# Recognizing COVID-19 from Lung CT Scan Images

Botond Gábor Barta CSGJK1 | Lilla Nóra Matyasi X986BV | Dorina Petra Lakatos FWH5YD

## 1. Introduction

As this year (2020) was all about the coronavirus we thought that this topic will be an appropriate choice for this year. Right now (11.12.2020) the coronavirus has infected more than 71 million people and unfortunately caused the death of approximately 1.6 million people all over the world, therefore this is a serious global crisis at the moment and will be in the next few years (or maybe months).

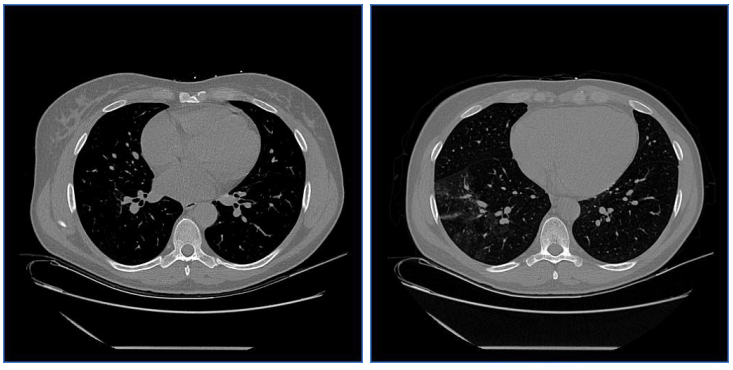
The application of Deep Learning in this case seems obvious, because once if the network has been trained well with the data, the model can tell in seconds if a patient is infected by the virus or not with a really high confidence. Our model can be really helpful for the doctors especially in times like this when they are really overwhelmed.

## 2. Background and Motivation

We chose this project because we had a personal motivation: one of our teammates' mother's had a lung CT in September and the doctors could not really decide if it was covid-19 or not. She had many symptoms, however all of her tests came back negative. One doctor said it was covid, and she had to go to the ICU immediately, however another said that it was nothing.

Therefore we decided to make a model which can decide with a really good accuracy whether the patient has Covid-19 infection or not.

The two type of pictures we want to test with our model:

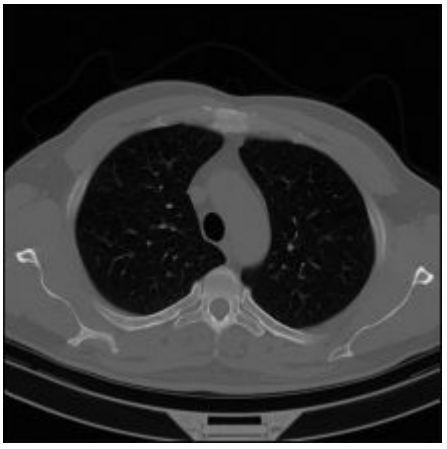


## 3. Data

We are using data from medRxiv provided by Mohammad Rahimzadeh. The covid CT dataset contains 9776 healthy and 2282 covid CT scans from 95 COVID-19

patients and 282 healthy people. The original images are 16-bit uint grayscale with 512*512 pixels resolution.

CT scan of a COVID-19 patient



CT scan of a healthy patient

# 4. Model

At first, we tried to build our own model based on general templates, that had some 2D convolutional layers as the input with maxpooling between them, after that, we wanted to flatten the data and use dense layers as the output. However we could not reach 85% with our model any the less we tried hyperparameter optimization. Therefore we decided to try with pretrained models.

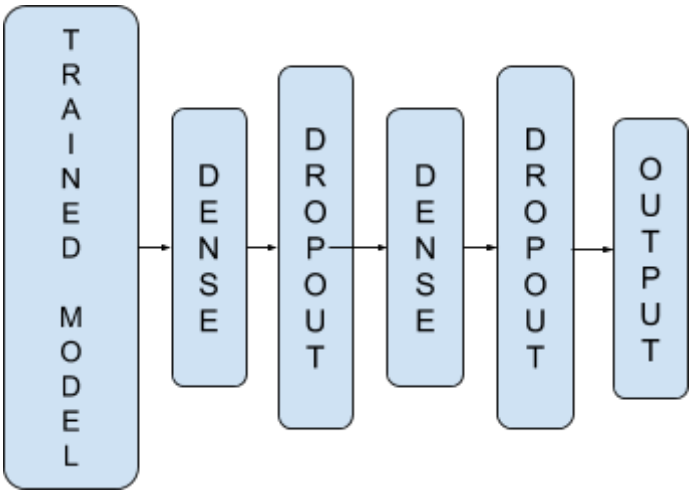
After that, we tried 4 different pretrained models (ResNet50V2, DenseNet169, VGG16 and InceptionV3) to find out which gives the best results.

After the trained model we put 2 Dense layers each followed by a DropOut layer for regularization purposes and the last Dense layer which is the output. The Dense layers are using ReLU activations and their number of neurons are between 16 and 128 and the parameter of the DropOut layer is between 0.1 and 0.4.

We used the hyperparameter optimization to find the best model. For this we used the HyperBand from keras tuner by Keras.

As the dataset has only 2 classes (COVID or Normal) we decided to use binary classification so that our model last Dense layer has only one neuron with sigmoid activation.

Our models structure:



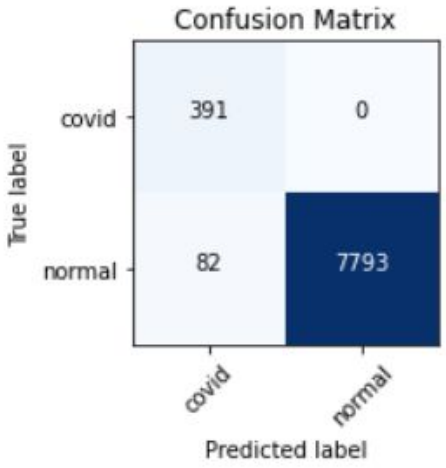
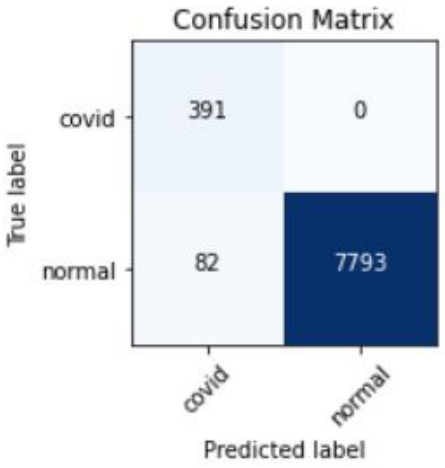
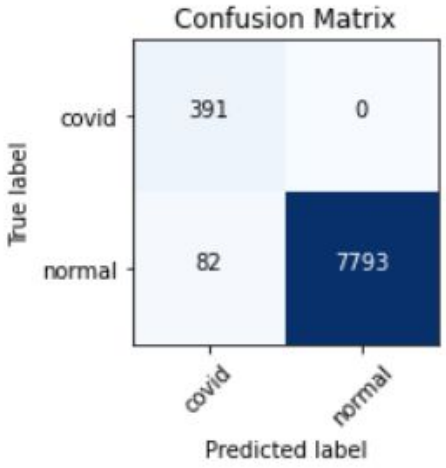
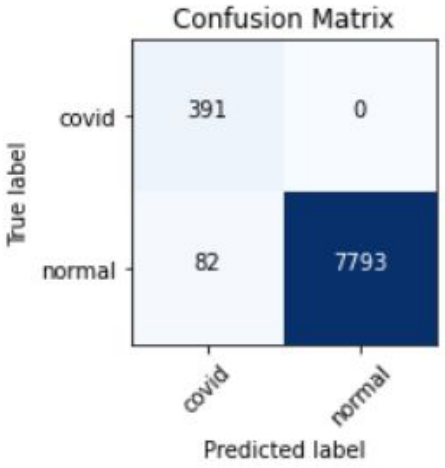

We did transfer learning, we set the base model's layers trainability to false, and hyperparameter optimized the new dense layers with Nadam optimizer. Afterwards we saved the model that gave the best validation accuracy. Then we fine tuned the model for some more epochs with SGD optimizer and early stopping.
The ResNet50V2 somehow did not work out as we thought it would no matter how hard we tried. So we decided to handle it another way: the base model's layers were trainable the whole time, this was the only way to achieve higher accuracy on the validation data.

# 5. Experimental Results

These are the best model's parameters and their results.

| | ResNet50V2 | DenseNet169 | VGG16 | InceptionV3 |
|---|---|---|---|---|
| dense_1_units | | | 80 | |
| dropout_1 | | | 0.2 | |
| dense_2_units | | | 112 | |
| dropout_2 | | | 0.1 | |
| val_accuracy | | | 0.9922 | |
| val_loss | | | 0.0183 | |
| test_accuracy | | | 0.9901 | |
| test_loss | | | 0.0260 | |
| False negative rate | | | 0 | |

We plotted the confusion matrices of all the trained models to find out which has the smallest false negative rate.
The confusion matrices of the models:

| | |
|---|---|
|  <br>TODO |  <br>TODO |
|  <br>VGG16 |  <br>TODO |

Overall we decided to use the _____ as the base model because this one had the best results on the validation dataset. We focused on the false negative rate as in our opinion it is more crucial to recognize positive results (even if they are false positive) than to miss the infected patients.

The best model has only predicted 0 patients as false negative out of 391.
Our model achieved an outstanding ___% accuracy on our validating dataset with a low false negative rate.

# 6. Conclusion

In conclusion we really think our model can decide with a really high accuracy whether a patient has Covid-19 infection or not, from a single lung CT. We believe that our model could be used in hospitals effectively to evaluate a CT automatically. Another thing to mention is to always try pretrained models as the base, because they have learned patterns for the images.

Last but not least according to our model it is 99% that our teammates' mother did not have Covid-19 infection.

# 7. References


The data we used:

Rahimzadeh, Mohammad and Attar, Abolfazl and Sakhaei, Seyed Mohammad, A Fully Automated Deep Learning-based Network For Detecting COVID-19 from a New And Large Lung CT Scan Dataset, 2020, 10.1101/2020.06.08.20121541, Cold Spring Harbor Laboratory Press, medRxiv
https://www.medrxiv.org/content/early/2020/06/12/2020.06.08.20121541

Sohrabi, C. et al. World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). *Int J. Surg.* 76, 71–76 (2020).