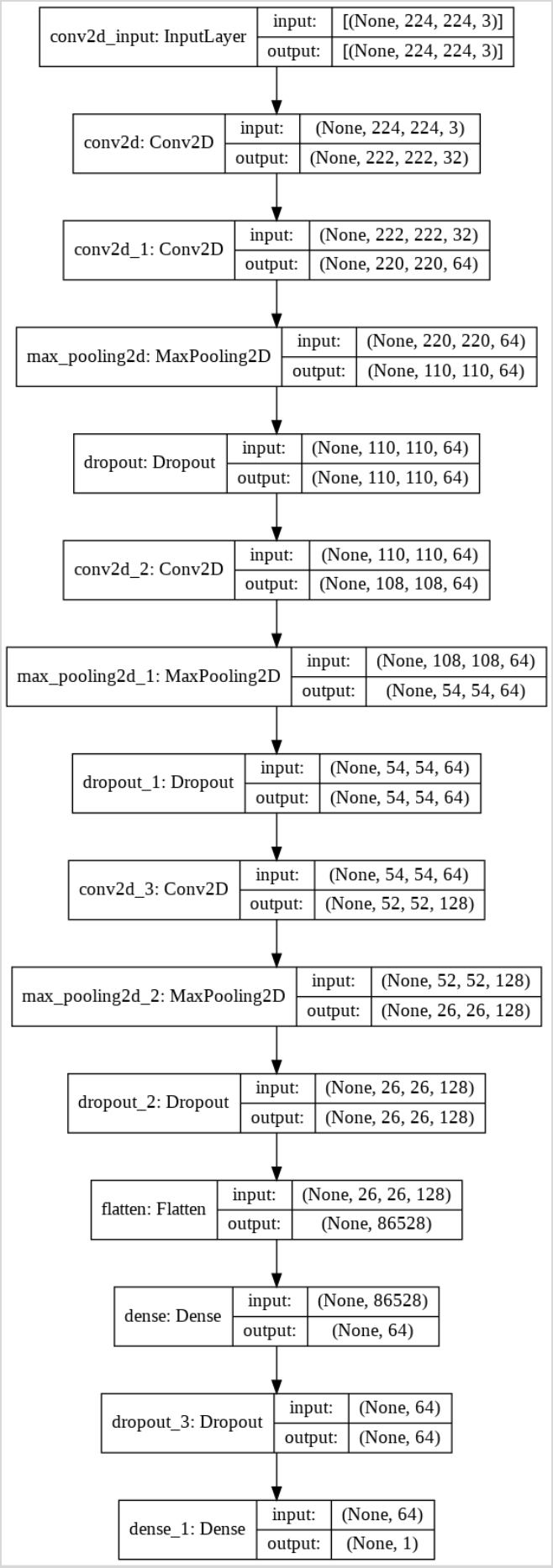


GRAD-CAM

We further pushed the model to highlight the features of a input by which the image is classified.

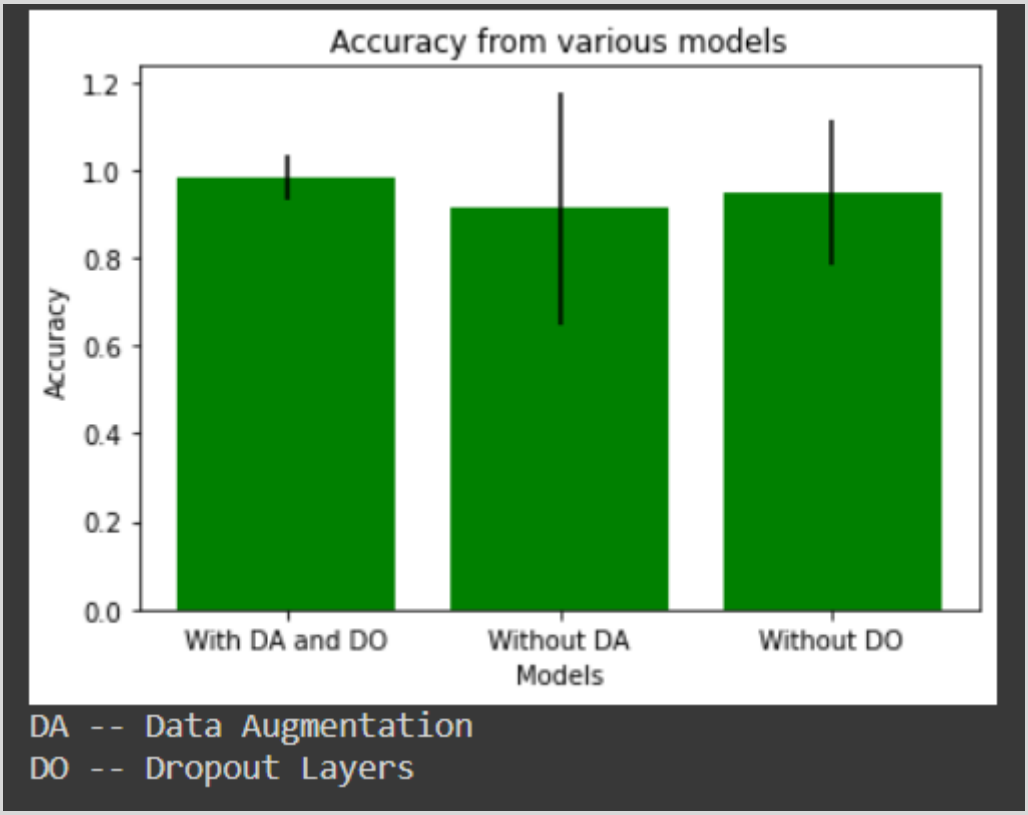


We overlapped the heatsink image with the lung image to get an overall picture of which part of the X-Ray is analysed to get the output.


conv2d_input: InputLayer
input: [(None, 224, 224, 3)]
output: [(None, 224, 224, 3)]
conv2d: Conv2D
input: (None, 224, 224, 3)
output: (None, 222, 222, 32)
conv2d_1: Conv2D
input: (None, 222, 222, 32)
output: (None, 220, 220, 64)
max_pooling2d: MaxPooling2D
input: (None, 220, 220, 64)
output: (None, 110, 110, 64)
dropout: Dropout
input: (None, 110, 110, 64)
output: (None, 110, 110, 64)
conv2d_2: Conv2D
input: (None, 110, 110, 64)
output: (None, 108, 108, 64)
max_pooling2d_1: MaxPooling2D
input: (None, 108, 108, 64)
output: (None, 54, 54, 64)
dropout_1: Dropout
input: (None, 54, 54, 64)
output: (None, 54, 54, 64)
conv2d_3: Conv2D
input: (None, 54, 54, 64)
output: (None, 52, 52, 128)
max_pooling2d_2: MaxPooling2D
input: (None, 52, 52, 128)
output: (None, 26, 26, 128)
dropout_2: Dropout
input: (None, 26, 26, 128)
output: (None, 26, 26, 128)
flatten: Flatten
input: (None, 26, 26, 128)
output: (None, 86528)
dense: Dense
input: (None, 86528)
output: (None, 64)
dropout_3: Dropout
input: (None, 64)
output: (None, 64)
dense_1: Dense
input: (None, 64)
output: (None, 1)

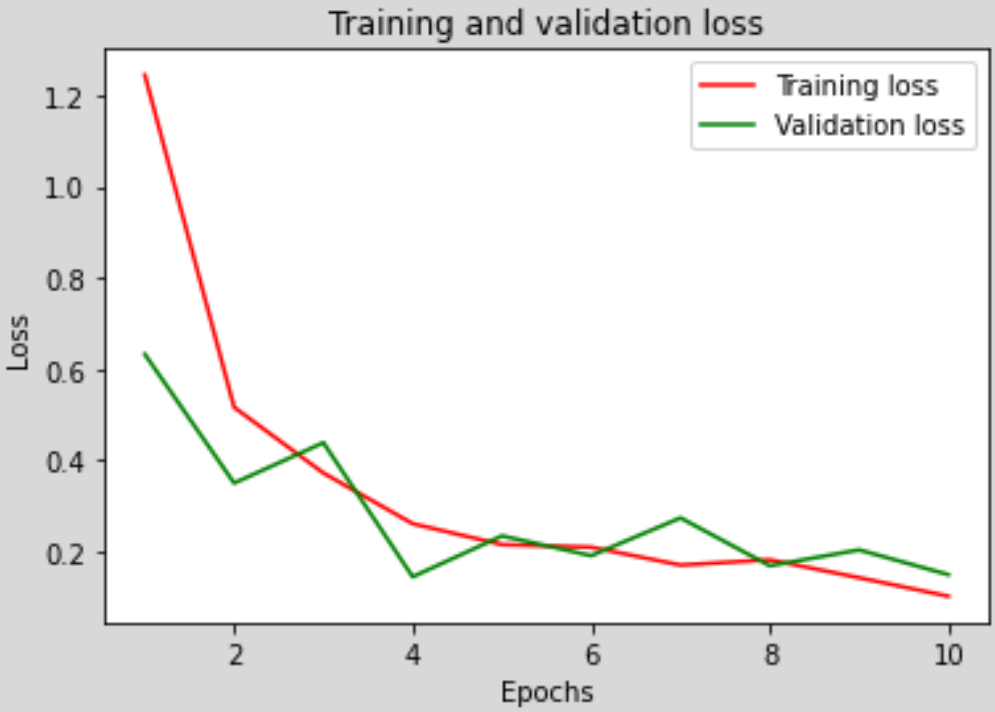
Model Comparison Based on

Data Augmentation and Dropout Regularization:



Conclusion:

On our project we created a traditional CNN model instead of applying the pre-trained VGG model due to its larger number of features. On coming to the application side of our project, it is not feasible as a decision medium, because the cost of error is too high. This model can be used only for educational purpose and also in the medical field to analyse the area of impact if affected by corona virus. The model is really simple and can be run on any standalone machine.



Confusion Matrix: