COVID-19 DETECTION USING DEEP LEARNING

A project report,
submitted to **Dr. Anuradha Singh**
in the subject of Big Data Analytics

by

Amal Shaji (2017BCS-010)
Chirag Jindal (2017BCS-013)
Manvi Gupta (2017BCS-017)
Sajal Verma (2017BCS-025)
Sandarbh Yadav (2017BCS-027)
Saumya Gupta (2017BCS-028)
Vaibhav Garg (2017BCS-038)
Vikram Chaudhary (2017BCS-039)

विश्वजीवनामृतं ज्ञानम्

ABV INDIAN INSTITUTE OF INFORMATION TECHNOLOGY AND MANAGEMENT, GWALIOR-474 015

TABLE OF CONTENTS

CHAPTER 1

Origin of the Proposal

Coronavirus disease 2019 (COVID-19) is a highly communicable disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It was first identified in November 2019 in Wuhan city of China. The World Health Organisation (WHO) declared COVID-19 as pandemic in March 2020. The pandemic has massively impacted daily lives of people and economy of countries. As of 25 November 2020, there have been more than 59.6 million confirmed cases with more than 1.4 million deaths [1]. Most of the countries are doing manual testing to detect the infected persons. But this manual testing has a lot of disadvantages. Testing kits are sparsely available. Blood tests are not only costly but also take 5-6 hours to give results. There is need of better and efficient detection techniques which can produce results as early as possible. This is where deep learning comes into the picture. Deep learning methods are easy to operate, less expensive and save a lot of time. Practically, we need an accuracy close to 100% as incorrect identification may lead to further spread of the disease. The proposed model performs binary classification on chest X-ray images of patients and labels them as COVID-19 or No-Findings.

CHAPTER 2

Review of status of Research and Development in the subject

Detection of COVID-19 is a very popular research topic nowadays. Many researchers are investigating different methods of detecting COVID-19. Deep learning based techniques are being extensively used by the researchers.

2.1 International Status

The global pandemic due to COVID-19 caused an increase in demand for testing. Currently, Polymerase Chain Reaction (PCR) tests are being used for COVID-19 but they have their own limitations. Deep learning based detection techniques are faster, more effective and less expensive. Most of these techniques use chest X-ray images of the patients while others use CT scans. Che et al [2] used ResNet-101 on lungs X-ray images and obtained an accuracy of 72%. Wang et al [3] used modified network inception model on CT scans of patients. They extracted 195 regions of interests using deep learning and obtained an accuracy of 73%. Xu et al [4] reported that PCR tests have low detection rate in the early stages of COVID-19. They developed an early screening model which used ResNet-18 on CT scans to distinguish COVID-19 patients from influenza patients. Their model produced a detection accuracy of 86.7%. Gozes et al [5] used deep learning to analyse CT scans of patients from USA and China. They developed 2 subsystems to analyse CT scans at 2 different levels. One subsytem performed 3D analysis while the other performed 2D analysis on each segment of the scan. Their system distinguished COVID positive patients from COVID negative patients with a specificity of 92%. Hemdan et al [6] developed COVIDX-Net which helps in identifying COVID positive patients using chest X-ray images. The COVIDX-Net used 7 different deep learning models: Inception V3, Xception, ResNet V2, Inception-ResNet V2, MobileNet, VGG19 and DenseNet. Their system produced a F-score of

91% for COVID cases. Hassanien et al [7] developed a classifier which uses multilevel thresholding and Support Vector Machines (SVM) to detect COVID-19 in chest X-ray images. The classifier achieved an accuracy of 97%.

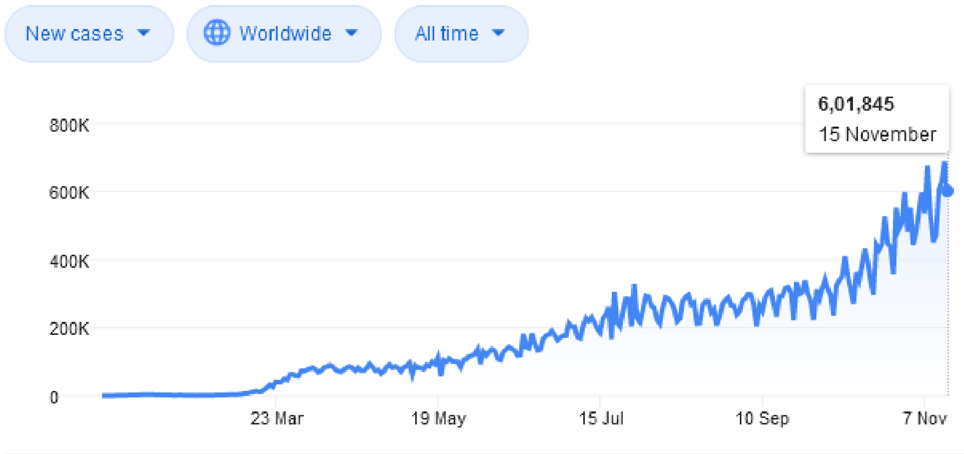

Figure 2.1: COVID cases worldwide

Narin et al [8] used ResNet-50 on chest X-ray images and achieved an accuracy of 98%. Yoo et al [9] used binary decision trees to detect COVID-19 in X-ray images. The decision trees achieved an accuracy of 95%. Alazab et al [10] used deep CNN on X-ray images to detect COVID. Their model produced a F-score of 95%. They also used Autoregressive Integrated Moving Average (ARIMA) model to make predictions about the number of coronavirus cases in Australia and Jordan. Their predictions were found out to be 94% accurate in Australia and 88% accurate in Jordan.

2.2 National Status

Although, PCR is used as the test for COVID-19 in India, researchers have made efforts to use deep learning based detection techniques as they are faster, more effective and less expensive. Sethy and Behera [11] used deep learning to extract features from chest X-ray images and using these features, classified the images as infected or healthy. They used 11 deep learning models: Xception, DenseNet, Inception V3, Inception-ResNet V2, ResNet 18, ResNet 50, ResNet 101, GoogleNet, AlexNet, VGG16 and VGG19. They achieved an accuracy of 95% using ResNet 50 and SVM. Jain et al [12] used 3 different deep learning models on chest X-ray images and compared their performances. The 3 models were based on Inception V3, Xception and ResNeXt frameworks. The Xception model produced the highest accuracy of 97%.

Figure 2.2: COVID cases in India

2.3 Importance of the proposed project in the context of current status

COVID-19 has spread very rapidly across the globe. The global pandemic has resulted in increase in demand for testing. Most of the countries are using PCR tests to detect the infected persons but this approach has its own disadvantages. Testing kits are sparsely available in many places. PCR tests are not only costly and but also take upto 24 hours to produce accurate results. There is need of better and efficient detection techniques which can produce results as quickly as possible. The proposed project aims to develop a deep learning based model which is faster, more effective and less expensive compared to current methods. Thus, the proposed project carries a lot of importance in context of current status. A successful breakthrough using deep learning might revolutionize the COVID detection methodology followed all around the world.

CHAPTER 3

Work Plan

This chapter discusses the methodology employed for the project and the results and conclusion.

3.1 Methodology

The basic idea of this project is to perform binary classification on chest X-ray images and label them as COVID-19 or No-Findings. Some researchers use CT scans but we have used X-rays as they are simpler, quicker and safer.

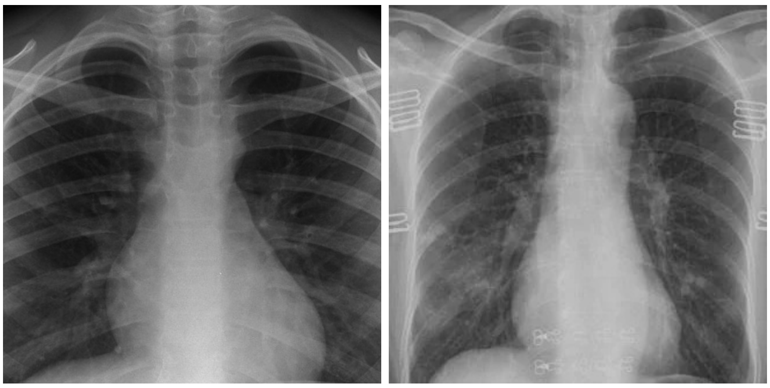

Figure 3.1: Chest X-ray of fit person vs COVID infected person

3.1.1 Tools and Technologies used

The project has been implemented in Python. It makes use of popular libraries like NumPy, PyTorch, Fast.ai etc.

3.1.2 Flow Diagram

The flow diagram is shown below. There are 4 major phases: Data Collection, Model Building, Training and Validation.

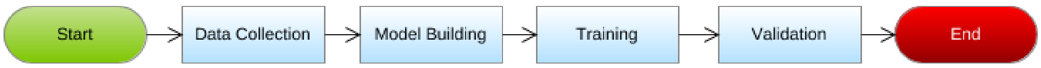

Figure 3.2: Flow Diagram

3.1.3 Data Collection

Deep learning techniques require a lot of data for training. In this project we have used 2 different datasets. Dr. Joseph Paul Cohen has provided an open-source database which contains chest X-ray images of patients infected by COVID-19 [13]. The database is continuously updated by new images. We have used this dataset for images of COVID infected patients. For chest X-ray images of fit persons, we have used another dataset provided by Wang et al [14].

3.1.4 Model Building

Deep learning is based on neural networks which mimic biological neurons. The term 'deep' in deep learning refers to the number of layers in the neural network. Convolutional Neural Networks (CNN) are a special type of neural networks which are mainly used for image recognition and image classification.

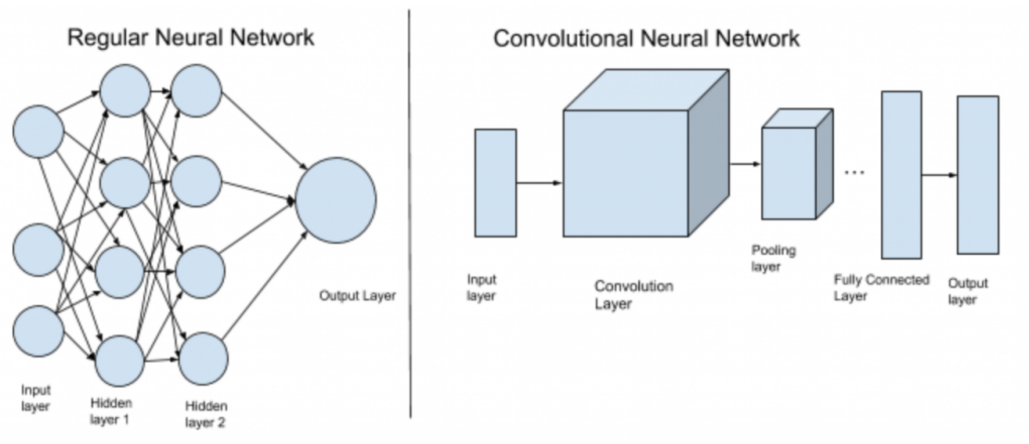

Figure 3.3: Regular Neural Network vs Convolutional Neural Network

CNN comprises of 3 types of layers:

- Convolution layer which uses filters to extract features from inputs
- Pooling layer which enhances computational performance by reducing size
- Fully-connected layer

A CNN model is built by combining one or more such layers. The internal parameters of the model are adjusted to perform a task like recognition or classification.

CNN image classifier takes an image as input, processes it and classifies it in one of classes. In this project, there are 2 classes: COVID-19 and No-Findings. There are 2 approaches of solving a problem using deep learning. First approach is to build a network from scratch. This approach is quite hectic and consumes a lot of time. The other approach is to use use pre-trained networks like AlexNet, GoogleNet, DarkNet etc. and fine-tune them according to our needs. In this project, we have used DarkNet-19 model.

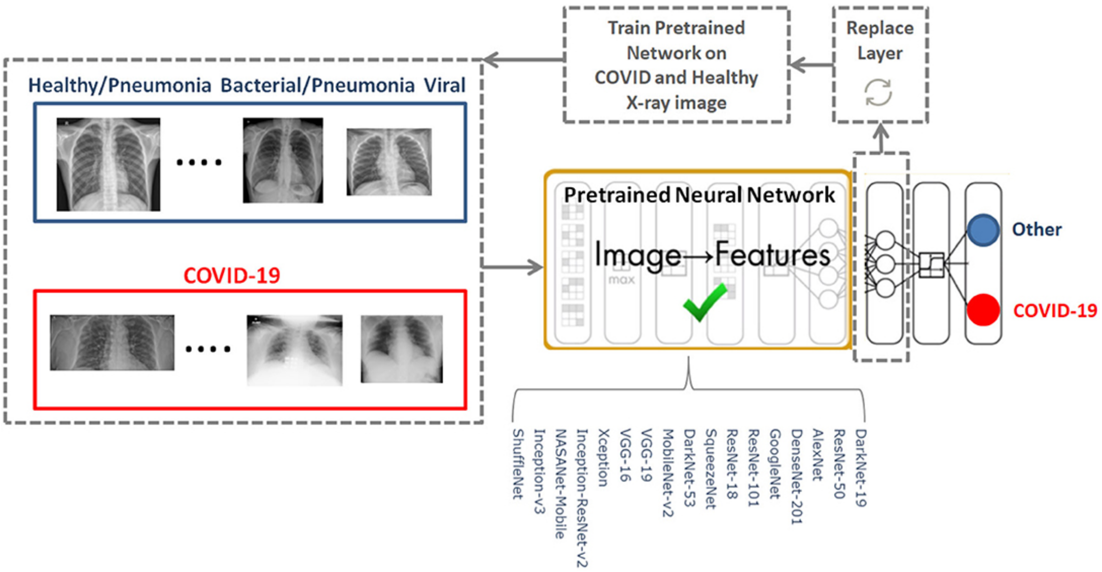

Figure 3.4: COVID detection using Deep Learning

There are 19 convolution layers and 5 pooling layers in DarkNet-19. The pooling layers implement Maxpool operation and are present after the 1^{st}, 2^{nd}, 5^{th}, 8^{th} and 13^{th} convolution layers. We have used a modified variation of DarkNet which has less number of layers and uses fewer filters. Avgpool and Softmax layers are present at the end of the model and produce the output. Figure 3.5 shows the operations in convolutional layer and Maxpool layer.

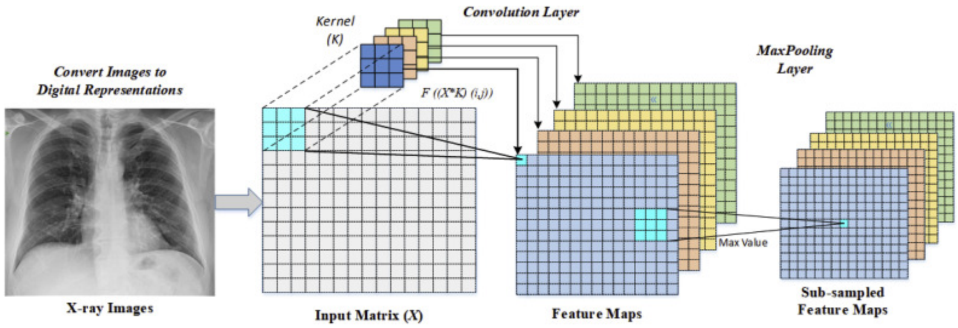

Figure 3.5: Convolution and Maxpool layer operations

Figure 3.6 shows the architecture of the model. There are 17 convolutional layers in total. Each DarkNet (DN) layer comprises of one convolution layer after which BatchNorm and LeakyReLU operations are performed in succession. BatchNorm operation standardizes the inputs and reduces the time consumed for training. LeakyReLU is a special version of ReLU operation and it prevents neurons from dying. In all the pooling layers, Maxpool operation is used. It determines the maximum of a region and downsizes the inputs. The proposed model has 1.16 million parameters. Adam optimizer is used for weight updation. Cross entropy is used as loss function and a learning rate of 0.003 is used.

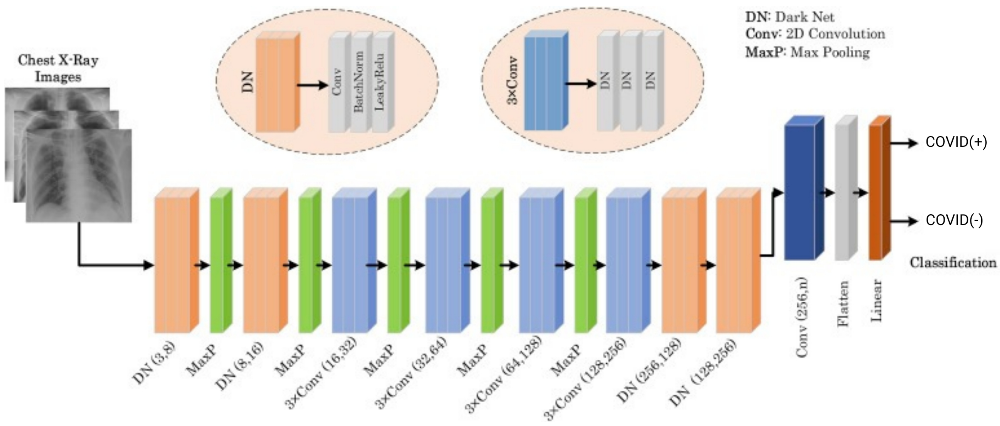

Figure 3.6: Architecture of proposed model

3.1.5 Training and Validation

K-fold cross validation is used to evaluate the performance of the model. We have set the value of k as 5. The images are divided into 5 folds. 80% of the images are used for training while 20% of the images are used for validation. This is repeated 5 times. The model is trained for 100 epochs. The model learns to classify the input images into 2 classes: COVID-19 and No-Findings.

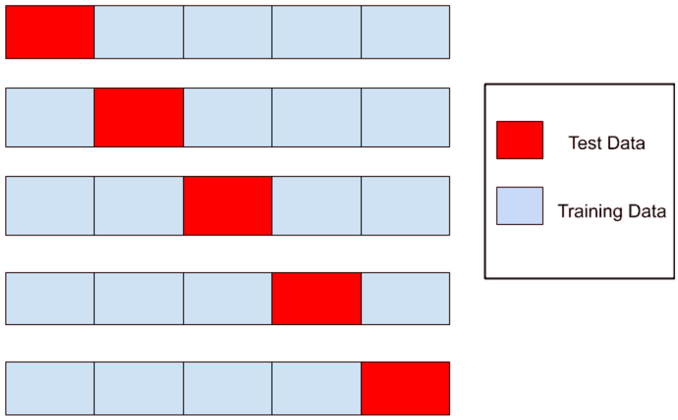

Figure 3.7: 5-fold cross validation

3.2 Results

Figure 3.8 shows the training and validation loss graphs of the model for first fold. Accuracy increases with increase in training.

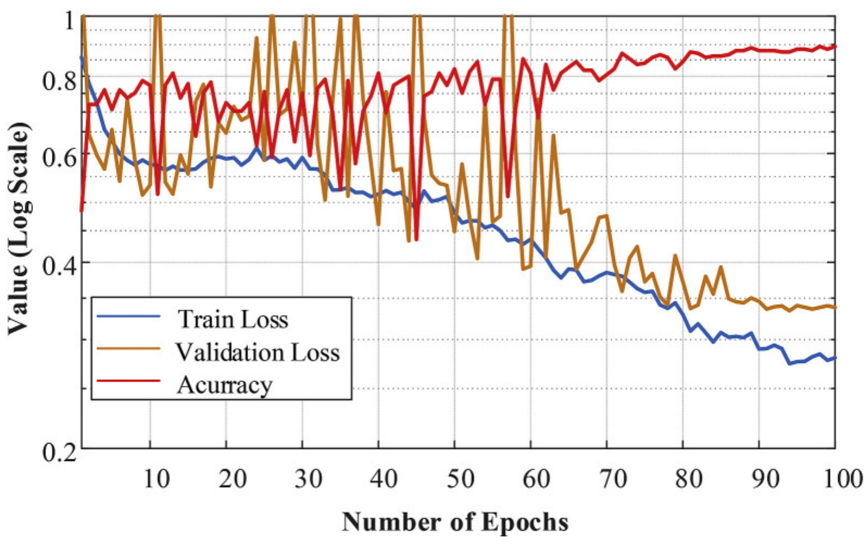

Figure 3.8: Training loss, validation loss and accuracy graphs

The confusion matrices of each fold along with overlapped confusion matrix is shown on next page. The overlapped confusion matrix is created by summing the confusion matrices of all 5 folds. The average accuracy of the proposed model is 98.08%.
Table 1 summarizes the details about the model like sensitivity, specificity, F1-score and precision.

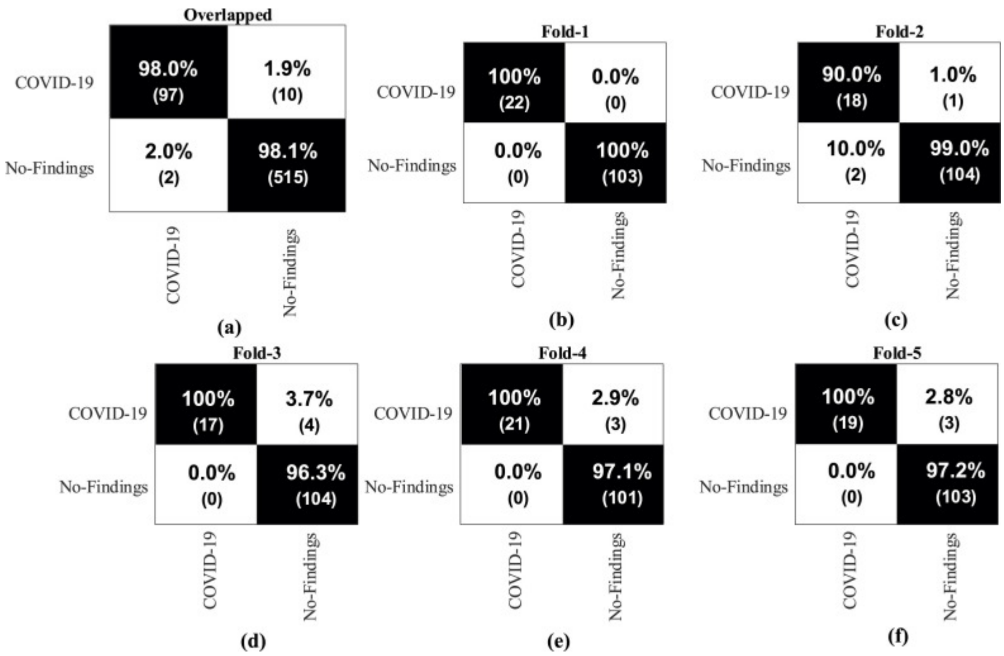

Figure 3.9: Confusion matrices

	Performance Metrics Values (in %)				
Folds	Specificity	Sensitivity	Precision	F-Score	Accuracy
Fold-1	100	100	100	100	100
Fold-2	96.42	96.42	94.52	95.52	97.60
Fold-3	90.47	90.47	98.14	93.79	96.80
Fold-4	93.75	93.75	98.57	95.93	97.60
Fold-5	93.18	93.18	98.58	95.62	97.60
		Overlapped			
COVID-19	90.65	99.61	97.97	94.17	98.07
No_Findings	99.61	90.65	98.09	98.84	98.07
Average	**95.13**	**95.3**	**98.03**	**96.51**	**98.08**

Table 1: Details of the model

3.3 Conclusion

COVID-19 has spread very rapidly across the globe. Most of the countries are doing manual testing to detect the infected persons but this manual testing has a lot of disadvantages. There was need of better and efficient detection techniques which can produce results as early as possible. Deep learning based techniques are faster, more effective and less expensive compared to manual testing methods of COVID-19. In this project, we have successfully developed a deep learning model which can detect COVID-19 with a pretty good accuracy of 98.08%. The model uses a modification of DarkNet-19 framework on chest X-ray images. However, cautionary measures should be taken while generalising the results to real world.

REFERENCES

[1] https://en.wikipedia.org/wiki/Coronavirus_disease_2019.

[2] M. Z. Che Azemin, R. Hassan, M. I. Mohd Tamrin, and M. A. Md Ali, "Covid-19 deep learning prediction model using publicly available radiologist-adjudicated chest x-ray images as training data: preliminary findings," *International Journal of Biomedical Imaging*, vol. 2020, 2020.

[3] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, *et al.*, "A deep learning algorithm using ct images to screen for corona virus disease (covid-19)," *MedRxiv*, 2020.

[4] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Q. Ni, Y. Chen, J. Su, *et al.*, "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, 2020.

[5] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, and E. Siegel, "Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis," *arXiv preprint arXiv:2003.05037*, 2020.

[6] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," *arXiv preprint arXiv:2003.11055*, 2020.

[7] A. E. Hassanien, L. N. Mahdy, K. A. Ezzat, H. H. Elmousalami, and H. A. Ella, "Automatic x-ray covid-19 lung image classification system based on multi-level thresholding and support vector machine," *medRxiv*, 2020.

[8] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *arXiv preprint arXiv:2003.10849*, 2020.

[9] S. H. Yoo, H. Geng, T. L. Chiu, S. K. Yu, D. C. Cho, J. Heo, M. S. Choi, I. H. Choi, C. Cung Van, N. V. Nhung, *et al.*, "Deep learning-based decision-tree classifier

for covid-19 diagnosis from chest x-ray imaging," *Frontiers in medicine*, vol. 7, p. 427, 2020.

[10] M. Alazab, A. Awajan, A. Mesleh, A. Abraham, V. Jatana, and S. Alhyari, "Covid-19 prediction and detection using deep learning," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 12, pp. 168–181, 2020.

[11] P. K. Sethy and S. K. Behera, "Detection of coronavirus disease (covid-19) based on deep features," *Preprints*, vol. 2020030300, p. 2020, 2020.

[12] R. Jain, M. Gupta, S. Taneja, and D. J. Hemanth, "Deep learning based detection and analysis of covid-19 on chest x-ray images," *Applied Intelligence*, pp. 1–11, 2020.

[13] `https://github.com/ieee8023/covid-chestxray-dataset`.

[14] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2097–2106, 2017.