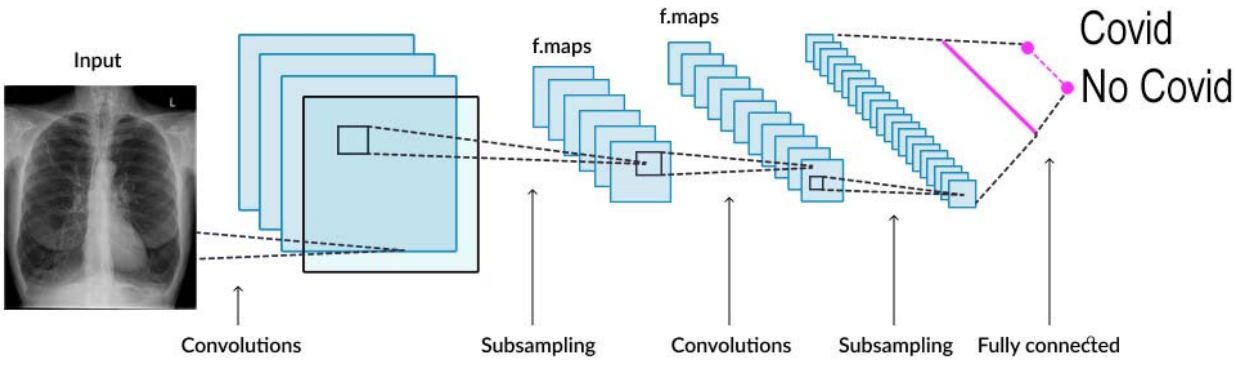
Alessandro Fabiani

May 2020

MANDS Project Work for Master in "Big Data Analytics"

A Deep Learning model for COVID-19 and Pneumonia detection using Convolutional Neural Networks on X-ray images



Context

In early 2020 major outbreaks of a new kind of pneumonia have been reported worldwide.

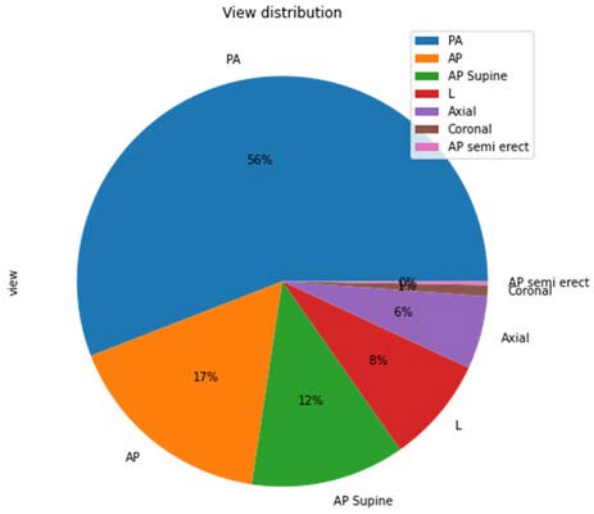
The causal agent has been identified as a **novel Coronavirus (COVID-19),** causing serious cases of pneumonia with a high number of patients in ICU and casualties.

The purpose of this project work is to apply deep learning techniques for image classification and try to classify **X-ray images in order to detect if the pneumonia is due to COVID.**

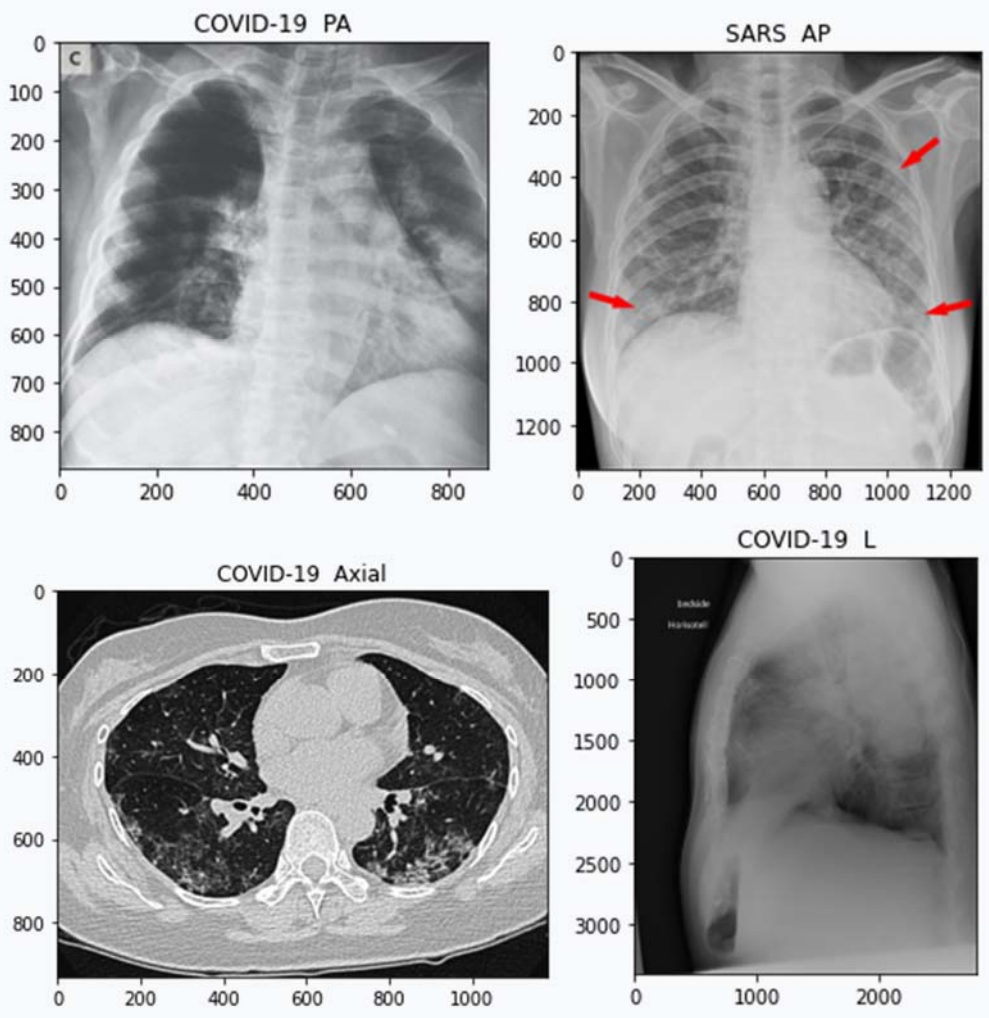
Dataset and data exploration

The first dataset is taken from a project created by Joseph Paul Cohen a PHD Fellow in the University of Montreal and can be found here https://github.com/ieee8023/covid-chestxray-dataset

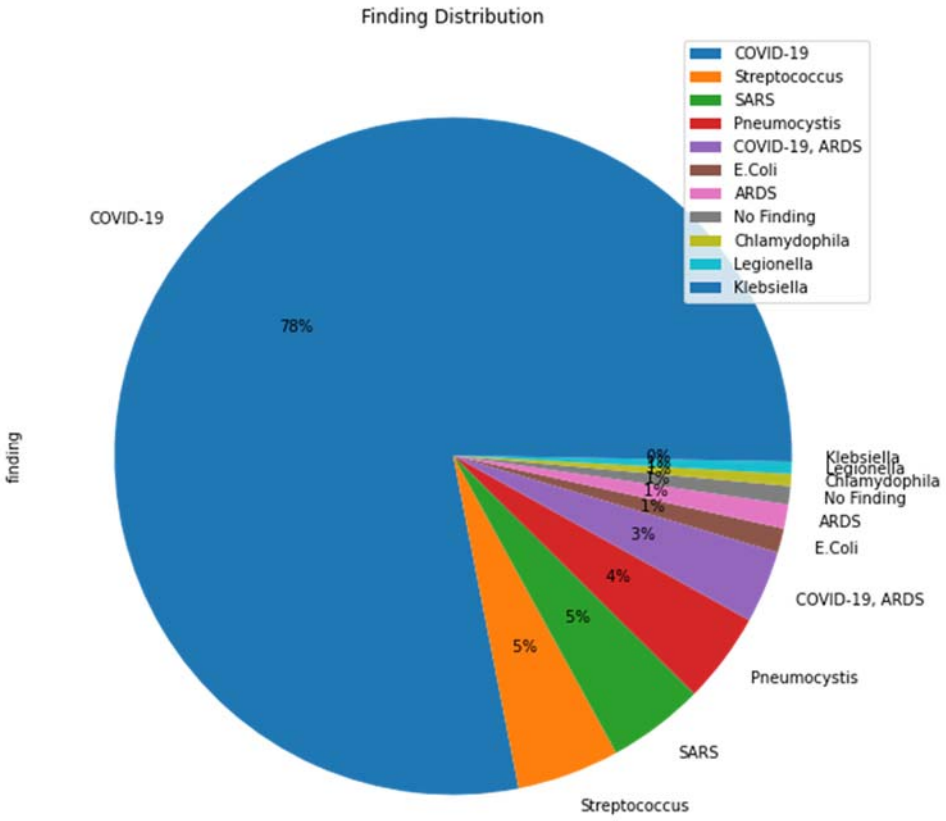
The image Dataset has been created scraping around the web from some public sources and from uploads from contributors and appears to be not very consistent as the 347 images are a mix of X-rays and CT scans with views mixed up between different views such as: **PA** (posteroanterior), **AP** (anteroposterior), **L** (lateral), **CT Scan** Axial, **AP supine**, **AP semierected**.



Images samples in the dataset



Diagnoses appears to be unbalanced as **of the 347 images only 76 are non-Covid cases.**



Along with the image dataset, a metadata file is given with datas such as : sex, survival, intubated, needed ICU, needed intubation, temperature, lymphocytes count etc but it appear not to be very consistent as a lot of values are missing.

Count of 'null' values in metadata file

patientid	0
offset	85
sex	43
age	54
finding	0
survival	235
intubated	277
intubation_present	270
went_icu	274
in_icu	308
needed_supplemental_O2	335
extubated	328

```
temperature                312
pO2_saturation             304
leukocyte_count            336
neutrophil_count           345
lymphocyte_count           337
view                         0
modality                     0
date                        77
location                   108
folder                       0
filename                     0
dtype: int64
```

So I decided to stick with image classification based on images files and "finding" labels.

First try – a deep learning model for binary classification of Covid Pneumonia Xray images

As a first try I decided to run a test with a **pretrained Convolutional Neural Network** on the whole images dataset to see if some recognition was possible.

The labels to predict are **"Covid" (1)** for COVID-19 cases and **"No Covid"** (0) for all other cases.

A **CNN** is built using **Keras** and **Tensorflow**, building a top layer made with:

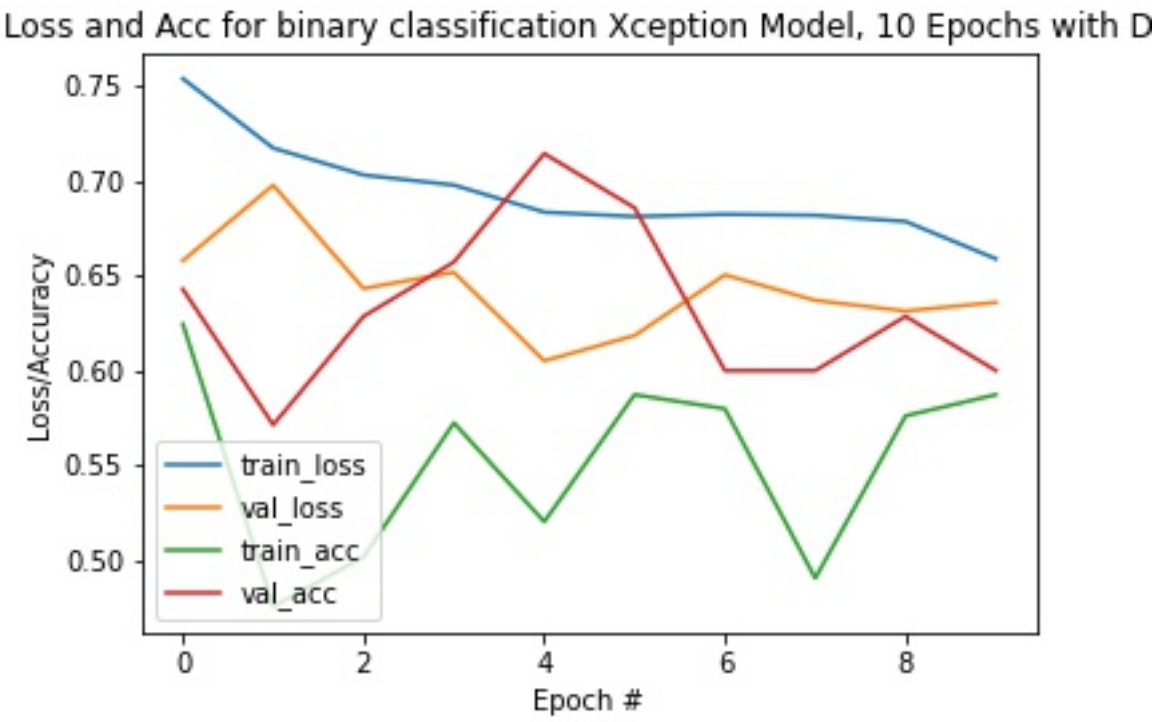
GlobalAveragePooling2D-> Fully connected layer->Drop-out->FC layer for binary classification on top of a **Xception** keras model pretrained with "**Imagenet**" weights.

Data augmentation (DA) is performed on images along with **weights** assignment labels to balance the unbalances dataset.

Classification report for the first try with Xception pretrained model

	precision	recall	f1-score	support
0	0.40	0.27	0.32	15
1	0.82	0.89	0.85	55
accuracy			0.76	70
macro avg	0.61	0.58	0.59	70
weighted avg	0.73	0.76	0.74	70

Accuracy and Loss plot on both train and validation set



Not very encouraging!

Train **loss** decreases, but **accuracy** on both and validation and train sets along with validation loss are very bumpy and unstable.

It seems that the model is **hard to train** on this dataset probably due to **lack of images** and their **quality and mix**.(lateral, Ct scans and more different views).

A new dataset is introduced

To fix the issue I decided to look for more datas and found another Xray datasets for pneumonia on Kaggle.

https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

This Dataset is particularly interesting because is a good Xray collection (**more than 5800 images**) of **healthy** lungs, lungs with **bacterial** pneumonia and lungs with **viral** pneumonia, and for the fact that was **last updated 2 years ago** we can be sure that the viral pneumonia is not COVID-19.

So I decided to take this project to the next level by:

1. Joining the 2 Datasets
2. Filtering only the **PA** (posteroanterior) Xrays as in medical literature it seems to be the **default standard** to check for pneumonia
3. Dividing the new enlarged Dataset in **COVID**, **BACTERIAL**, **VIRAL** (NO COVID), and **NORMAL** to perform a multi-class deep learning classification

2nd Model – a 4 classes Xray image classification for pneumonia

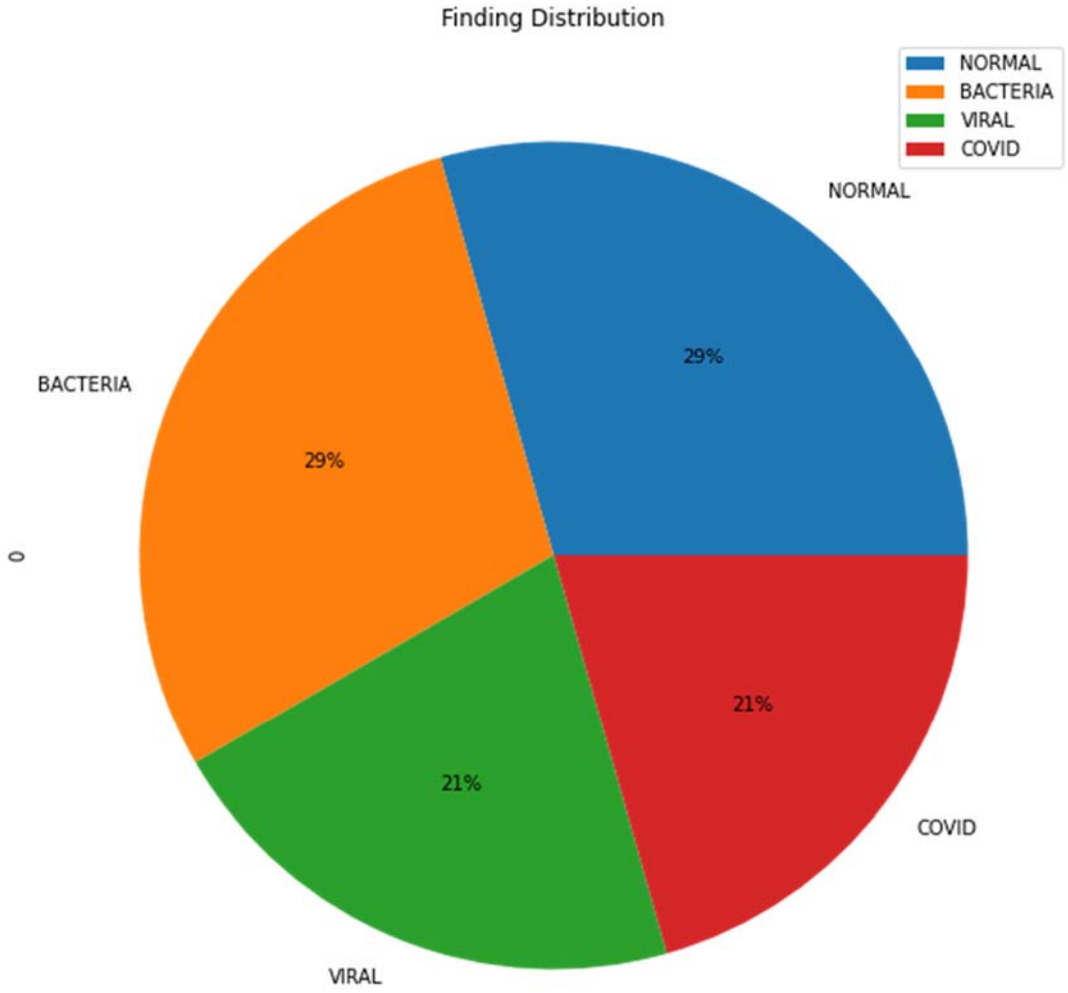
The new Dataset consists in **680 Xray images** with PA view split into:

Viral : 143

Bacterial: 197

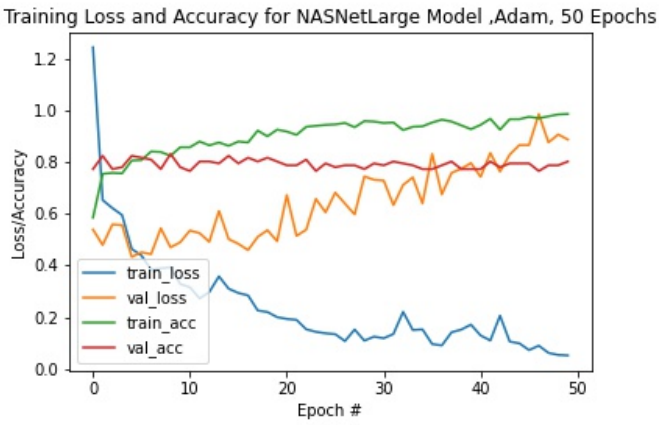
Covid: 140

Normal: 200



Now the dataset looks pretty **balanced**!

As a first try I used a **NASNetLarge** pretrained model (Keras) with the same custom top layer as in the first dataset try, **without Data Augmentation**.



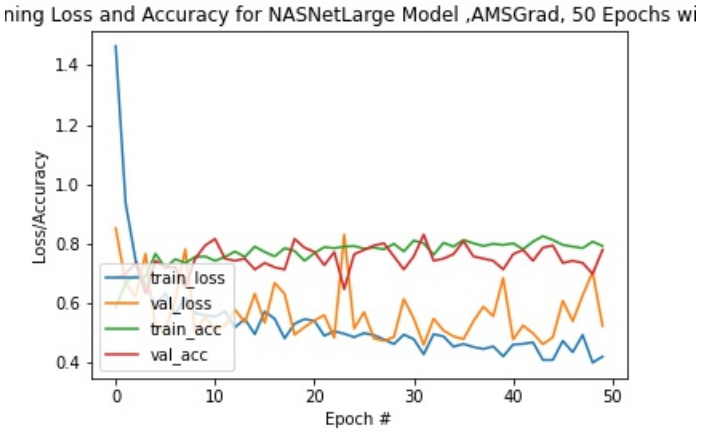
The result is that the model seems to **overfit** (increasing training accuracy and increasing validation loss) so in the next training I always used **DA**.

To find the best model, 4 **pretrained CNN models from Keras** were trained for 50 epochs, performing data augmentation. i.e.

1. **NASNetLarge**
2. **VGG16**
3. **Xception**
4. **InceptionResnetV2**

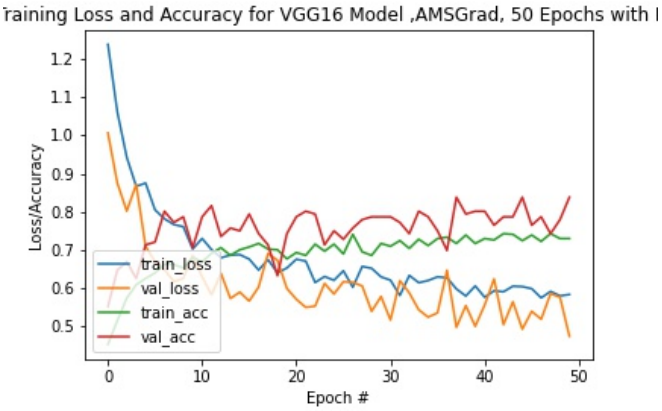
Here are the results of Accuracy and Losses plot along with Classification reports

NASNetLarge



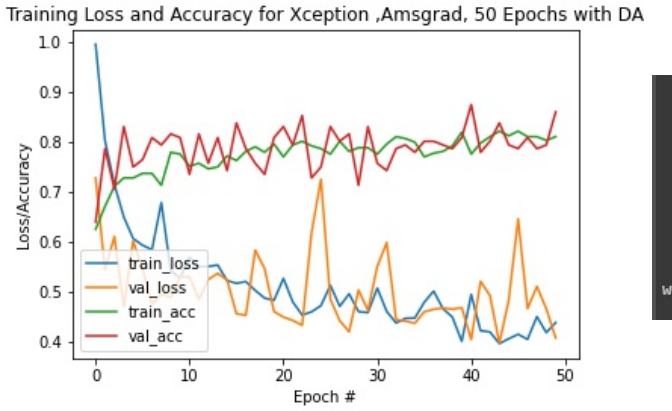
	precision	recall	f1-score	support
BACTERIA	0.70	0.67	0.68	39
COVID	1.00	0.93	0.96	28
NORMAL	0.81	0.95	0.87	40
VIRAL	0.62	0.55	0.58	29
accuracy			0.78	136
macro avg	0.78	0.77	0.78	136
weighted avg	0.78	0.78	0.78	136

VGG16



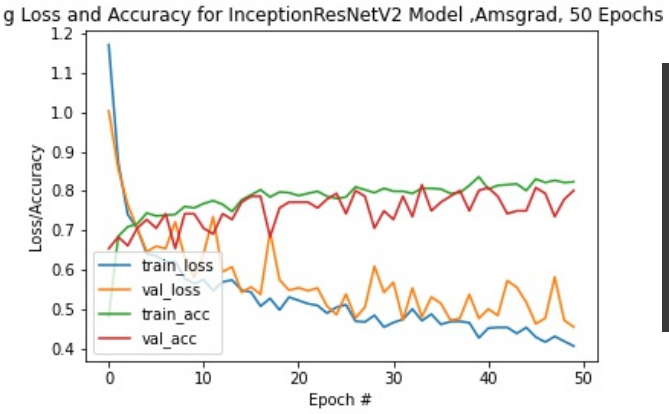
	precision	recall	f1-score	support
BACTERIA	0.72	0.92	0.81	39
COVID	0.96	0.96	0.96	28
NORMAL	0.86	0.95	0.90	40
VIRAL	0.93	0.45	0.60	29
accuracy			0.84	136
macro avg	0.87	0.82	0.82	136
weighted avg	0.86	0.84	0.83	136

Xception



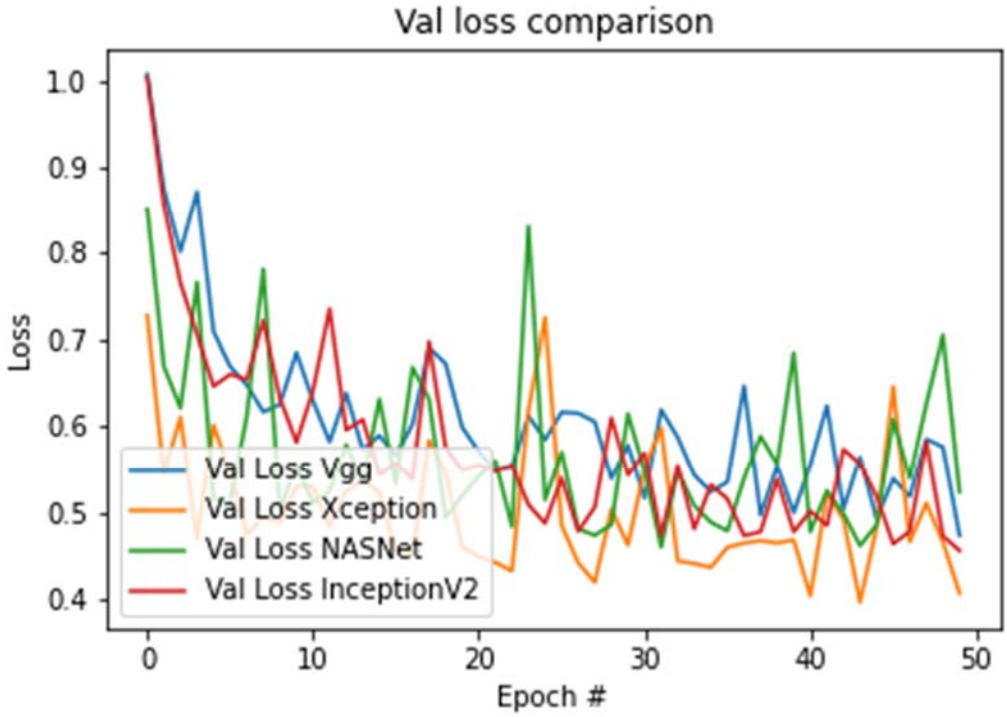
	precision	recall	f1-score	support
BACTERIA	0.76	0.90	0.82	39
COVID	0.93	1.00	0.97	28
NORMAL	0.93	0.93	0.93	40
VIRAL	0.85	0.59	0.69	29
accuracy			0.86	136
macro avg	0.87	0.85	0.85	136
weighted avg	0.86	0.86	0.85	136

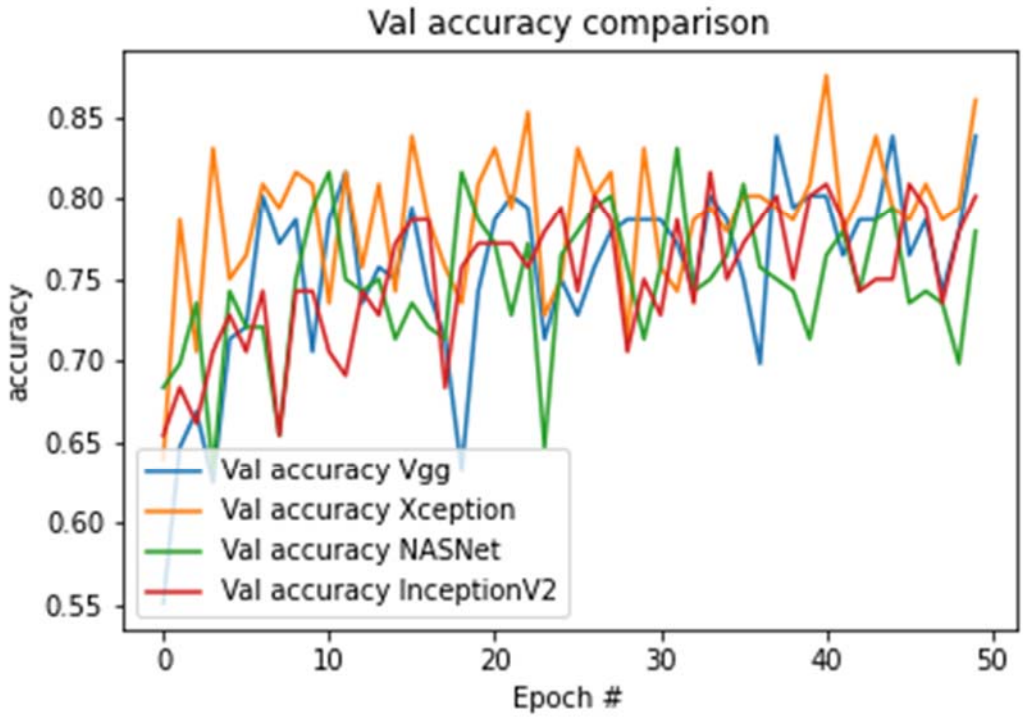
InceptionResnetV2



	precision	recall	f1-score	support
BACTERIA	0.72	0.72	0.72	39
COVID	1.00	1.00	1.00	28
NORMAL	0.80	1.00	0.89	40
VIRAL	0.68	0.45	0.54	29
accuracy			0.80	136
macro avg	0.80	0.79	0.79	136
weighted avg	0.79	0.80	0.79	136

Losses and accuracies Comparison





Definitely better!

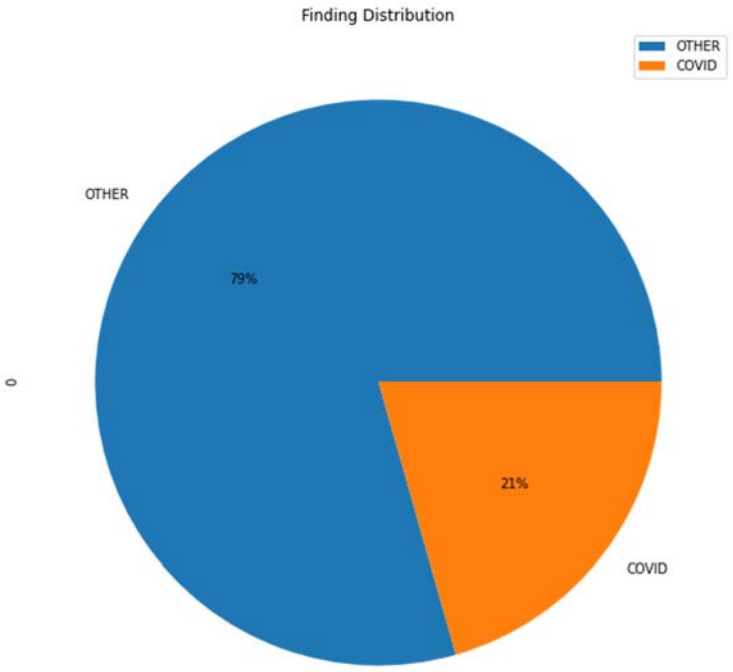
Compared to the first try this enlarged dataset seems to train the network. In all trainings did with data augmentation **both train set and validation set losses decrease** as **accuracies increase** and the classification reports show that **COVID and healthy cases are classified** better than viral non-covid pneumonias and bacterial pneumonias.

This encouraging results led to the idea of make a third try, going back to **binary classification** but this time with the enlarged dataset of **PA Xray images**.

This time the CNN will try to classify **COVID-19** Xray images **Vs** all other cases (Viral, Bacterial, Normal), labeled as "**OTHER**".

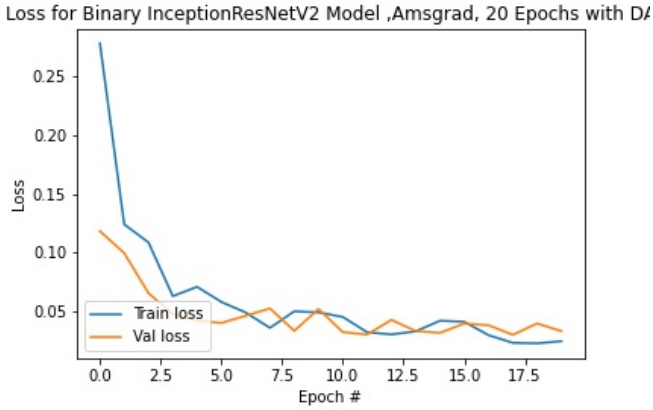
3rd Model – a COVID-NO COVID binary classification model for Xray images

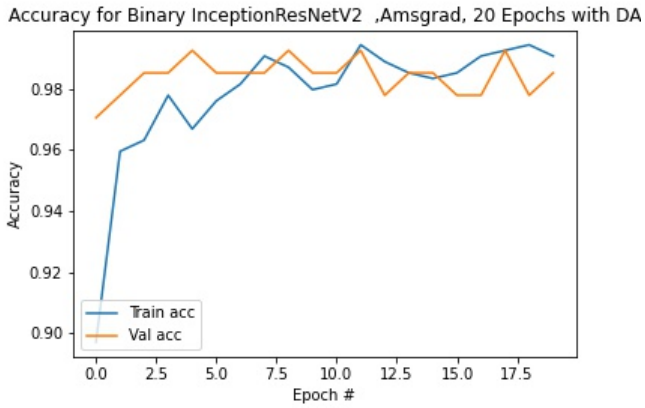
This time we merged **viral, bacterial** and **normal** cases in a class named "**OTHER**". Obtaining a dataset split in this way:



As for the model, we decided to use the **InceptionResNetV2** from the previous case, as it appeared to be the one with best performance in COVID cases classification. The model has been modified for the classification with two classes.

Here below the plots for losses and accuracies





Classification report for the CNN for binary Covid Xray classification

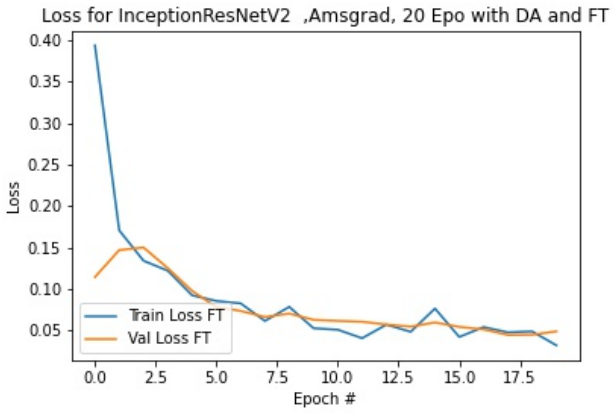
	precision	recall	f1-score	support
COVID	1.00	0.93	0.96	28
OTHER	0.98	1.00	0.99	108
accuracy			0.99	136
macro avg	0.99	0.96	0.98	136
weighted avg	0.99	0.99	0.99	136

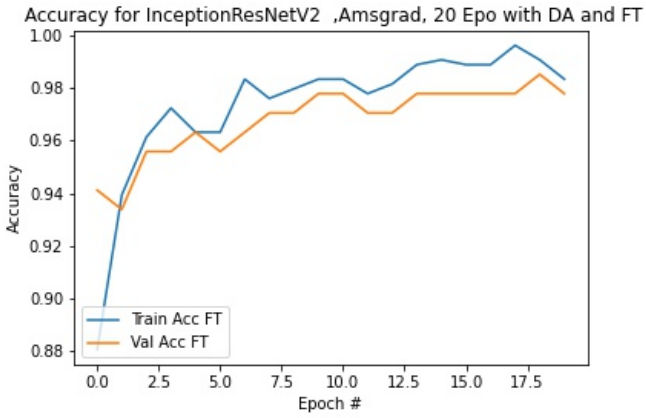
Not bad! In this case, the model seems to predict with fair accuracy the COVID cases.

Let's take a little step further – Fine Tuning

So far we trained only the top layer of the model that we placed on top of Keras **InceptionResNetV2**, importing and freezing all pretrained weights.

Now we will unfreeze last two layers of the **InceptionResnetV2** and retrain the network to see if this **fine tuning** will improve performance.





Classification report

	precision	recall	f1-score	support
COVID	0.97	1.00	0.98	28
OTHER	1.00	0.99	1.00	108
accuracy			0.99	136
macro avg	0.98	1.00	0.99	136
weighted avg	0.99	0.99	0.99	136

Both precision and recall improved! Suggesting that this model is **fairly accurate** when it comes to classify the X-ray images from the two datasets.

Conclusions and further investigations

- The trained model seems to be accurate on this dataset but what is really detecting? Remember that the Xray images come from 2 different datasets for what concerns Covid and No-Covid, is there any chance the model is detecting this difference maybe in the way the X-ray images were taken?

- The dataset is still too small for this purpose it would be useful to make the same training on a larger dataset

- In medical literature It seems that **CT scans** are more precise when it comes to detect pneumonia, training the same model on CT scans would be worth to try