COVID-19 Detection from Chest Radiography Images: A Comparison of Deep Learning Algorithms

Li Nguyen, ID: 934644485
Department of Informatics
Technical University of Munich
li.nguyen@tum.de

Alexander Koenig, ID: 918254061
Department of Informatics
Technical University of Munich
awc.koenig@tum.de

***Abstract*—The ongoing COVID-19 pandemic challenges the world to treat and reduce the growing number of daily infections. Effectively cutting off transmission routes requires broad testing. Methods using chest X-ray images and deep learning methods could help to satisfy this growing demand for rapid testing. This work compares three deep learning approaches for classifying chest radiography images into the categories "normal", "pneumonia" and "COVID-19 pneumonia". Our first method performs transfer learning with a pre-trained ResNet-50. Further, we propose a model that relies on anomaly detection with the U-Net architecture and a ResNet-50 classifier. Our third approach is based on multitask learning where a modified U-Net performs a reconstruction task and a classification task simultaneously. We obtain the best results from the multitask learning method (average sensitivity of 80.0% and average precision of 81.6%). The code repository of this project can be accessed via https://github.com/axkoenig/dl4mi**

Keywords: COVID-19 Detection, Chest X-ray, Deep Learning, U-Net, Autoencoder, Anomaly Detection, Transfer Learning, Multitask Learning

I. Introduction

The COVID-19 pandemic puts countries' medical systems under challenging pressure, not only due to the number of available intensive care beds and respirators but also due to the speed at which the population can be tested. The faster the testing can occur the faster virus transmission routes can be cut off [29]. Currently, the diagnosis is based on Polymerase Chain Reaction (PCR), which is a time-consuming process. Moreover, test capacities for PCR are sometimes still too low to test a large number of people. Especially under the essential need to correctly diagnose the infection to cut off transmission routes, analyzing X-ray images can help in further solidifying the diagnosis.

Radiologists found that X-ray images show ground-glass opacity as a COVID-19-specific visual indicator for an infection [24]. By leveraging deep learning methods for analyzing and visualizing these visual indicators we aim to support the clinical process of diagnosing COVID-19. In comparison to equipment of PCR tests, X-ray machines are more frequently available in hospitals and even smaller medical practices. Due to ethical concerns and laws, a diagnosis completely based on a deep learning model will not be possible shortly. However, these models can help to visualize the infected lung tissues and even estimate the progress state of the disease via a severity scoring [38]. In this project, we focus on developing deep learning models that can classify chest X-ray images as healthy, non-COVID pneumonia, or COVID-19 infection.

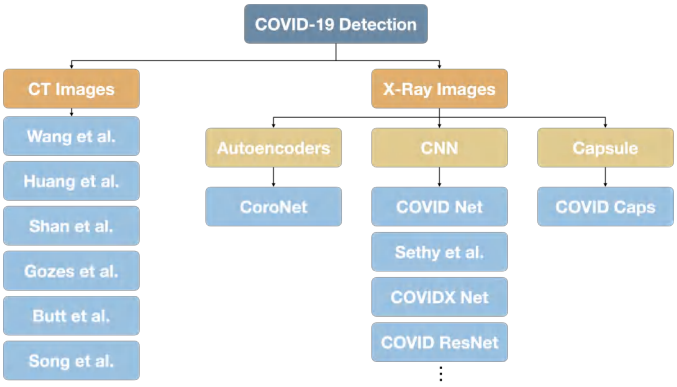

Fig. 1. Overview of related work

This report is structured as follows: section I gives an overview of previous work done in the area of deep learning for diagnosing COVID-19 which is grouped into the types of approaches the research groups used. In section III we present our network architectures and we evaluate the results in chapter IV-D. We provide a critical discussion of the results in chapter V-D. Lastly, we conclude and give ideas on future work in section VI.

II. Related Work

Researchers put a strong effort into understanding the disease better to cope with its spread. The artificial intelligence research community has leveraged machine and deep learning methods to aid the diagnosis of COVID-19. There are approaches using CT Images [25] [32] [15] [37] [34] [19] [7] and ones that use X-ray images for classification, see Figure 1. X-ray machines are more widely distributed in both hospitals and smaller practices and since our project uses X-ray images we are not going into detail about approaches using CT images. Most of the related projects use a type of CNN architecture [35] [26] [33] [1] [18] [14] [16] and [12]. The convolutional neural network (CNN) approaches get a good accuracy for classifying into healthy, non-COVID and COVID classes. However, any of them do not provide a sophisticated visualization that helps to support the radiologists' diagnosis.

A. COVID-Net by Wang et al.

The earliest and most prominent work at the moment is the COVID-Net by Wang et al. [35] who proposed the first version of this COVID-19 detection network in only 7 days. Their "generative synthesis" approach generates several architectures and compares them on a test set. The best performing neural network is characterized by selective long-range connectivity and densely-connected layers. An architectural advantage of the synthesized COVID-Net is that it contains intermediate layers that act as central hubs and compensate computational complexity and memory consumption of classical densely-connected architectures. Furthermore, they assembled the COVIDx dataset which is currently the biggest and most broadly used baseline dataset. Several other research groups have tested their models on the COVIDx [18] [12] [23]. The dataset by J. Cohen [9] which is one of the largest COVID-19 chest X-ray image collections is one of the building blocks of COVIDx.

B. Combining CNNs and SVM

An approach which differs slightly from the other CNN-based approaches is the one of Sethy et al. They compare the performance of different CNN architectures combined with a support vector machine (SVM) classifier [31]. The features of a deep layer are extracted and fed to the SVM classifier. They compare the following architectures: AlexNet, DenseNet201, GoogLeNet, InceptionV3, ResNet18, ResNet50, ResNet101 VGG16, VGG19, XceptionNet and InceptionResNetV2. The ResNet50 outperforms the other models with an accuracy of 95.38% and a specificity of 93.47%. They combined the dataset of Cohen et al. [9] and the Kaggle Pneumonia Dataset [22].

C. Anomaly Detection

Zhang et al. [39] focus on differentiating viral pneumonia from non-viral pneumonia and healthy lungs and formulate this as a one-class classification-based anomaly detection. Their anomaly detection model consists of a shared feature extractor, an anomaly detection module, and a confidence prediction model. An X-ray image then counts as an outlier if it lies above a defined anomaly score. The advantage of this approach is that all known diseases are treated as one class and detected outliers can either be COVID-19 or a harbinger for newly arising diseases. They use an in-house dataset combined from the X-VIRAL and the X-COVID dataset which are collected from 6 institutions and consist of 106 COVID-19 and 107 normal cases. The highest accuracy they achieve is 80.65% and the highest specificity is 79.87%.

Khobani et al. [23] use autoencoders to encode the latent representation of healthy and non-COVID pneumonia. For the classification of an X-ray image at hand, the sample is fed through both autoencoders. Ideally, only the healthy and non-COVID specific features are reconstructed. They build a residual tensor from the two reconstructed images which is then fed into a ResNet18 classifier. This allows them to exploit the limited available data of COVID-19 and also provides tools that can explain the diagnosis. They reach an accuracy of 93.50% and a precision of 93.63%.

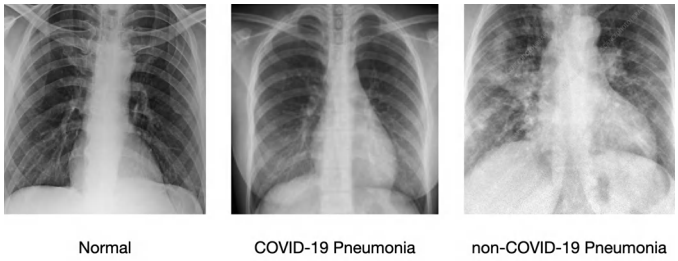

Fig. 2. Goal: discriminate between the three classes

D. Capsule Networks

Afshar et al. [2] leverage capsule networks that were first proposed by Sabour et al. [28]. The advantage of capsule networks is that spatial information such as object pose and orientation is preserved better than in CNNs. Furthermore, CNNs require larger datasets and a higher number of training parameters than capsule networks. Furthermore, Afshar et al. apply the concept of transfer learning by using images of the same domain. In comparison to pre-training on natural images, they were able to improve their accuracy from initially 95.7% to 98.3% and their specificity from 95.8% to 98.6%. Wang et al. for example performed transfer learning using natural images from ImageNet [35] [11].

Generally, all existing projects deal with the problem of an unbalanced dataset since COVID-19 is a newly emerging disease and data has yet to be collected. There are plenty of X-ray images of non-COVID pneumonia and healthy subjects. Most of the CNN-based projects deal with the unbalanced dataset by over- or undersampling or a weighted loss function. The last approach of Khobani et al. tackles this issue architecturally: at an initial stage one autoencoder is trained for the healthy and non-COVID pneumonia images each. In a second stage all images (i.e. healthy, pneumonia and COVID-19) are processed by a second classifier module which uses the autoencoders trained in the first stage.

III. Dataset

We use the COVIDx3 dataset of Wang et al. [35] which is comprised of the following openly accessible datasets.

1) COVID-19 Image Data Collection by Cohen et al. [9]
2) Figure 1 COVID-19 Chest X-ray Dataset Initiative [8]
3) RSNA Pneumonia Detection Challenge dataset [22]
4) ActualMed COVID-19 Chest X-ray Dataset Initiative [3]
5) COVID-19 radiography database [21]

The COVIDx3 dataset is unbalanced. The training set constists of only 253 COVID-19 cases, while there are 7,966 healthy cases and 5,451 non-COVID pneumonia images. The test set consists of 300 samples (i.e. 100 samples for each class). It is important to note that only train and test sets but no validation set are provided. Figure 2 shows samples images from the dataset.

IV. METHODS

A. ResNet-50 Baseline Classifier

The first approach is motivated by the recent success of transfer learning. Transfer learning refers to the process of using a pre-trained neural network and fine-tuning its parameters to a specific task. We use the ResNet-50 architecture [17] which was pre-trained on 1000 classes from the ImageNet dataset [11]. The network hereby learned to extract low-level features such as patterns and shapes in its first layers. As shown in figure 3 we use these first layers of the ResNet-50 as a pattern extractor. We keep all of its parameters fixed and replace the last layer with our own fully connected layer, which maps to three classes (i.e. "Normal", "Pneumonia", "COVID-19") instead of the original 1000 classes. Therefore, the high-level features in the last layer that are specific to the ImageNet dataset are discarded.

This approach serves as a baseline for comparison to the other two algorithms. With this baseline classifier f_θ we also evaluate which approach handles the imbalance of the COVIDx dataset the best. Our experiments showed that weighting the loss function provided slightly better results than oversampling the under-represented classes. Consequently, we choose the weighted cross-entropy function in equation 1 as our classification loss $\mathcal{L}_c$. Higher weights w_i are given to samples that occur less often in the dataset. To ensure comparability, the classification loss is the same for all presented approaches unless stated otherwise.

$$\mathcal{L}_c = \frac{1}{N}\sum_{n=1}^{N} w_i \mathbf{y_i} \log(f_\theta(\mathbf{X_i})) \tag{1}$$

B. Anomaly Detection with U-Net

The second approach is based on anomaly detection using deep autoencoders. Such approaches provided good results in tasks such as lesion segmentation in brain magnetic resonance images [5]. Due to its recent successes we chose the U-Net [27] autoencoder and used an implementation from the work by Buda et al. [6] which was originally used for abnormality segmentation in brain MRI. We slightly modified the network to output three channels instead of one since we want to reconstruct a full image instead of a binary segmentation map. The training of our method follows a two-step approach.

Stage 1: In this stage the autoencoder g_θ is trained to reconstruct non-anomalous data (i.e. healthy chest radiography images) by minimizing the reconstruction loss in equation 2. The network specializes in the task of reconstructing healthy chest radiography images in a fully unsupervised manner.

$$\mathcal{L}_r = \frac{1}{N}\sum_{n=1}^{N} \left(g_\theta(\mathbf{X_i}) - \mathbf{X_i}\right)^2 \tag{2}$$

Stage 2: Consequently, the network g_θ is used to run inference on images of all three classes (i.e. "Normal", "Pneumonia", "COVID-19"). Since the autoencoder was only trained on healthy data it should not be able to reproduce those image features that are specific to the disease. Thereby, the autoencoder produces an image that resembles a healthy image. To isolate the difference between the reconstructed image $g_\theta(\mathbf{X_i}$ and the original image $\mathbf{X_i}$ we subtract the reconstruction from the original to obtain a so-called anomaly map (also see figure 4). We then train a ResNet-50 classifier to discriminate between the three classes based on the anomaly map. The ResNet-50 follows the same architecture as our baseline classifier from section IV-A and was also pre-trained on ImageNet. While the parameters of the last layer of the ResNet-50 are trained from scratch on the new classification task, the parameters of the autoencoder g_θ are left unchanged.

C. Multitask Learning

The third approach for COVID-19 detection is based on multitask learning. The network is guided to solve a reconstruction task and a classification task simultaneously.

We augment U-Net with a classification network that is appended to U-Net's bottleneck (see figure 5). The architecture of the appended network is inspired by the work of Frid-Adar et al. [13] who used this approach to classify and segment endotracheal tubes in chest radiographies. The appended network consists of an average pooling layer which is used to condense the information in the bottleneck. It is followed by two fully connected layers and a softmax operation which map to the three classes. The overall loss function in equation 3 is a weighted combination between the classification loss $\mathcal{L}_c$ and the reconstruction loss $\mathcal{L}_r$. This joint optimization of the reconstruction and classification objective is motivated by the idea that both tasks can assist each other by producing more meaningful encodings in the bottleneck. The scaling parameter for the reconstruction loss is α and is experimentally determined.

$$\mathcal{L} = \mathcal{L}_c + \alpha\mathcal{L}_r \tag{3}$$

D. Implementation Details

The presented methods were implemented in the PyTorch deep learning framework and trained using an NVIDIA Tesla K80 graphics processor. We suspect that other works such as the one by Wang et al. [36] indirectly overfit to the COVIDx test set. The approach by Wang et al. is based on "generative synthesis" – a machine-driven design exploration strategy. Wang et al. did not mention validation data in their publication while COVIDx only provides a train and a test set. Therefore, we suspect that hyper-parameter tuning was done on the test set, which suggests that the estimated performance of their approach on unseen data may be too optimistic. To avoid indirectly overfitting on the test set we performed k-fold cross-validation (with $k = 10$). All classifiers were trained for 20 epochs in total (2 epochs per fold) with batch size 16. The U-Net in *Stage 1* of the anomaly detection approach was trained for 10 epochs and batch size 16. The Adam optimizer with standard hyperparameters was used (learning rate $s = 0.0002$, $\beta_1 = 0.9$, $\beta_2 = 0.999$).

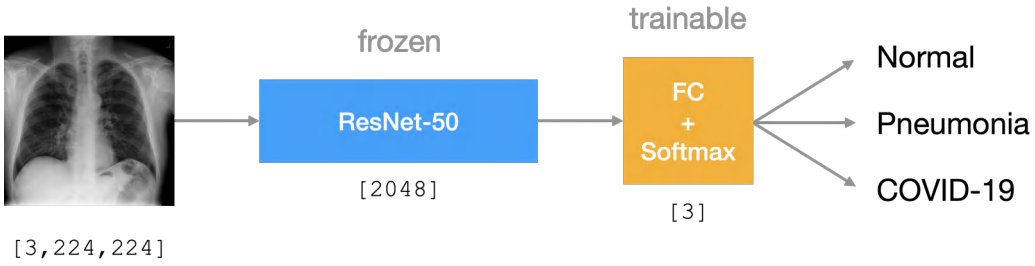

Fig. 3. Architecture of ResNet-50 Baseline Classifier

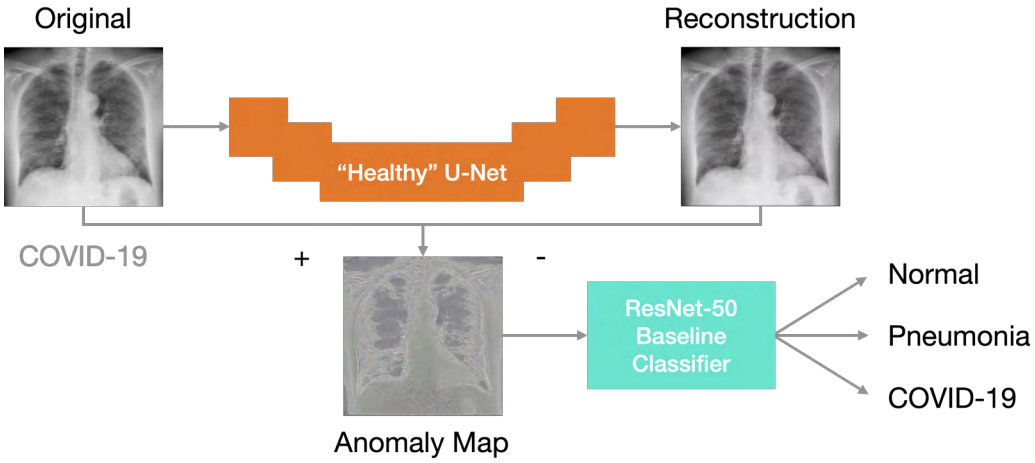

Fig. 4. Architecture of Anomaly Detector

Furthermore, the training data was augmented using the following methods in a pre-processing step.

- Random horizontal flip with probability $p = 0.5$
- Random rotation with range $r = \pm 5°$
- Random crop to [3,224,224] with scale $s \in [0.75, 1.0]$
- Random brightness change with factor $f = 0.2$
- Random contrast change with factor $f = 0.1$

V. Experimental Results

The performance of the three approaches on the test set is compared. We show confusion matrices for each approach, where the ground truth results are represented by the rows, and the algorithmic predictions are shown in the columns. Refer to tables IV and V to see a direct comparison of the sensitivity and precision of the approaches.

A. ResNet-50 Baseline Classifier

The ResNet-50 baseline provides good sensitivity for images that come from the normal and pneumonia class (both 88%). However, the sensitivity for COVID-19 images is much lower (only 42%). This comes from the unbalanced dataset which provides much fewer cases for COVID-19. What should be noted is that the precision of the COVID-19 is very high (i.e. if COVID-19 is predicted it is likely true).

TABLE I
Confusion Matrix of ResNet-50 Baseline Classifier

	Normal	Pneumonia	COVID-19
Normal	88	12	0
Pneumonia	11	88	1
COVID-19	18	40	42

B. Anomaly Detection with U-Net

The anomaly classifier provides worse results in the healthy class than the baseline classifier. However, the anomaly maps provide a benefit for the classes the U-Net was not trained on (i.e. 1% more sensitive for pneumonia images and 16% more sensitive for COVID-19 images).

An interesting byproduct of this approach are the anomaly maps. Figure 6 shows some sample images from all three classes. It can be noted that the anomaly maps highlight any haziness that is present in the original image. The "normal" anomaly maps in the left of figure 6 show a clear view of both lungs whereas for pneumonia cases the anomaly maps show more hazy patterns across the lungs. For the COVID-19 cases, the anomaly maps highlight the diffuse ground-glass opacity in both lower lobes, which aligns with previous pathological findings [10].

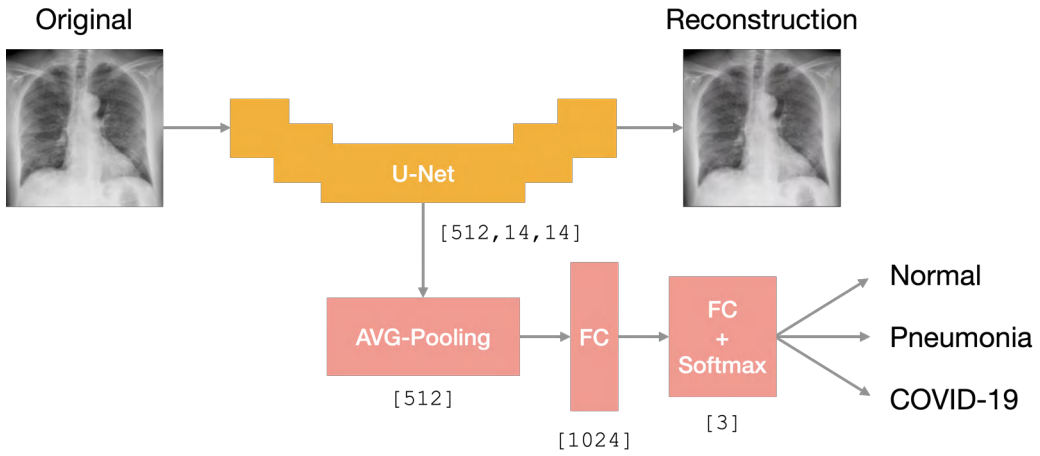

Fig. 5. Architecture of Multitask U-Net

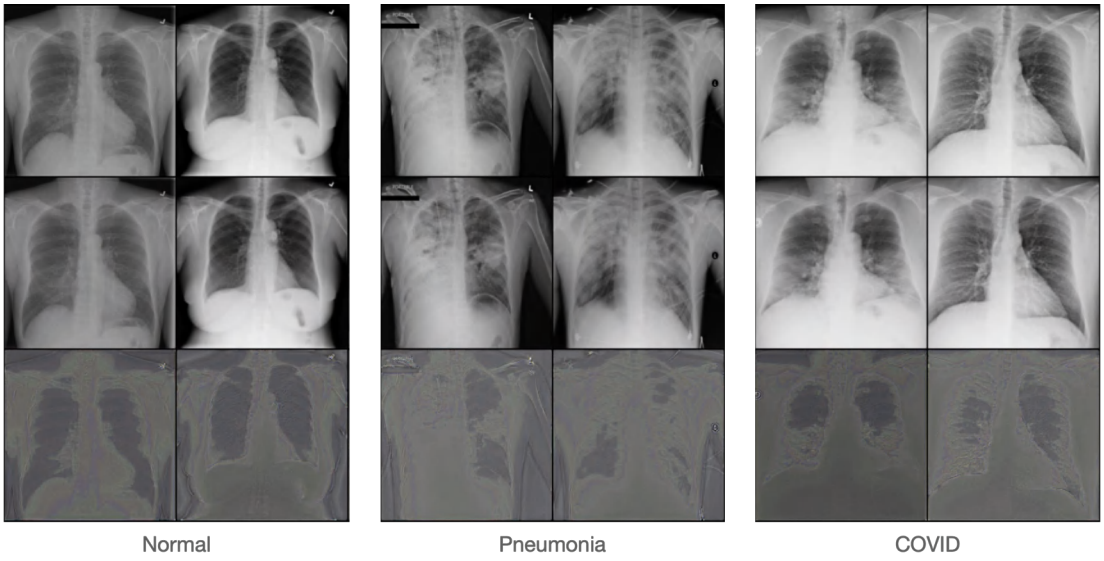

Fig. 6. Anomaly maps generated for all three classes

TABLE II
CONFUSION MATRIX OF ANOMALY DETECTION WITH U-NET

	Normal	Pneumonia	COVID-19
Normal	46	47	7
Pneumonia	3	89	8
COVID-19	7	35	58

TABLE III
CONFUSION MATRIX OF MULTITASK LEARNING

	Normal	Pneumonia	COVID-19
Normal	87	11	2
Pneumonia	4	88	8
COVID-19	7	28	65

C. Multitask Learning

The multitask learning approach provided the most promising results of all presented methods (see table III). The parameter $\alpha = 0.5$ performed the best in several experiments.

D. Comparison

From table IV it becomes clear that the multitask learning approach provided the highest sensitivity on the test set on average. For the normal and the pneumonia class the multitask learning method was not the most sensitive, although the sensitivity difference to the best performing method is only 1% each. It should be noted, that for the COVID-19 class the multitask learning approach outperformed the other approaches by a large margin. Furthermore, multitask learning also provides the best average precision while only being outperformed by the baseline classifier in the COVID-19 class.

In general, it is unsurprising that the performance of all approaches for the COVID-19 class is significantly worse than for the normal and the pneumonia class. This is due to the severe lack of data in the COVID-19 class. Interestingly, we

TABLE IV
AVERAGE AND CLASS SENSITIVITY OF APPROACHES (IN %)

	Normal	Pneumonia	COVID-19	Average
ResNet-50	**88.0**	88.0	42.0	72.7
Anomaly	46.0	**89.0**	58.0	64.3
Multitask	87.0	88.0	**65.0**	**80.0**

TABLE V
AVERAGE AND CLASS PRECISION OF APPROACHES (IN %)

	Normal	Pneumonia	COVID-19	Average
ResNet-50	75.2	62.9	**97.7**	78.6
Anomaly	82.1	52.0	79.5	71.2
Multitask	**88.8**	**69.3**	86.7	**81.6**

obtain high average precision scores for COVID-19 images, but less so in the pneumonia class.

VI. DISCUSSION AND OUTLOOK

One of the main challenges we addressed in this project was handling the imbalance of the dataset. We pre-processed the data using data augmentation, used a weighted loss function, and trained the models using cross-validation. Furthermore, we reduced dependency on COVID-19 images by using an autoencoder as an architectural building block of the anomaly detection approach. We verified that ResNet-50 is a good baseline model and compared different approaches with each other. We believe the multitask learning approach has a lot of potential for future development.

In future work, we would like to combine the weighted loss function with over- or undersampling. Furthermore, to deal with the unbalanced data other loss functions could be used, e.g. the cosine loss by Barz et al. [4] or the perceptual loss by Johnson et al. [20].

The project can further be extended by augmenting the original X-ray scans with an overlay that highlights the affected areas of the lungs. This could either be done by using Grad Gam by Selvaraju et al. [30] or by using the anomaly maps we presented in figure 6. It could be observed that the anomaly maps created for COVID-19 cases show more dense patterns in the lower lobes, while non-COVID pneumonia shows an overall haziness and healthy lungs are mostly clear. By overlaying colored anomaly maps onto the original X-ray scan a visualization tool for radiologists could be created. However, it should be investigated in cooperation with medical experts if the anomaly maps obtained from our second method are medically meaningful. A sanity check for the anomaly detection method should be performed by calculating the pixel-wise distance between the original and the reconstruction of healthy X-ray scans and sick images. If our assumption is correct, the average distance of healthy scans to their reconstruction should be much lower than the ones of sick scans, since the autoencoder is trained solely on healthy images and should be less good in reconstructing sick samples. The anomaly detection approach can further be extended as an early warning system when trained on all existing lung diseases. Outliers can be detected and further scrutinized to see newly arising diseases earlier.

VII. CONCLUSION

Our approach focuses on COVID-19 detection using Chest X-ray images and leverages the methods of deep learning. A big challenge is the imbalance of the dataset because only a few COVID-19 X-ray scans are available due to the novelty of the disease in comparison to other known pneumonia and healthy X-ray scans. To deal with this issue we try alternative architectures compared to other research groups that often implement classical CNNs. In our work, we compare three different approaches with each other of which two use an autoencoder. We found that the last architecture using multitask learning works best. However, for pure classification into COVID-19, non-COVID pneumonia, and healthy cases this does not outperform the state of the art classification task. On the other hand, our approaches bring other notable advantages such as visualization in the form of anomaly maps.

ACKNOWLEDGMENT

The authors would like to thank Dana and Hayit for their ongoing support during the Deep Learning in Medical Imaging lecture at Tel Aviv University. We hope to stay in touch in the future and extend this project further in collaboration with you.

REFERENCES

[1] Asmaa Abbas, Mohammed M Abdelsamea, and Mohamed Medhat Gaber. "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network". In: *arXiv preprint arXiv:2003.13815* (2020).

[2] Parnian Afshar et al. "Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images". In: *arXiv preprint arXiv:2004.02696* (2020).

[3] Chung et al. *Actualmed COVID-19 chest x-ray data initiative*. 2020. URL: https://github.com/agchung/Actualmed-COVID-chestxray-dataset.

[4] Bjorn Barz and Joachim Denzler. "Deep learning on small datasets without pre-training using cosine loss". In: *The IEEE Winter Conference on Applications of Computer Vision*. 2020, pp. 1371–1380.

[5] Christoph Baur et al. "Deep Autoencoding Models for Unsupervised Anomaly Segmentation in Brain MR Images". In: *Lecture Notes in Computer Science* (2019), 161fffdfffdfffd169. ISSN: 1611-3349. DOI: 10.1007/978-3-030-11723-8_16. URL: http://dx.doi.org/10.1007/978-3-030-11723-8_16.

[6] Mateusz Buda, Ashirbani Saha, and Maciej A Mazurowski. "Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm". In: *Computers in Biology and Medicine* 109 (2019). DOI: 10.1016/j.compbiomed.2019.05.002.

[7] Charmaine Butt et al. “Deep learning system to screen coronavirus disease 2019 pneumonia”. In: *Applied Intelligence* (2020), p. 1.

[8] A Chung. *Figure 1 COVID-19 chest x-ray data initiative*. 2020.

[9] Joseph Paul Cohen et al. “Covid-19 image data collection: Prospective predictions are the future”. In: *arXiv preprint arXiv:2006.11988* (2020).

[10] Diletta Cozzi et al. “Chest X-ray in new Coronavirus Disease 2019 (COVID-19) infection: findings and correlation with clinical outcome”. In: *La Radiologia Medica* (2020), p. 1.

[11] Jia Deng et al. “ImageNet: A Large-Scale Hierarchical Image Database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2009, pp. 248–255.

[12] Muhammad Farooq and Abdul Hafeez. “Covid-resnet: A deep learning framework for screening of covid19 from radiographs”. In: *arXiv preprint arXiv:2003.14395* (2020).

[13] Maayan Frid-Adar, Rula Amer, and Hayit Greenspan. *Endotracheal Tube Detection and Segmentation in Chest Radiographs using Synthetic Data*. 2019. arXiv: 1908.07170 [eess.IV].

[14] Biraja Ghoshal and Allan Tucker. “Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection”. In: *arXiv preprint arXiv:2003.10769* (2020).

[15] Ophir Gozes et al. “Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis”. In: *arXiv preprint arXiv:2003.05037* (2020).

[16] Lawrence O Hall et al. “Finding covid-19 from chest x-rays using deep learning on a small dataset”. In: *arXiv preprint arXiv:2004.02060* (2020).

[17] Kaiming He et al. *Deep Residual Learning for Image Recognition*. 2015. arXiv: 1512.03385 [cs.CV].

[18] Ezz El-Din Hemdan, Marwa A Shouman, and Mohamed Esmail Karar. “Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images”. In: *arXiv preprint arXiv:2003.11055* (2020).

[19] Lu Huang et al. “Serial quantitative chest ct assessment of covid-19: Deep-learning approach”. In: *Radiology: Cardiothoracic Imaging* 2.2 (2020), e200075.

[20] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. “Perceptual losses for real-time style transfer and super-resolution”. In: *European conference on computer vision*. Springer. 2016, pp. 694–711.

[21] Kaggle. *COVID-19 radiography database*. 2020. URL: https://www.kaggle.com/tawsifurrahman/covid19-radiography-database.

[22] Kaggle. *Radiological Society of North America. RSNA pneumonia detection challenge*. 2019. URL: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data.

[23] Shahin Khobahi, Chirag Agarwal, and Mojtaba Soltanalian. “CoroNet: A Deep Network Architecture for Semi-Supervised Task-Based Identification of COVID-19 from Chest X-ray Images”. In: *medRxiv* (2020).

[24] Weifang Kong and Prachi P Agarwal. “Chest imaging appearance of COVID-19 infection”. In: *Radiology: Cardiothoracic Imaging* 2.1 (2020), e200028.

[25] Lin Li et al. “Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT”. In: *Radiology* (2020).

[26] Ali Narin, Ceren Kaya, and Ziynet Pamuk. “Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks”. In: *arXiv preprint arXiv:2003.10849* (2020).

[27] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV].

[28] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. “Dynamic routing between capsules”. In: *Advances in neural information processing systems*. 2017, pp. 3856–3866.

[29] Marcel Salathé et al. “COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation.” In: *Swiss medical weekly* 150.11-12 (2020), w20225.

[30] Ramprasaath R Selvaraju et al. “Grad-cam: Visual explanations from deep networks via gradient-based localization”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 618–626.

[31] Prabira Kumar Sethy and Santi Kumari Behera. “Detection of coronavirus disease (covid-19) based on deep features”. In: *Preprints* 2020030300 (2020), p. 2020.

[32] Fei Shan et al. “Lung infection quantification of covid-19 in ct images with deep learning”. In: *arXiv preprint arXiv:2003.04655* (2020).

[33] Vishal Sharma and Curtis Dyreson. “COVID-19 detection using Residual Attention Network an Artificial Intelligence approach”. In: *arXiv preprint arXiv:2006.16106* (2020).

[34] Ying Song et al. “Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images”. In: *medRxiv* (2020).

[35] Linda Wang and Alexander Wong. “COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images”. In: *arXiv preprint arXiv:2003.09871* (2020).

[36] Linda Wang and Alexander Wong. *COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images*. 2020. arXiv: 2003.09871 [eess.IV].

[37] Shuai Wang et al. “A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)”. In: *MedRxiv* (2020).

[38] Melissa A Warren et al. “Severity scoring of lung oedema on the chest radiograph is associated with

clinical outcomes in ARDS". In: *Thorax* 73.9 (2018), pp. 840–846.

[39] Jianpeng Zhang et al. "Viral pneumonia screening on chest X-ray images using confidence-aware anomaly detection". In: *arXiv: 2003.12338* (2020).