Automatic detection of pneumonia in chest X-rays using Lobe deep residual network

Daniel Kvak[1], Karolína Kvaková[2]

May 2021

Abstract:

Chest radiography is one of the critical tools for early detection and subsequent evaluation of the incidence of lung diseases. At a time when the speed and reliability of results, especially for COVID-19 positive patients, is important, the development of applications that would facilitate the work of untrained staff involved in the evaluation is also crucial. Our model takes the form of a simple and intuitive application, into which you only need to upload X-rays: tens or hundreds at once. In just a few seconds, the physician will determine the patient's diagnosis, including the percentage accuracy of the estimate. While the original idea was a mere binary classifier that could tell if a patient was suffering from pneumonia or not, in this paper we present a model that distinguishes between a bacterial disease, a viral infection, or a finding caused by COVID-19. The aim of this research is to demonstrate whether pneumonia can be detected or even spatially localized using a uniform, supervised classification.

Keywords: *automatic detection, chest X-ray, convolutional neural network, COVID-19, deep learning, feature extraction, image classification, pneumonia.*

[1] Faculty of Arts, Masaryk University, Brno. kvak@mail.muni.cz

[2] Faculty of Science, Masaryk University, Brno. 460405@mail.muni.cz

Introduction

Computer-aided diagnostics (CAD) appeared in mammography as early as the end of the 1990s, when it was used for the second reading of the image using the methods of support vector machines (Singh et al. 2006). At that time, it encountered a rather negative approach. However, this was not due to the low accuracy of the estimate or the distrust of medical staff, as some machine learning experts incorrectly believe. The main problems are rather practical: the user environment was often not very friendly, but above all, radiologists fell victim to confirmatory bias. Instead of using the reading as double validation, the situation forced them to trust machine judgment. We can only monitor the turnover in recent years. An important milestone is the study by Wang et al., which used a convolutional neural network for computer vision to detect colorectal adenomas and other polyps during colonoscopy (2019). The difference that contributed to a successful practical deployment is not related to the model used – earlier models and today's deep neural networks use methods of feature extraction from image data and subsequent classification – but rather a different approach that allows us to view computer assistance as a recommendation system. Thus, endoscopic examination using artificial intelligence did not focus on a mere visual decision but was linked to the actual performance of a biopsy. A similar parallel can be found in radiology, where the system based on artificial intelligence monitors artifacts that represent deviations in gradient-weighted map measurements.

1. Background

Chest radiography is an important tool for early detection and subsequent verification of lung diseases. These are devices that are easily accessible, and the price of scanning is only in the order of few dollars (Du et al. 2021). However, in the current pandemic situation of COVID-19, we encounter a lack of radiologists and trained staff to analyze the vast number of images taken. The limited availability of high-resolution computed tomography and polymerase chain reaction (RT-PCR) in countries with high patient turnover underlines the importance of chest X-rays as a suitable tool for screening and diagnosis.

To date, the most widely used method for detecting COVID-19 pneumonia is the RT-PCR test, which takes approximately 4 hours to evaluate (Gandhi et al. 2020). With the time needed for collection and transport, we can thus range in the range of days, with not every region having an evaluation center; not to mention the price of up to several hundred dollars (Du et al. 2021). Reliability converging to 100% does not appear until the third repetition (Gandhi et al. 2020). Another method is represented by antigen tests, which offer cheaper and faster detection, but a positive result might be shown only if the individual is at the stage of greatest risk of transmission. An increasingly common procedure is the detection of antibodies, but these are found in the body one to three weeks after the infection (Mina et al. 2021). The aim of this research is to present a deep learning approach that would eliminate the time and resources needed to develop new technologies and related algorithms. The results presented in this text suggest that the Lobe application based on the deep residual network ResNet-50 could represent an optimal solution for the detection of X-ray findings of pneumonia with a success rate of up to 99%.

2. Summary of previous research

Similar research can be found, for example, in Elgendi et al. (2020) who compared 17 available deep learning algorithms for easier, faster, and most importantly cheaper detection of COVID-19 using chest radiography and the DarkNet-19 deep neural network. As a result, DarkNet-19 proved to be the optimal pre-trained network for detecting radiographic images of COVID-19 pneumonia with 94.28% accuracy in 5,854 X-rays available from the Kaggle database and another dataset from Vancouver Hospital. The input images of the network, showing posteroanterior and anteroposterior projection, were verified by radiologists with many years of experience. For validation, the networks were provided with a million different images and a balanced set of healthy individuals and those with pneumonia and COVID-19, with DarkNet-19 showing the most appropriate balance between speed and reliability of detection. Further for comparison, e.g., Narin et al. (2020) report an accuracy of 98% using a convolutional network on a balanced dataset, Sethy and Behera (2020) report an accuracy of 93% on a balanced dataset, Zhang et al. (2020) also report 93%, but on an unbalanced dataset, Hemdam et al. (2020) 90% on a balanced or Apostolopoulos and Bessiana (2020) 98% on an unbalanced data sample.

The study by Elgendi et al. worked with various architectures in the MATLAB environment. Although the results were promising, the complete dataset included only

50 COVID-19 positive patients, so it was severely limited, with less confusion of infection with bacterial or viral pneumonia. Narin et al. as well as Sethy and Behera list ResNet50 as the most suitable network: this architecture is also used in our study. Sethy and Behera combined the ResNet50 and SVM convolutional network with histograms to achieve 93.4% accuracy in the validation set. Zhang et al. collected 100 chest X-rays in the GitHub database, of which 70 had confirmed COVID-19 cases, thousands more with different pneumonia infections and healthy patients, but as a result, the model shows too high a percentage of false-positive findings. As well as other studies, they are aware of the need for subsequent validation of the collected clinical data. Hemdan et al. also worked with only a small dataset comprising 50 images, of which 25 were confirmed as COVID-19 positive. Apostopoulos and Bessiana had a larger database, including 224 images of COVID-19 positive patients.

2. Datasets and methods

2.1 Proposed model and construction

It is the unpretentious and user-friendly form of the used model that is the decisive factor why we believe that the application of pulmonary findings detection mentioned below has the potential for use in practice. Despite the simple design, the program hides one of the most modern solutions for pattern recognition in image data (Papers with Code 2021). The ResNet-50 deep neural network, first presented by He et al. (2015), introduces *state-of-the-art* computer vision technology. This architecture uses a method of extracting features from image files, thanks to which the network learns according to which perceptions it can sort individual images into different classes. Because the findings falling under the same class show similar deviations, the application uses a prediction capability, which then divides the new images into previously segmented classes.

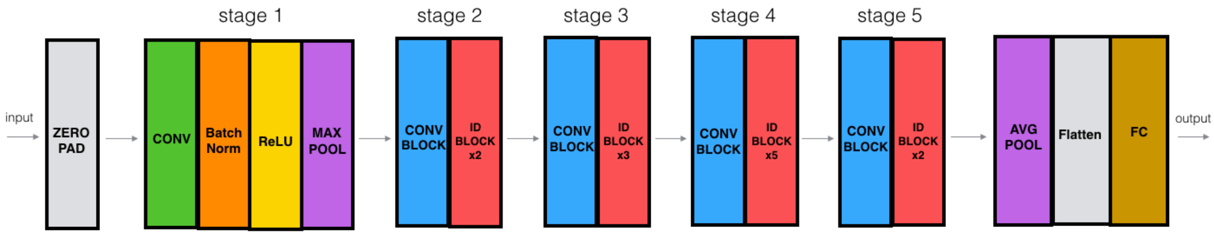

Figure 1. *Simplified visualization of the used ResNet-50 architecture.*

Deep convolutional networks are an ideal solution for extracting features from image data, while stacking multiple layers improves prediction capabilities to some extent. Here, however, we encounter a known limitation that prevents the convergence of networks with dozens of layers: *the problem of vanishing and exploding gradient*. Adding additional layers to the model can increase the error value for the training set (Feng 2017). The ResNet-50 architecture consists of five stages, each with a convolutional block and an identity block, which is primarily used as a *shortcut connection* (He et al. 2015). The convolution block contains three convolution layers, each identity block operating with three convolution layers. The shortcut is used to skip one or more layers, e.g., from the first layer it is possible to skip to the third layer using the shortcut.

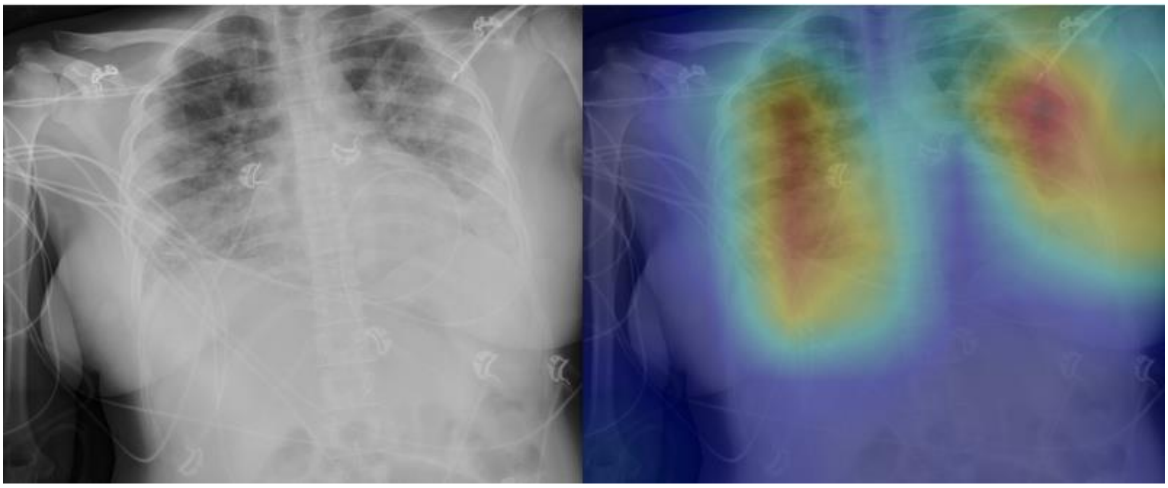

Figure 2. *Localization of the gradient-weighted activation map for COVID-19 finding.*

2.2. Dataset collection and preparation

Dataset 1 comes from the open-source Mendeley Data repository. The data set was collected by authors from three different medical facilities in Uttar Pradesh and Rajasthan. 188 images (28 COVID-19, 83 with findings, 77 healthy images) were taken at *King George's Medical University* in Uttar Pradesh. 68 images of patients with COVID-19 were collected from the *University of Medical Sciences* in Uttar Pradesh. 543 X-rays (326 COVID-19, 217 with findings) come from *Government Medical College* in Rajasthan.

Dataset 2 was obtained from the Kaggle public data repository. Chest X-rays were selected from retrospective cohorts of pediatric patients aged one to five years from

Women and Children's Medical Center in Guangzhou. All images were performed as part of routine clinical patient care. Prior to chest x-ray analysis, all images were first subjected to quality control and removal of any poor quality or illegible scans. The diagnoses were subsequently verified by two specialists. The data set was also checked by a third-party expert to take account of possible errors in the evaluation. The dataset consists of 2530 images with bacterial pneumonia, 1345 images with viral disease and 1341 images without findings.

Dataset 3 is a collection of three large datasets that have been collected on the Kaggle datastore. Following the needs of our research, the individual classes were reduced to a simple distinction between images with and without pneumonia findings. The motivation of this dataset can be defined as a binary classification, where the input is an X-ray of the chest, while the output is a binary label indicating the absence or presence of a finding of pneumonia.

2.3. Data augmentation

In order to properly train our model, it is useful to increase the size of the data set using data enhancement that reduces noise and preserves the original quality. This process is performed during the loading and training process, so the performance of the model can be improved by solving the problem of overfitting. The use model automatically crops the imported data to the center square, scale your image and automatically create small variations of the images to reflect the noisiness of real-world data. During training, the model makes five variations of your images with randomly varied:

- Brightness
- Contrast
- Saturation
- Hue
- Rotation
- Zoom
- JPEG encoding noise

Each option has its ability to represent images in different ways to provide important features during the training phase and thus enhances the model's performance better.

2.4. Classification performance

Dataset	Type of classification	Number of data for each class	Prediction accuracy on a test set
Dataset 1[3]	Multi-class classification	COVID-19 (422), Non-COVID-19 (302), Normal (77)	99 %
Dataset 2[4]	Multi-class classification	Bacterial (2530), Normal (1341), Viral (1345)	92 %
Dataset 3[5]	Binary classification	Normal (3293), Pneumonia (8352)	99 %

Table 1. *Predictive accuracy of the ResNet-50 neural network on individual datasets*.

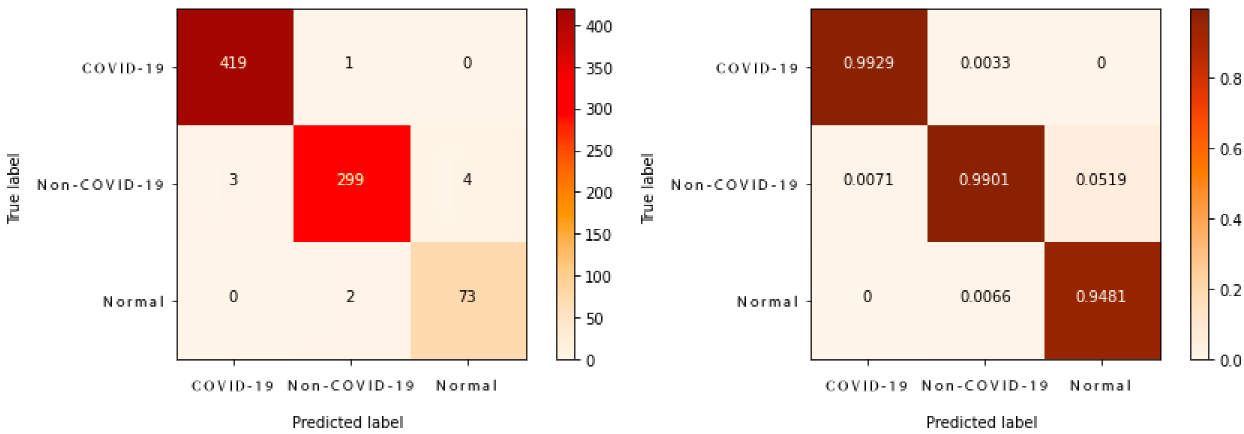

Figure 3. *Confusion matrix for dataset 1.*

An F_1-score becomes a critical evaluation tool to determine False Positive and False Negative rates yielded through a discriminating threshold in a similar situation with unbalanced dataset samples. For an overall identification of the ResNet-50

[3] Dutta, M. K. (2021). *COVID Chest X-Rays*. *1*. https://doi.org/10.17632/4n66brtp4j.1

[4] Mooney, P. (b.r.). *Chest X-Ray Images (Pneumonia)*. Retrieved 6 April 2021, from https://kaggle.com/paultimothymooney/chest-xray-pneumonia

[5] Kvak, D. (2021). Sběr RTG snímků plic z volně přístupných databází.

performance, the use of the following metrics evaluated its critical points. The following includes the accuracy, precision, recall, and F_1-score, calculated based on the following equations below:

$$Accuracy = \frac{TP + TN}{Total\,Samples}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Figure 4. *Evaluation metrics formulas for classification problems.*

Class	Accuracy (%)	Precision	Recall	F_1-score
COVID-19	99,5	1,0	0,99	1,0
Non-COVID-19	98,75	0,98	0,99	0,98
Normal	99,25	0,97	0,95	0,96

Table 2. *Performance measure of multiclass classifier for dataset 1.*

5. Limitations of the used methodology

Although radiological imaging is a widely available, as well as affordable method, due to the dangers of ionizing radiation from CT and X-ray equipment, it is not suitable for frequent use, and there are other contraindications to this procedure. It is necessary to consider imaging pregnant women, other complications can be caused by the presence of metal objects, the inability of the patient to take a deep full breath or also to adapt the equipment for skinny or obese patients (Uppot 2018).

The disadvantage of all similar studies, apart from the ongoing validation that awaits this research, may be the excessive technical complexity of projects (e.g., the use of MATLAB or Python). The primary goal of this study is therefore to provide a user-friendly environment suitable for inexperienced assessors under stress, and thus help early detection and faster action in patients with more severe stages of the disease. Imaginary barriers may include the impossibility of importing through the traditional DICOM imaging format used for the transmission of biomedical image data within

PACS systems. This limitation can be circumvented by implementing the *med2image* or *pydicom* libraries.

Discussion

Chest radiography is a widely available and affordable tool for screening patients with symptoms of lower respiratory tract infection or suspected pneumonia. Automatic detection using X-rays can act as an early diagnosis of the disease; this is especially true in areas facing a shortage of trained radiologists. Based on preliminary results, we demonstrate that the use of deep learning methods, hence the ResNet-50 architecture, can be a functional solution for automatic feature extraction from X-rays related to the diagnosis of viral and bacterial pneumonia, or directly to the detection of diseases caused by COVID- 19.

However, it is necessary to investigate whether the extracted features segmented by a deep neural network represent reliable biomarkers that help detect lung infection. Follow-up research should focus on the sensitivity of imaging patients with mild symptoms: these symptoms may not be accurately imaged by X-rays or may not even be visible at all. It is also necessary to accept the fact that each medical or experimental workplace scans X-ray images in a different way, including image quality, angle or brightness. Although the research is conceived primarily as a source of initial estimation, we believe that using a sufficiently general dataset, this solution could present great promise for the future. While datasets often contain, for example, only one age group, our follow-up goal is to create a comprehensive database that can be used to generalize the findings.

Ethical procedure

- *We hereby declare that this research article meets all applicable standards with regards to the ethics of experimentation and research integrity.*
- *This work did not involve humans, animals, and other living specimens during experiments.*

Acknowledgment

We would like to take this opportunity to thank MUDr. Stanislav Kašík from the Radiodiagnostic Department of the Regional Hospital in Kolín, Ing. Dalibor Kačmář, Ph.D. from Microsoft and doc. Radim Burget, Ph.D. from the Faculty of Electrical Engineering and Communication VUT in Brno for their willingness and valuable advice.

References:

Apostolopoulos, I. D., & Bessiana, T. (2020). Covid-19: Automatic detection from X-Ray images utilizing Transfer Learning with Convolutional Neural Networks. *Physical and Engineering Sciences in Medicine*, *43*(2), 635–640. https://doi.org/10.1007/s13246-020-00865-4

Du, Z. et al. (2021). Comparative cost-effectiveness of SARS-CoV-2 testing strategies in the USA: A modelling study. *The Lancet Public Health*, *6*(3), e184–e191. https://doi.org/10.1016/S2468-2667(21)00002-5

Dutta, M. K. (2021). *COVID Chest X-Rays*. *1*. https://doi.org/10.17632/4n66brtp4j.1

Elgendi, M. et al (2020). The Performance of Deep Neural Networks in Differentiating Chest X-Rays of COVID-19 Patients From Other Bacterial and Viral Pneumonias. *Frontiers in Medicine*, *7*. https://doi.org/10.3389/fmed.2020.00550

Feng, V. (2017, July 17). *An Overview of ResNet and its Variants*. Medium. https://towardsdatascience.com/an-overview-of-resnet-and-its-variants-5281e2f56035

Gandhi, D. et al. (2020). Clinical Landscape of COVID-19 Testing: Difficult Choices. *arXiv:2011.04202 [q-bio]*. http://arxiv.org/abs/2011.04202

He, K. et al. (2015). Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]*. http://arxiv.org/abs/1512.03385

Hemdan, E. E.-D. et al. (2020). COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images. *arXiv:2003.11055 [cs, eess]*. http://arxiv.org/abs/2003.11055

Ji, Q. et al. (2019). Optimized Deep Convolutional Neural Networks for Identification of Macular Diseases from Optical Coherence Tomography Images. *Algorithms*, *12*(3), 51. https://doi.org/10.3390/a12030051

Mina, M. J. (2021). Clarifying the evidence on SARS-CoV-2 antigen rapid tests in public health responses to COVID-19. *The Lancet*, *0*(0). https://doi.org/10.1016/S0140-6736(21)00425-6

Mooney, P. (n.d.). *Chest X-Ray Images (Pneumonia)*. Retrieved 6 April 2021, from https://kaggle.com/paultimothymooney/chest-xray-pneumonia

Narin, A. et al. (2020). Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks. *arXiv:2003.10849 [cs, eess]*. http://arxiv.org/abs/2003.10849

Papers with Code—ImageNet Benchmark (Image Classification). (b.r.). Retrieved 6 April 2021, from https://paperswithcode.com/sota/image-classification-on-imagenet

Rahman, T. et al. (2021). Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. *Computers in Biology and Medicine*, *132*, 104319. https://doi.org/10.1016/j.compbiomed.2021.104319

Sethy, P. K., & Behera, S. K. (2020). *Detection of Coronavirus Disease (COVID-19) Based on Deep Features*. https://doi.org/10.20944/preprints202003.0300.v1

Singh, S. et al. (2006). SVM based system for classification of microcalcifications in digital mammograms. *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2006*, 4747–4750. https://doi.org/10.1109/IEMBS.2006.259320

Uppot, R. N. (2018). Technical challenges of imaging & image-guided interventions in obese patients. *The British Journal of Radiology*, *91*(1089). https://doi.org/10.1259/bjr.20170931

Wang, P et al. (2019). Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: A prospective randomised controlled study. *Gut*, *68*(10), 1813–1819. https://doi.org/10.1136/gutjnl-2018-317500

Zhang, J. et al. (2020). Viral Pneumonia Screening on Chest X-ray Images Using Confidence-Aware Anomaly Detection. *arXiv:2003.12338 [cs, eess]*. http://arxiv.org/abs/2003.12338