Lung Xray Images Analysis for COVID-19 Diagnosis

Anna Kloska[1[0000−0002−5776−5448]], **Martyna Tarczewska**[2], **Agata Giełczyk**[2[0000−0002−5630−7461]], **Beata Marciniak**[2[0000−0003−1092−1279]]

[1]*Faculty of Medicine Ludwik Rydygier Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Torun, Bydgoszcz, Poland*
[2]*Bydgoszcz University of Science and Technology, Bydgoszcz, Poland*
agata.gielczyk@pbs.edu.pl

DOI:10.34658/9788366741928.77

Abstract. *Background: The SARS-CoV-2 pandemic began in early 2020. It paralyzed human life all over the world and threatened our security. Thus, proposing some novel and effective approaches to diagnosing COVID-19 infections became paramount. Methods: This article proposes a method for the classification of chest X-ray images based on the transfer learning. We examined also different scenarios of dataset augmentation. Results: The paper reports accuracy=98%, precision=97%, recall=100% and F1-score=98% in the most promising approach. Conclusion: Our research proofs that machine learning can be used in order to support medics in chest X-ray classification and implementing augmentation can lead to improvements in accuracy, precision, recall, and F1-scores.*
Keywords: *COVID-19, image processing, augmentation, artificial intelligence*

1. Introduction

A recently observed SARS-CoV2 pandemic has infected millions of people all over the world and caused numerous problems: health, mental, social and economical. It was confirmed by World Health Organization that more than 200 countries have been affected by the coronavirus pandemic. The infection caused some disease symptoms (cough, fever, fatigue, and respiratory distress), but the pandemic lead to an unprecedented failure in the health services due to the lack of medical staff or overloading entire healthcare systems. However, latest scientific reports suggest that artificial intelligence (AI) could be used to overcome the pandemic crisis including by its usage in: diagnosis, drug development and treatment. In this article we present the baseline COVID-19 detection system based on lung Xray images and its further improvements by selected dataset augmentation techniques.

2. Related work

Two important issues may be raised after the state-of-the-art review: the necessity of using segmentation and the strong tendency to use various transfer learning in lung Xray images analysis. Authors in [1] prepared the comparison of various ResNet architectures: ResNet18, ResNet34, ResNet50, ResNet101, and ResNet152 providing the highest F1-score=94%. In [2] Narin A. et al. implemented some CNN-based transfer learning models like: ResNet50, ResNet101, ResNet152, InceptionV3, and Inception-ResNetV2. The ResNet50 architecture was the most promising that provided the accuracy over 96%. Bargshady G. et al. [3] used Inception V3 for Xray images analysis. They proposed a novel approach to the dataset augmentation based on GAN. Even though it seemed to be very computationally demanding, it enabled to reach the accuracy exceeding 94%. In [4] the role of segmentation was emphasized. Authors proposed COVID-19 Lung segmentation network (CLSeg) inspired by the popular U-Net architecture, consisting the encoder and decoder parts. They proved that segmentation is a crucial step using the heat maps.

3. Proposed method

Figure 1 presents all steps of the proposed pipeline: data, augmentation, preprocessing, and classification. Finally, the method gives the answer of *'true'* for the COVID-19-positive sample and *'false'* for the healthy sample.

In the research we used the dataset `https://www.kaggle.com/datasets/andyczhao/covidx-cxr2` described in [5] for training and validating. For evaluation we used our own dataset available at `https://github.com/UTP-WTIiE/Xray_data.git` which was previously described in [6]. We used also a novel architecture based on two ResNet networks. The first one performs segmentation and is a model pre-trained for lung segmentation whereas the second performs the classification. The parameters for training classification model, ResNet18, were: optimizer – SGD (stochastic gradient descent), loss function – cross entropy, 200 epochs, and batch size 16. Whole experiments were conducted using the PyTorch library.

The augmentation is the key point of this article. We proposed and assessed 5 different approaches to augmentation: 1. **None** – no augmentation methods were used in the baseline approach. It was essential since we want to research the influence of augmentation on the model's prediction; 2. **Blurring** – slightly blurring the images in order to reduce noises and get rid of some unwanted details: motion blur, median blur and Gaussian blur; 3. **Contrast and brightness** – manipulating the contrast and brightness of the image: CLAHE (Contrast Limited Adaptive Histogram Equalization), random brightness modification, random con-

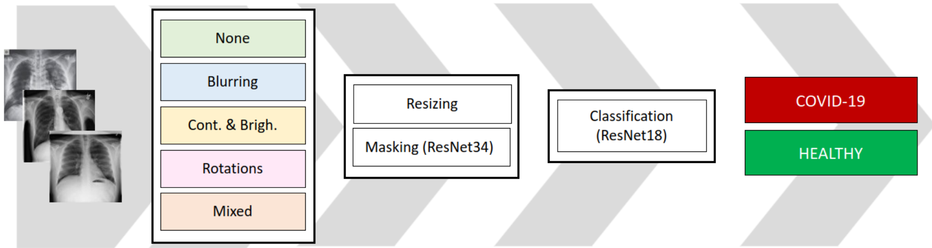

Figure 1. The pipeline of the proposed system (from left side): data used in the research, augmentation, pre-processing (resize and masking by ResNet34), classification (ResNet18) and the final result: COVID-19/Healthy. Source: own work.

trast modification, sharpening; 4. **Rotations** – rotating an image by a fixed angle to simulate different orientations of the image; images were rotated by angle in range $< -3, 3 >$ expressed in degrees; 5. **Mixed** – a mix of all mentioned augmentation methods.

4. Evaluation and results

Each model in this research was evaluated using four validation metrics as follows: Accuracy (Eq. 1), Precision (Eq. 2), Recall (Eq. 3), and F1-score (Eq. 4), which use the below mentioned measures TP – *true positives* – COVID-19 patient classified as sick, FP – *false positives* – healthy patient classified as sick, FN – *false negatives* – COVID-19 patient classified as healthy, and TN – *true negatives* – healthy patient classified as healthy. These metrics can be treated as a golden standard in ML-based studies. They can help also in comparison of the proposed method to the state-of-the-are results. The obtained results were presented in Table 1. The most promising results were highlighted in bold.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{F1-score} = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{4}$$

Table 1. The results obtained for the selected augmentation methods.

Augmentation	Accuracy	Precision	Recall	F1-score
None	0.9032	0.9063	0.9063	0.9063
Blurring	0.9516	0.9677	0.9375	0.9524
Contrast and Brightness	0.9355	0.9667	0.9063	0.935
Rotations	0.9516	0.9677	0.9375	0.9524
Mixed	**0.9839**	**0.9697**	**1.0000**	**0.9846**

5. Conclusions

In this paper we proposed the baseline architecture for lung Xray images-based COVID-19 detection using two ResNet networks. We presented and examined also some powerful augmentation techniques. In the proposed schema the most promising approach was to join all described types of augmentation and to implement them together. Obviously, the proposed schema can be extended in the future. It is possible to add some additional augmentations and, what seems to be more important, to add some explainability. It could help the medics in understanding how does the ML-based system really work.

References

[1] Showkat S., Qureshi S., *Efficacy of transfer learning-based resnet models in chest x-ray image classification for detecting covid-19 pneumonia*, *Chemometrics and Intelligent Laboratory Systems*, 2022, vol. 224, p. 104534.

[2] Narin A., Kaya C., Pamuk Z., *Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks*, *Pattern Analysis and Applications*, 2021, vol. 24, pp. 1207–1220.

[3] Bargshady G., Zhou X., Barua P.D., Gururajan R., Li Y., Acharya U.R., *Application of cyclegan and transfer learning techniques for automated detection of covid-19 using x-ray images*, *Pattern Recognition Letters*, 2022, vol. 153, pp. 67–74.

[4] Zhao H., Fang Z., Ren J., MacLellan C., Xia Y., Li S., Sun M., Ren K., *Sc2net: A novel segmentation-based classification network for detection of covid-19 in chest x-ray images*, *IEEE Journal of Biomedical and Health Informatics*, 2022, vol. 26, no 8, pp. 4032–4043.

[5] Wang L., Lin Z.Q., Wong A., *Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images*, *Scientific Reports*, 2020, vol. 10, no 1, p. 19549, ISSN 2045-2322, doi: 10.1038/s41598-020-76550-z.

[6] Giełczyk A., Marciniak A., Tarczewska M., Kloska S.M., Harmoza A., Serafin Z., Woźniak M., *A novel lightweight approach to covid-19 diagnostics based on chest x-ray images*, *Journal of Clinical Medicine*, 2022, vol. 11, no 19, doi: 10.3390/jcm11195501.