

On the Detection of SARS-CoV-2 induced Pneumonia in X-Ray Thorax Images with Convolutional Neural Networks

Patrick Kurz¹, Paul Kaufmann¹, Roman Kalkreuth², Jannis Born³, Roman Klöcker⁴, Felix Hahn⁴, Felix Döllinger⁵, Timo A. Auer⁵

¹Chair for Computational Intelligence

Johannes Gutenberg University Mainz, Germany

E-Mail: pakurz@students.uni-mainz.de, paul.kaufmann@uni-mainz.de

²Department of Computer Science

TU Dortmund University, Germany

Email: roman.kalkreuth@tu-dortmund.de

³Department of Biosystems Science and Engineering

ETH Zürich, Zürich, Switzerland

Email: jborn@ethz.ch

⁴Department of Diagnostic and Interventional Radiology

University Medical Center of the Johannes Gutenberg-University Mainz,
Germany

Email: roman.kloeckner@unimedizin-mainz.de, fhahn@uni-mainz.de

⁵Department of Radiology

Charité Universitätsmedizin Berlin, Germany

Email: felix.doellinger@charite.de, timo-alexander.auer@charite.de

DOI: 10.58895/ksp/1000124139-7 erschienen in:

Proceedings – 30. Workshop Computational Intelligence: Berlin, 26. - 27. November 2020

DOI: 10.58895/ksp/1000124139 | <https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000124139/>

Abstract

SARS-CoV-2 is a highly contagious virus that can induce pulmonary complications like viral pneumonia and acute respiratory distress syndrome (ARDS). In order to support RT-PCR testing, chest X-rays are used to identify the presence of COVID-19 in lungs. Radiologists proficient in chest X-Ray (CXR) interpretation are scarce, motivating our work of examining the performance of convolutional neural networks (CNNs) on this task. CNNs are the state-of-the-art image classification method. In this work we classify X-rays into four classes (COVID-19, other lung opacity, other diseases and normal). We find, that MobileNetV1 is the best CNN for this task and achieve an overall accuracy of 70% and a COVID-19 accuracy of 83% on a test data set. By increasing the number of images of the unrestricted classes normal, lung opacity and other, the COVID-19 accuracy can be increased to 95% and the overall accuracy to 78%. We leverage model interpretability techniques and provide attention heatmaps that can assist in validating the model's decision process.

1 Introduction

The novel coronavirus disease 2019 (COVID-19) has resulted in an ongoing outbreak of viral pneumonia all over the world. By now more than 12.9 million people have been infected of which more than 570 thousand have died [1]. The containment of the disease is based on the identification of infected people, backtracing of infections, and isolation of contagious people. Due to the fact, that COVID-19 is easily confused with influenza, specific tests are necessary. There are different methods to test for the existence of SARS-CoV-2 such as an RNA-based assay using nasopharyngeal swabs (RT-PCR testing). Unfortunately, there were already shortages of swabs in the past and testing capacities reached their limits in various regions of the world [2]. Therefore, supporting methods for COVID-19 identification are needed. One alternative way of testing COVID-19 is thorax X-Ray imaging which can give immediate diagnostic information. Additionally, the stage of the disease and the risk of a severe course can be seen [3]. While thoracic X-Ray can be conducted

rapidly, the scarce availability of analyzing radiologists depict a key bottleneck toward accelerating automatic differential diagnosis. Therefore, the study of automatic image classification approaches for the identification of COVID-19 is an emerging research topic. This work focuses on convolutional neural networks (CNNs). Our contributions can be summarized as follows: Using transfer learning, different CNN base models for image classification tasks such as Xception, ResNet50, InceptionV3, MobilenetV1, and MobilenetV2 are compared with well-known classification performance metrics. The analysis finds, that MobileNetV1 is the best CNN for this task and achieved an overall accuracy of 70% and a COVID-19 accuracy of 83% on a test data set. These results are based on a relatively small training data set (each class $n=1000$, except COVID-19 $n=538$). By increasing the number of images of the unrestricted classes normal, lung opacity and other, the COVID-19 accuracy can be increased to 95% and the overall accuracy to 78%. This work also indicates that there may be unique features in thorax X-Ray images of COVID-19 infected people which differ from other forms of pneumonia. This raises new questions for further research such as why some cases of COVID-19 are relatively easy to identify and what are the key characteristics on which a CNN determines the presence of COVID-19. Future work on these research questions might be helpful to obtain more understanding of COVID-19 which could contribute to the development of further treatments.

This work is structured as follows: We give an explanation of convolutional neural networks and transfer learning in Chapter 2. In Chapter 3 we survey relevant work in the field of medical image recognition with focus on the detection of COVID-19. Chapter 4 presents an overview of the data which we used for our experiments. Chapter 5 revolves around the structure and results of our experiments. Chapter 6 discusses our results and findings. Finally, Chapter 7 summarizes the outcome of this work, and future research is proposed.

2 Related Work

2.1 Convolutional Neural Networks

Convolutional neural networks (CNN) are a state-of-the-art method for image classification. The advantage of CNNs to standard (deep) feed-forward networks is that they can work with two-dimensional input data. This allows CNNs to capture spatial dependencies of an image and reduce parameters as well as improve the reusability of weights [4]. CNNs receive matrices as input. Different layer types process the information. The most important ones are the convolutional layer, the fully connected layer, and the pooling layer.

The convolutional layer is the primary building block of CNNs. In the convolutional step, an input image is multiplied (dot product) with a weight matrix (filter) to generate a feature map representing one feature of the input image. Subsequent pooling steps reduce the size of feature maps while keeping the essential information. Pooling is a filter that strides over the given feature map and returns only a single value.

The last layers of a CNN are usually fully connected. Within these layers, every neuron is connected to every other neuron. Because fully connected layers can only process vectors, the last feature maps are flattened. The structure of a typical CNN is shown in Figure 2.

2.2 Transfer Learning

Many medical image classification tasks suffer from a low number of cases. Transfer learning can be used in such situations. Transfer learning describes the process of not training CNNs from scratch but to employ pre-trained models instead. The training time reduces dramatically as the weights of the most layers are rendered immutable, and only the last layers are re-trained. The results are excellent since most image features (e.g., edge detection) can be reused on related problems without modifications [5, 6, 5].

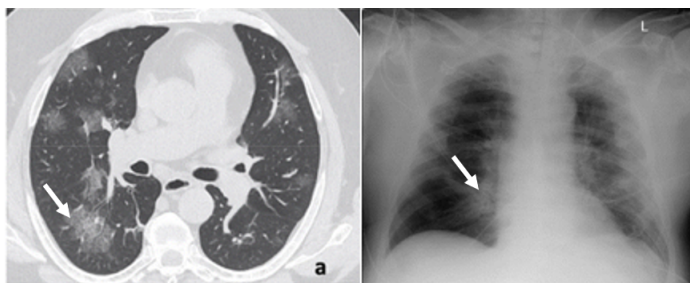


Figure 1: Comparison of COVID-19 glass-opacity in CT scan (left) and X-ray (right). Left image from Caruso et al. [16] right image from actual data set

3 CNN for Medical Image Recognition

The past few years have witnessed a rise of deep learning methods in medical image analysis [7]. CNNs now match or surpass the performance of humans in disease detection across a rich set of tasks [8] such as mammography detection [9], diagnosing pulmonary conditions from X-Ray [10], and computed tomography (CT) data [11].

COVID-19 induces patterns of viral pneumonia that can be identified through imaging techniques such as chest X-rays, CT, lung ultrasound. The most prevalent patterns are ground-glass opacity (GGO) and patchy consolidations in CT and X-Ray, and B-lines and pleural line abnormalities in ultrasound (for a detailed review see [12]). Throughout 2020, several hundred publications attempting to automatically detect COVID-19 from imaging data appeared [13]; overviews about individual methods can be found in [14, 15]). In the following, results of all three methods are briefly summarized.

3.1 COVID-19 Detection in X-rays

In current research there are two publications outstanding to the others. One was written by Narin et al. [18], who achieved the overall best score of 97% accuracy (COVID-19 sensitivity = 96%) with a pretrained ResNet50 model on a binary classification task (COVID-19 / healthy). As data set they used 50 X-rays of healthy and 50 of COVID-19 infected people [18]. The other

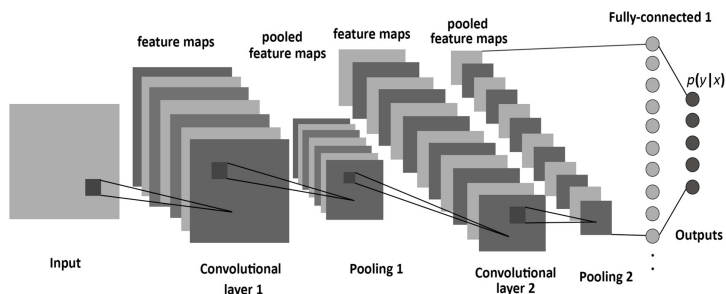


Figure 2: A typical convolutional network (Source: Albelwi and Mahmood [17], p. 5)

is related to this seminar. Apostolopoulos and Mpesiana [19] conducted a comparison of multiple CNNs for classification into three classes (common pneumonia, COVID-19, and normal incidents). They compared pretrained VGG19, MobileNetV1, Inception, Xception and Inception ResNetV2 on a data set containing 224 COVID-19, 700 common pneumonia and 500 healthy X-rays. They found VGG19 with the best overall accuracy of 92.85% (COVID-19 sensitivity = 92.85%) and MobileNetV1 with the best COVID-19 sensitivity of 99.10% (accuracy = 92.85%) [19].

3.2 COVID-19 Detection in CT Scans

With CT Scans it is possible to create cross-sectional images of the human body and thereby create a more holistic view of a specific body region. To achieve this, multiple X-rays are conducted from different angles, where detectors compute the image. Due to this procedural, more images from one patient can be generated. On the downside the amount of radiation exposure is much higher. In recent publications the biggest data set was generated by Chen et al. [20], who used 46.096 CT images from just 106 patients (51 COVID-19 positive). With that data, a U-Net++ model was trained to determine the presence of COVID-19. The model achieved an overall accuracy of 95.24% (COVID-19 sensitivity = 100%) on the test set. This model was also tested in combination with radiologists and improved their reading time of a CT scan by 65% (Chen et al. 2020). Another publication worth mentioning is

written by Butt et al. [21]. The authors achieved the best accuracy with 86.7% (COVID-19 sensitivity = 86.7%) on a three classes classification task (COVID-19, Influenza-A, irrelevant-to-infection). This accuracy was achieved by using a ResNet23-based model on a data set with 618 images (CT samples from 110 patients with COVID-19, 224 with Influenza-A and 175 healthy people) [21]. We show that in case of binary and tertiary classification an accuracy and sensitivity of around 90% on both image types is possible.

3.3 COVID-19 Detection in Lung Ultrasound (LUS)

LUS has been repeatedly recommended by clinical authorities [22] but has been neglected by the ML community [13], partially due to data heterogeneity resulting from higher operator dependency. However, public databases with LUS recordings of different pathologies are emerging [23] and the authors achieve a specificity of 0.91 and sensitivity of 0.98 on COVID-19 detection in a 3-class classification incorporating also bacterial pneumonia and healthy controls [24]. Others focused on severity assessment of COVID-19 infection and achieve recall of 0.6 and positive predictive value of 0.7 [25].

4 Preparation of data

The different classification models need a common data set to make the results comparable. Therefore, a training set and a test set are built. Both sets consist of front view chest X-ray images divided in four different classes: COVID-19, other lung opacities, other diseases, and healthy lungs (Figure 3). All but the COVID-19 images are randomly drawn from the Kaggle RSNA Pneumonia Detection Challenge (Kaggle 2018). The COVID-19 data set was built by Kalkreuth and Kaufmann [26] which contains multiple sources (162 images). This data set is expanded with images provided by radiologists of the Johannes Gutenberg-University Mainz (73 images), by the Charité – Universitätsmedizin Berlin (9 images) and by the Cohen database [27] (389 images). The structure of the training set (85% of the whole data set) and test set (15% of the whole dataset) is shown in Table 1. As visible the number of COVID-19 images is

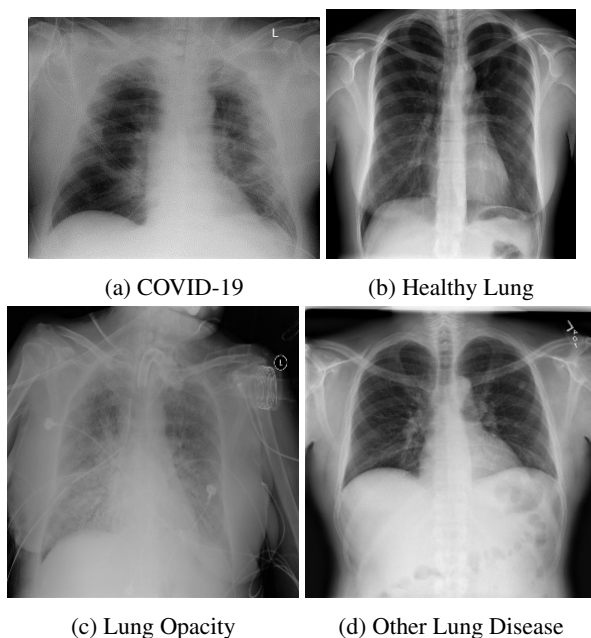


Figure 3: Example images from the data set

much smaller than the number of images in the other classes. On the given set the CNNs are trained and compared to identify the best one which is then optimized. On the one hand this produces an overview of the performance of different CNN and on the other hand it shows which model is the most promising for further fine tuning.

5 Experiments

5.1 Overview of Experiments

Figure 4 gives an overview of all experiments carried out. The first experiment was conducted to identify suitable hyperparameters. MobileNetV2 was chosen as the reference model due to its relatively small size. This introduces bias to the experiments because a good learning rate might differ between different

Table 1: Structure of the training and test dataset

Class	Training Set	Validation Set
COVID-19	538	95
Other lung opacity	1000	176
Other lung disease	1000	176
Healthy lung	1000	176

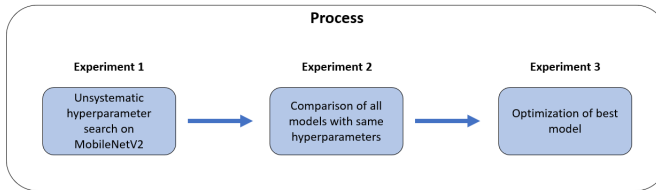


Figure 4: Process of experiments

models. In the second experiment the different models are compared with the given hyperparameters of the first experiment. The last experiment revolves around the optimization of the best model from the second experiment. In the following, each chapter describes the experiment in detail as well as its results.

5.2 Identification of Hyperparameter (Experiment 1)

The first experiment was used to identify the hyperparameter which work good for initial comparison. For this experiment the pretrained MobileNetV2 was used as the base model expanded by some custom layers which is shown in Figure 5.

Each set of hyperparameter settings (as well as the experiments later) should be conducted with a cross validation of at least five. Unfortunately, at the time of this work there was not enough computational power available. Thereby, the experiments should be re-evaluated with cross validation when possible. The results have been compared on the best validation error through

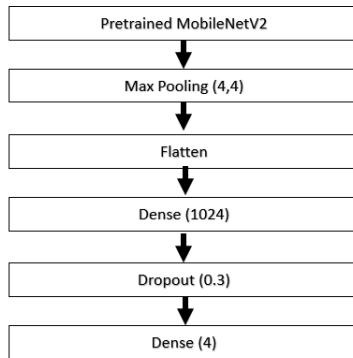


Figure 5: Additional layers on top of the pretrained model

training. Following hyperparameter have been sighted: the Adam learning rate, number of epochs, data augmentation, dropout rate and unfreezing of layers. The outcome indicates that a learning rate of 0.001, 60 epochs with a completely frozen pretrained model and nearly no data augmentation are good starting values to compare the different CNNs. Even though just little optimization has been conducted, the MobileNetV2 already has an accuracy of 64.8% on the validation set.

5.3 Benchmark of CNNs for COVID-19 identification (Experiment 2)

As a next step the five CNNs (Xception, ResNet50, InceptionV3, MobilenetV1, MobilenetV2) are compared with following Hyperparameters:

Just as in the experiments before the validation accuracy was used to compare the performance of the different base models. Not until experiment three, we have a look at the actual performance on the test set. All experiments were conducted with a cross validation of five and the mean is plotted. Figure 6 - Figure 10 show the training and validation accuracy for each model with Figure 11 as a summary. The three best performing models, regarding maximum validation accuracy over all epochs are MobileNetV1 (68.46%), InceptionV3

Table 2: Hyperparameter for Comparison

Hyperparameter	Value
Optimizer	Adam
learning rate	0.001
dropout rate	0.3
epochs	60
pretrained	ImageNet
freeze	Completely frozen
data augmentation	Rotation range = 10

(66.17%), and Xception (60.54%). Despite the max validation accuracy, we decided to optimize the MobileNetV2 due to its low training accuracy. It seems like an increase in the amount of epochs will also increase the overall accuracy which makes it much easier to optimize. Another side effect is the small size of the network which might be able to run on mobile x-ray devices. Thereby, MobileNetV2 is selected as the most promising CNN to optimize its hyperparameter systematically in experiment three.

5.4 Optimization of MobileNetV2 (Experiment 3)

In the last round of experiments, we try to optimize the MobileNetV2 as far as possible regarding the overall accuracy and then have a look at the performance on the test set. Equally to the previous experiment, cross validation of five was used. All hyperparameter were examined one after each other, consequently it can be possible, that other combinations might even increase the accuracy further, but this topic remains as a subject for following research. First, the learning rate of the adam optimizer was optimized, followed by the dropout rate and the number of frozen layers. The results are summarized in the following table, with best parameters in bold (Table 3 - Table 5):

These hyperparameter where used to train the final model and test it on the never seen test set. Due to the best performance in the second experiment, we also test the MobileNetV1 on the test set. For testing the overall best

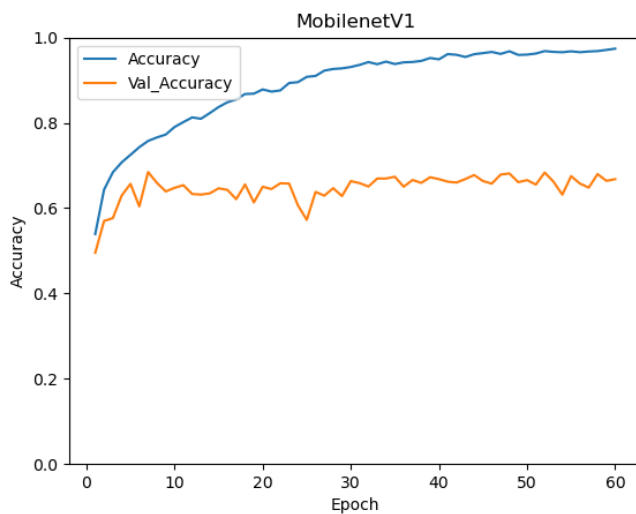


Figure 6: Performance result of MobilenetV1

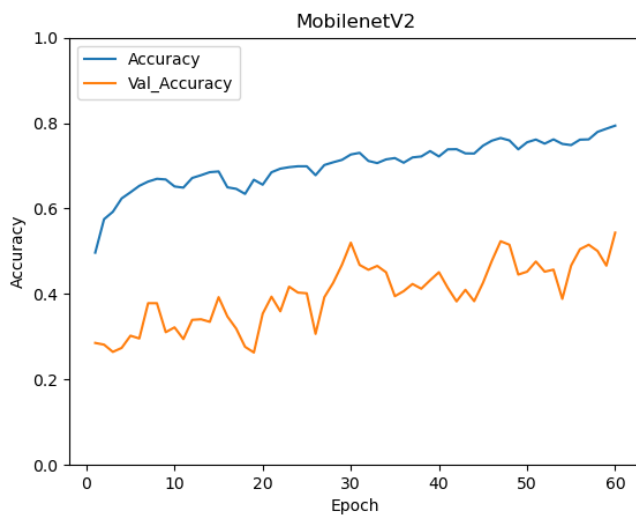


Figure 7: Performance result of MobilenetV2

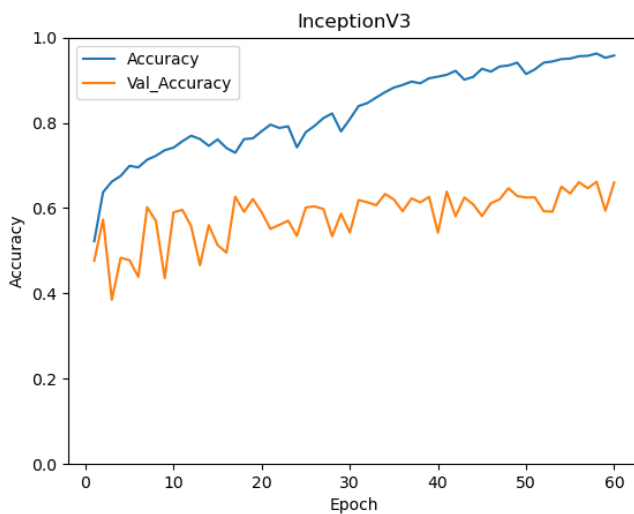


Figure 8: Performance result of InceptionV3

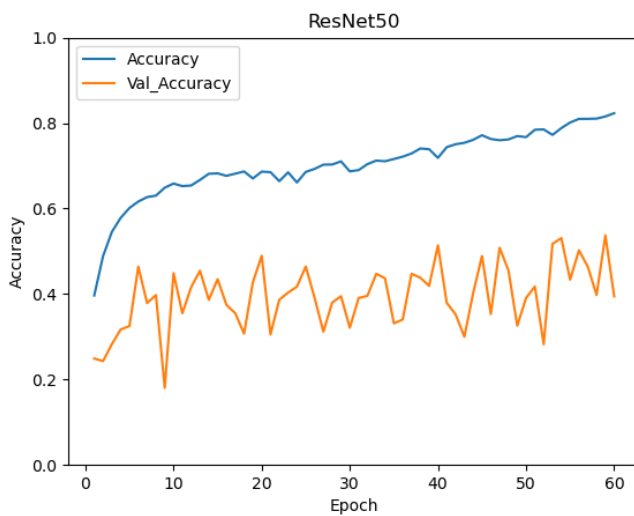


Figure 9: Performance result of ResNet50

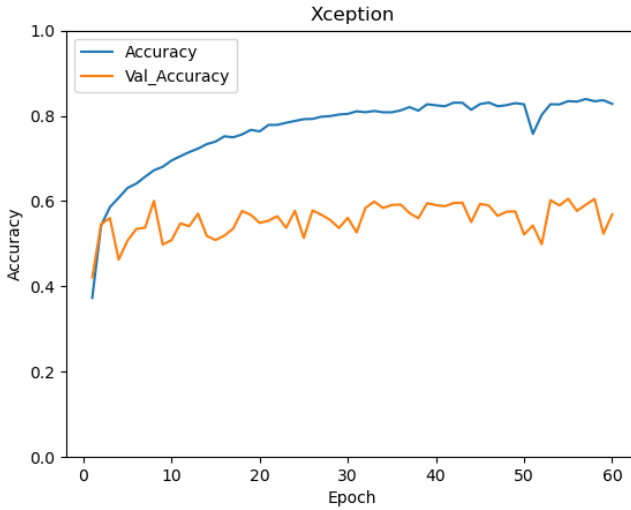


Figure 10: Performance result of Xception

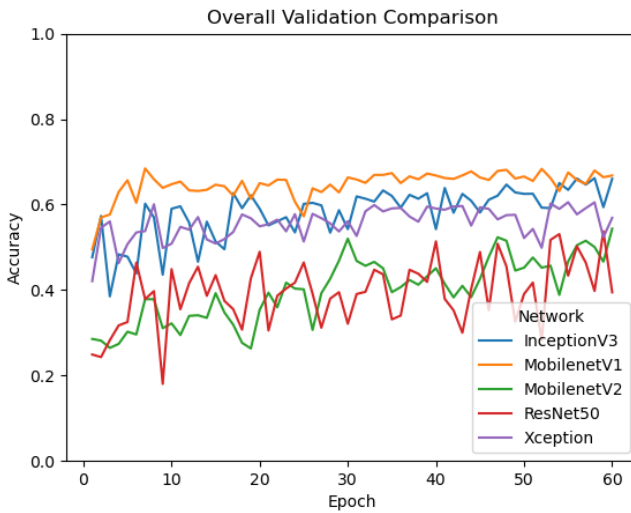


Figure 11: Overall performance of models

Table 3: Results from the optimization of the learning rate

Learning Rate	Val. Accuracy
0.001	0.5848
0.0025	0.3946
0.005	0.3206
0.0075	0.2784
0.00075	0.606
0.0005	0.582

Table 4: Results from the optimization of the dropout rate

Dropout Rate	Val. Accuracy
0.25	0.5467
0.2	0.5416

epoch was used. Figure 12 and Figure 13 show that even though we optimized MobileNetV2 it performs slightly worse than MobileNetV1. Both CNN were able to distinguish COVID-19 and normal lungs from the other classes peaking in a COVID-19 accuracy of 83% (MobileNetV2 78%). Both networks had major problems to differentiate lung opacities from other lung problems. Which leads to lower average accuracy of 70% for the MobileNetV1 and 64% for the MobileNetV2.

On the upside, the problematic classes are not restricted in terms of images therefore the same models were trained again on a bigger training data set. For each class except COVID-19, 2000 images instead of 1000 images were used. After training the new models were tested on the same test set. The results are summarized in Figure 14 and Figure 15.

The increase in all classes except COVID-19 leads to an increase in all KPIs and raises the accuracy for COVID-19 detection up to 95% (MobileNetV2 91%).

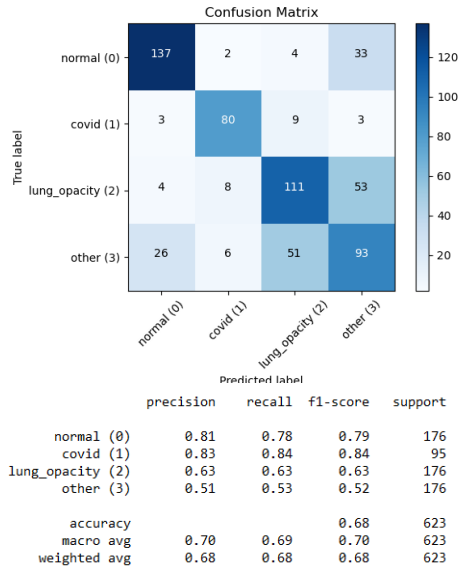


Figure 12: Classification result of MobileNetV1 using 1000 images in each Non-COVID-19 class

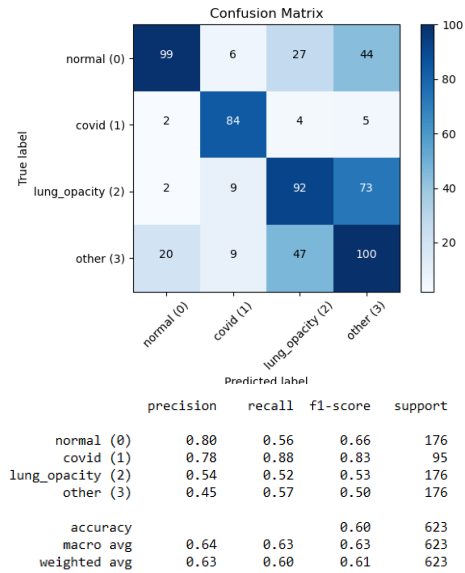


Figure 13: Classification result of MobileNetV2 using 1000 images in each Non-COVID-19 class

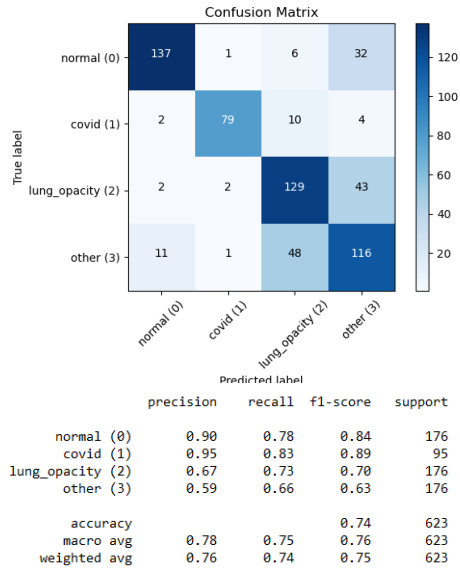


Figure 14: Classification result of MobileNetV1 using 2000 images in each Non-COVID-19 class

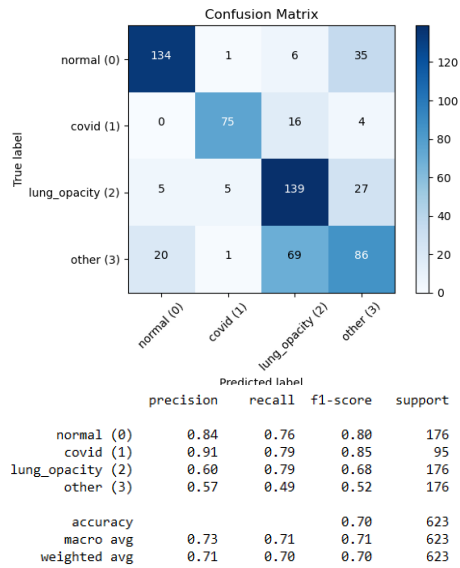


Figure 15: Classification result of MobileNetV2 using 2000 images in each Non-COVID-19 class

Table 5: Results from the optimization of the unfreeze level

Unfreeze	Val. Accuracy
1	0.4087
2	0.4008
3	0.3847

5.5 Interpretability analysis

Model interpretability is paramount in healthcare applications, because in standard medical treatment, an explanation, accompanying the diagnosis, is required from a physicians. As an instance of post-hoc, gradient-based interpretability methods, GradCAM [28] is a technique that uses the gradients of any target layer (usually the last convolutional layer) to compute a localization heatmap highlighting the most informative parts of an input image given a class label.

Following the training of the MobileNetV2 model, we computed CAMs of all images in the test dataset and provide a few exemplary maps in Figure 16. While it can be seen that the model successfully learned to highlight pulmonary markers like GGOs in, e.g., the COVID-19 example, it is evident from the negative examples that the model is prone to artifacts and occasionally focuses on irrelevant regions such as borders.

6 Discussion

Previous experiments show that the chosen CNNs can identify COVID-19 infected lungs rather good on the test set, despite the small training set. This chapter is dedicated to point out different possible reasons for the examined behaviour. First, COVID-19 has some unique features to other infections and thereby can be identified easier than other classes. This theses would be supported by the fact that also normal lungs have a quite good accuracy compared to the classes lung opacity and other where the images are much harder to distinguish. Another reason for the good identification of COVID-19 might be

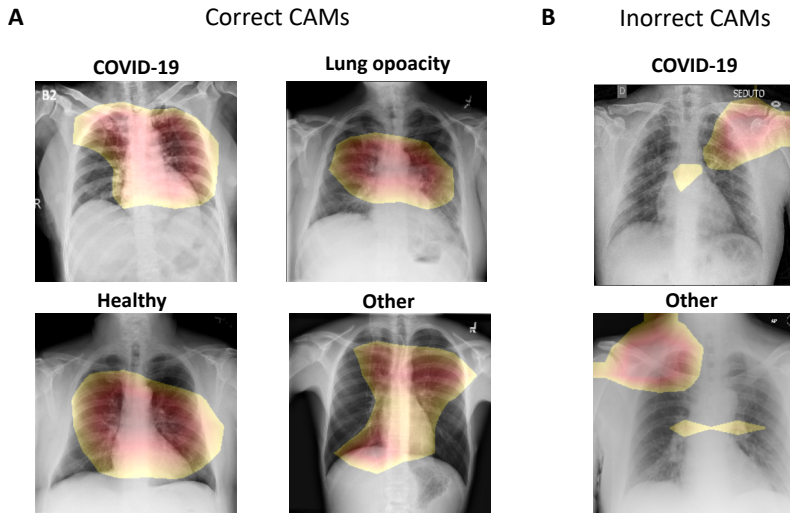


Figure 16: Visualizations of class activation maps with GradCAM [28]. **Subfigure A** shows four samples from the test dataset with CAMs focused on the chest. For example, the heatmap in the COVID-19 is highlighting GGOs. **Subfigure B** depicts two cases where the CAMs are not aligned to pathologically relevant areas.

the underlying data. Due to its novelty COVID-19 images are rare and available data is more homogeneous than the data in the other classes. This can lead to a bias in terms of accuracy where the CNN predicts COVID-19 always when the pictures differ from the homogeneous data from all other classes combined. Then it might not learn the features of COVID-19 but other differences such as the quality of the image to classify. The presence of this problem at least to some degree is supported by the last experiment with increased samples for all classes but COVID-19. Even though, the amount of COVID-19 samples stayed the same, its accuracy increased by 12% (MobileNetV1). Since this finding has been a major problem for our work, we started to analyze the predictions with the help of GradCA and found that our model can, to some extent, learn to focus on the relevant regions in the image. Although we also found a significant amount of samples with a mislocated attentional focus, it is encouraging to see the model extracting spatial biomarkers in a self-supervised way and without any segmentation masks. However, in a next step, lung segmentation with a U-



Figure 17: Lung segmentation

Net segmentation model [29] will be used to exclude identification by features other than the lung. An example is shown in Figure 17.

7 Conclusion and future work

In this work different CNN's were compared and the most promising (MobileNetV2) was optimized. Unexpectedly, MobileNetV1 still performed better and peaked in an overall accuracy of 78% and a COVID-19 accuracy of 95% the model was able to detect COVID-19 infected lungs. These scores are relatively high especially for COVID-19. This increases the risk of the probability that the underlying data mainly generated from two sources introduced some bias. This work raises multiple questions, which can be examined in further research. One topic might be why the cases of COVID-19 are relatively easy to identify and in combination with that what are the key characteristics on which a CNN determines the presence of COVID-19. This research could help to further understand the virus and ultimately might help to find a cure. Another topic might be to compare the performance of CNNs with support vector machines or other image classification methods.

References

- [1] World Health Organization et al. Coronavirus disease 2019 (covid-19): situation report, 176. 2020.
- [2] Zhengtu Li, Yongxiang Yi, Xiaomei Luo, Nian Xiong, Yang Liu, Shaoqiang Li, Ruilin Sun, Yanqun Wang, Bicheng Hu, Wei Chen, Yongchen Zhang, Jing Wang, Baofu Huang, Ye Lin, Jiasheng Yang, Wensheng Cai, Xuefeng Wang, Jing Cheng, Zhiqiang Chen, Kangjun Sun, Weimin Pan, Zhifei Zhan, Liyan Chen, and Feng Ye. Development and clinical application of a rapid igm-igg combined antibody test for sars-cov-2 infection diagnosis. *Journal of Medical Virology*.
- [3] Mahmud Mossa-Basha, Carolyn C. Meltzer, Danny C Kim, Michael J Tuite, K. Pallav Kolli, and Bien Soo Tan. Radiology department preparedness for covid-19: Radiology scientific expert panel. *Radiology*, 0(0):200988, 0. PMID: 32175814.
- [4] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences, 2014.
- [5] Rajat Raina, Alexis Battle, Honglak Lee, Benjamin Packer, and Andrew Y. Ng. Self-taught learning: Transfer learning from unlabeled data. ICML '07, page 759–766, New York, NY, USA, 2007. Association for Computing Machinery.
- [6] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '14*, page 580–587, USA, 2014. IEEE Computer Society.
- [7] Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H Schwartz, and Hugo JWL Aerts. Artificial intelligence in radiology. *Nature Reviews Cancer*, 18(8):500–510, 2018.
- [8] Xiaoxuan Liu, Livia Faes, Aditya U Kale, Siegfried K Wagner, Dun Jack Fu, Alice Bruynseels, Thushika Mahendiran, Gabriella Moraes, Mohith

- Shamdas, Christoph Kern, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *The lancet digital health*, 1(6):e271–e297, 2019.
- [9] Ayelet Akselrod-Ballin, Michal Chorev, Yoel Shoshan, Adam Spiro, Alon Hazan, Roie Melamed, Ella Barkan, Esma Herzel, Shaked Naor, Ehud Karavani, et al. Predicting breast cancer by applying deep learning to linked health records and mammograms. *Radiology*, 292(2):331–342, 2019.
- [10] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silvana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghighoo, Robyn Ball, Katie Shpanskaya, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 590–597, 2019.
- [11] Diego Ardila, Atilla P Kiraly, Sujeeth Bharadwaj, Bokyung Choi, Joshua J Reicher, Lily Peng, Daniel Tse, Mozziyar Etemadi, Wenxing Ye, Greg Corrado, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nature medicine*, 25(6):954–961, 2019.
- [12] Di Dong, Zhenchao Tang, Shuo Wang, Hui Hui, Lixin Gong, Yao Lu, Zhong Xue, Hongen Liao, Fang Chen, Fan Yang, et al. The role of imaging in the detection and management of covid-19: a review. *IEEE reviews in biomedical engineering*, 2020.
- [13] Jannis Born, David Beymer, Deepta Rajan, Adam Coy, Vandana V Mukherjee, Matteo Manica, Prasanth Prasanna, Deddeh Ballah, Pallav L Shah, Emmanouil Karteris, et al. On the role of artificial intelligence in medical imaging of covid-19. *medRxiv*, 2020.
- [14] Feng Shi, Jun Wang, Jun Shi, Ziyang Wu, Qian Wang, Zhenyu Tang, Kelei He, Yinghuan Shi, and Dinggang Shen. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. *IEEE reviews in biomedical engineering*, 2020.

- [15] Anwaar Ulhaq, Asim Khan, Douglas Gomes, and Manoranjan Pau. Computer vision for covid-19 control: A survey. *arXiv preprint arXiv:2004.09420*, 2020.
- [16] Damiano Caruso, Marta Zerunian, Michela Polici, Francesco Pucciarelli, Tiziano Polidori, Carlotta Rucci, Gisella Guido, Benedetta Bracci, Chiara de Dominicis, and Andrea Laghi. Chest ct features of covid-19 in rome, italy. *Radiology*, page 201237, 2020.
- [17] Saleh Albelwi and Ausif Mahmood. A framework for designing the architectures of deep convolutional neural networks. *Entropy*, 19(6):242, 2017.
- [18] Ali Narin, Ceren Kaya, and Ziyne Pamuk. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks, 2020.
- [19] Ioannis D. Apostolopoulos and Tzani Bessiana. Covid-19: Automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, 2020.
- [20] Jun Chen, Lianlian Wu, Jun Zhang, Liang Zhang, Dexin Gong, Yilin Zhao, Shan Hu, Yonggui Wang, Xiao Hu, Biqing Zheng, Kuo Zhang, Huiling Wu, Zehua Dong, Youming Xu, Yijie Zhu, Xi Chen, Lilei Yu, and Honggang Yu. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study. *medRxiv*, 2020.
- [21] Charmaine Butt, Jagpal Gill, David Chun, and Benson A Babu. Deep learning system to screen coronavirus disease 2019 pneumonia. *Applied Intelligence*, page 1, 2020.
- [22] Danilo Buonsenso, Davide Pata, and Antonio Chiaretti. Covid-19 outbreak: less stethoscope, more ultrasound. *The Lancet Respiratory Medicine*, 8(5):e27, 2020.
- [23] Jannis Born, Gabriel Brändle, Manuel Cossio, Marion Disdier, Julie Goulet, Jérémie Roulin, and Nina Wiedemann. Pocovid-net: Automatic

- detection of covid-19 from a new lung ultrasound imaging dataset (pocus). *arXiv preprint arXiv:2004.12084*, 2020.
- [24] Jannis Born, Nina Wiedemann, Gabriel Brändle, Charlotte Buhre, Bastian Rieck, and Karsten Borgwardt. Accelerating covid-19 differential diagnosis with explainable ultrasound image analysis. *arXiv preprint arXiv:2009.06116*, 2020.
 - [25] Subhankar Roy, Willi Menapace, Sebastiaan Oei, Ben Luijten, Enrico Fini, Cristiano Saltori, Iris Huijben, Nishith Chennakeshava, Federico Mento, Alessandro Sentelli, et al. Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound. *IEEE Transactions on Medical Imaging*, 2020.
 - [26] Roman Kalkreuth and Paul Kaufmann. Covid-19: A survey on public medical imaging data resources, 2020.
 - [27] Joseph Paul Cohen, Paul Morrison, and Lan Dao. Covid-19 image data collection. *arXiv 2003.11597*, 2020.
 - [28] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017.
 - [29] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells III, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer, 2015.