Comprehensive Analysis of the Screening of COVID-19 Approaches in Chest X-ray Images from Portable Devices

Daniel I. Morís[1, 2], J. de Moura[1, 2], J. Novo[1, 2] and M. Ortega[1, 2, *]

1- Centro de Investigación CITIC, Universidade da Coruña, 15071, A Coruña, Spain

2- Grupo VARPA, Instituto de Investigación Biomédica de A Coruña (INIBIC), Universidade da Coruña, 15006, A Coruña, Spain

Abstract. Computer-aided diagnosis plays an important role in the COVID-19 pandemic. Currently, it is recommended to use X-ray imaging to diagnose and assess the evolution in patients. Particularly, radiologists are asked to use portable acquisition devices to minimize the risk of cross-infection, facilitating an effective separation of suspected patients with other low-risk cases. In this work, we present an automatic COVID-19 screening, considering 6 representative state-of-the-art deep network architectures on a portable chest X-ray dataset that was specifically designed for this proposal. Exhaustive experimentation demonstrates that the models can separate COVID-19 cases from NON-COVID-19 cases, achieving a 97.68% of global accuracy.

1 Introduction

The COVID-19 is a highly infectious disease that was first detected in Wuhan, China at the end of 2019. This disease is caused by the novel coronavirus SARS-CoV-2 and due to the rapid spread of this pathogen, the World Health Organization (WHO) declared the global pandemic in March, 2020 [1]. The symptomatology of this disease can be difficult to distinguish from other common pulmonary infections as those caused by the flu or the pneumonia [2]. The most common symptoms are tiredness, dry cough and fever. Moreover, the most severe cases could lead patients to need mechanical ventilation and ICU admission.

In order to diagnose and study the evolution of this relevant disease in every patient, it is very important to have fast and reliable methods to do so. In the context of the pulmonary diseases, the chest X-ray imaging is a well-established image modality typically used to assess the diagnosis of common pathologies as tuberculosis, pneumonia or lung nodules, among many others [3]. In the global

pandemic context, the expert clinicians are strongly encouraged to prioritize the use of chest radiography for early evaluation over any other image modality.

As the COVID-19 diagnosis is a very critical task in the current moment, several works have demonstrated the potential of the automatic COVID-19 screening in chest X-ray images [4, 5, 6, 7, 8, 9, 10]. However, it should be noted that these works used datasets composed of images that were acquired with fixed chest X-ray devices. In contrast, due to the importance of cutting the chains of transmission, critical to overcome the high spread of the SARS-CoV-2, the American College of Radiology (ACR) recommends the use of portable chest X-ray devices, as this kind of devices can reduce the risk of cross-contamination [11]. Nevertheless, these acquisition devices provide images with a lower quality, due to a poorer contrast, and a lower level of detail in comparison with the fixed machinery, an aspect that hinders the COVID-19 diagnosis. Consequently, this also makes the development of a reliable automatic method a more challenging process. To the best of our knowledge, only De Moura *et al.* [12] have addressed the problem of automatic COVID-19 screening in chest X-ray images that were acquired with portable devices. However, despite that this work shows promising results, this problem is partially addressed, as only a superficial analysis of the problem is performed using a single convolutional network architecture with a small dataset only composed of 240 COVID-19 cases.

In this work, we propose a fully automatic COVID-19 screening methodology on a portable chest X-ray imaging dataset. The images were provided using portable acquisition devices in a real clinical context to separate among COVID-19 cases and NON-COVID-19 cases (*i.e.*, patients that have pulmonary pathologies others than COVID-19 or patients without pleural or pulmonary affectations). Despite a lower level of detail, the portable chest X-ray images are commonly used in healthcare services as these are more suitable to reduce the risk of cross-contamination. To achieve the objectives of this work, we tested 6 representative state-of-the-art deep pretrained network architectures to comprehensively evaluate the performance of automatic COVID-19 screening using a dataset designed specifically for this purpose. For the validation, we consider a more heterogeneous, representative and extensive chest X-ray dataset with a significant amount of COVID-19 cases.

2   Materials and Methods

In this section, we present the portable chest X-ray dataset used in this work (Subsection 2.1) as well as the proposed methodology (Subsection 2.2). This methodology is illustrated in Fig. 1. Particularly, 6 different representative deep architectures were trained. For each architecture, we evaluate the separability between the COVID-19 class and the NON-COVID-19 class.

2.1   Dataset

The chest X-ray samples were provided by the Radiology Service of the Complexo Hosipitalario Universitario de A Coruña (CHUAC), obtained with portable de-
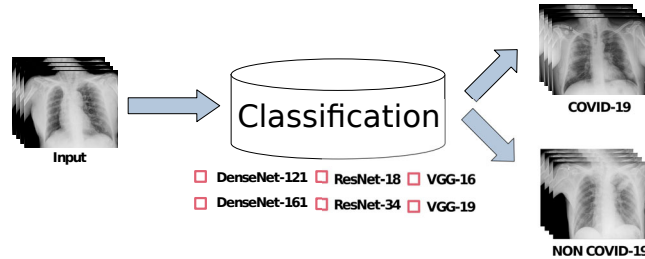
Fig. 1: A diagram of the proposed methodology.

vices. Particularly, the acquisition device models were Agfa dr100E and Optima Rx200. This kind of devices are more versatile and flexible to aid in the diagnosis in the context of the healthcare services. Thus, the dataset is composed of 3584 chest X-ray images divided into 2071 COVID-19 cases, 716 pathological cases and 797 normal cases. It is remarkable that the dataset is then composed of samples that present COVID-19 affectation (COVID-19 class), that present other pulmonary or pleural affectations (pathological class) and samples without shows of this type of affectations that, however, could present pathologies that affect other regions (normal class).

2.2   Analysis of automatic COVID-19 screening

To exhaustively assess the performance of the automatic COVID-19 screening, we considered 6 different representative deep network architectures: 2 Dense Convolutional Network Architectures [13] (a DenseNet-121 and a DenseNet-161, with a depth of 121 and 161, respectively), 2 Residual Network Architectures [14] (a ResNet-18 and a ResNet-34, with 18 and 34 deep layers, respectively) as well as 2 VGG network architectures [15] (a VGG-16 and a VGG-19, both with batch normalization as well as with 16 and 19 weighted layers, respectively). They were selected due to their suitability in similar problems. For every network architecture, we evaluate the separability among COVID-19 and NON-COVID-19 chest X-ray images. In this way, the approach herein proposed is designed for distinguishing COVID-19 not only from normal patients but also to differentiate them from patients with pulmonary conditions others than COVID-19. On the other hand, we respect a proportion of $\frac{2}{3}$ and $\frac{1}{3}$ for NON-COVID-19 and COVID-19, respectively, in order to keep a balance between negative and positive classes. Thus, we consider a random subset of 716 samples from the COVID-19 class alongside a random subset of 1432 samples from the NON-COVID-19 class.

In this context, the dataset was randomly partitioned in 3 different subsets, with the 60% of the images for training, 20% of the images for validation and the remaining 20% for test. For the network initialization, all the models were pretrained in the ImageNet dataset. Moreover, to optimize the network weights, the algorithm of Stochastic Gradient Descent (SGD) was considered, with a constant learning rate of $\alpha = 0.01$, a mini-batch size of 4, a first-order momentum

of 0.9 and the cross-entropy was considered as the loss function. It is important to note that, in order to exhaustively obtain representative values of the performance of the models, each training process was repeated 5 times.

2.3 Evaluation

The performance evaluation of the models was obtained by comparing the manual annotations that were made in the dataset against the outputs predicted by the models. Regarding the values of True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN), we considered the following commonly used evaluation metrics: precision, recall, F1-Score and accuracy.

3 Experimental results

The evolution of the training and validation accuracy values is depicted in Fig. 2. These values represent the mean accuracy obtained after 5 independent repetitions. It is clearly observed that globally all the models achieve satisfactory convergences, since the accuracy reaches values higher than 90% in all the cases.
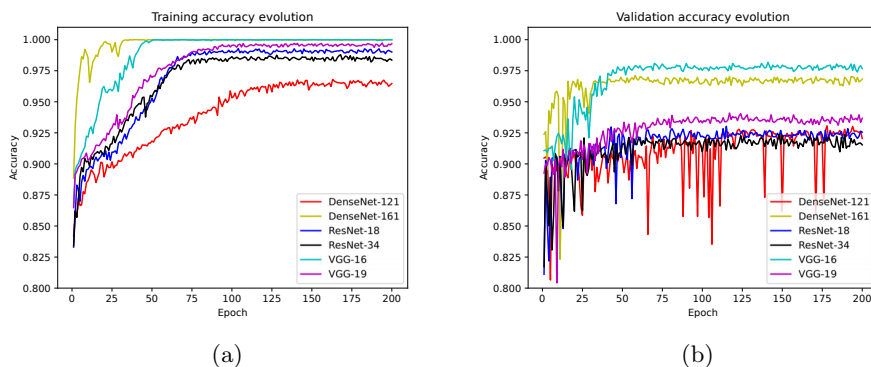

(a)   (b)

Fig. 2: Evolution of the accuracy values for 6 deep network architectures. *(a) Training accuracy evolution. (b) Validation accuracy evolution.*

Table 1 shows the results of the test step in terms of precision, recall and F1-Score for each class, after the proposed experiments were conducted. It can be clearly seen that the VGG-16 model achieves the best values in metrics of precision, recall and F1-Score for both classes, as all of these metrics have values greater than 95.60%. Additionally, these results reflect that all models are able to separate accurately the samples from both classes. It is also remarkable that the performance of the models is generally higher when classifying NON-COVID-19 samples, as all the values surpass 94.00% in terms of F1-Score for this class, while for COVID-19 samples these values are between 87.40% and 96.60%. Moreover, the experimental validation clearly depicts the robustness of the models, with a standard deviation always lower than 6.29%. Particularly, the VGG-16 results

Architecture	Class	Precision	Recall	F1-Score
DenseNet-121	NON-COVID-19	93.80%±2.71%	97.20%±1.72%	95.20%±1.47%
	COVID-19	93.80%±3.43%	86.00%±6.29%	89.80%±3.71%
DenseNet-161	NON-COVID-19	97.00%±1.41%	98.60%±0.49%	97.60%±0.80%
	COVID-19	96.60%±1.02%	94.00%±2.45%	95.40%±1.02%
ResNet-18	NON-COVID-19	94.40%±1.74%	95.20%±2.64%	94.80%±0.98%
	COVID-19	89.80%±4.53%	88.20%±4.02%	88.80%±1.33%
ResNet-34	NON-COVID-19	92.80%±1.72%	95.40%±1.85%	94.00%±0.63%
	COVID-19	90.40%±3.38%	85.60%±3.14%	87.40%±0.80%
VGG-16	NON-COVID-19	**98.00%±0.63%**	**98.60%±1.02%**	**98.40%±0.49%**
	COVID-19	**97.20%±1.47%**	**95.60%±1.02%**	**96.60%±1.02%**
VGG-19	NON-COVID-19	95.00%±2.28%	96.60%±1.50%	95.60%±1.85%
	COVID-19	92.20%±3.06%	89.60%±4.84%	91.00%±3.35%

Table 1: Mean precision, recall and F1-Score values obtained on each experiment. The best model results are highlighted in bold.

are notably robust, as all the standard deviation values are lower than 1.47%. It is important to remark that the training, validation and test accuracy values are very similar and, therefore, this depicts the absence of overfitting. In the same way, these results are certainly obtained thank to the used dataset, that has a representative amount of chest X-ray samples with COVID-19 affectation. In terms of global performance, the models achieved the following mean accuracy values: a $93.60\% \pm 2.24\%$ of mean accuracy for the DenseNet-121, a $96.84\% \pm 0.78\%$ for the DenseNet-161, a $92.81\% \pm 1.03\%$ for the ResNet-18, a $92.06\% \pm 0.54\%$ for the ResNet-34, a $97.68\% \pm 0.49\%$ for the VGG-16 and a $94.06\% \pm 2.29\%$ for the VGG-19. Therefore, it can be observed that all models obtain an accuracy higher than 90%, concluding that the network architecture with the best accuracy value is the VGG-16. Results also demonstrate the robustness of the models. Particularly, the model with the best robustness is the VGG-16 with a standard deviation lower than 0.5%.

4   Conclusion

This work proposes a fully automatic methodology for COVID-19 screening using chest X-ray images that were provided by portable devices from a real clinical scenario. The used dataset was designed specifically for this analysis in collaboration with the Radiology Service of the Complexo Hospitalario Universitario de A Coruña (CHUAC) with a significant amount of COVID-19 samples. As portable chest X-ray images provide a lower quality in contrast with the fixed machinery, this represents a challenging scenario for the automatic screening methods. Particularly, several experiments were conducted considering 6 representative state-of-the-art deep pretrained network architectures to analyze the degree of separability among COVID-19 and NON-COVID-19 samples. The experimental validation shows accurate and robust results, obtaining high values for all the considered metrics. Particularly, we achieved a best mean global accuracy of 97.68% using a VGG-16 architecture. In the same way, this idea is reinforced as precision, recall and F1-Score values are higher than 95% in this particular case for both classes. Despite the obtained results of the VGG-16

were very satisfactory, it could be interesting to test with a bigger dataset to explain this behavior more effectively. As future works, this analysis could be also performed in relevant and useful scenarios others than COVID-19 vs NON-COVID-19. In the same way, we could test other aspects, as the computational cost of each approach, their ease of use and other elements that could be relevant in a practical context.

References

[1] Casey A Pollard, Michael P Morran, and Andrea L Nestor-Kalinoski. The covid-19 pandemic: a global health crisis. *Physiological Genomics*, 52(11):549–557, 2020.

[2] Souheil Zayet, Quentin Lepiller, Hajer Zahra, Pierre-Yves Royer, Lynda Toko, Vincent Gendrin, Timothée Klopfenstein, et al. Clinical features of covid-19 and influenza: a comparative study on nord franche-comte cluster. *Microbes and infection*, 22(9):481–488, 2020.

[3] AM dos Santos, B de B Pereira, JM de Seixas, FCQ Mello, and AL Kritski. Neural networks: an application for predicting smear negative pulmonary tuberculosis. In *Advances in statistical methods for the health sciences*, pages 275–287. Springer, 2007.

[4] Jianpeng Zhang, Yutong Xie, Yi Li, Chunhua Shen, and Yong Xia. Covid-19 screening on chest x-ray images using deep learning based anomaly detection. *arXiv preprint arXiv:2003.12338*, 27, 2020.

[5] Sanhita Basu, Sushmita Mitra, and Nilanjan Saha. Deep learning for screening covid-19 using chest x-ray images. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 2521–2527. IEEE, 2020.

[6] Xin Li, Chengyin Li, and Dongxiao Zhu. Covid-mobilexpert: On-device covid-19 screening using snapshots of chest x-ray. *arXiv preprint arXiv:2004.03042*, 2020.

[7] Linda Wang, Zhong Qiu Lin, and Alexander Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports*, 10(1):1–12, 2020.

[8] Dilbag Singh, Vijay Kumar, Vaishali Yadav, and Manjit Kaur. Deep neural network-based screening model for covid-19-infected patients using chest x-ray images. *International Journal of Pattern Recognition and Artificial Intelligence*, 35(03):2151004, 2021.

[9] Ankita Shelke, Madhura Inamdar, Vruddhi Shah, Amanshu Tiwari, Aafiya Hussain, Talha Chafekar, and Ninad Mehendale. Chest x-ray classification using deep learning for automated covid-19 screening. *medRxiv*, 2020.

[10] Joaquim de Moura, Jorge Novo, and Marcos Ortega. Fully automatic deep convolutional approaches for the analysis of covid-19 using chest x-ray images. *medRxiv*, 2020.

[11] Soheil Kooraki, Melina Hosseiny, Lee Myers, and Ali Gholamrezanezhad. Coronavirus (covid-19) outbreak: what the department of radiology should know. *Journal of the American college of radiology*, 17(4):447–451, 2020.

[12] Joaquim De Moura, Lucía Ramos García, Plácido Francisco Lizancos Vidal, Milena Cruz, Laura Abelairas López, Eva Castro Lopez, Jorge Novo, and Marcos Ortega. Deep convolutional approaches for the analysis of covid-19 using chest x-ray images from portable devices. *IEEE Access*, 8:195594–195607, 2020.

[13] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[15] Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv 1409.1556*, 09 2014.