Convolutional Neural Networks Evaluation for COVID-19 Classification on Chest Radiographs

Felipe André Zeiser [1] **Cristiano André da Costa** [1] **Gabriel de Oliveira Ramos** [1]

Abstract

Early identification of patients with COVID-19 is essential to enable adequate treatment and to reduce the burden on the health system. The gold standard for COVID-19 detection is the use of RT-PCR tests. However, due to the high demand for tests, these can take days or even weeks in some regions of Brazil. Thus, an alternative for the detection of COVID-19 is the analysis of Chest X-rays (CXR). This paper proposes the evaluation of convolutional neural networks to identify pneumonia due to COVID-19 in CXR. The proposed methodology consists of an evaluation of six convolutional architectures pre-trained with the ImageNet dataset: InceptionResNetV2, InceptionV3, MovileNetV2, ResNet50, VGG16, and Xception. The obtained results for our methodology demonstrate that the Xception architecture presented a superior performance in the classification of CXR, with an Accuracy of 85.64%, Sensitivity of 85.71%, Specificity of 85.65%, F1-score of 85.49%, and an AUC of 0.9648.

1. Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is a new beta-coronavirus first identified in December 2019 in Wuhan province, China (Andersen et al., 2020). SARS-CoV-2 spread rapidly across the globe, challenging health systems (Shereen et al., 2020). Currently, the gold standard for the diagnosis of COVID-19 is the Reverse Transcription Polymerase Chain Reaction (RT-PCR) (Marson, 2020). However, due to the difficulties in the purchase and increase in the prices of materials and equipment, the lack of laboratories and qualified professionals, and the high demands for tests of the RT-PCR test, the diagnosis can take days or even weeks in some cities in Brazil (Marson, 2020).

As alternatives to RT-PCR, radiological exams such as Computed Tomography (CT) and Chest X-Ray (CXR) are being used as valuable tools for the detection and definition for treatment of patients with COVID-19 (Cohen et al., 2020). Studies point out sensitivities equivalent to the RT-PCR test using CT images (Ai et al., 2020; Xie et al., 2020). Pulmonary changes can be observed even in asymptomatic patients of COVID-19, indicating that the disease can be detected by CT even before symptoms appear (Lee et al., 2020).

CXR has lower sensitivity indices compared to CT. However, due to some challenges in CT, health systems have adopted CXR in the front line for the screening and monitoring of COVID-19, as in the United Kingdom and Italy (Borghesi & Maroldi, 2020). Among the main challenges in the use of CT compared to CXR are: (i) cost of the exam; (ii) increased risk for cross-infection; and (iii) more significant exposure to radiation (Wong et al., 2020). In addition, in underdeveloped countries, the infrastructure of health systems generally does not allow the use of RT-PCR tests or obtaining CT images for all suspected cases. However, devices for obtaining radiographs are already more widespread and can serve as a fundamental instrument in combating the epidemic in these countries (Cohen et al., 2020).

In this sense, this work aims to evaluate Convolutional Neural Networks (CNN) architectures to classify chest X-rays. In particular, we intend to provide a set of CNN models to serve as screening and diagnostic support systems for suspected or confirmed patients with COVID-19. Thus, we evaluated six convolutional architectures: InceptionResNetV2 (Szegedy et al., 2017), InceptionV3 (Szegedy et al., 2016), MovileNetV2 (Sandler et al., 2018), ResNet50 (He et al., 2016), VGG16 (Simonyan & Zisserman, 2014), and Xception (Chollet, 2017). The main scientific contribution of this study is the performance comparison of CNN models for pneumonia detection due to COVID-19. These pre-trained models can serve as a basis for future studies and indicate a second opinion to the radiologist during the CXR analysis.

[*]Equal contribution [1]Software Innovation Laboratory (SOFTWARELAB), Graduate Program in Applied Computing, Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo, Brazil. Correspondence to: Felipe André Zeiser <felipezeiser@edu.unisinos.br>, Cristiano André da Costa <cac@unisinos.br>, Gabriel de Oliveira Ramos <gdoramos@unisinos.br>.

Proceedings of the LatinX in AI (LXAI) Research workshop at ICML 2021.

2. Related Works

Since the initial outbreak of COVID-19, studies using CNN have been used to detect COVID-19 in CXR images. However, at the beginning of the outbreak, the lack of positive CXR images was a problem. In (Hemdan et al., 2020), the performance of seven CNNs were compared using fine-tuning in a set of 50 CXR, with 25 positive cases and 25 negative cases for COVID-19. However, the few images used, the lack of use of a test set, and the learning charts presented indicate that the models could not generalize the problem. The use of several pre-trained CNNs stacked, working together to extract characteristics, can be a solution for small datasets (Gupta et al., 2021).

Another alternative for detecting COVID-19 in CXR is the detection in levels using the VGG-16, as adopted in (Brunese et al., 2020). At the first level, CXRs are analyzed for the detection of pneumonia or not. In the second level, the classification of the CXR in COVID-19 or not is carried out. Finally, on the third level, the activation heat maps for the COVID-19 CXRs are presented. The accuracy of the study, according to the authors, is 99%.

In summary, several recent studies have investigated the use of CNNs for the CXR classification of COVID-19. However, for the most part, these studies were based on small datasets or carried out evaluations directly on the validation sets. Therefore, these studies lack evidence about their ability to generalize the problem, which prevents its use as a system to assist radiologists. Thus, our contribution to these gaps is the proposal to evaluate six convolutional models in a dataset with more than five thousand CXR. We evaluated the models in a set of CXR not used in training and validation.

3. Method and Results

We present an overview of the methodology proposed in this work in Figure 1. The proposed methodology is divided into three stages: (i) preprocessing with resizing and normalization of the images; (ii) data augmentation to generate synthetic images; and (iii) training and testing of convolutional models.

3.1. Datasets

The CXR used were obtained from the Covid ChestXray Dataset (Cohen et al., 2020) and Labeled Optical Coherence Tomography and Chest X-Ray Images for Classification (Kermany et al., 2018). The Covid ChestXray Dataset is a public and open dataset. The images are collected from public sources, and through indirect collection from hospitals and doctors (Cohen et al., 2020). The Labeled Optical Coherence Tomography and Chest X-Ray Images for Classification is a public dataset with CXR images of normal, viral, and bacterial pneumonia (Kermany et al., 2018). Table 1 shows the number of images used for each database.

Table 1. The number of images used for each dataset.

CLASS	(COHEN ET AL., 2020)	(KERMANY ET AL., 2018)
COVID	468	-
NORMAL	-	1583
VIRAL	38	1493
BACTERIAL	46	1454

3.2. Preprocessing

In the preprocessing step, the images were resized to a size of 512×512 pixels. To avoid distortions, we apply a proportional reduction to each dimension of the images and adding black padding to complement the size of 512 pixels to the smaller dimension. We applied to the resulting images Contrast-Limited Adaptive Histogram Equalization (CLAHE) to enhance the images. After the contrast enhancement, the images are normalized by the mean and standard deviation of the overall pixel intensity (Spanhol et al., 2016). We then divided the datasets into training (70%), validation (10%), and testing (20%).

3.3. Data Augmentation

One technique that helps in the convergence and learning of CNN is the use of data augmentation. We applied data augmentation to the COVID-19 CXR using horizontal rotations and inversions. The final number for the COVID-19 images was 1434.

3.4. Convolutional Models

We used six pre-trained convolutional models for the ImageNet dataset to perform the CXR classification: Inception-ResNetV2 (Szegedy et al., 2017), InceptionV3 (Szegedy et al., 2016), MovileNetV2 (Sandler et al., 2018), ResNet50 (He et al., 2016), VGG16 (Simonyan & Zisserman, 2014), and Xception (Chollet, 2017). For the training of the models, we adjusted some hyperparameters to avoid overfitting and improve the results. In Tab. 2, we present the hyperparameters for each of the models.

For training, we used categorical cross-entropy to measure the error at the end of each epoch and the Adam algorithm to optimize the weights. We measured the accuracy and obtained the best weights set at the end of each epoch based on the accuracy. We trained each of the convolutional models for 100 epochs in the training set and saved the best weights for each model based on the metrics for the validation set. The results for each model in the test set are shown in Table 3.

Analyzing the Tab. 3, the results shows relative stability

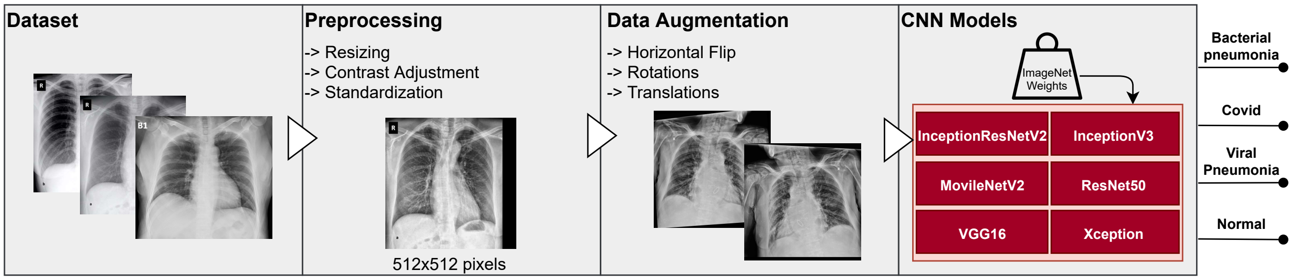

Figure 1. Proposed methodology.

Table 2. Hyperparameters used for each CNN model.

MODEL	LEARNING RATE	BATCH SIZE	TRAINABLE PARAMS	NON-TRAINABLE PARAMS	DEPTH
INCEPTIONRESNETV2	5×10^{-7}	4	57 816 420	60 544	572
INCEPTIONV3	5×10^{-7}	4	26 488 228	34 432	159
MOVILENETV2	5×10^{-7}	4	7 468 036	34 112	88
RESNET50	1×10^{-6}	16	31 909 252	45 440	50
VGG16	5×10^{-7}	16	48 231 684	38 720	23
XCEPTION	5×10^{-7}	4	29 196 844	54 528	126

Table 3. Results for the test set for each model. *Acc* is the Accuracy, *Sen* is the Sensitivity, *Spe* is Specificity, *F1* is the F1-Score, and *AUC* is the area under the ROC curve.

MODEL	ACC	SEN	SPE	F1	AUC
INCEPTIONRESNETV2	84.16%	84.24%	84.10%	84.16%	0.9596
INCEPTIONV3	83.42%	83.50%	83.40%	83.45%	0.9590
MOVILENETV2	79.37%	79.53%	79.32%	79.41%	0.9413
RESNET50	84.24%	84.32%	84.08%	84.14%	0.9634
VGG16	84.16%	84.15%	84.46%	84.14%	**0.9663**
XCEPTION	**85.64%**	**85.71%**	**85.65%**	**85.49%**	0.9648

in the performance metrics for each model, with a maximum variation of 0.30% for the VGG16 model. These results indicate an adequate generalization of each model for the detection of pneumonia due to COVID-19. As for sensitivity, which measures the ability to classify positive classes correctly, we have a 6.34% variation. The maximum variation between models for specificity is 6.33%.

In general, the Xception model presented the best results. This better performance may be associated with the organization of the convolutional blocks at Xception. Xception is a modification of Inception, with the replacement of the original inception blocks by depthwise separable convolutions (Chollet, 2017). Adopting this concept, the Xception architecture can improve the results, compared to the Inception architecture, reducing the computational cost during training without decreasing trainable parameters.

4. Conclusion and Future Work

In this article, we compared six convolutional architectures to detect pneumonia due to COVID-19 in CXR images. In order to improve the generalization of the results, we apply a set of preprocessing techniques. We use several models with pre-trained weights in ImageNet dataset, and we propose the classification in normal cases, viral pneumonia, bacterial pneumonia, or COVID-19. According to the results obtained, CNN Xception presented a better performance.

As future work, we intend to analyze the influence of datasets in learning of CNN models for CXR COVID-19 classification, exploring whether CNNs can generalize characteristics for different datasets. In addition, we intend to investigate Explainable Artificial Intelligence approaches to show to specialists the features present in the images that were used to form the diagnosis suggestion. Finally, the use of multimodal methodologies, for example, using clinical data and images, can be helpful in the transparency of diagnostic suggestions.

Acknowledgments

The authors would like to thank CAPES (C.F. 001), CNPq (No. 309537/2020-7), and the NVIDIA GPU Grant Program for supporting this work.

References

Ai, T. et al. Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: A report of

1014 cases. *Radiology*, 296(2):E32–E40, 2020. doi: 10.1148/radiol.2020200642. URL https://doi.org/10.1148/radiol.2020200642.

Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C., and Garry, R. F. The proximal origin of sars-cov-2. *Nature medicine*, 26(4):450–452, 2020.

Borghesi, A. and Maroldi, R. Covid-19 outbreak in italy: experimental chest x-ray scoring system for quantifying and monitoring disease progression. *La radiologia medica*, 125(5):509–513, 2020.

Brunese, L., Mercaldo, F., Reginelli, A., and Santone, A. Explainable deep learning for pulmonary disease and coronavirus covid-19 detection from x-rays. *Computer Methods and Programs in Biomedicine*, 196:105608, 2020. ISSN 0169-2607. doi: https://doi.org/10.1016/j.cmpb.2020.105608. URL https://www.sciencedirect.com/science/article/pii/S0169260720314413.

Chollet, F. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.

Cohen, J. P. et al. Covid-19 image data collection: Prospective predictions are the future, 2020.

Gupta, A., Anjum, Gupta, S., and Katarya, R. Instacovnet-19: A deep learning classification model for the detection of covid-19 patients using chest x-ray. *Applied Soft Computing*, 99:106859, 2021. ISSN 1568-4946. doi: https://doi.org/10.1016/j.asoc.2020.106859.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Hemdan, E. E.-D., Shouman, M. A., and Karar, M. E. Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. *arXiv preprint arXiv:2003.11055*, 2020.

Kermany, D., Zhang, K., and Goldbaum, M. Labeled optical coherence tomography (oct) and chest x-ray images for classification, 2018. URL https://data.mendeley.com/datasets/rscbjbr9sj/2.

Lee, E. Y., Ng, M.-Y., and Khong, P.-L. Covid-19 pneumonia: what has ct taught us? *The Lancet Infectious Diseases*, 20(4):384–385, 2020.

Marson, F. A. L. Covid-19 – 6 million cases worldwide and an overview of the diagnosis in brazil: a tragedy to be announced. *Diagnostic Microbiology and Infectious Disease*, 98(2):115113, 2020. ISSN 0732-8893. doi: https://doi.org/10.1016/j.diagmicrobio.2020.115113. URL http://www.sciencedirect.com/science/article/pii/S0732889320304909.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

Shereen, M. A., Khan, S., Kazmi, A., Bashir, N., and Siddique, R. Covid-19 infection: Origin, transmission, and characteristics of human coronaviruses. *Journal of Advanced Research*, 24:91 – 98, 2020. ISSN 2090-1232. doi: https://doi.org/10.1016/j.jare.2020.03.005. URL http://www.sciencedirect.com/science/article/pii/S2090123220300540.

Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Spanhol, F. A., Oliveira, L. S., Petitjean, C., and Heutte, L. Breast cancer histopathological image classification using Convolutional Neural Networks. In *Proceedings of the International Joint Conference on Neural Networks*, volume 2016-Octob, pp. 2560–2567. Institute of Electrical and Electronics Engineers Inc., 2016. doi: 10.1109/IJCNN.2016.7727519.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017.

Wong, H. Y. F. et al. Frequency and distribution of chest radiographic findings in patients positive for covid-19. *Radiology*, 296(2):E72–E78, 2020. doi: 10.1148/radiol.2020201160. URL https://doi.org/10.1148/radiol.2020201160.

Xie, X. et al. Chest ct for typical coronavirus disease 2019 (covid-19) pneumonia: Relationship to negative rt-pcr testing. *Radiology*, 296(2):E41–E45, 2020. doi: 10.1148/radiol.2020200343. URL https://doi.org/10.1148/radiol.2020200343.