Analysis of Deep Learning Models for COVID-19 Diagnosis from X-Ray Chest Images

Francisco Lacruz Sisamón
pacolacruzsisamon@gmail.com

Raul Lorenzo Vidarte
rlorvint@gmail.com

ABSTRACT

The coronavirus disease (COVID-19) has become a pandemic disease on 11th March 2020, affecting the whole world. Currently, there are more than 6,5M cases confirmed and 390,000 deaths. The objective of this paper is to test the performance of different convolutional neural network architectures on Chest X-Ray images for COVID-19 detection. The public dataset used in this research is Kaggle's COVID-19 Radiography dataset and it contains three different classes of images: (1) COVID-19, (2) viral pneumonia and healthy people chest X-Ray images. Their dataset contains 219 COVID-19, 1345 viral pneumonia, and 1341 normal chest X-Ray images. First of all, a data augmentation was done to the COVID-19 set of images due to the huge difference between the number of images in COVID-19 and the other two classes. The architectures used in the research were VGG16, Xception, ResNet-50, ResNet-101, DesNet-201 and InceptionResNetV2. The best results were obtained by using the Xception Net with an accuracy value of 97.34% during test evaluation.

As can be seen, the performance given by these methods is accurate enough and they can be highly useful in this pandemic situation.

Keywords
COVID-19; Deep Learning; CNN; Xception Net; X-Ray

1. INTRODUCTION

Coronavirus disease (COVID-19) is an infectious disease and it has been spreading during 2020 throughout the whole world, reaching out to be considered a pandemic disease by World Health Organization (WHO) on 11th March 2020 [3]. The last pandemic disease before coronavirus was the H1N1 virus and it was declared in 1918, so this is the first pandemic caused in the last 100 years and also the first pandemic disease related to coronaviruses. This situation has raised the rate alarm of the governments and the concern about COVID-19. Due to it is a global health crisis, the countries are imposing social distancing, border restrictions, canceling international and domestic flights, and maximizing the conscience of their citizens regarding hygiene. At this point, 23rd April 2020, there are 2.7 million COVID-19 cases confirmed, almost 200,000 deaths, 720,000 recovered, 1.7 million mild (97%) and 58,189 (3%) critical cases [1]. Attending to experts, most of the people infected with COVID-19 face the following symptoms: sore throat, cough, fever, and respiratory illness and even causing deadly pneumonia in the most severe cases. This virus is really dangerous mostly for elderly people and for those who have cardiovascular diseases, diabetes, chronic respiratory disease, renal o hepatic diseases, and cancer [2].

In order to stop the spreading of COVID-19 disease and reduce the possible new cases, it is required to screen the patients and give them the medical response needed. For the time being, the unique screening method used for testing the COVID-19 patients is the Reverse Transcription Polymerase Chain Reaction (RT-PCR) test [4]. This technique has been using in all countries since the situation has become critical and it is a nuclear-derived method for detecting the presence of specific genetic material from any pathogen, including virus [5]. To make this test, a sample is collected for the patient where the coronavirus gathers, like throat or nose. The problem related to this kind of test is that it is manual, complicated, laborious and it takes around 24h to know the results. Besides that, this test is not manufactured is every country so many of them have scarcity. Due to this situation, not all the patients are being tested and that is why the results of the number of cases given by the country are incorrect [6]. If a large number of infected people are not being tested, they can be in contact with healthy people and transmit them the virus, spreading the disease and getting worse the situation.

There are other tools for COVID-19 diagnosis such as positive radiographic images (computed tomography o chest radiograph). As it said before, in severe cases the infection can become pneumonia-causing acute respiratory distress syndrome (ARDS) [7 - 10]. Therefore, Chest Radiograph (CXR) images can be really useful to detect the severe cases of COVID-19, however, this kind of image of positive COVID-19 cases do not differ too much from the cases of viral pneumonia and they overlap with other kinds of lung diseases. This makes it difficult to distinguish COVID-19 between another type of disease that affects the lungs. Besides, considering the current situation where it is necessary to make a diagnosis as soon as possible to take measurements, a radiologist can make an incorrect diagnosis due to be working round the clock. So, that incorrect diagnosis can expose a positive COVID-19 case to healthy people, resulting in more people infected and overloading more the hospitals with consequent more efforts and costs and risks. [11],[12]. Due to this type of diagnosis work with images, it is possible to introduce Artificial Intelligence (AI) and make a diagnosis with high accuracy and help in three ways: (1) make a better diagnosis, (2) make a faster diagnosis and (3) help to unload the hospitals.

Deep learning techniques can extract features from given images and classify them regarding some labels. Inside deep learning techniques, there are multiple types of architectures of neural networks that have shown to be useful to extract features from images introduced in the network and generalize the results for new data. The most common architecture for feature extraction and widely used for classifying X-Ray images are the Convolutional Neural Networks (CNN) [13]. The responsible to

Table 1: Details of Kaggle's COVID-19 Radiography Dataset [23]. The table is divided into two parts, with and without data augmentation. After introducing the data augmentation the number of images in each class is equalized.

Types	Database	Number of Images	Without data augmentation			With data augmentation		
			Total Images per class	*Train-Val set*	*Test set*	Total Images per class	*Train-Val set*	*Test set*
COVID -19	SIRM-ITALIAN [21]	70	219	152	44	1314	1051	263
	Novel Corona Virus 2019 Dataset [22]	60						
	Database from different architecture [23]	89						
Normal	Chest X-Ray Images (pneumonia) [24]	1341	1341	1073	268	1341	1073	268
Viral Pneumonia		1345	1345	1076	269	1345	1076	269

extract the features in this type of architecture are the convolutional layers and then are led to the fully connected layers in charge of classifying the given image between different classes. In research like [14], [15], [16] and [17] deep learning architectures have been implemented to detected disease with X-Ray images.

Recently, researches like [18] and [19], have reported deep-learning-based COVID-19 pneumonia detection techniques. In both pieces of research said before used the same dataset [20] and it will be the same dataset used in this research. In the [18] research, they create a Convolutional Neural Network from scratch and introducing data augmentation they obtained 95.64% of accuracy on the test set. On the other hand, in [19] they tried four pre-trained architectures (Resnet18, AlexNet, DenseNet201, and SqueezeNet) in two different classification problems: (1) classify COVID-19 and normal X-Ray images and (2) classify normal, viral and COVID-19 pneumonia images.

In this paper, we are going to follow the work done by the previous research and try to combine them. We are going to introduce data augmentation because as can be seen in [18] research the results obtained are improved notably and then different architecture is going to be implemented to classify COVID-19 disease cases from normal and viral pneumonia cases. In section (II), a brief description of the dataset and the method used in the research will be provided. In section (III), we present the experiments done to compare the different architectures implemented and the discussion of the results obtained in those experiments. Finally, in section (IV) we will present the conclusions of the research.

2. DATASET
The dataset that it is going to be used in this research is Kaggle's COVID-19 Radiography Database [23]. The dataset is public and it can be downloaded on Kaggle's website. This dataset is composed of three different types of images: (1) Chest X-Ray images belonging to COVID-19 infected patients, (2) Chest X-Ray images of healthy people, and (3) Chest X-Ray images of patients with viral pneumonia. Some of the images of this dataset are shown in Figure 1.

These images were collected from different sources. Chest X-Ray images belonging to a patient infected with COVID-19 were taken from Italian Society of Medical and Interventional Radiology (SIRM) COVID1-19 Database [21], Novel Corona Virus 2019 Dataset developed by Joseph Paul Cohen and Paul Morrison and Lan Dao in Github [22], and images extracted from 43 different publications. The normal and viral pneumonia images were taken from Kaggle's Chest X-Ray pneumonia dataset [24]. Currently, 20th April 2020, there are 219 COVID-19 positive images, 1341 normal images, and 1345 viral pneumonia images. There are not a large amount of COVID-19 positives images yet, but after seeing the impact of this pandemic disease, they might continue updating this dataset as they collect more images from new sources.

Figure 1. Examples of the images in Kaggle's COVID-19 Radiography Database [23]. Left. Picture of a patient infected with COVID-19. **Middle.** Picture of a healthy patient. **Right.** Picture of a patient with viral pneumonia

3. METHODOLOGY
To execute the training of the model for deep feature extraction, it is necessary to develop a good preprocessing of the images dataset. As the amount of COVID-19 X-ray images is much lower than the two other classes, data augmentation is implemented to the COVID-19 class. The preprocessing is required for two main reasons. First of all, it is important to deal with the imbalanced data as it might lead to a bad fitting of the model [25]. Secondly, having a larger dataset is crucial to improve the performance of the Deep learning model. In order to implement data augmentation, the original images from the COVID-19 class are being processed by the following methods; rotation, noise addition, crop, flip and rescale. This preprocessing multiplies the amount of original COVID-19 images by five, balancing this amount with the rest of the classes. The data augmentation has been implemented using functions from the *Imgaug* libraries [26].
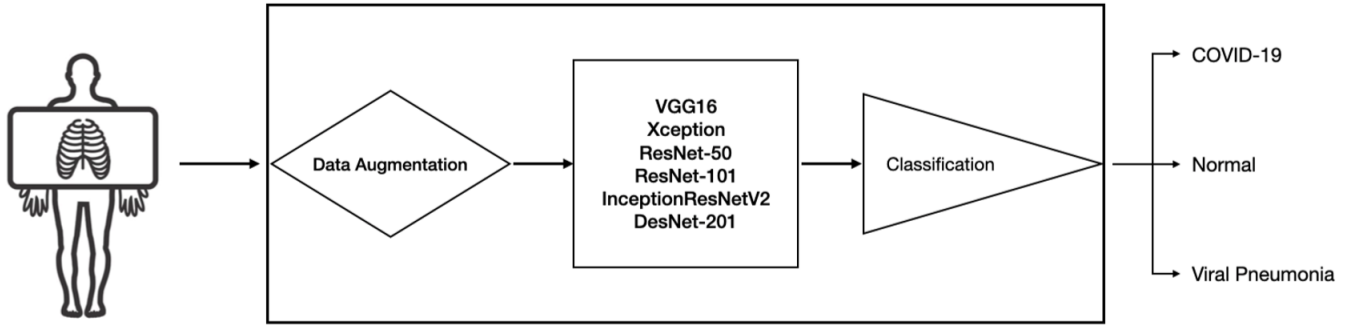
Figure 2. Block diagram of the Block Structure. The Chest X-Ray images are introduced in the Deep Learning Architecture, data augmentation is applied, then images are introduced into the CNN, and finally, the images are classified.

Table 2. Comparison of the performance of different architectures on *train, val,* and *test sets*. Each of these architectures is obtained after doing the optimization of the hyperparameters in order to obtain the best performance of each architecture.

	Optimizer	Training set		Validation set		Test set	
		Loss	Accuracy	Loss	Accuracy	Loss	Accuracy
VGG16	SGD	0.0278	0.9906	0.1783	0.9391	0.1512	0.9562
Xception	**SGD**	**0.0001**	**1.0000**	**0.1084**	**0.9688**	**0.1207**	**0.9737**
ResNet-50	SGD	0.0237	0.9961	0.2231	0.9515	0.1714	0.9587
ResNet-101	SGD	0.0298	0.9926	0.3570	0.9375	0.2480	0.9500
InceptionResNetV2	SGD	0.0010	0.9992	0.3066	0.9344	0.2114	0.9612
DenseNet201	SGD	0.0003	1.0000	0.2531	0.9468	0.1987	0.9525

these architectures, a final dense layer with 512 number of units and Relu activation, followed by batch normalization and dropout regularization, is added. Finally, the last dense layer with 3 number of units and SoftMax as the activation function Will give the predicted probabilities to classify the image. Usually, when implementing transfer learning from a pre-trained model, finetuning is applied in order to reduce the training time of the model, this means freezing some weights of the pre-trained CNN as they have already been trained to extract common specific features from the images. It is considered that finetuning is not worth to be applied in the Project as the images from the dataset are non-conventional and the features to be extracted from them might differ from a typical image classification model.

The value of the batch size for the training process is 32 and the number of epochs required to converge the model is 20. The optimizer used for compiling the model is SGD with a learning rate of 0.003 and a momentum of 0.9, the loss function is categorical cross-entropy and the metric used to evaluate the performance of the model is the accuracy. Additionally, the callback used to visualize the internal states, and the statistics of the model during training is ModelCheckpoint, which saves the model after every epoch the validation accuracy improves. Finally, the evaluation of each model is performed using the test data obtained from the data generator.

One of the main problems of applying Deep Learning technology to the biomedical field is that the doctors see the convolutional networks as a "black box" that cannot explain how they make their decisions. "If doctors do not understand why the algorithm made a diagnosis, then why should patients trust the recommended course of treatment?" [28]. To help with the study of understanding these "reasons", the Class Activation Map (CAM) can be very useful. The CAM is a simple technique to get the discriminative image regions used by the CNN to classify the input image. The implementation to obtain the CAM is performed on the Xception architecture, as it is the one that obtained the best accuracy. To create the CAM, the outputs from the last convolutional layer are taken, multiplied by their assigned weights and sum.

4. RESULTS AND DISCUSSION

The experimental studies were implemented using Python Jupiter Notebook. All applications were run on an MSI laptop equipped with NVIDIA GeForce GTX 1660 Ti GPU (Memory size: 6 GB | Base clock speed: 1500 MHz). The python environment is GPU enabled with CUDA and CuDNN so Tensorflow can be executed on the GPU to speed the training process up. The measurement of the performance of each classifier is measured in terms of accuracy. The results obtained after training and testing the model can be seen in Table 2 down below.
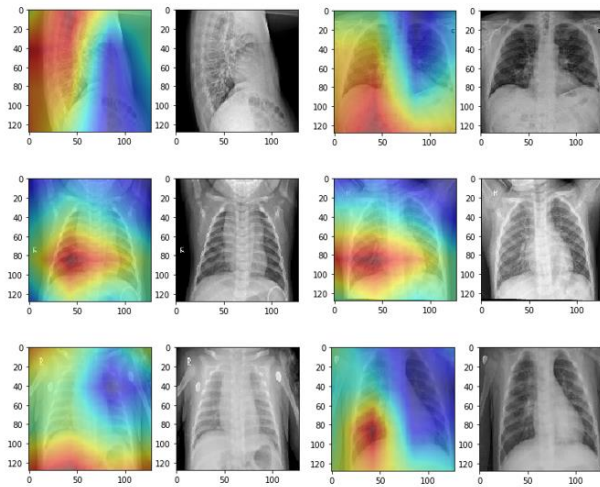
Figure 3. Class Activation Map examples. The red-colored area of the map shows where the model brings its attention to predict the images.

As can be seen in Table 2, the architecture that offered the best performance in both terms of loss and accuracy is the Xception Net. It is worth mentioning that during the search of the optimal hyperparameters for the training process, even accuracy of 0.98 was obtained when training the Xception Net with a batch size of 64. Unfortunately, training all the experiments with a batch size of 64 was not possible due to the lack of RAM memory on the GPU, giving TensorFlow allocated memory errors.

Xception Net is considered a modified version of the Inception-v3 Net. Xception was proposed by François Chollet, the creator and chief maintainer of the Keras library. Xception is an extension of the Inception architecture which replaces the standard Inception modules with depthwise separable convolutions [29]. The original depthwise separable convolution is the depthwise convolution followed by a pointwise convolution. The first depthwise convolution is the channel-wise n x n spatial convolution. As the images from the dataset have 3 channels (128x128x3) the architecture will have 3 n x n spatial convolution. Finally, the pointwise convolution is a 1x1 convolution to change the dimension [30].

The final part of the project is to create a Class Activation Map (CAM) in order to give information to the doctors and scientists that may help to understand how the model makes the classification. As the Xception Net is the one that gave the better performance, the CAM was applied to that specific architecture. The results obtained for six different well-classified images can be seen in Figure 5 - 7 down below. The first row of images corresponds to two well classified COVID images, the second row corresponds to Normal patients and the last one corresponds to Pneumonia images.

It is very difficult for the human eye to see what are the features extracted by the model that makes it classify the images. It can be appreciated that the right-inferior area of the thorax is where the model brings all its attention. These CAM images could be analyzed by an expert in order to find what are the exactly important features that can help to diagnose a patient with rather COVID-19 or Pneumonia.

5. CONCLUSION

Throughout this project, the impact of the Corona Virus Disease has been explained to aware of the importance of finding a feasible and efficient diagnosis predictor. Early diagnosis is crucial due to the limited amount of resources available during the pandemic. Implementing innovative technologies into the biomedical field has always been a challenge for humanity, and nowadays this innovative technology is Deep Learning. During this project, it has been demonstrated that it is feasible to predict Covid-19 using X-Ray images from the thorax. An accuracy of 97.37 % has been achieved with a limited amount of data of 229 Covid-19 diagnosed images. These results could be improved a lot by collecting new data and using more computational resources in order to search and find more optimal hyperparameters such as batch size. Apart from the great results obtained, the implementation of a class activation map can be very helpful, not only in this project but on every CNN application to the biomedical field. However, the results obtained by the CAM should always be analyzed by scientists, doctors, or experts with good background knowledge about the topic of interest and case study.

6. REFERENCES

[1] Worldometer International Team. (2020, April 20). "COVID-19 Coronavirus pandemic". [Online]. Available: https://www.worldometers.info/coronavirus/

[2] (2020). Coronavirus Disease 2019 (COVID-19). [Online]. Available: https://www.cdc.gov/coronavirus/risk.html

[3] (2020). WHO Director-General′s opening remarks at the media briefing on COVID-19 - 11 March 2020. [Online]. Available: https://www.who.int/dg/speeches/detail/

[4] W. Wang, Y. Xu, R. Gao, R. Lu, K. Han, G. Wu, et al., ″Detection of SARS-CoV-2 in Different Types of Clinical Specimens,″ Jama, 2020.

[5] N. Jawerth. IAEA Office of Public Information and Communication (2020, March 27). [Online]. Available: https://www.iaea.org/newscenter/news/

[6] A. J. NEWS. (2020). India′s poor testing rate may have masked coronavirus cases. [Online]. Available: https://www.aljazeera.com/news/2020/03/india-poor-testing-rate-masked-coronavirus-cases-200318040314568.html

[7] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, et al., ″Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China,″ Jama, 2020.

[8] N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, et al., ″Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study,″ The Lancet, vol. 395, pp. 507-513, 2020.

[9] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, et al., ″Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia,″ New England Journal of Medicine, 2020.

[10] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, et al., ″Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China,″ The Lancet, vol. 395, pp. 497-506, 2020.

[11] A. M. Tahir, M. E. Chowdhury, A. Khandakar, S. Al-Hamouz, M. Abdalla, S. Awadallah, et al., ″A systematic approach to the design and characterization of a smart insole

for detecting vertical ground reaction force (vGRF) in gait analysis,″ Sensors, vol. 20, p. 957, 2020.

[12] M. E. Chowdhury, A. Khandakar, K. Alzoubi, S. Mansoor, A. M Tahir, M. B. I. Reaz, et al., ″Real- Time Smart-Digital Stethoscope System for Heart Diseases Monitoring,″ Sensors, vol. 19, p. 2781, 2019.

[13] S. Shelly, A. Ben-Cohen, O. Shimon, M. M. Amitai, H. Greenspan, and E. Klang. "Convolutional neural networks for radiologic images: a radiologist's guide." Radiology 290, no. 3 (2019): 590-606.

[14] C. Vikash, S. K. Singh, A. Khamparia, D. Gupta, P. Tiwari, C. Moreira, R. Damaševičius, and V. H. C. de Albuquerque. "A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images." Applied Sciences 10, no. 2 (2020): 559.

[15] D. Gershgorn. (2017). The data that transformed AI research—and possibly the world. [Online]. Available: https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/

[16] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. Summers, ″Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases,″ in IEEE CVPR, 2017.

[17] P. Lakhani and B. Sundaram, ″Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks,″ Radiology, vol. 284, pp. 574-582, 2017.

[18] P. Lázaro and R. Barco, "COVID-19 Detection from chest X-Ray images using Convolutional Neural Network". (2020)

[19] C. Muhammad EH, T. Rahman, A. Khandakar, R. Mazhar, M. Abdul Kadir, Z. B. Mahbub, K. R. Islam et al. "Can AI help in screening Viral and COVID-19 pneumonia?." arXiv preprint arXiv:2003.13145 (2020).

[20] C. Muhammad EH, T. Rahman, A. Khandakar, R. Mazhar, M. Abdul Kadir, Z. B. Mahbub, K. R. Islam et al. "Can AI help in screening Viral and COVID-19 pneumonia?." arXiv preprint arXiv:2003.13145 (2020).

[21] Societa Italiana di Radiologia Medica e Interventistica. (2020). COVID-19 Database. [Online]. Available: https://www.sirm.org/category/senza-categoria/covid-19/

[22] J. C. Monteral. (2020). COVID-Chestxray Database. [Online] Available: https://github.com/ieee8023/covid-chestxray-dataset

[23] T. Rahman, M. E. H. Chowdhury, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. Al-Emadi, M. B. Ibne Reaz. (2020). COVID-19 CHEST RADIOGRAPHY DATABASE. [Online] Available: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database

[24] P. Mooney. (2018). Chest X-Ray Images (Pneumonia). [Online] Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

[25] J. Brownlee (2015). "8 Tactics to Combat Imbalanced Classes in Your Machine Learning Dataset" [Online]. Available: https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/

[26] A. Deis (2019) "Data Augmentation for Deep Learning" [Online]. Available: https://towardsdatascience.com/data-augmentation-for-deep-learning-4fe21d1a4eb9

[27] I. Michlin (2019) "Keras data generators and how to use them" [Online]. Available: https://towardsdatascience.com/keras-data-generators-and-how-to-use-them-b69129ed779c

[28] E. Bender. (2019). "Unpacking the black box in Artificial Intelligence for Medicine" [Online]. Available: https://undark.org/2019/12/04/black-box-artificial-intelligence/

[29] A. Rosebrock. (2017) "ImageNet: VGGNet, ResNet, Inception, and Xception with Keras" [Onlines]. Available: https://www.pyimagesearch.com/2017/03/20/imagenet-vggnet-resnet-inception-xception-keras/

[30] S. H. Tsang (2018) "Review: Xception — With Depthwise Separable Convolution, better than Inception-v3 (Image Classification)" [Online]. Available: https://towardsdatascience.com/review-xception-with-depthwise-separable-convolution-better-than-inception-v3-image-dc967dd42568