

COVID-19 diagnosis based on chest X-ray images

Agata Gielczyk¹[0000-0002-5630-7461], Anna Marciniak^{1,2}[0000-0002-9900-3951],
and Martyna Tarczewska¹

¹ University of Science and Technology, Bydgoszcz, Poland
agata.gielczyk@pbs.edu.pl

² Faculty of Medicine Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Torun, Bydgoszcz, Poland

Abstract. The SARS-CoV-2 pandemic began in early 2020 and paralyzed human life all over the world. Due to the overstretching of the healthcare system, the rapid and efficient tools for COVID-19 diagnosis and distinguishing became especially needed. In this article we implement a ML-based approach with some pre-processing methods, and verify whether they can improve the COVID-19 positive vs. negative classification using the chest X-ray images. We provide also some promising results: Accuracy > 97%, Precision > 96%, Recall > 96% and F1-score > 96%.

Keywords: SARS-CoV-2 · COVID-19 · Machine learning · Image processing · Pre-processing

1 Introduction

The beginning of the COVID-19 pandemic in 2020 has shaken up the modern world. It has caused societies to close, crowded streets to become deserted, pubs and clubs to be silenced, and popular meeting places to die down. Currently, people all over the world are doing their best to overcome the pandemic's impact on the social, medical, psychological, economic, and industrial aspects of society. In this moment, the reverse transcription polymerase chain reaction (RT-PCR) testing is the main screening method for detecting COVID-19 infections. However, patients suffering from COVID-19 can also present with abnormalities on chest X-ray (CXR) images that are characteristic of infection [5]. This imaging modality is highly available and accessible in many clinical locations, and it is considered standard equipment in most healthcare systems. Moreover, CXR imaging is more widely available than CT (computed tomography) imaging, especially in developing countries due to high equipment and maintenance costs. X-ray analysis can be time-consuming and requires highly educated specialists to interpret. But, the use of machine learning (ML)-based methods can improve efficiency, support medics in the diagnosis of COVID-19, speed up the time to diagnosis, and lighten the already burdened health care system.

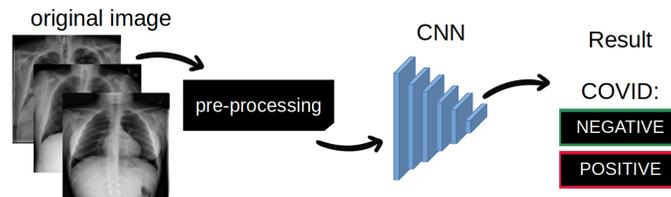
In this article we propose a ML-based method for COVID-19 infection detection using chest X-ray images analysis with different pre-processing methods. The presented article will be possibly extended by adding more lung diseases e.g., pneumonia.

The intensification of the COVID-19 pandemic caused that more and more researchers became working on the automatic disease recognition. Some of them focus strictly on the image pre-processing. It has numerous advantages: reduces noises, enhance image and by resizing image can reduce the computation time. In [6, 7] the segmentation problem was raised as an emerging issue. Medical data suffers from weak annotations, as the description of the CT and X-ray images is complicated and time consuming. Authors in [2] proposed a novel, hybrid, multimodal deep learning system. With the use of Contrast-Limited Adaptive Histogram Equalization (CLAHE) and a Butterworth bandpass filter, the authors were able to enhance the contrast of X-ray images and eliminate the noise leading to an accuracy of 99.93%. Mahdy et al. in [4], proposed a method to automatically classify COVID-19 chest X-rays using a multi-level threshold based on the Otsu algorithm and support vector machine (SVM).

2 Proposed method

In this research, we used the dataset available for the public at [1]. It consists of images collected from various sources: GitHub repository, Kaggle, Radiopedia, Italian Society of Radiology (SIRM), and Figshare data repository websites. The dataset contains almost 7.000 samples presenting posteroanterior (PA) chest X-ray images.

Fig. 1. The overview of the proposed method: original CXR image, black box of pre-processing, CNN and the final result



To verify how the selected pre-processing method affects the final classification result, we proposed a baseline system. The general overview of this system is presented in Fig. 1. The black box visible in Fig. 1 marks the selected pre-processing method. The pre-processing step is an important element in the image analysis schema. It can enhance the original image and reduce noise or unwanted details. In our research, we examined 6 different approaches to pre-processing:

1. None — in this method the only one element is size reduction.

2. Histogram equalization — this method extends the pixel’s intensity range from the original range from 0 to 255. As the result, the enhanced image has a wider range of intensity and slightly higher contrast.
3. Hist. eq. + Gaussian blur — this filter reduces some noise and unwanted details that can be confusing for the neural network; the filter kernel size was experimentally set to 5×5 size.
4. Hist. eq. + bilateral filter — this filter also reduces some noise and unwanted details that can be confusing for the neural network, but its main feature is to preserve edges; the experimentally set up parameters of the filter: $diameter = 5$, $\sigma_{color} = \sigma_{space} = 75$.
5. Adaptive masking — in [3] the authors proved that by removing the diaphragm from the sample it is possible to improve the classification results. In this proposed pre-processing method, we first found the maximum (max) and minimum (min) intensity of pixels and then applied the binary thresholding using the threshold expressed in Eq. 1. The next step used morphologic closing. This creates the adaptive mask that after bitwise operation removes the diaphragm from the source image.
6. Adaptive masking + hist. eq. + Gaussian blur — this method joins adaptive masking with histogram equalization and Gaussian blur (with kernel size 5×5 again).

$$threshold = min + 0.9 \cdot (max - min) \quad (1)$$

In a classification step we implemented a convolution network (CNN). It can be described as a chain of convolution layers, with rectified linear unit activation functions, pooling layers, and batch normalization operations. In this specific approach we used 12 layers. The output from the neural network is the classification, whether the patient is COVID-19 positive or negative. All the experiments were executed using the online Kaggle notebook. There were almost 7,000 samples in the dataset. We decided to divide the dataset into three disjoint subsets: training-65%, validating-15%, and testing-20%. All of the experiments were executed 3 times to prove their independence from the learning data. Thanks to the balance in the dataset, we did not need any sample augmentation.

3 Results

The above-mentioned experiments provided some promising results. We used 4 parameters for the evaluation methods - accuracy, precision, recall, and F1-score. The parameters were calculated using a confusion matrix reporting the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The obtained results are presented in Table 1.

4 Conclusion

In this article, we proposed a novel approach for the fully automated analysis of COVID-19 chest X-ray images using a convolutional neural network. We

Table 1. The obtained results for all pre-processing methods

No	Method	Accuracy	Precision	Recall	F1-score
1.	None	0.9385	0.8566	0.9806	0.9144
2.	Histogram equalization	0.9602	0.9340	0.9482	0.9411
3.	Hist. eq. + Gaussian blur	0.9609	0.9199	0.9676	0.9432
4.	Hist. eq. + Bilateral filter	0.9725	0.9436	0.9762	0.9597
5.	Adaptive masking	0.9544	0.9739	0.8877	0.9288
6.	Ada. mask. + hist. eq. + Gauss. blur	0.9761	0.9614	0.9676	0.9645

also presented an improvement in the proposed method by introducing the pre-processing part of the ML-based system. In this early step of image analysis, a few crucial operations are performed: adaptive masking (very light parts of the image are removed), histogram equalization (widens the range of pixels intensity), and Gaussian blur (removes noise and some unwanted details). The presented results proved that the proposed pre-processing method increases the efficiency of the system as the F1-score raised from 91% to over 96%. Our results are comparable to other similar ML-based approaches in the literature, but there are plenty of pre-processing methods that can improve the efficiency of the system and be implemented in future work. Other possible extension of this method is to add more classes and more diseases to classification.

References

1. COVID19 Pneumonia NormalChest Xray PA Dataset. www.kaggle.com/94amanullahasraf/covid19-pneumonia-normal-chest-xray-pa-dataset, [Online; accessed 03.01.2022]
2. Al-Waisy, A., Mohammed, M.A., Al-Fahdawi, S., Maashi, M., Garcia-Zapirain, B., Abdulkareem, K.H., Mostafa, S., Kumar, N.M., Le, D.N.: Covid-deepnet: hybrid multimodal deep learning system for improving covid-19 pneumonia detection in chest x-ray images. *Computers, Materials and Continua* **67**(2), 2409–2429 (2021)
3. Heidari, M., Mirniaharikandehei, S., Khuzani, A.Z., Danala, G., Qiu, Y., Zheng, B.: Improving the performance of cnn to predict the likelihood of covid-19 using chest x-ray images with preprocessing algorithms. *International journal of medical informatics* **144**, 104284 (2020)
4. Mahdy, L.N., Ezzat, K.A., Elmousalami, H.H., Ella, H.A., Hassanien, A.E.: Automatic x-ray covid-19 lung image classification system based on multi-level thresholding and support vector machine. *MedRxiv* pp. 2020–03 (2020)
5. Ng, M.Y., Lee, E.Y., Yang, J., Yang, F., Li, X., Wang, H., Lui, M.M.s., Lo, C.S.Y., Leung, B., Khong, P.L., et al.: Imaging profile of the covid-19 infection: radiologic findings and literature review. *Radiology: Cardiothoracic Imaging* **2**(1), e200034 (2020)
6. Saood, A., Hatem, I.: Covid-19 lung ct image segmentation using deep learning methods: U-net versus segnet. *BMC Medical Imaging* **21**(1), 1–10 (2021)
7. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X.: Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis* **63**, 101693 (2020)