International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 12 | Dec - 2021 ISSN: 2582-3930

Detection of Covid-19 cases using X-ray images

Ms Neelam Sharma (Assistant Professor)
Maharaja Agrasen Institute of Technology, GGSIPU, India

Himanshu Singh Rajawat
Maharaja Agrasen Institute of Technology, GGSIPU (India)

Hardik Bachhan
Maharaja Agrasen Institute of Technology, GGSIPU(India)

Rajat Goel
Maharaja Agrasen Institute of Technology, GGSIPU (India)

Abstract

As we know, the COVID-19 pandemic, also known as the coronavirus pandemic, is an ongoing pandemic of coronavirus disease 2019. It was first identified in December 2019 in Wuhan, China. Symptoms of COVID-19 are highly variable, ranging from none to severe illness. It is critical to detect the positive cases as early as possible so as to prevent the further spread of this epidemic and to quickly treat affected patients. The need for auxiliary diagnostic tools has increased as there are no accurate automated toolkits available. A clinical study of COVID-19 infected patients has shown that these types of patients are mostly infected from a lung infection after coming in contact with this disease. Chest x-ray (i.e., radiography) and chest CT are a more effective imaging technique for diagnosing lung- related problems. Still, a substantial chest x-ray is a lower cost process in comparison to chest CT. Deep learning is the most successful technique of machine learning, which provides useful analysis to study a large amount of chest x-ray images that can critically impact the screening of Covid-19.

So, this study aims to investigate if applying machine learning and deep learning approaches on chest X-ray images can detect cases of coronavirus. The chest X-ray datasets were obtained from Kaggle and pre-processed into a single dataset using random sampling. We applied several machine learning and deep learning methods including Convolutional Neural Networks (CNN) and also evaluated specificity and fall out rate along with accuracy to identify non-COVID-19 individuals more accurately. As a result, our new models might help to early detect COVID-19 patients and prevent community transmission compared to traditional methods.

1) Introduction

Our project is based on the CNN (Convolutional Neural Network) Model which is an artificial neural network (ANN) with multiple layers between the input and output layers. Using the Convolutional Neural Network (CNN) we will be layering Dataset X-ray images to get better accuracy of trained models.

Dataset will contain X-Ray images of lungs of people infected by Covid-19 as well as those not infected by covid-19. Images will be chosen manually and the entire model will be trained on more than over 400 images.

After the model is trained, we will do the prediction of Input X-Ray image, and set some threshold value for detecting positive and negative cases.

We will be taking Input on the web page, which will be created using HTML/CSS for better user experience and then output will be displayed as positive and negative on the web page with the help of API's we will be using in the project.

A) Image Processing

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be an image or characteristics/features associated with that image.

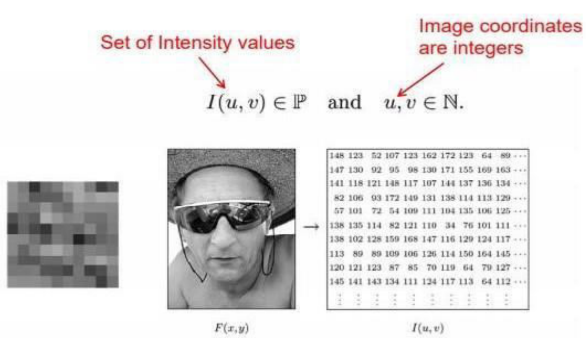


Figure 1 : 2-dimensional matrix of Intensity (grey or color) values

B) Digital Image

A digital image is an image composed of picture elements, also known as pixels, each with finite, discrete quantities of numeric representation for its intensity or grey level that is an output from its two-dimensional functions fed as input by its spatial coordinates denoted with x, y on the x-axis and y-axis, respectively.

C) Image Format

- Values per point/pixel (B&W or Grayscale)
- Values per point/pixel (Red, Green, and Blue)
- Values per point/pixel (Red, Green, Blue, + "Alpha" or Opacity)

Figure 2 : Image types

D) Imaging System

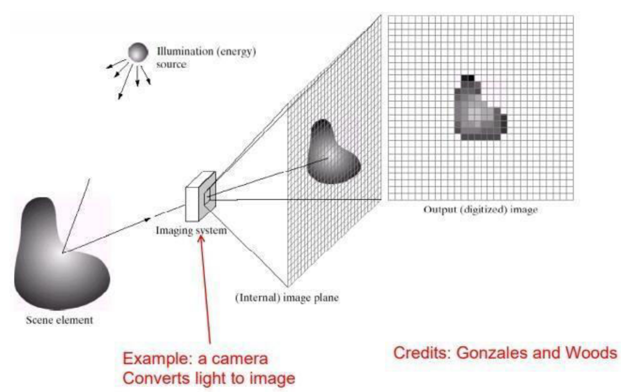


Figure 3 : Conversion of light to image

E) Image Processing

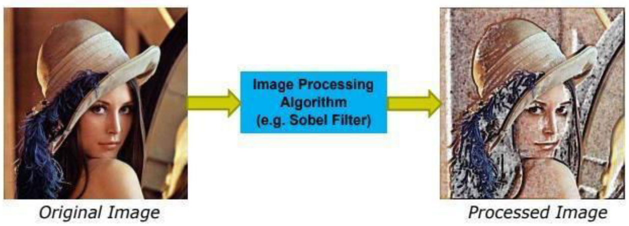


Figure 4 : Algorithms that alter an input image to create new image * Input is image, output is image.

2) Literature Survey

Recently, several studies investigated COVID-19 medical images using various deep learning-based methods. Wang et al. [25] introduced an open source deep learning model and realized a large benchmark dataset called COVIDx with 13,975 patient's chest X-ray images to explore COVID-19. This model not only ensured greater insights of COVID-19 critical factors but also extracted relevant information from the experimented images. Their study generated an accuracy of 93.3% over the COVIDx dataset. Maghdid et al. [3] applied deep convolutional neural network (CNN) and transfer learning using AlexNet into X-ray and CT scan image datasets. They identified 94% accuracy (specificity 88%) for CNN and 95% for AlexNet (specificity 92%). Then, Ali Narin et al.
[21] analyzed chest X-Ray images using three CNN models such as ResNet50, InceptionV3 and InceptionResNetV2 for COVID-19 automatic detection. Their study showed the highest 98% accuracy using 5-fold cross validation. Abbas et al. [1] proposed the DeTraC deep CNN model that aimed to give a solution by transferring knowledge from generic object recognition to domain-specific tasks. Their algorithm showed 93.21% accuracy (specificity of 91.87%, and a precision of 93.36%). Apostolopoulos and Mpesiana [3] implemented transfer learning using CNNs into a small medical image dataset. Their dataset contained 1427 X-ray images including 224 confirmed COVID-19 images.

Our best-performing CNN was also tested on another dataset prepared by Chowdhury et al. [21] to verify its generalizability and robustness further. The dataset contains 1341 healthy, and 423 COVID-19 X-ray images. All the images used are from the same source as the dataset used in this study. Our method obtained an accuracy of over 96% on this dataset after training.

The promising deep learning models used for the detection of COVID-19 from x-ray images indicate that deep learning likely still has untapped potential and can possibly play a more significant role in fighting this pandemic. There is definitely still room for improvement, through other processes such as increasing the number of images and implementing preprocessing techniques (i.e., data augmentation and/or image enhancement).

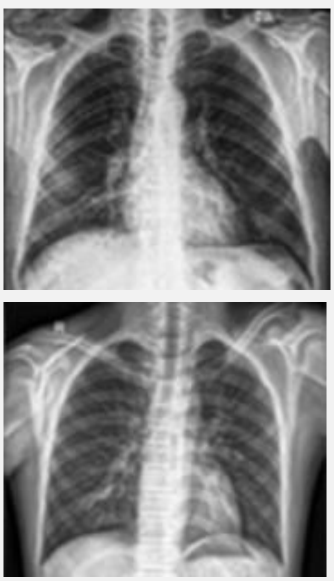
3) Proposed Architecture

A) Dataset and Input Preprocessing

In this work, the chest X-ray image dataset was downloaded from kaggle which was prepared by the authors of [21], who undertook the tedious task of collecting and indexing the X-ray images.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 12 | Dec - 2021 ISSN: 2582-3930

This dataset consists of X-rays from 1200 individuals with COVID-19, 1341 X-rays from healthy individuals.All the images are in the Portable Network Graphics (PNG) file format, and with a resolution of either 1024-by-1024 pixels or 256-by-256 pixels. It must be noted that the dataset is divided into 3575 training and 311 test images. In the training phase, the dataset was prepared and verified as reliable by reviewing it with chest specialists. Before passing the images into a pretrained model for feature extraction, we resize all images to a size of 224 x 224 x 3 pixels. All images were normalized according to the pretrained model standards.



a) COVID-19

b) Healthy

Figure 5 : Example of the X-ray images in this training set that were used in this study.

B) <u>Performance Metrics</u>

This subsection describes the evaluation of the performance of different deep learning models for classifying the CXR images. The trained models were validated using tenfold cross-validation, and the performance metrics derived from the confusion matrix were used for experimental analyses. The confusion matrix provides a guideline to the four outcomes of false negative (FN), false positive (FP), true negative (TN), and true positive (TP). The presence of both FNs and FPs could affect medical decisions negatively. An FP result is produced when an individual is inaccurately assigned to a class, such as when a healthy individual is incorrectly categorized as a COVID-19 patient. An FN occurs when an individual who is supposed to fall into a given class is instead excluded from this group. The performance of the different networks was evaluated on the test set by computing the macro average of accuracy (Acc), F1 score, precision (PPV), specificity (Spc), sensitivity (Sen), and Matthew Correlation Coefficient (MCC) [34] as quantitative evaluation indices. These are defined as :-

$$\text{Accuracy (Acc)}_i = \frac{\text{TP}_i + \text{TN}_i}{\text{TP}_i + \text{FP}_i + \text{TN}_i + \text{FN}_i}, \tag{1}$$

$$F1\ \text{score}_i = 2 \times \frac{\text{PPV}_i \times \text{Sen}_i}{\text{PPV}_i + \text{Sen}_i}, \tag{2}$$

$$\text{Precision (PPV)}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i}, \tag{3}$$

$$\text{Specificity (Spc)}_i = \frac{\text{TN}_i}{\text{FP}_i + \text{TN}_i}, \tag{4}$$

$$\text{Sensitivity (Sen)}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i}, \tag{5}$$

$$\text{MCC}_i = \frac{(\text{TP}_i \times \text{TN}_i) - (\text{FP}_i \times \text{FN}_i)}{\sqrt{(\text{TP}_i + \text{FP}_i)(\text{TP}_i + \text{FN}_i)(\text{TN}_i + \text{FP}_i)(\text{TN}_i + \text{FN}_i)}}, \tag{6}$$

where i refers to the class of COVID-19 and healthy.

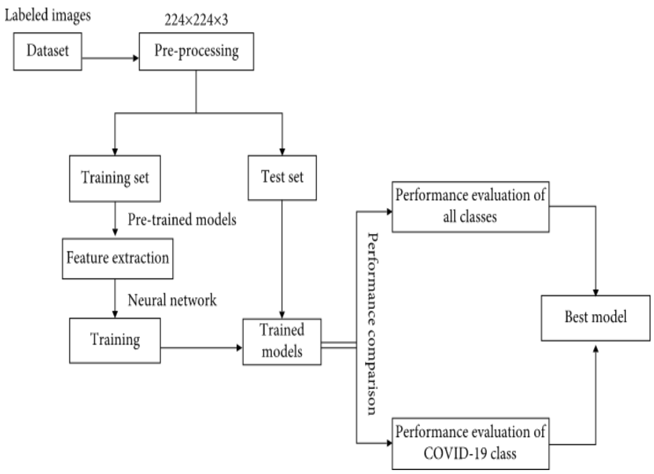


Figure 6 : CNN architecture

It is a 2 tier architecture model, It consists of front as User Interface and backend where all the processing is done. So, when a user visits the application, there will be an option to upload the image of X-Ray. When a user clicks on the predict button, A post API will be called which will take the image to the backend and predict the results based on the model which we have trained on the backend side. Model is trained using CNN algorithms. When the image is sent to predict class, a response is returned from the predict class method which is again using post API return back to the user from backend to frontend i.e., on the webpage.

Our project is based on the CNN Model, the various layers used in our CNN model are described below:-

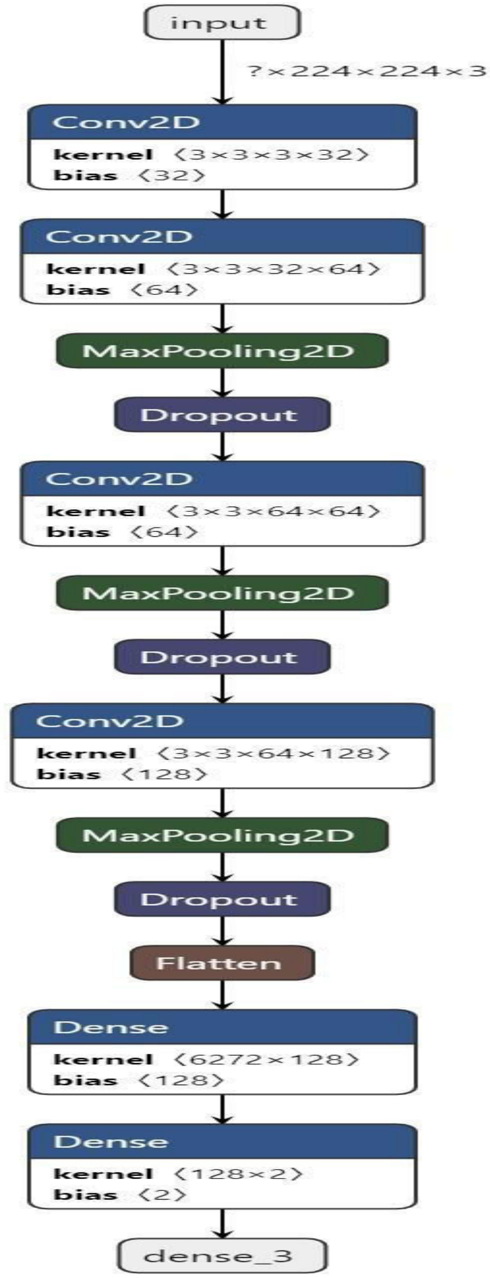


Figure 7 : CNN model flowchart

The various layers used for the CNN model are described below.

- **Conv2D Layer (3,3,3,32):** Conv2D Layer contains 32 kernels, size (3,3) ====>Output: 32 images of size (222,222).

- **Conv2D Layer (3,3,32,64):**

Conv2D Layer contains 64 kernels, size (3,3)
====> Output: 64 images of size (220,220).

- **MaxPool2D Layer (3,3):** MaxPool2D Layer is applied with (3,3) pool-size
====> Output: 64 images of size (73,73).

- **Dropout Layer:** It helps in improving the accuracy of the model.

- **Conv2D Layer (3,3,64,64):** Conv2D Layer contains 64 kernels, size (3,3)
====> Output: 64 images of size (71,71).

- **MaxPool2D Layer (3,3):** MaxPool2D Layer is applied with (3,3) pool-size
====> Output: 64 images of size (23,23).

- **Dropout Layer:** It helps in improving the accuracy of the model.

- **Conv2D Layer (3,3,64,128):** Conv2D Layer contains 128 kernels, size (3,3)
====> Output: 128 images of size (21,21).

- **MaxPool2D Layer (3,3):** MaxPool2D Layer is applied with (3,3) pool-size
====> Output: 128 images of size (7,7).

- **Dropout Layer:** It helps in improving the accuracy of the model.

- **Flatten Layer**: This will flatten all the (7,7,128) images **====>** Output: (6272).

- **Dense Layer (6272,128):** Here the Flattened array will be passed to 128 neurons in the Hidden Layer by applying 'ReLu' Activation Function ====> Output: (128).

- **Dense Layer (128,2):** Output of 128 Neurons is passed to 2 Neurons with Activation Function
'SoftMax' which will help in deciding the final output. ====> Output: (2)

C) Training the CNN model

This step is the main step where we fit our images in the training set and the test set to train our sequential model, which we built using the Keras library.

We have trained the model for 20 epochs (iterations). However, we can train for a greater number of epochs to attain higher accuracy lest there occurs *over-fitting*.

After the model is trained, we have done the prediction of Input X-Ray images using the predict_classes method of the model.

4) Result

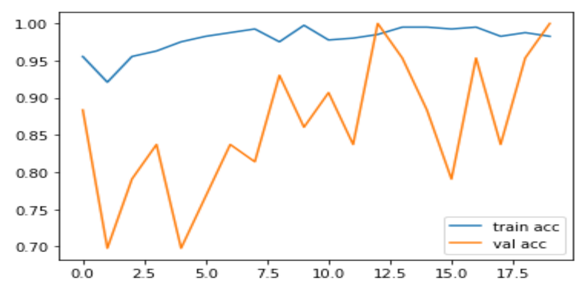


Figure 8 : Accuracy: Chart depicting train accuracy vs validation accuracy, validation accuracy is approaching train accuracy till last iteration.

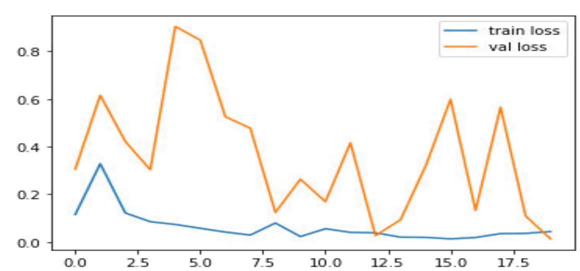


Figure 9 : Loss: Chart depicting train loss vs validation loss, validation loss and train loss should decrease with iterations.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 12 | Dec - 2021 ISSN: 2582-3930

A) Loss & Accuracy

We have achieved accuracy of around 90 – 96%, which justifies that our model prediction will be accurate and consistent.


```
Epoch 13/20
26/26 [==============================] - 18s 684ms/step - loss: 0.0370 - accuracy: 0.9852 - val_loss: 0.0262 - val_accuracy: 1.0000
Epoch 14/20
26/26 [==============================] - 18s 686ms/step - loss: 0.0191 - accuracy: 0.9951 - val_loss: 0.0923 - val_accuracy: 0.9535
Epoch 15/20
26/26 [==============================] - 18s 678ms/step - loss: 0.0190 - accuracy: 0.9951 - val_loss: 0.3165 - val_accuracy: 0.8837
Epoch 16/20
26/26 [==============================] - 18s 677ms/step - loss: 0.0121 - accuracy: 0.9926 - val_loss: 0.5985 - val_accuracy: 0.7907
Epoch 17/20
26/26 [==============================] - 18s 679ms/step - loss: 0.0177 - accuracy: 0.9951 - val_loss: 0.1315 - val_accuracy: 0.9535
Epoch 18/20
26/26 [==============================] - 18s 682ms/step - loss: 0.0336 - accuracy: 0.9827 - val_loss: 0.5654 - val_accuracy: 0.8372
Epoch 19/20
26/26 [==============================] - 18s 677ms/step - loss: 0.0345 - accuracy: 0.9877 - val_loss: 0.1085 - val_accuracy: 0.9535
Epoch 20/20
26/26 [==============================] - 18s 677ms/step - loss: 0.0421 - accuracy: 0.9827 - val_loss: 0.0126 - val_accuracy: 1.0000
```

Figure 10 : Accuracy obtained

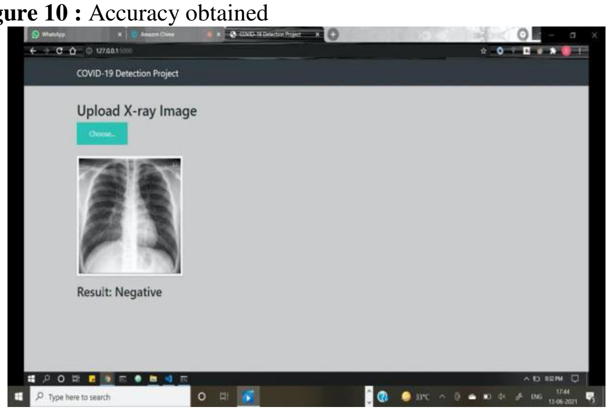


Figure 11 : When the X-ray image uploaded belongs to covid infected person, the result came Positive.

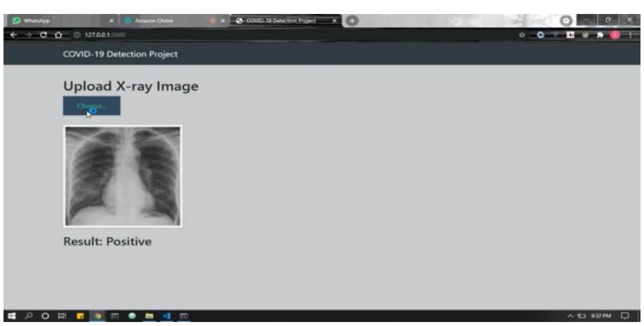


Figure 12 : When the X-ray image uploaded belongs to a normal person, the result came negative.

5) Conclusion

We have achieved accuracy of around 90 – 96%, which justifies that our model prediction will be accurate and consistent when compared to other models.

On a sample input X-ray, we detected whether the person is Covid infected or not. For better results we have shown the output on the UI (Web Page). However, this study does have its shortcomings. In particular, a more detailed analysis requires a larger amount of patient data, especially COVID-19 data. After all, effective deep learning models are usually trained on more than a million images, a number that is difficult to obtain in the medical domain. Besides, there is a possibility that training deep neural networks on a limited dataset results in overfitting and hinders its generalization. Visual ablation studies can be performed along with deep transfer learning, which will significantly improve the detection of COVID-19 manifestations in the X-Ray images.

6) Future Scope

Covid-19 pandemic is a growing manifold daily. With the ever-increasing number of cases, bulk testing of cases swiftly may be required. In this work, we experimented with multiple CNN models in an attempt to classify the Covid-19 affected patients using their chest X-ray scans. We have successfully classified covid-19 scans, and it depicts the possible scope of applying such techniques in the near future to automate diagnosis tasks. The high accuracy obtained may be a cause of concern since it may be a result of overfitting. This can be verified by testing it against new data that is made public shortly. In the future, the large dataset for chest X-rays can be considered to validate our proposed model on it.

7) References

1) Lee E. Y., Ng M.-Y., and Khong P.-L., "COVID-19 pneumonia: what has CT taught us?," The Lancet Infectious Diseases, vol. 20, pp. 384–385, 2020. pmid:32105641

2) Pan F., Ye T., Sun P., Gui S., Liang B., Li L., et al., "Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia," Radiology, 2020.

3) Bernheim A., Mei X., Huang M., Yang Y., Fayad Z. A., Zhang N., et al., "Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection," Radiology, p. 200463, 2020. pmid:32077789

4) Long C., Xu H., Shen Q., Zhang X., Fan B., Wang C., et al., "Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT?," European journal of radiology, p. 108961, 2020.

5) Kong W. and Agarwal P. P., "Chest imaging appearance of COVID-19 infection," Radiology: Cardiothoracic Imaging, vol. 2, p. e200028, 2020.

6) Li Y. and Xia L., "Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management," American Journal of Roentgenology, vol. 214, pp. 1280–1286, 2020. pmid:32130038

7) A. Z. Khuzani, M. Heidari, and S. A. Shariati, *COVID-Classifier: an automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images*, medRxiv, 2020.

8) G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012

9) M. A. Elaziz, K. M. Hosny, A. Salah, M. M. Darwish, S. Lu, and A. T. Sahlol, "New machine learning method for image-based diagnosis of COVID-19," *PLoS One*, vol. 15, no. 6, article e0235187, 2020.

10) D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," 2020, https://arxiv.org/abs/2010.16061.

11) L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest X-ray images," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

12) S. Toraman, T. B. Alakus, and I. Turkoglu, "Convolutional CapsNet: a novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks," *Chaos, Solitons & Fractals*, vol. 140, article 110122, 2020.

13) V. Chouhan, S. K. Singh, A. Khamparia et al., "A novel transfer learning based approach for pneumonia detection in chest X-ray images," *Applied Sciences*, vol. 10, no. 2, p. 559, 2020.

14) H. Shi, X. Han, N. Jiang et al., "Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study," *The Lancet Infectious Diseases*, vol. 20, no. 4, pp. 425–434, 2020.

15) A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using X-ray images and deep convolutional neural networks," 2020

16) T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Computers in Biology and Medicine*, vol. 121, p. 103792, 2020.