DOI:10.34918/85104

Wiktoria ŚLIWIŃSKA[1,*], Aleksandra SUWALSKA[1], Michał MARCZYK[1,2,*]

Chapter 4. INFLUENCE OF VARIOUS PREPROCESSING TECHNIQUES AND MODEL PARAMETERS ON NETWORK PERFORMANCE IN COVID-19 DETECTION

4.1. Introduction

SARS-CoV-2 virus, which belongs to the *coronaviridae* family, can attack human organisms [1]. The result of potential infection is COVID-19 disease. Despite the fact, that illnesses due to coronaviruses had occurred for at least 50 years [2], SARS-CoV-2 was the first one that caused a pandemic. In the majority of infected people, COVID-19 causes symptoms similar to a cold, however in some cases it can lead to much more dangerous health changes in the lungs, like consolidations or ground-glass opacities (Fig. 1). The SARS-CoV-2 infection could be detected by Nucleic Acid Amplification Tests (NAAT) or antigen tests. An alternative to confirm the disease is by recognizing the specific changes in the lungs on the chest X-Ray (CXR) or Computed Tomography (CT) images. Since CXR are much cheaper and easier to obtain, this imaging procedure is performed more often [3]. Too intense increase in infections resulting in COVID-19 disease led to the need of studying CXR images on a large scale, so scientists started to find a way of automatization the process, mostly using deep learning-based algorithms.

[1] Department of Data Science and Engineering, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, Gliwice, Poland.
[2] Yale Cancer Center, Yale School of Medicine, 06511 New Haven, CT, USA.
* Corresponding authors: michal.marczyk@polsl.pl, wiktsli527@student.polsl.pl, ul. Akademicka 16, 44-100 Gliwice, PL.

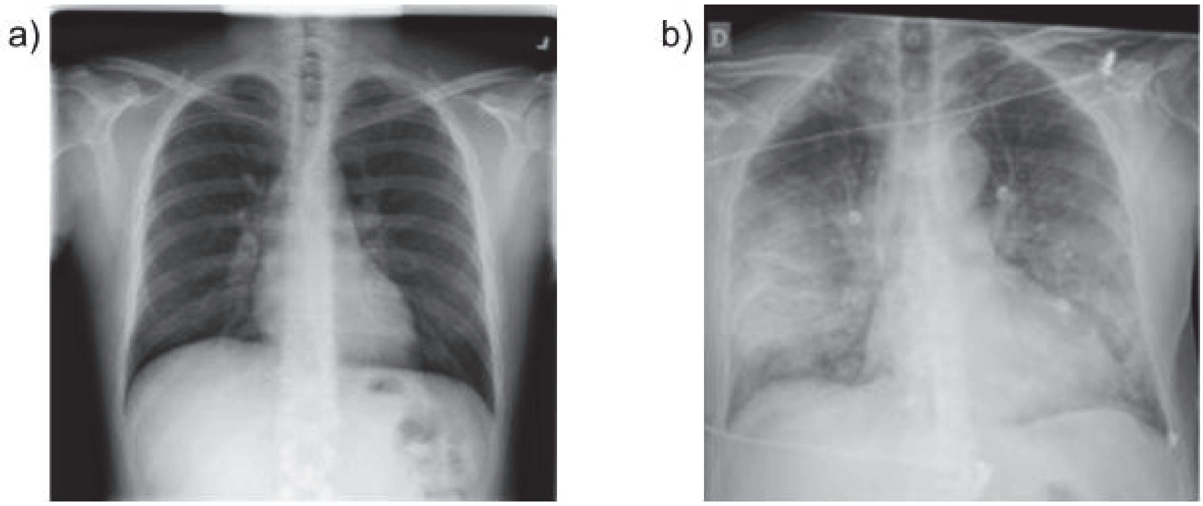

Fig. 1. Comparison of healthy patient lungs (a) and lungs of a patient with COVID-19 (b)
Rys. 1. Porównanie płuc zdrowego pacjenta (a) oraz płuc osoby chorującej na COVID-19 (b)

Deep learning methods are becoming more and more popular, due to their high efficiency and growing collections of well-annotated data. These methods are using artificial neural networks, which are comprised of artificial neurons, that learn from a huge number of data examples. The learning process is choosing the optimal parameters to model the formula, so the cost function would obtain minimal value. Images are mostly analyzed using convolutional neural networks, which can recognize patterns. The performance of the network is primarily evaluated by accuracy, but there are more measures used, like recall and sensitivity.

Deep learning methods were applied in many works concerning patients with pulmonary diseases. Some of them extended the approach to non-viral pneumonia [4], and other solves both, binary and multiclass problems [5, 6]. Researchers in [6] compared different image preprocessing methods, however, the best results were obtained for the network learned with original images. In [7-9] the combination of deep and machine learning methods was checked. The model described in [8] uses ResNet-50. The features were collected from the first fully connected layer and were fed as input to four machine learning algorithms, which made decisions by voting.

The goal of this study is to find the best technique and model parameters to improve the performance of the deep networks for COVID-19 detection. We analyzed the effect of changing the number of convolution layers, the influence of applying the Contrast Limited Adaptive Histogram Equalization algorithm, and the segmentation of the lung region before modeling. Many experiments were performed, from which we selected two models with the best results: one with and one without the application of lung segmentation.

4.2. Materials and Methods

4.2.1. Data

Models were trained with the use of the COVIDx online dataset [10] which contains other publicly available sets with CXR images. The number of images was equal to 10 316. 8 751 of them were images taken on healthy people (no COVID-19 symptoms) and 1 565 were taken on COVID-19 patients, so there was a disproportion between classes. Some of the images had artifacts induced by medical equipment or preliminary assessment, associated with adding arrows indicating changes in lungs.

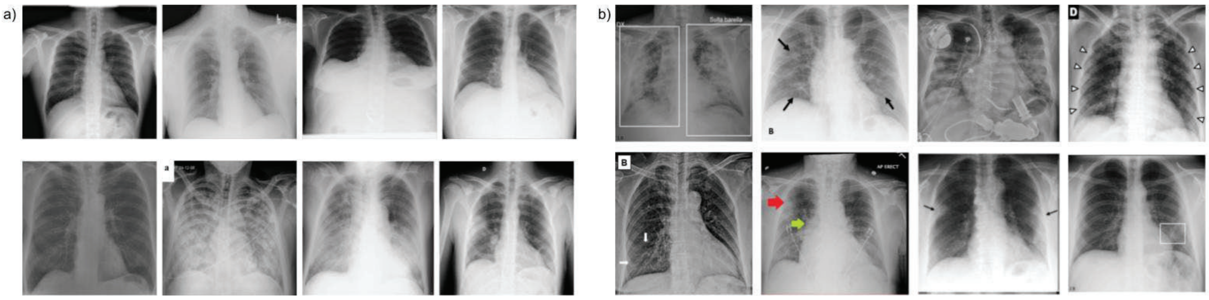

Fig. 2. Example images from COVIDx database without (a) and with artifacts (b)
Rys. 2. Przykładowe obrazy z bazy COVIDx bez (a) oraz z artefaktami (b)

Some networks were trained with input data after lung segmentation. In most cases, it allowed the removal of artifacts from images, because a lot of them were located outside the lung region. Binary image masks were generated with the use of the U-NET network, which is a part of the CIRCA system [11]. In the case of the positive class, where lungs are affected by changes, the difference between them and soft tissues is not sharp, resulting in troubles with segmentation. Therefore, masks with low quality appeared mostly in the COVID-19 class.

To check model generalization, apart from the COVIDx test set, networks were also evaluated using the independent, Spanish clinical dataset (SIIM-Covid19) [12]. We have used 6 333 CXR images. All of them belonged to the COVID-19 class, however, 4 subclasses were indicating the level of lung occupation: 'Negative for Pneumonia', 'Typical Appearance', 'Indeterminate Appearance', and 'Atypical Appearance'.

4.2.2. Preprocessing

All data was loaded as grayscale images and resized to 300 x 300 pixels. The COVIDx dataset was split into training, validation, and test set. In the beginning, 100 images of each class were randomly chosen for the test set. The rest was divided into training (70%) and validation (30%) sets.

In some experiments, the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm was applied. This technique is used to enhance the contrast of images, so the changes in the lungs could be more visible. CLAHE is an improvement of standard Histogram Equalization (HE), which only stretches the histogram to obtain more varying pixel intensities. Nevertheless, it works globally, so HE may result in too bright or too dark areas. In the case of CLAHE, equalization is adaptive – it works locally, so the problem mentioned above does not occur. Additionally, the limitation of contrast enables control of the height of the histogram, which helps with noise reduction [13].

To enhance the mask quality, the Convex Hull algorithm was used. It finds the smallest convex set, that contains all points so that the lungs have softer edges and do not have any gaps. It allows for preserving important information from the inside of the lungs. The use of this method was necessary because a lot of patients with COVID-19 disease had advanced changes in their lungs that made segmentation difficult and resulted in low-quality masks.

The data were standardized with the mean and standard deviation of the training set. Spanish set was previously scaled from 32-bit to 8-bit images because networks were trained on 8-bit images from the COVIDx dataset.

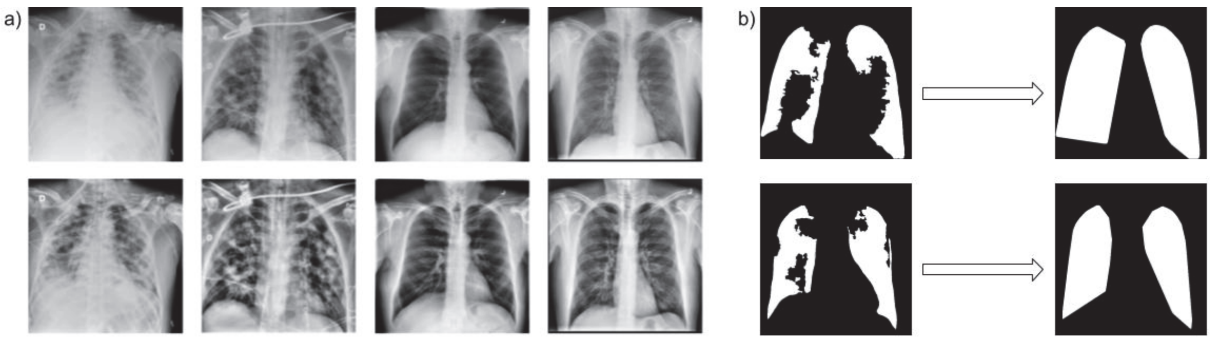

Fig. 3. (a) The effect of using the CLAHE algorithm. Original images are presented in the first row, the second row shows images after CLAHE. (b) The effect of using the convex hull algorithm on damaged masks

Rys. 3 (a) Efekt zastosowania algorytmu CLAHE. Obrazy oryginalne przedstawiono w pierwszym wierszu, wiersz drugi przedstawia obrazy po zastosowaniu CLAHE. (b) Wpływ zastosowania algorytmu convex hull na uszkodzone maski

4.2.3. Experiments

The conducted experiments were divided into two categories: those which were done using input images without segmentation and with it. In total, 96 models were created, 48 for both categories. Their architecture differs in the number of layers, activation function, learning rate, dropout coefficient, and the use of the CLAHE algorithm. Adam Optimizer was used in all cases. In each experiment, the augmentation technique was applied with horizontal flip and rotation up to 15°. The output of every convolutional layer was batch normalized and then passed to the pooling layer that calculates the maximum (max pooling). In the convolutional and pooling layers, padding was set to 'same', the step in the convolutional layers was equal to one, and in the pooling layers to two. As final models, two networks with the best performance were selected, one with and one without segmentation.

4.2.4. Evaluation

For network comparison, two metrics were used: accuracy and recall. The first one is the most used and gives us information about the percentage of cases correctly classified. A recall is a metric that informs how good the model is in classifying positive cases. It is especially important in medical image classification because it is better to double-check a patient that we are not sure about than leave him with the disease.

For selected models, Grad-CAM (Gradient-weighted Class Activation Mapping) maps were prepared. Grad-CAMs allow for the visualization of areas that are important for the network during the prediction [14]. For a time, Neural Networks were considered black boxes, since the process of learning was not clear to researchers. The input data were just entered, and the result was recorded, however, we did not know why the network classified this entry into a specific class. With methods like Grad-CAM, it is possible to understand better how the model works, as they enable the verification of network decisions. The influence of changing the number of convolutional layers and applying CLAHE was checked with the use of statistical tests. Additionally, the effect size was checked. In the same way, the dependency between mask quality and the model prediction was verified.

4.3. Results

The architectures of selected networks are presented in Fig. 4. In panel (a) architecture of the network without lung segmentation is presented. The CLAHE algorithm with a *clipLimit* parameter equal to 2 was used to transform the input data. The network uses the Adam optimizer with a learning rate equal to 0.001. It consists of 3 convolutional layers with a sigmoidal activation function. After each convolutional layer, a max-pooling layer was applied. There are 3 fully connected layers, two of them use a sigmoidal function, and the last one uses a softmax activation function. This network obtained an accuracy and recall of 95%. The second network shown in panel (b), which works on images after lung segmentation, also uses the CLAHE with the *clipLimit* equal to 2. This network contains 4 convolutional, 4 max-pooling, and 4 fully connected layers. The activation function in all layers, except the last one with the softmax function, is ReLU. The optimizer used in this network is Adam with a 0.0001 learning rate. The accuracy obtained by the second network was equal to 93% whereas recall was equal to 87%.

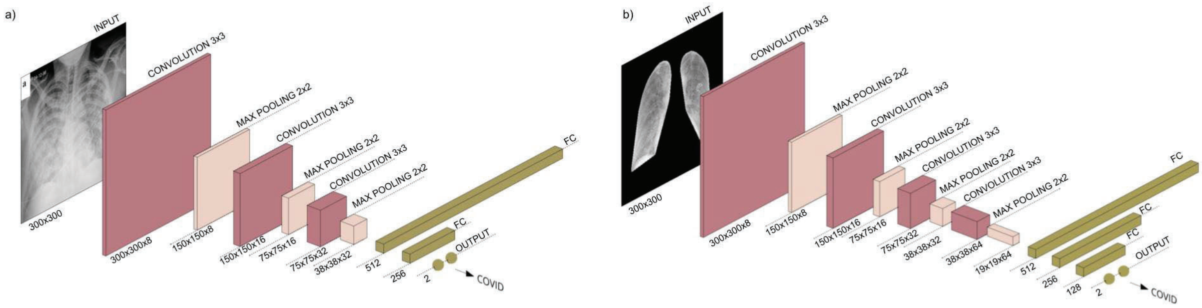

Fig. 4. (a) Architecture of network trained without the use of segmentation. (b) Architecture of network trained with the use of lung segmentation

Rys. 4. (a) Architektura sieci uczonej bez segmentacji płuc. (b) Architektura sieci uczonej na obrazach po segmentacji płuc

Grad-CAM activation maps are shown in Fig. 5. The first column presents original images and the second one Grad-CAM maps for this image created by the first model. In the third column are images after lung segmentation, and in the last column, are Grad-CAM maps for the second model.

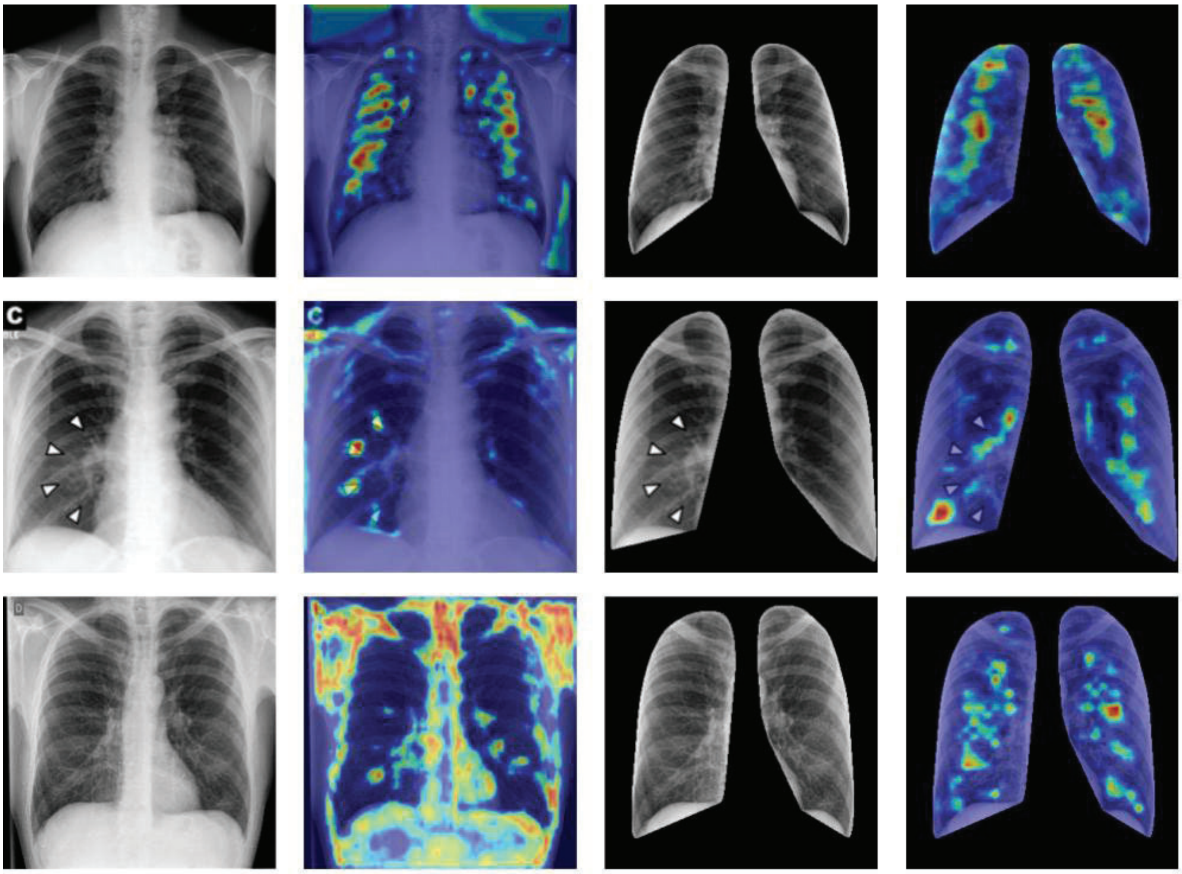

Fig. 5. Grad-CAM activation maps. The first column contains original images, the second one Grad-CAM maps for the first model, the third column images after lung segmentation and the fourth column – Grad-CAM maps for the second model

Rys. 5. Mapy aktywacji Grad-CAM. Pierwsza kolumna zawiera obrazy oryginalne, druga mapy aktywacji dla pierwszego modelu, kolumna trzecia przedstawia obrazy po segmentacji płuc, a czwarta mapy aktywacji dla drugiego modelu

The difference in results for 3 and 4 convolutional layers and results for networks trained with and without the use of CLAHE algorithm was tested by U-Mann-Whitney statistical tests with a 5% significance level and the effect size of those relationships was checked with Glass rank biserial correlation. The same tests were used to verify if the model prediction depends on the mask quality (Table 1). There is no statistically significant difference in medians of the results for 3 and 4 convolutional layers for COVIDx and SIIM-Covid19 databases. Nevertheless, the effect size for SIIM--Covid19 showed a small, negative correlation between the number of convolutional layers and the accuracy of the model. There is a statistically significant difference in medians of the results for networks with the use of CLAHE and without for SIIM--Covid19. The effect size showed a large, positive correlation. However, applying the CLAHE algorithm did not change the accuracy of the model in the COVIDx dataset. The relationship between mask quality and model prediction (Table 1), showed that in both datasets there is a difference between mask quality median for healthy and COVID-19 predictions. The effect size for the COVIDx dataset is medium and for SIIM-Covid19 small.

Table 1

Results of conducted statistical tests

Tested parameter	Dataset	P-value	Effect size
No. of convolutional layers	COVIDx	0.595	-0.0634
	SIIM-Covid19	0.05966	-0.224
The use of CLAHE	COVIDx	0.6494	-0.0543
	SIIM-Covid19	0.00001176	0.52
The mask quality	COVIDx	0.0000172	0.355
	SIIM-Covid19	0.01863	0.112

4.4. Discussion

Artifacts are more likely to occur in the positive group and as a result, the network can learn that this additional information suggests a positive class, which is not desirable. We used the activation maps to enable the visual verification of selected models. The network trained on the original images considered background, body, and artifacts during model prediction, which may lead to erroneous findings (Fig. 5). After segmentation, only the lungs remained on the image, so the network could not learn anything from the background or other body parts. Since a lot of artifacts occurred outside the lungs, the segmentation also removed them. After that, there were too few artifacts for the network to learn them as a class-specific pattern. This situation can be observed in the second row of Fig. 5.

CLAHE algorithm improved results for the SIIM-Covid19 dataset, but there was no statistical dependency for the COVIDx dataset. The effect size of 0.52 for SIIM--Covid19 shows, that there was a large difference – after using the CLAHE, network performance was better. Applying the CLAHE algorithm in most cases improves results [15-17] but as presented in [18], where the accuracy dropped from 93.87% to 92.04% and recall from 92.13% to 89.56% after using the CLAHE as a pre-processing step, the improvement is not the rule. Results for networks with 3 and 4 convolutional layers did not show a statistically significant difference. It may be that the difference of a single layer is not enough to show the dependency and more experiments should be performed. However, there is a small negative correlation for the SIIM-Covid19 dataset, which indicates better results for networks with 3 convolutional layers. In [19] comparison of some ready architectures for the problem of COVID-19 detection was made. VGG16 model was the one with the best performance of 95.88% accuracy. This network contains 13 convolutional layers. Architectures with more convolutional layers obtained the worst accuracy, like ResNet152V2 which consists of 152 layers and obtained an accuracy of 61.76%.

Masks with low quality appeared mostly in the COVID-19 class, so there was a possibility that the network will make decisions based on this information. U-Mann--Whitney statistical tests were conducted in order to verify this hypothesis. They showed that at the 5% significance level in both datasets there is a difference in quality mask medians for classes. However, Glass rank biserial correlation showed that for SIIM-Covid19 this dependency is small, and for COVIDx is medium. The poorly visible dependence on the Spanish set could be caused by a small number of masks of poor quality in relation to the entire collection, the number of which was over 6 000. In the COVIDx set, the relationship was checked on a test set of 200 images.

4.5. Conclusion

Lung segmentation is a very important preprocessing step as it can remove biologically unimportant information from the image. As a result, the network focuses on essential areas of the image. The CLAHE improved the generalization of our models, but there was no dependency on the dataset on which the network was trained. There was no relationship between the number of convolutional layers and network results. Therefore, the use of a network with 3 layers appears to be a better option as the network is less susceptible to overfitting and at the same time faster due to lower computational complexity. Additionally, it is important to check if there are no other dependencies, like in our case, the relationship between mask quality and model prediction.

Bibliography

1. Y.A. Malik: Properties of coronavirus and sars-cov-2, *The Malaysian journal of pathology* (2020) **42(1)**:3–11.
2. Z. Pejsak, K. Tarasiuk, B. Tokarz-Deptuła: Wybrane dane na temat zakażeń koronawirusami, ze szczególnym uwzględnieniem sars-cov-2, *Medycyna Weterynaryjna* (2020) **76(5)**:258–262.
3. A. Jacobi, M. Chung, A. Bernheim, C. Eber: Portable chest x-ray in coronavirus disease-19 (covid-19): A pictorial review, *Clinical imaging* (2020) **64(8)**:35–42.
4. R. Jain, M. Gupta, S. Taneja, D.J. Hemanth, Deep learning based detection and analysis of covid-19 on chest x-ray images, *Applied Intelligence* (2021) **51(3)**:1690–1700.
5. T. Ozturk, M. Talo, E.A. Yildirim, U.B. Baloglu, O. Yildirim, U.R. Acharya: Automated detection of covid-19 cases using deep neural networks with x-ray images, *Computers in biology and medicine* (2020) **121(8)**:103792.
6. C. Doglioni, C. Ravaglia, M. Chilosi, G. Rossi, A. Dubini, F. Pedica, S. Piciucchi, A. Vizzuso, F. Stella, S. Maitan, V. Agnoletti, S. Puglisi, G. Poletti, V. Sambri, G. Pizzolo, V. Bronte, A. Wells, V. Poletti: Covid-19 interstitial pneumonia: Histological and immunohistochemical features on cryobiopsies, *Respiration* (2021) **100(5)**:369–379.
7. L. Hussain, T. Nguyen, H. Li, A.A. Abbasi, K.J. Lone, Z. Zhao, M. Zaib, A. Chen, T.Q. Duong: Machine-learning classification of texture features of portable chest x-ray accurately classifies covid-19 lung infection, *BioMedical Engineering OnLine* (2020) **19(1)**:1–18.
8. P. Saha, M.S. Sadi, M.M. Islam: Emcnet: Automated covid-19 diagnosis from x-ray images using convolutional neural network and ensemble of machine learning classifiers, *Informatics in medicine unlocked* (2021) **22**:100505.
9. D. Wang, J. Mo, G. Zhou, L. Xu, Y. Liu: An efficient mixture of deep and machine learning models for covid-19 diagnosis in chest x-ray images, *PloS one* (2020) **15(11)**:e0242535.
10. L. Wang, Z.Q. Lin, A. Wong, https://github.com/lindawangg/covidnet/blob/master/docs/covidx.md [accessed: 22.05.2020 r.].
11. https://circa.aei.polsl.pl/ [accessed: 15.10.2021 r.].
12. https://www.kaggle.com/c/siim-covid19-detection/data [accessed: 15.10.2021].

13. S. Sahu, A.K. Singh, S.P. Ghrera, M. Elhoseny: An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE, *Optics & Laser Technology* (2019) **110**:87–98.
14. R.R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, D. Batra: Grad-CAM: Why did you say that?, *arXiv preprint arXiv:1611.07450* (2016).
15. S.S. Narli, G. Altan, CLAHE-based Enhancement to Transfer Learning in COVID-19 Detection, *Gazi Mühendislik Bilimleri Dergisi* (2022) 1–11.
16. D. Reynaldi, B.S. Negara, S. Sanjaya, E. Satria: COVID-19 Classification for Chest X-Ray Images using Deep Learning and Resnet-101, *2021 International Congress of Advanced Technology and Engineering (ICOTEN)* IEEE (2021) 1–4.
17. F. Saiz, I. Barandiaran: COVID-19 detection in chest X-ray images using a deep learning approach, *Journal of Interactive Multimedia and Artificial Intelligence*. (2020) 1.
18. A. Tahir, Y. Qiblawey, A. Khandakar, T. Rahman, U. Khurshid, F. Musharavati, M.T. Islamd, S. Kiranyaza, S. Al-Maadeed, M.E.H. Chowdhury: Deep Learning for Reliable Classification of COVID-19, MERS, and SARS from Chest X-ray Images, *Cognitive Computation* (2022) **14(5)**:1752–1772.
19. A. Makris, I. Kontopoulos, K. Tserpes: COVID-19 detection from chest X-Ray images using Deep Learning and Convolutional Neural Networks, *11th Hellenic conference on artificial intelligence* (2020) 60–66.

INFLUENCE OF VARIOUS PREPROCESSING TECHNIQUES AND MODEL PARAMETERS ON NETWORK PERFORMANCE IN COVID-19 DETECTION

Abstract

The COVID-19 pandemic caused a need for an efficient tool to predict patients infected with SARS-CoV-2. One of the possibilities for recognizing the disease is by analyzing chest X-ray (CXR) images, as COVID-19 can cause lung consolidations or ground-glass opacities. However, due to the huge number of infected people, radiologists could not perform real-time analyses, so computer-assisted methods are necessary. This work aims to compare the impact of several techniques on the performance of a deep learning-based approach for COVID-19 prediction. Networks were trained using 10 316 CXRs images. Different preprocessing methods and model parameters were tested, such as the number of convolutional layers, the influence of applying the Contrast Limited Adaptive Histogram Equalization algorithm, and the usage of lung segmentation before model building. Finally, two best-performing networks were selected (one with and one without lung segmentation), which achieved an accuracy of 95% and 93%, respectively. Additionally, networks were evaluated on the independent set of 6 333 images reaching an accuracy of 92.9% and 97.6%, respectively. Grad-CAM and statistical tests were used to analyze the model performance scores. In conclusion, CLAHE mostly improves model generalization. According to the number of convolutional layers, there were no differences in results. During lung segmentation, unnecessary information is removed, which increases the probability of learning only truly important patterns.

Keywords: deep learning, classification, COVID-19, X-ray