2020 2nd International Workshop on Human-Centric Smart Environments for Health and Well-being (IHSH) | 978-1-6654-4084-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/IHSH51661.2021.9378703

A Comparative Study of Deep Learning Networks for COVID-19 Recognition in Chest X-ray Images

Sabrina Nefoussi, Abdenour Amamra, Idir Amine Amarouche
Ecole Militaire Polytechnique
Bordj El-Bahri BP 17, Algiers, Algeria

***Abstract*—The COVID-19 pandemic is devastatingly affecting the health and well-being of the worldwide population. A basic advance in the battle against it resides in effective screening of infected patients, with one of the key screening approaches such as radiological imaging based on chest radiography. Faced with this challenge, various artificial intelligence (AI) frameworks, mostly based on deep learning, have been proposed and results have been getting better and very promising as the precision of positive cases recognition is constantly refined. In the light of previous work on automated X-ray image screening, we train several deep convolutional networks for the classification of chest pathologies into : *normal*, *pneumonia*, and *COVID-19*. We use three open-source and one private dataset for the validation of our findings. Unfortunately, data scarcity remains a big challenge hurdling COVID-19 automatic recognition research. In our case, we used a total of 518 COVID-19 positive X-ray images. We evaluate different architectures for COVID-19 recognition with different deep neural architecture.**

***Index Terms*—COVID-19, Coronavirus Pneumonia, Chest X-ray, Deep learning, Convolutional Neural Netwroks, Transfer Learning**

I. Introduction

Coronavirus has become a real medical threat all around the globe. The number of infected people and deaths are expanding quickly as per the updated reports of the World Health Organization. It is affirmed that this infection has taken over 1.26 million lives until the date of writing this paper. Since the start of its spreading, numerous Artificial Intelligence specialists proposed frameworks and strategies for foreseeing the infection's evolution or recognizing the disease. One of the potential methods for deciding whether a patient is contaminated with COVID-19 is achieved through the analysis of the chest X-ray images. As there is a large number of patients in hospitals, it would be tedious and hard to manually process a large number of X-ray images, therefore emerges the need for an automated solution.

The performance of automatic detection systems depends mainly on the training phase. The latter requires good quality and sufficient quantity of data to reach its full potential. The development of research on automatic detection of the COVID-19 in chest X-ray images remains modest, however, due to the insufficient amount of publicly available COVID-19 annotated data. In this study, we compare the performance of different architectures that have proven their performance on the natural recognition task.

The remainder of the paper is structured as follows. In Section, II we provide a review of the state-of-the-art deep convolutional neural architectures for COVID-19 chest X-ray classification. In Section III, we lay out the theoretical background of the tools used in our study. We later explain our methodology in Section IV, where we shed light on the utilized data and the tested models. Experimental results and comparative performance assessment deep learning classifiers are provided in Section V. Finally, the paper is concluded with Section VI where we summarize our contributions and we give some future directions.

II. Related Work

As highlighted by the World Health Organization, the COVID-19 pandemic is extremely challenging the best healthcare systems, worldwide, under enormous pressure. The early identification of this infection will help in alleviating the tension of healthcare organizations. Chest X-rays data have been playing a central role in the determination of infections like Pneumonia. As COVID-19 is a type of flu, it is conceivable to recognize it through chest X-ray screening. With rapid development in the area of Machine Learning (ML) and Deep learning, several methods had been proposed to distinguish pneumonia and normal cases.

Several works have been contributed to COVID-19 detection. The authors of [1] proposed a neural network for classifying X-ray images into three classes: normal, pneumonia, and COVID-19, using two public datasets. Their data contains only 180 X-ray COVID-19 samples. They also proposed a concatenation of Xception and ResNet50V2 networks to improve their accuracy.

In [2], the authors proposed a machine learning-based classification of COVID-19 and Pneumonia chest X-ray images using ResNet152. Only 62 COVID-19 X-ray images were used. Another work [3] introduced COVID-Net, a deep convolutional neural network for the detection of COVID-19 cases in chest X-ray images. Other authors [4] proposed COVIDX-Net, a deep learning framework that can be used to assist radiologists to automatically diagnose COVID-19 in X-ray images. They validated the study with only 50 Chest X-ray images with 25 confirmed positive COVID-19 cases. The proposed model included seven different architectures such as the modified VGG19 and the second version of Google MobileNet. The authors in [5] demonstrated how to detect the COVID-19 infection with chest X-rays using iteratively pruned deep learning model ensembles. To improve the performance of their approach, they used transfer learning and fine-tuning approaches.

In all cases, the application of deep learning strategies to distinguish and identify novel COVID-19 in x-ray data is still challenging because of the scarcity of publicly available data, ready and adequate for training learning algorithms. Also, besides its tediousness, the preparation of the training data in hospitals, under current global health conditions, remains of lower priority. Therefore, in this paper we extend the prior contributions of the different existing deep neural architectures in recognizing COVID-19 in X-ray imagery.

III. Pneumonia Chest X-ray Image Databases and Deep learning image classifiers

In this section, we discuss the most used pneumonia chest X-ray datasets in the literature for medical image analysis. Then, we describe some of the existing state-of-the-art deep learning image classifiers.

A. *Chest X-ray datasets*

The creation of a large annotated private image dataset is tedious and prone to error, reason for which researchers rely on the publicly available databases. Several studies have tried to create small COVID-19 datasets in order to encourage works on the automatic screening of the pathology. Among the public well-known datasets for viral (COVID-19 as well as other viruses) pneumonia detection and classification we mention:

- **COVID-19 Radiography Database** [6]: A team of researchers in collaboration with medical doctors has created a database of chest X-ray images for COVID-19 positive cases along with Normal and Viral Pneumonia images. In their current release, there are 219 COVID-19 positive images, 1341 normal images, and 1345 viral pneumonia images.
- **COVID Chest X-ray Dataset Master** [7]: This work is a database of confirmed or suspicious COVID-19 cases or other viral and bacterial pneumonia (MERS, SARS, and ARDS). The project is approved by the University of Montreal's Ethics Committee.
- **Mendeley Chest X-ray Images (Pneumonia)** [8] : it consists of 5856 X-Ray images (in JPEG image format) organized into two categories (Pneumonia/Normal), it only contains 1 to 5 years old children X-ray images. The particularity of this dataset is that the pneumonia type (viral or bacterial) is provided.

B. *Deep Image classification models*

In this section, we review some of the existing state-of-the-art deep learning image classifiers that have proven their performance.

- **VGG**: Visual Geometry Group Network was introduced in [9], Very Deep Convolutional Networks for Large Scale Image Recognition. In order to have improved image extraction functionality, the VGGNet used only small 3×3 convolutional layers stacked on top of each other in increasing depth. There are two versions of this deep network architecture; namely, VGG16 and VGG19, The "16" and "19" stand for the number of weight layers in the network. this network is painfully slow to train due to its heavy classifier.
- **Inception**: Inception network or GoogLeNet micro-architecture was first introduced in [10]. This network is based on a multi-level feature extractor by computing 1×1, 3×3, and 5×5 convolutions within the same module of the network. Later versions are referred to as Inception VN where N is the version number.
- **ResNet50**: This architecture [11] relies on micro-architecture modules, or elementary building blocks of the network. Even though ResNet is much deeper than VGG16 and VGG19, the model size is smaller due to the usage of global average pooling rather than fully-connected layers in the classifier.
- **Xception**: Was proposed in [12], which is an extension of the Inception architecture. In this architecture, the standard Inception modules were replaced with depthwise separable convolutions.
- **Mobilenet**: This network [13] was designed to maximize accuracy while minimizing the computational burden common to most modern deep neural networks. This network has the particularity of running on smartphones.

Fig. 1. Example of COVID-19 chest X-ray image

IV. Methodology

In this section, we present our strategy, the different architectures tested on our selected dataset and the performance of each of the above-mentioned models. We study the behavior of each network on chest X-ray images with COVID-19. Our primary focus is to discover the most adequate models, for COVID screening, by testing different combinations.

A. *Data description:*

Since COVID-19 image datasets are scarce, we decided to combine data from three open-source pneumonia chest X-rays datasets and an Algerian private COVID-19 chest X-rays dataset. The first dataset dubbed *COVID Chest X-Ray Dataset Master* [7] available from a GitHub repository, consists of X-ray and CT scan images of patients infected by COVID-19 and other types of pneumonia. We only considered the X-ray

Fig. 2. Example of viral pneumonia chest X-ray image

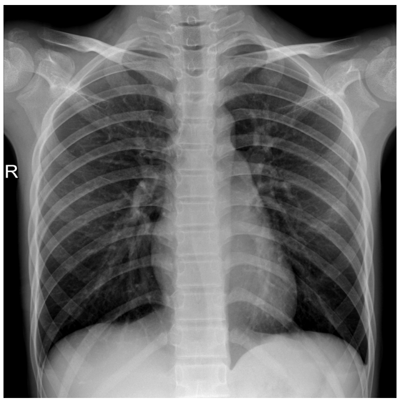

Fig. 3. Example of normal chest X-ray image

TABLE I
EXPERIMENTAL DATASETS

Classes	COVID-19	Normal	Viral Pneumonia
[6][1]	219	1341	1345
[7][2]	261	-	-
[8][3]	-	1583	1493
Algerian Hospital data	38	-	-
Total	**518**	**2924**	**2838**

[1]https://www.kaggle.com/tawsifurrahman/covid19-radiography-database
[2]https://github.com/ieee8023/covid-chestxray-dataset
[3]https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

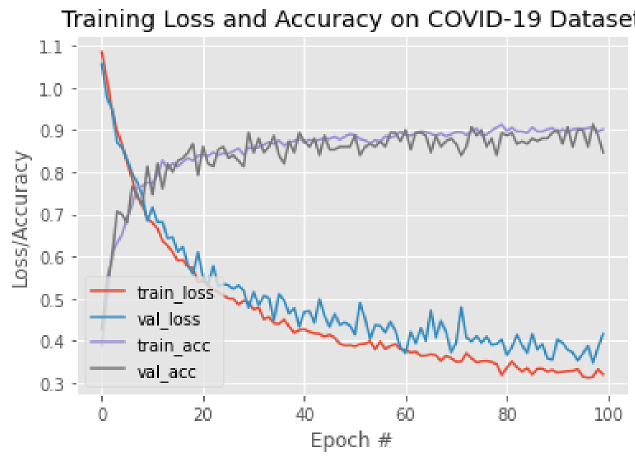

Fig. 4. VGG19 training and validation performance progress on COVID-19 dataset

images, that contain a total of 261 COVID-19 samples. The second dataset *COVID-19 Radiography Database* [6], contains 219 COVID-19 positive images, 1341 normal images and 1345 viral pneumonia images. The third dataset considered was *Mendeley ChestX-Ray Images(Pneumonia)* [8], which contains X-Ray images divided into 2 categories : Pneumonia (non COVID-19) and Normal. The latter has an important role in increasing the quality of classification with respect to the sensitivity between COVID-19 and non-COVID-19 pneumonia. The last dataset is private, it contains 38 chest X-ray images collected in an Algerian hospital from COVID-19 patients. The details of the datasets are given in Table. I.

Models trained on imbalanced dataset will show bias in favor of majority class, and may ignore the minority class altogether, in extreme cases [14]. In our study, we considered only balanced data classification, meaning that all classes contain the same number of samples. We will consider the following three classes of chest X-ray images: *Normal* for healthy cases without any pulmonary pathological traces, *COVID-19* corresponding to patients affected by COVID-19 related pneumonia, and *Viral Pneumonia* for patients suffering from non-COVID-19 pneumonia. Since the COVID-19 class, our class of interest, contains the minimum number of samples, we apply random under-sampling (RUS) to discard random samples from alternative classes in order to balance the sample number among categories; thus resulting in 518 samples for each class. 363 images were randomly selected from the 518 for training, amounting to 70% of the total number of considered images. the remaining data was divided into 10% for validation, amounting to 51 images, and 20%, 104 images, for the test. It is worth mentioning that all the training, validation, and test data were selected randomly.

B. Architecture:

In order to properly achieve our study, we selected eight of the best performing deep learning models in the image recognition domain. We evaluated the performance of the following networks pre-trained that have been trained on the ImageNet challenge dataset : (1) VGG16, (2) Inception ResNet v2, (3) Incepetion v3, (4) NasNet Large, (5) ResNet50, (6) Xception, (7) Mobilenet and (8) VGG19. Each deep neural network is able to analyze the chest X-ray image in order to classify a sample as Normal, positive COVID-19, or positive Viral Pneumonia. All images are resized to 400×400 pixels. We applied different augmentation on the training data before feeding it to the models, such as rotation, flipping, shearing transformations and zoom. As preprocessing function, we used contrast limited adaptive histogram equalization technique to improve contrast in chest X-rays images. According to the Pareto principle, 80% of data were used for training and validation of the eight models. We trained each model for 100 epochs, Fig. 4 illustrates an example of the training performances of VGG19.

We test the trained model on 20% of data, which should class each sample into one of the three classes. Then, we compute the evaluation metrics to assess Their respective performances on the test data.

V. Results and discussion

We used four performance metrics derived from the confusion matrix to evaluate the models :

Accuracy: the ratio of correct predictions to the total number of predictions,

$$Acc = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

Precision or positive predictive value is the ratio of good positives from the total detected positives:

$$Pr = \frac{TP}{(TP + FP)} \tag{2}$$

Recall or sensitivity : corresponds to the probability of correctly classifying pneumonia sample in our case,

$$R = \frac{TP}{(TP + FN)} \tag{3}$$

F-score: is a measure that combines both precision and recall as determined in the formula:

$$F_{score} = \frac{2 \times Pr \times R}{(Pr + R)} \tag{4}$$

Table. II shows the accuracy results for each experiment conducted through the eight different models and on the test set. We recall that we evaluate the models' classification performance on a balanced dataset, we used a test support of 518 for each class among the three classes. VGG16, Xception, and Mobilenet models gave the best result in terms of accuracy with a value of 0.94, then comes the Inception model with an accuracy of 0.93, VGG19 with 0.92, for the Inception Resnet V2 and Nasnet Large models, they reached the value of 0.91, the least accurate model was ResNet50 with only 0.85 of accuracy.

The precision performance analysis for each deep learning classifier is illustrated in Table. III. For our class of interest (COVID-19), the most precise networks are Inception Resnet V2, Xception, and MobileNet. However, the Resnet50 model got the worst value of 0.83. VGG models recorded the highest precision of 0.89 for Normal class, as opposed to Nasnet Large network with only 0.79. Regarding Viral Pneumonia class, Nasnet Large network had the best precision and ResNet50 the worst one, compared to the other networks.

Table. IV shows a comparison of the recall achieved by the different models across the three classes. It can be seen that the Xception model achieves the best result for the COVID-19 class, while VGG19 was the worst with 0.93. Nasnet Large model gave the best value for the Normal case, as opposed to Resnet model with only 0.82. For the Viral Pneumonia, the models had relatively inferior recall results which were below 0.95, reached by Inception Resnet V2, and greater than 0.74 recorded by Nasnet Large.

TABLE II
Accuracy Results of different tested models

Accuracy			
VGG16	**Inception resnet v2**	**Incepetion v3**	**Nasnet Large**
0.94	0.91	0.93	0.91
ResNet50	**Xception**	**Mobilenet**	**VGG19**
0.85	0.94	0.94	0.92

TABLE III
Precision for each infection type over the different models.

Precision			
Classes	**COVID-19**	**Normal**	**Viral Pneumonia**
VGG16	0.99	0.89	0.95
Inception resnet v2	1.00	0.84	0.92
Incepetion v3	0.99	0.85	0.96
Nasnet Large	0.99	0.79	1.00
RESNET50	0.83	0.88	0.85
Xception	1.00	0.86	0.97
Mobilenet	1.00	0.86	0.99
VGG19	0.95	0.89	0.91

The last metric considered in the evaluation is the F-Score criterion, Table. V summarizes the results of the different models. The Xception model achieved the best results for COVID-19 class, compared with VGG19 model that was less efficient in terms of F-Score with 0.94. The VGG16 model had the best F-score on the two other classes (Normal, Viral Pneumonia) with 0.93 and 0.92 respectively. 0.85, 0.81 are F-Score results of ResNet50 for Normal and Viral Pneumonia cases, as the worst results in these classes.

The models that achieved the best results of accuracy are the most capable of classifying images properly, such as VGG16, Xception, and MobileNet. The highest precision is correlated with correct identification of the positive cases, for example for the class of interest COVID-19, all samples in the test set were classified as positive cases by Xception model were originally positive. The recall expresses the performance of detecting all positive cases correctly, in other terms the 518 COVID-19 cases were all identified by the Xception model, and this is the most important case because of the high contamination speed of this virus.

VI. Conclusion

In this study, we trained several deep convolutional networks for classifying X-ray images into three classes: Normal, Viral pneumonia, and COVID-19, on four different datasets. This multitude of data sources mainly contributes to better generalization properties for the trained networks. Unfortunately, the data of COVID-19 is still limited in terms of quantity, which presents a huge challenge for the development of automatic classification systems. Our data contains more than 500 X-ray images that belong to COVID-19 positive cases, which presents a significant number compared with previous work in this field. This research mainly involves comparing the performance of several networks on COVID-19 classification task. The presence of the viral pneumonia class plays an

TABLE IV
RECALL FOR EACH INFECTION TYPE OVER THE DIFFERENT MODELS.

Recall			
Classes	**COVID-19**	**Normal**	**Viral Pneumonia**
Vgg16	0.98	0.96	0.88
Inception resnet v2	0.95	0.94	0.95
Incepetion v3	0.99	0.96	0.84
Nasnet Large	0.99	1.00	0.74
RESNET50	0.97	0.82	0.77
Xception	1.00	0.97	0.84
Mobilenet	0.98	0.99	0.86
VGG19	0.93	0.94	0.88

TABLE V
F-SCORE FOR EACH INFECTION TYPE OVER THE DIFFERENT MODELS.

F-Score			
Classes	**COVID-19**	**Normal**	**Viral Pneumonia**
Vgg16	0.99	0.93	0.92
Inception resnet v2	0.98	0.89	0.88
Incepetion v3	0.99	0.90	0.89
Nasnet Large	0.99	0.89	0.85
RESNET50	0.90	0.85	0.81
Xception	1.00	0.91	0.90
Mobilenet	0.99	0.92	0.92
VGG19	0.94	0.92	0.89

important role in increasing the sensitivity of the model. Among all the eight models tested, we can assert that Xception model achieved the best results according to all the evaluation criteria. Our work contributes to the new line of research aimed at the early detection of COVID-19 cases thus helping doctors in the diagnosis of this disease. As perspective, we will carry on with the detection of the pulmonary areas affected with the virus in order to drive the focus of the radiologists to the region of interest, with labeled tagging indicating placement of abnormalities in the chest x-ray. Also, we will examine cross-Validation that helps when we are dealing with medical challenges.

REFERENCES

[1] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting covid-19 and pneumonia from chest x-ray images based on the concatenation of xception and resnet50v2," *Informatics in Medicine Unlocked*, vol. 19, p. 100360, 2020. [Online]. Available: http://dx.doi.org/10.1016/j.imu.2020.100360

[2] R. Kumar, R. Arora, V. Bansal, V. J. Sahayasheela, H. Buckchash, J. Imran, N. Narayanan, G. N. Pandian, and B. Raman, "Accurate prediction of covid-19 using chest x-ray images through deep feature learning model with smote and machine learning classifiers," *medRxiv*, 2020. [Online]. Available: https://www.medrxiv.org/content/early/2020/04/17/2020.04.13.20063461

[3] L. Wang and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," 2020.

[4] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," 2020.

[5] S. Rajaraman, J. Siegelman, P. O. Alderson, L. S. Folio, L. R. Folio, and S. K. Antani, "Iteratively pruned deep learning ensembles for covid-19 detection in chest x-rays," *IEEE Access*, vol. 8, pp. 115 041–115 050, 2020.

[6] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz, and M. T. Islam, "Can ai help in screening viral and covid-19 pneumonia?" *IEEE Access*, vol. 8, pp. 132 665–132 676, 2020. [Online]. Available: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database/data

[7] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "Covid-19 image data collection: Prospective predictions are the future," *arXiv 2006.11988*, 2020. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset

[8] G. M. Kermany Daniel, Zhang Kang, "Large dataset of labeled optical coherence tomography (oct) and chest x-ray images,mendeley data, v3," 2018.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[12] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," 2017.

[13] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017.

[14] J. Van Hulse, T. M. Khoshgoftaar, and A. Napolitano, "Experimental perspectives on learning from imbalanced data," in *Proceedings of the 24th International Conference on Machine Learning*, ser. ICML '07, New York, NY, USA, 2007, p. 935–942. [Online]. Available: https://doi.org/10.1145/1273496.1273614