COVID-19 detection on lung X-ray and CT images using Deep Convolutional Networks

(COVID-19 detektálás tüdőröntgen- és CT-felvételeken mély konvolúciós hálózatok segítségével)

Zs. Szalay, F. Jakab, B. Szegedi

Abstract— Nowadays the COVID-19 virus spreads at a frightening speed, apparently in all around the world. To help to ease the enormous load from the health system, many data scientists try to develop methods with different devices. One of the applicable fields is deep learning usage for detection during patience examination. More precisely the convolutional neural networks as great tools for this task. In this university project, we try to re-create similar attempts, where the network has to binary classify between CT and X-ray scan images for COVID infection and improve it further. Starting with many better-known models with transfer learning on a smaller dataset, then we selected the best one according to the produced scores. The VGG-16 showed the best potential, thus we trained on larger data and fine-tuned it with hyperparameter optimaziation.

Absztrakció— Manapság a COVID-19 vírus ijesztő sebességgel terjed, látszólag az egész világon. Az egészségügyi rendszer hatalmas terhelésének megkönnyítése érdekében sok kutató megpróbál különböző eszközökkel kidolgozni módszereket. Az egyik alkalmazható terület a mély tanulás használata a páciens vizsgálat során. Köztudottan a konvolúciós hálózatok az egyik legalkalmasabb eszközök ezen a területen. Ebben az egyetemi projektben megpróbálunk rekreálni hasonló kísérleteket, ahol a tanított hálózatnak binárisan kell osztályoznia a CT és a röntgen képeket a COVID fertőzés szempontjából, továbbá javítani próbálunk az eddig ismert eredményeken. Sok ismertebb modellből kiindulva, egy kisebb adathalmazon történő transzfertanulással, majd az előállított pontszámok alapján kiválasztottuk a legjobbat. A VGG-16 mutatta a legjobb potenciált, így nagyobb adatokra tanítottuk, és a hiperparaméterek optimalizálásával finomhangoltuk őket.

Index Terms— Chest X-ray, Covid-19, Deep Learning, Medical Image Classification, Transfer Learning

1 Introduction

If we had to introduce 2020 with one word, it would be „coronavirus”. Most experts are working on solutions and innovations all around the world, to save more human life from the virus. One of the most common complications of the coronavirus is pneumonia. When the patient produces the symptoms, an X-ray or CT scan aimed to lung are the best way to reveal the illness. The expert radiologists can decide if the patient has got pneumonia. It is caused by COVID-19 or not, but it takes a while. The medical workers are doing their best and working all day, every day. It could help them a lot, if computers could analyze the images instead of them.

The X-ray and CT images are very divergent, the angle, the distance, the patient, the machine makes them highly different. It is nealy impossible to classify them with algorithms. Then what to do? Maybe the answer is the AI. As we know Convolutional Neural Networks are used all around the world to classify images. There are a lot of network, and databases on the internet, most experts are researching topics like this as mentioned before. We tried to read every source we could and implement them into our research to add and create something new in this topic.

Our goal is to create a network which analyzes these specific images. The input is the X-ray or CT captures, the output is 0 to 1. If we recognise pneumonia caused by COVID-19 the output is 1, otherwise 0 (other diseases and healthy lungs). The idea behind the network is that if the patients have got coronavirus, we must separate them into quarantine immediately. We can say it does a preprocess before the real disease analysis, very reliably and fast.

2 Review of Previous works

From the start of 2020 many experts, scientists started to develop the solution for this problem. Numerous other methods used, and significant works have been published too, not only for this specific challenge, but related areas either. The reason why Convolutional Neural Networks

- *Zs. Szalay is with the Mechatronics, Optics, and Mechanical Engineering Informatics Department, Budapest University of Technology and Economics, Budapest, HUN 1111. E-mail: szalay2318@gmail.com*
- *J. Farkas is with the Mechatronics, Optics, and Mechanical Engineering Informatics Department Budapest University of Technology and Economics, Budapest, HUN 1111. E-mail: jakab.farkas00@gmail.com*
- *B. Szegedi is with the Mechatronics, Optics, and Mechanical Engineering Informatics Department, Budapest University of Technology and Economics, Budapest, HUN 1111. E-mail: balint.szegedi1@gmail.com*

 Published by the IEEE Computer Society

(CNN) shows great potential for this task is, because their strength in image classification and object detection. The convolution layer(s) with pooling layer(s) are a well-known combination for feature extraction from pictures. Since they appeared in the fields of data science, their architectures have changed a lot, and some basic efficient model design gained more recognition than others.

When we started to work on this project, we searched for different articles, previous works. The first impression was that this is a 'hot' topic. As the introduction part mentions, the world is now in a competition against the virus. Many people try to help their professionals in any way to save lives. In the perspective of a data scientist, there were too many attempts to process in the frame of this writing. Therefore, there is a great possibility that what we thought for great results, is already acknowledged and improved by someone else.

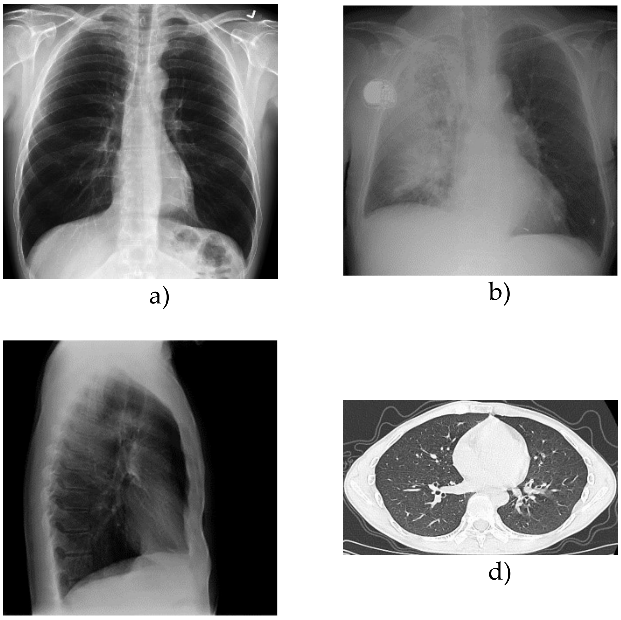


Figure 1. The variants from the dataset's images a) healthy; b) Covid infected (also pacemaker user); c) side scanned; d) CT section scan

For most similar solutions, the neural networks are used on chest X-ray or computed tomography (CT) scan images. These images were collected from all around the world's hospitals. Many specialized groups or universities started to systematize and convert them into databases, which can be used free of charge now. More detailed introduction of the datasets can be read in the later part.

We realized that there are two main directions in development. One is using transfer learning and fine-tuning on well-proven models. The other is designing a new special network for only this sole purpose. Take a look at some of the more outstanding work, from both categories.

In the paper [1], the authors chose a different ResNet variant, InceptionV3 and a mix between model, Inception-ResNetV2. This work was done during the first outbreak, in 2019. At that time, there were not too many resources to work with, hence they used pre-trained models. This means they did not need massive hardware and can train with smaller data-sample sizes. Their goal was binary classification: 1 - COVID, 0 - not-COVID (including pneumonia, other diseases and healthy lung image too). According to their results, the ResNet50 can produce the highest accuracy on three different datasets. Abbas, Abdelsamea and Gaber in [2] validated and adapted a new model, called DeTraC (after Decompose, Transfer, and Compose). One of its greatness is handling irregularities and high accuracy. Their main idea came from article [3], where a similar approach has been made, but earlier. The proposed network in this case, called COVID-NET. It was one of the first Covid-specific networks. The model design has two levels, a human-machine collaboration, because during the time of its creation there was literally no data in hand for training. A non-expert person in medical scan image viewing is nearly impossible to tell the difference between corona and other diseases. This caused a useful byproduct, the COVIDx dataset. Which the authors collected and labeled for their project, and later let the other developers use it, and maintain it. The other stage is machine-driven design exploration one.

On [4], Farooq and Haafez proposed a specialized ResNet, called COVID-ResNet. Its basis is a fine-tuned ResNet50 network. It can produce multiclass choices between three different infections and a healthy person's lung with a high accuracy rate, only after 41 epochs training. This result may cause the 'too good to be true' illusion. The reason for COVID-ResNet success is in its special three-step design stages. In [5] another extraordinary model was proposed, called M-Inception. The difference between this and the vanilla version lies in the last fully - connected layer. The developers reduced the dimension of the features, before it was sent to the final layer. They did not achieve prominent accuracy with their method. On the other hand, this article is worth mentioning, because it is more detailed in all its aspects of the corona X-ray detection task, which provides a great basis for understanding the process behind it. Paper [6] details the transfer learning and fine tuning of ResNet18, ResNet-50, SqueezeNet and DenseNet-121 for the Covid-non-Covid-healthy classification. The SqueezeNet produced the best overall scores.

Makris, Kontopoulos and Tserpes [7] trained and examined different network structures with coronavirus X-ray dataset. These models are basically, which are available in Keras application as pre-trained models. Contrary to what articles have been collected here so far, they got highly different results. They concluded that the VGG16 and 19 variants are much better for this task, rather than ResNet and its versions. The network's task was to use a multiply classification on three types of images. These are Covid, pneumonia, and normal. Of course, there are a lot of variables at a setup of deep neural network's training. We cannot and won't state this shows a fundamental truth in the topic. Another group [8] used VGG16 with a small corona specific dataset. The speciality was, in this case, they used merely 122 images and all of them were taken from children under age 5. The accuracy is high

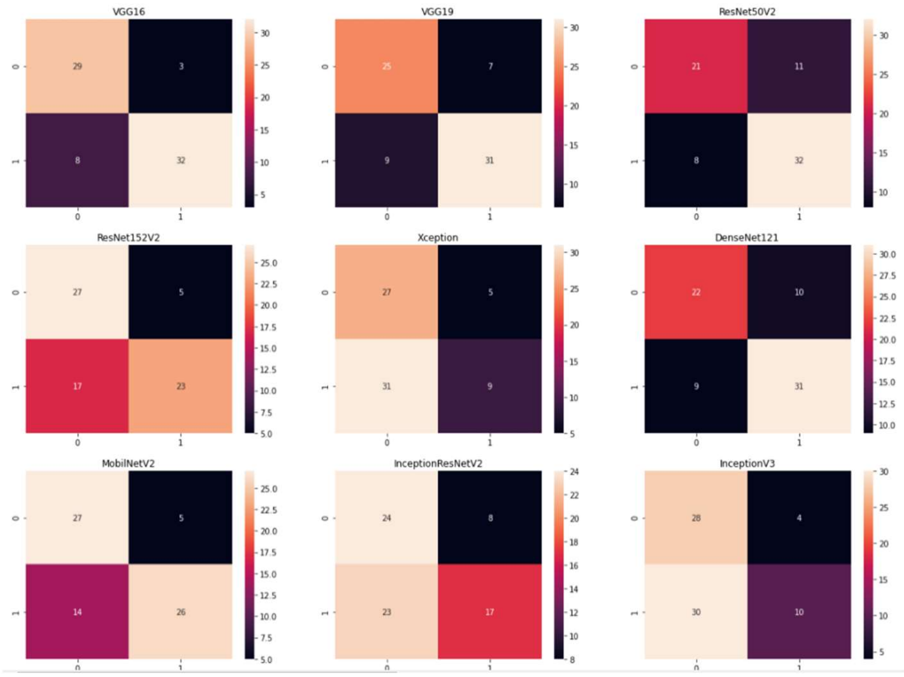


Figure 2. The confusion matrices of different models after transfer learning

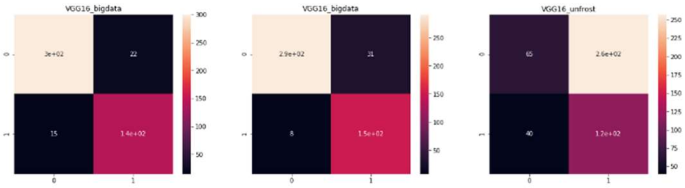
enough for using the model for medical intention. The only problem, which the authors mention too, is that the trained dataset consists of that age group which is not concerned by the virus. Mohammed Y. Kamil [9] likewise chose VGG19 for the task. The choice was not random, he applied many transfer learnt models, and after compared them to each other. Then fine tuned the one with best results. He used a 1000 image large dataset. The final network can make a binary decision if the input scanned image has Covid or not. In paper [10] the authors created three self-designed models and compared them with VGG16, VGG19, ResNet50. The unique networks here follow the basic CNN principles in means of architecture. The differences between Model 1, 2, 3 (as the designers cite them), are from the number of the convolutional layers. Each has one more. All the models trained for binary classification (Covid or not). The self created models can bypass the others according to the creators. Zebin and Rezvy [11] compared VGG16, ResNet50, and EfficientNetB0 models on the task. They functioned as three categorical classifiers (Covid, pneumonia, or normal). It is worth mentioning that from a lot of authors, only in their work used EfficientNet. Yet it produced the highest accuracy. Additionally, they trained a generative adversarial network (CycleGAN) to generate and augment the minority of the Covid class. In paper [12], the same experience can be read, using the well-known models to transfer learning, with one exception. The numbers of layers which they unfroze for finetune. As there is no golden rule for this parameter, it is interesting to see what totally different values can achieve in this aspect. Three type classification was done here, with the best result of VGG-16. [13] paper shows a VGG16, VGG19, MobilNet, and InceptionResNetV2 pre-trained transfer learning for Covid or no results. The Visual Geometry Group nets produce the highest scores. [14] likewise shows a VGG-16, and VGG19 comparison with other common networks, for a three-type classification.

We focused more on the transfer learning, fine-tuning articles, because from the start this was our main intention. We saw that likewise, there are many variants for this task, with varying setups and models. The two prime groups maybe the VGG and ResNet/Inception would be.

3 Datasets

As mentioned before, the first well-known attempt to image based Covid filtering [3] induced the idea to collect these data in a structured form and let other developers use it as an open-source dataset. They created the COVIDx dataset, which can be found on github [15]. It is built up from different subsets and maintained up to date. The data consists of real-life cases from both, healthy and diseased sections, not just coronavirus infected ones from different setups and directions. For this project we used four different building sets from the whole. The reason behind this decision was that the overall dataset is too huge to work with without adequate hardware setup.

The images also have paired tags, so it can be easily used for supervised learning. These are stored in a .csv file and contain varying information not just about the diagnosis itself, but the patience and other medical data too. For example, the columns separated by a unique ID, which every examined person given, sex, age, disease finding, PCR test positivity, survival, X-ray or CT scanned, the date of the checking, and geological location. It is important to note that medical experts decided about the diagnosis results. The decision is not trivial at all, for unqualified eyes. By receiving only already examined images, people trained in data scientist fields can start to work with them immediately. In Figure 1; some distinct examples can be seen from the dataset.



a) b) c)

Figure 3. The confusion matrices of VGG16 with different sigmoid treshholds on bigger dataset. a) original 0.5 b) modified 0.3 c) learnt with the original smaller dataset

4 Implementation and Evaluation

In the beginning we used a smaller part of the full database, only one source to determine which pre-trained network is the best, in our case (719 X-ray images). In those images the COVID +/- rate is about 50 - 50 so this database is quite balanced we can say. First, we made our self-designed simple and basic CNN with Keras's sequential, the accuracy went very close to 1, but the test results were dire. Our network over-learned. The next step was to implement a pre-trained network from Keras Applications. But which one?

We used 9 different Keras CNN for transfer learning: DenseNet121, ResNet15V2, VGG19, VGG16, ResNet50V2, MobileNetV2, InceptionV3, InceptionResNetV2 and Xception. The top layers of networks were always the same MLP layers (flatten-256-512-1). Of course, those are not the best hyper parameters, we optimised them later when we have already found the best CNN for our task. The output layer's activation was sigmoid, as it is common among all binary classification networks. All the CNN weights were frozen in order to accelerate the learning processes. The error was determined according to binary cross entropy.

The best predictions to the test images came from the VGG networks, they made constantly good predictions in every learning. Sometimes other networks like ResNet and DenseNet also gave low losses, but they were unreliable, if all the learnings are included, we have made. The results can be seen in Figure 2.

From here, we only dealt with the VGG16. In the next learning we made the upper 2D convolutional layers trainable. The result improved a little bit, but not as much as we expected. The resource investment was not worthy at all.

We could focus on the MLP's hyper parameter optimization. We have used the Hyperas python library. Both of the tested activation functions (ReLU, Leaky ReLU) produced the same efficiency, but as we know the ReLU is simpler, so we used that afterward. The average pooling between the pre-trained CNN and the MLP caused no difference, so we used it, because it reduces the number of the trainable weights. The best batch size was 32, we stick with that afterward in all learning processes. The greatest optimizer was Adam by a lot. The most outstanding neuron number in the layer one was 256 and 1024. In the second layer it was 512 and 256. This way we do not had to change the MLP parameters, just the training hyper parameters.

We had loaded all our data at this point and used them to teach further the network. Previously we had only used one subset, but from now on we implement all the other three sources too, we found during the source research. It has a great impact, as we said the images are divergent in one source, when we talk about all the sources this statement is exponentially true. In the big database there is no balance in COVID +/-. Most of the images are from people whose PCR test was negative (or did not even had PCR test). It is veritable, because there were taken many X-ray images before the coronavirus

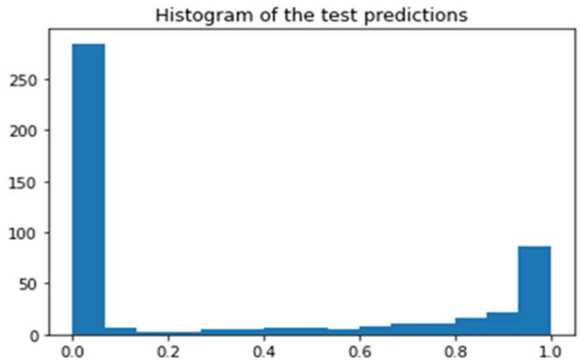


Figure 4. Histogram of the test predictions

even appeared. We can examine how the unbalanced input data had effect on the learning on Figure 3.

As we expected in the test results there are pretty much false negative classifications, despite a larger chunk of the test dataset is COVID negative (as we can see in Figure 4.) just like in the train dataset. We can see that the balance in the input data is very important. But all in all, our network became much better from the bigger dataset. If check the predictions of the previous network to this new test database, the result is disappointing (Figure 3. a). Interesting if we define COVID +|- border to 0.3 (Figure 3. b) from the original 0.5 (Figure 3. c) the results are better; it is maybe because of the unbalanced input.

To be able to evaluate the capabilities of the model, as a part of the process we collected the number of true negatives (TN), true positives (TP), false negatives (FN) and false positives (FP). The metrics we calculated from these numbers are can be seen on equation (1) (2) (3) (4). We did not invent these parameters; it is used basically in all the articles which written in this topic.

$$\text{Accuracy} = (TN+TP)/(TN+TP+FN+FP) \quad (1)$$
$$\text{Recall} = TP/(TP+FN) \quad (2)$$
$$\text{Specificity} = TN/(TN+FP) \quad (3)$$
$$\text{Precision} = TP/(TP+FP) \quad (4)$$

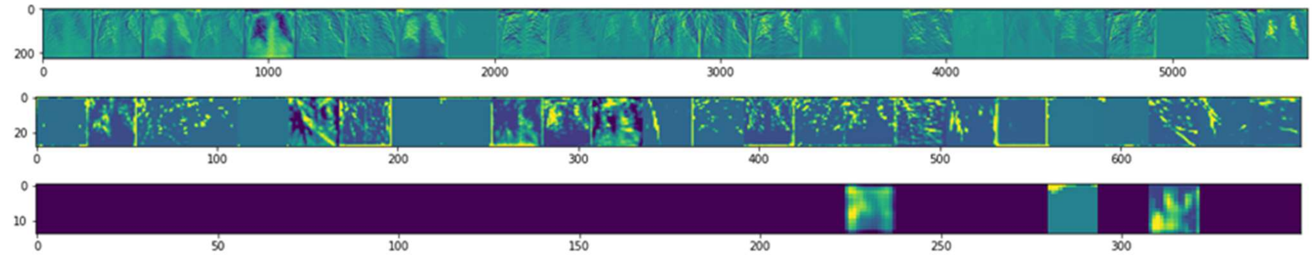


Figure 5. Feature map of the best true positive result

Maybe if we would check some of the networks feature-maps we will have a more informative result. Figure 5. shows a feature maps of the network. We have checked the surest true positive and the surest true negative, to see which features are the most influential.

In the feature map is where are lot of light pixels, those kernels are the most important. These are the determinative features. More detailed datas about the results can be found on the code notebook.

5 Future Plans

The model could be trained with even more data to be more accurate in detecting COVID-19. On the other hand, we couldn't accomplish a comprehensive hyperparameter optimization, due to our limited resources. Maybe more top layers or layers consisting of even more neurons could help to achieve better accuracy, but unfortunately, we were short of both hardware and human resources to complete these studies to the deadline of the project.

6 Summary

It has become clear to us that machine learning and CNNs can achieve unbelievable results, if they are implemented properly, using the right parameters and a decent database. However, we must dig deeper if we would like to achieve as good results as other researchers did in this subject. Although we could produce a roughly accurate network for COVID-19 detection, it became clear that we could enhance this model with more optimization and larger databases. We hope that we can continue to work on this project until we achieve even more compelling results.

References

[1] C. K. Z. P. Ali Narin, "Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks," *arXiv:2003.10849.*

[2] M. M. A. M. M. G. Asmaa Abbas, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," *arXiv:2003.13815.*

[3] A. W. Linda Wang, "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images," *arXiv:2003.09871.*

[4] A. H. Muhammad Farooq, "COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs," *arXiv:2003.14395.*

[5] B. K. J. M. X. Z. M. X. J. G. M. C. J. Y. Y. L. X. M. B. X. Shuai Wang, "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)," *medRxiv,* vol. https://doi.org/10.1101/2020.02.14.20023028.

[6] R. K. M. S. S. Y. G. J. S. Shervin Minae, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep," *ScienceDirect.*

[7] I. K. K. T. Antonios Makris, "COVID-19 detection from chest X-Ray images using Deep Learning and Convolutional Neural Networks," *medRxiv,* vol. https://doi.org/10.1101/2020.05.22.20110817 .

[8] R. P. D. B. G. G. M. G. Lawrence O. Hall, "Finding Covid-19 from Chest X-rays using Deep Learning on a Small Dataset," *arXiv:2004.02060.*

[9] M. Y. Kamil, "A deep learning framework to detect Covid-19 disease via chest," *International Journal of Electrical and Computer Engineering (IJECE),* vol. 11, no. 1, p. pp. 844~850, 2021.

[10] A. A. Khandaker Foysal Haque, "A Deep Learning Approach to Detect COVID-19 Patients from Chest X-ray Images," *MDPI.*

[11] T. Z. S. Rezvy, "COVID-19 detection and disease progression visualization: Deep learning on chest X-rays for classification and coarse localization," *AppliedIntelligence,* Vols. https://doi.org/10.1007/s10489-020-01867-1.

[12] N. Z. T. A. A. Mundher Taresh, "Transfer learning to detect COVID-19 automatically from X-ray images, using convolutional neural networks," *medRxiv,* vol. https://doi.org/10.1101/2020.08.25.20182170 .

[13] S. M. ,. G. H. R. A. ,. Mohammadi R., "Transfer Learning-Based Automatic Detection of Coronavirus Disease 2019 (COVID-19) from Chest X-ray Images," *Journal of Biomedical Physics and Engineering,* pp. 10-15, 2020.

[14] S. C. M. P. A. U. B. P. M. S. N. S. Michael J Horry, "X-Ray Image based COVID-19 Detection using Pre-trained Deep Learning Models," *engrXiv.*

[15] "https://github.com/lindawangg/COVID-Net/blob/master/docs/COVIDx.md," [Online].