SHREYAA JAISWAL

Technical Report on Binary Classification - Covid-19 detection using X-ray images

SUMMER TRAINING PROGRAM

JUNE – JULY

2022

MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNIOLOGY, PRAYAGRAJ

UNDER THE SUPERVISION OF:

DR. RANVIJAY

-ASSISTANT PROFESSOR

COMPUTER SCIENCE DEPARTMENT

MNNIT, PRAYAGRAJ

COVID-19 detection from chest X-Rays using deep learning models

A technical report submitted during
internship in Machine Learning at
Motilal Nehru National Institute of Technology,
Prayagraj in partial fulfilment of the requirements for
the degree of
Bachelor of Technology in
Computer Science Engineering by SHREYAA
JAISWAL

Under the Guidance of
Dr. RANVIJAY

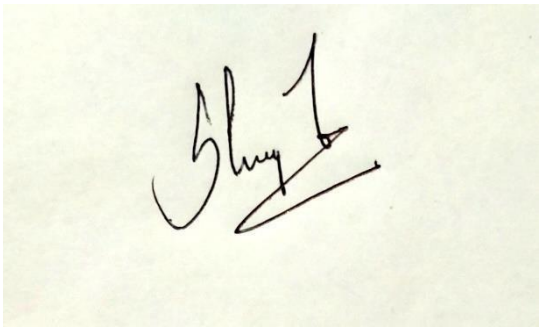
COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY ALLAHABAD
PRAYAGRAJ – 211004, INDIA

June-July, 2022

UNDERTAKING

I declare that the work presented in this report titled "*COVID-19 detection from chest X-Rays using deep learning models*", submitted to the Computer Science and Engineering Department, Motilal Nehru National Institute of Technology Allahabad, for the award of the internship certificate in Computer Science Engineering (Machine Learning), is my original work. I have not plagiarized or submitted the same work for the award of any other program.

June-July, 2022
Prayagraj



(Shreyaa Jaiswal)

CERTIFICATE

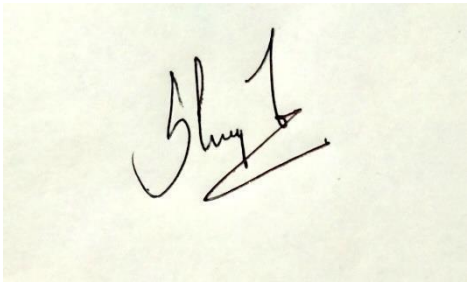
Certified that the work contained in the report titled "*COVID-19 detection from chest X-Rays using deep learning models*", by *Shreyaa Jaiswal* has been carried out under my supervision and this work has not been submitted elsewhere for any program.

(Dr. Ranvijay)
Computer Science and Engineering Dept.
MNNIT Allahabad

June, 2022

Acknowledgment

With great pleasure, I would like to express my sincere gratitude and thanks to my parents and professors for their sincere support they have given me. I would also like to thank the competent authorities of Motilal Nehru National Institute of Technology for the care and support they have given me all throughout my Summer Training Program. I again thank my supervisor Dr. Ranvijay for his guidance and cooperation during the entire training program. I would also like to thank Dr. Divya Kumar for the support he ha provided me throughout the training. At the end, I would like to express my deep sense of gratitude to all who helped me directly or indirectly for my project work.



SHREYAA JAISWAL

Biographical Sketch

SHREYAA JAISWAL

Mutthiganj, Prayagraj, Uttar Pradesh, India

E-Mail: shreyaajaiswal82326@gmail.com , Contact. No. +91-9076867800

Father's Name: Mr. Ashish Jaiswal

Mother's Name: Mrs. Sanchita Jaiswal

Education

- Shreyaa Jaiswal is the student of B.Tech. in Computer Science Engineering at Dehradun Institute of Technology (DIT University), Dehradun, Uttrakhand.

Abstract

COVID-19, initially discovered in Wuhan, China, in December 2019, has spread over the whole world like a pandemic, causing massive human and economic losses. COVID-19 is complicated to detect at an early stage so that it can be cured in the initial stage only, and the further spread of COVID-19 can be controlled and stopped. After doing some research and deep study on COVID-19, it was found that the chest XRay and CT scan technique are useful to identify whether a patient is infected with COVID-19 or not. The lungs are the most affected part of the human body by the coronavirus. The characteristics generally found in the lungs of a person infected with COVID-19 patients are multifocal, bilateral ground-glass opacities(GGO), consolidation, and vascular dilation in lesions. Artificial intelligence techniques could be used with radiological imaging (X-Ray, CT) to detect disease accurately. It will be possible to overcome the lack of specialized doctors in remote villages. With the help of a machine learning model, it will be easy and fast to determine whether a person is suffering from COVID-19. Our model produced an accuracy of 98% for binary classification.

Introduction

A contagious disease called COVID-19 is caused on by the rapidly proliferating SARSCoV-2 viruses. We all are aware that COVID-19 is spreading very fast like a pandemic and causing massive losses in the world. Although we are currently getting fewer new cases, it has affected many people's lives. There are some tests, such as the RT-PCR, available to check whether a person is affected by coronavirus or not.

The majority of virus-infected individuals may feel abnormal respiratory system and that can be controlled without extra medical treatment. However few individuals will get quite sick and need to get treated. Serious illnesses are more likely to strike older adults and those with other health issues, including diabetes, cancer, chronic lung diseases, or cardiovascular disease. Anyone with less immunity power or any medical issues could be easily affected by a coronavirus and may die at any time. So, to prevent itself from corona, it was advised that don't stay close to any person, use masks while going outside, and use alcohol-based sanitizer to wash your hands frequently.

It is also advised to get properly vaccinated to break the chains of coronavirus infection. Detection of COVID-19 at its early stages is very important so that the affected person cannot spread the virus to others and get treatment so that they can be cured. Thus, a tool is required for fast detection that an individual has been infected with covid or not at the earliest. A machine learning model that takes an X-Ray of the person's chest as input can be used to detect the COVID-19. Various research has been carried out in this direction in the recent past.

Problem statement

The project contains the work on the automatic detection of COVID-19 by resorting chest X-ray pictures and deep structured leaning.

Model used

- A machine learning model named Convolutional Neural Network has been used in this report to offer precise diagnostics for binary classification.
- This model obtained an accuracy of approximately 98% for binary classification.

Challenges

1. The main challenge faced by me was to collect a good X-Ray data-set to train the model. The more the number of images in a particular data-set, the more the chances of getting accurate answers.
2. The model was not very effective in feature extraction in case of bad quality Xray images.
3. In the model, it was required to identify the total number of Convolutional layers, since said layers needed in the model depends on the level of feature extraction required, which would be used for the different classification purposes by the final fully connected layer.

Dataset collection

The dataset includes the chest X-Rays images of COVID-19-infected patients and healthy individuals and has been gathered from the Kaggle website. The dataset obtained is divided into two parts: the training and testing datasets. And each set is divided into sets of covid and normal. In the training dataset, there were 288 images and in the testing dataset, there were 60 images.

Approaches to the Solutions

After going through the different research papers, it is found that only test kits and expert doctors would not be able to overcome the problems. Because these resources are not available in good amounts and also not reachable to all. So, there must be a system which could take the chest X- Rays and visualize the new patterns which are found different from normal patients to covid patients and according to this a machine learning model is developed to detect covid. Later sections contain the features seen in the X-rays of COVID-19 patients that are utilized to identify the disease.

In this project report, the Sequential model for the covid detection process has been proposed. For which the dataset has been collected from kaggle website and performed pre-processing on this so that it could be fed to the convolutional layer and with the help of this a feature map is generated which is given as input to the final layer of the model for classification.

Why chest X-Ray

These are some of the reasons behind the selection of chest X-Ray for the detection of COVID-19 using ML models:

- X-Rays are inexpensive and simple to understand, making them more accessible to rural and remote healthcare practitioners.
- chest X-Ray of COVID-19 affected patients contains some patterns that CNN models could use.
- CNN could be used for COVID detection using X-Ray as it is very powerful for image processing and image classification.

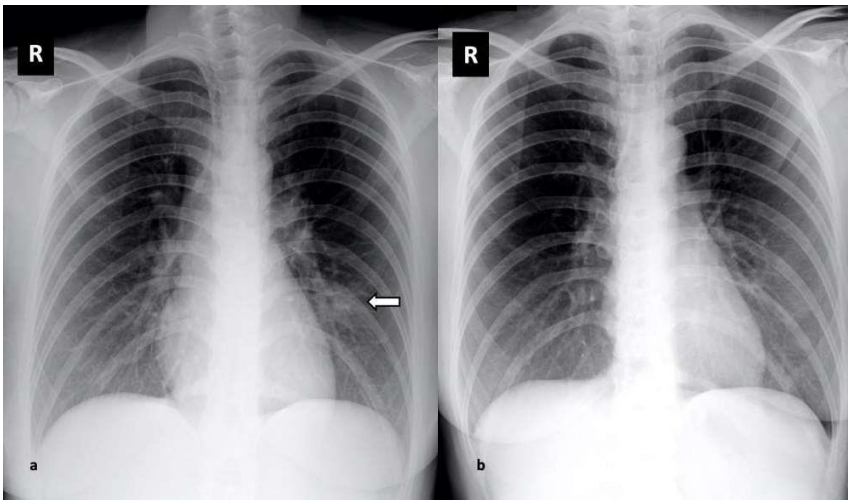
Observed features

The following characteristics are useful in identifying the infection in the X-Ray image of COVID-19.

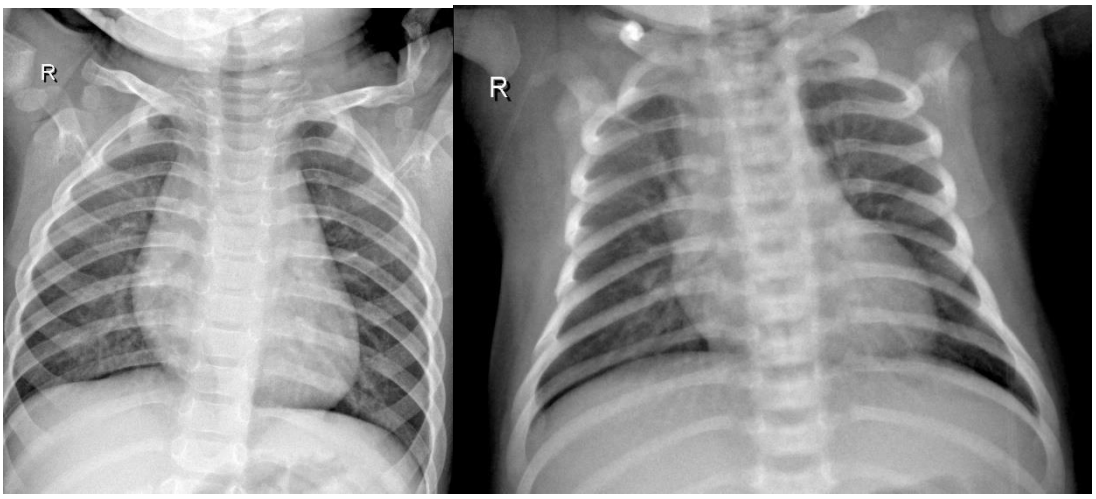
- Reticular interstitial pattern with a peripheral predominance.
- Ground-glass opacities (GGO) (bilateral, multifocal, subpleural, peripheral, posterior, medial, and basal).
- Bilateral peripheral opacities in upper, middle, and lower fields.
- Air space consolidation.
- Broncho vascular thickening (in the lesion)

Training Dataset:

Covid:



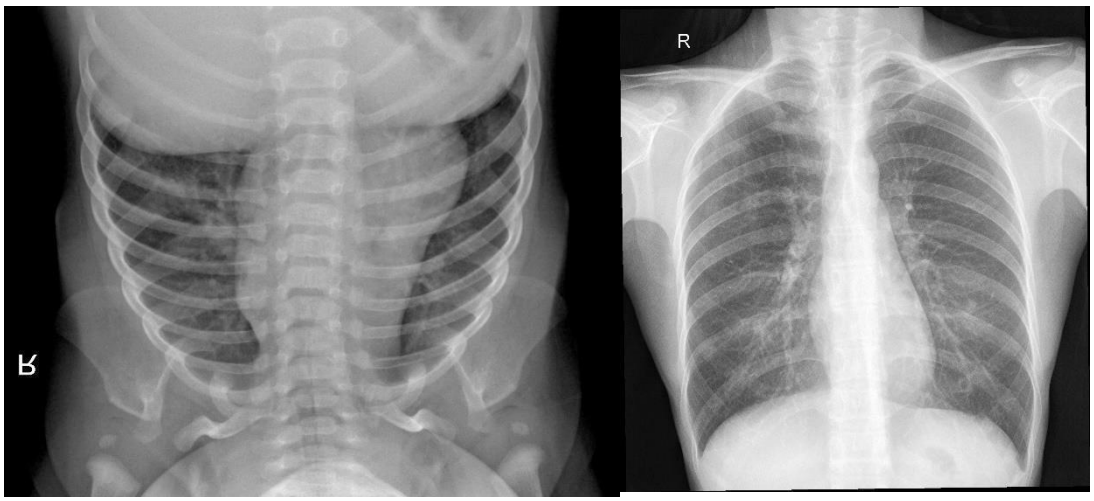
Non-Covid :



Testing Dataset:

Covid:

Non-Covid:



CODE:

1-Dependencies:

```
import numpy as nmp
import matplotlib.pyplot as pltt
import keras
from keras.layers import Dense, Conv2D, MaxPool2D, Dropout, Flatten
from keras.models import Sequential
from keras.preprocessing import image
```

2-Connecting to Google Drive to load the dataset:

```
from google.colab import drive
drive.mount("/content/drive")
```

Output:

```
Mounted at /content/drive
```

3-Dataset Loading:

```
trn_dtgen = image.ImageDataGenerator(

    rescale = 1/255, horizontal_flip = True, zoom_range = 0.2, shear_
range = 0.2
)
trn_dt     =     trn_dtgen.flow_from_directory(directory="/content/dr
ive/MyDrive/Dataset/Train",    target_size=(256,256),    batch_size=16,
class_mode = 'binary')
```

Output:

```
Found 288 images belonging to 2 classes.
```

```
trn_dt.class_indices
```

Output:

```
{'Covid': 0, 'Normal': 1}
```

```
tst_dtgen = image.ImageDataGenerator(
rescale = 1/255
                                                    )
tst_dt = tst_dtgen.flow_from_directory(directory="/content/driv
e/MyDrive/Dataset/Val", target_size=(256,256), batch_size=16,
class_mode = 'binary'
                                                )
```

Output:

```
Found 60 images belonging to 2 classes.
```

4-CNN Model:

```
modl = Sequential()
 modl.add(Conv2D(filters = 32, kernel_size=(3,3), activation='relu',
input_shape= (256,256,3)))
 modl.add(Conv2D(filters = 64, kernel_size=(3,3),
activation='relu'))
modl.add(MaxPool2D())
modl.add(Dropout(rate=0.25))
 modl.add(Conv2D(filters = 128, kernel_size=(3,3),
activation='relu')
)  modl.add(MaxPool2D())
modl.add(Dropout(rate=0.25))
modl.add(Flatten())
modl.add(Dense(units=64,
activation='relu'))
```

```
modl.add(Dropout(rate=0.50))
modl.add(Dense(units=1, activation='sigmoid'))
 modl.compile(loss = keras.losses.binary_crossentropy, optimizer =
'adam', metrics = ['acc'])
modl.summary()
```

Output:

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 254, 254, 32)      896

 conv2d_1 (Conv2D)           (None, 252, 252, 64)      18496

 max_pooling2d (MaxPooling2D  (None, 126, 126, 64)     0
 )

 dropout (Dropout)           (None, 126, 126, 64)      0

 conv2d_2 (Conv2D)           (None, 124, 124, 128)     73856

 max_pooling2d_1 (MaxPooling  (None, 62, 62, 128)      0
 2D)

 dropout_1 (Dropout)         (None, 62, 62, 128)       0

 flatten (Flatten)           (None, 492032)            0

 dense (Dense)               (None, 64)                31490112

 dropout_2 (Dropout)         (None, 64)                0

 dense_1 (Dense)             (None, 1)                 65

=================================================================
Total params: 31,583,425
Trainable params: 31,583,425
Non-trainable params: 0
_________________________________________________________________
```

```
modl.fit_generator(trn_dt, steps_per_epoch=8, epochs=20, validat
ion_steps=2, validation_data=tst_dt)
```

Output:

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: UserWarning: `Model.fit_generator` is deprecated and will be removed in a future version. Please use `Model.fit`, which supports
  """Entry point for launching an IPython kernel.
Epoch 1/20
8/8 [==============================] - 75s 9s/step - loss: 2.6789 - acc: 0.6172 - val_loss: 0.6427 - val_acc: 0.8750
Epoch 2/20
8/8 [==============================] - 59s 7s/step - loss: 0.6071 - acc: 0.6328 - val_loss: 0.6297 - val_acc: 0.9688
Epoch 3/20
8/8 [==============================] - 58s 7s/step - loss: 0.4007 - acc: 0.8594 - val_loss: 0.4365 - val_acc: 0.9062
Epoch 4/20
8/8 [==============================] - 59s 7s/step - loss: 0.3509 - acc: 0.8125 - val_loss: 0.4692 - val_acc: 0.9062
Epoch 5/20
8/8 [==============================] - 57s 7s/step - loss: 0.3280 - acc: 0.8438 - val_loss: 0.1398 - val_acc: 0.9688
Epoch 6/20
8/8 [==============================] - 59s 7s/step - loss: 0.3663 - acc: 0.8594 - val_loss: 0.2933 - val_acc: 0.9062
Epoch 7/20
8/8 [==============================] - 58s 7s/step - loss: 0.2285 - acc: 0.9219 - val_loss: 0.1889 - val_acc: 0.9688
Epoch 8/20
8/8 [==============================] - 60s 8s/step - loss: 0.2621 - acc: 0.9375 - val_loss: 0.2336 - val_acc: 0.9688
Epoch 9/20
8/8 [==============================] - 57s 7s/step - loss: 0.3163 - acc: 0.9141 - val_loss: 0.2401 - val_acc: 0.9375
Epoch 10/20
8/8 [==============================] - 57s 7s/step - loss: 0.2334 - acc: 0.9375 - val_loss: 0.1533 - val_acc: 0.9375
Epoch 11/20
8/8 [==============================] - 58s 7s/step - loss: 0.1819 - acc: 0.9531 - val_loss: 0.1700 - val_acc: 0.9375
Epoch 12/20
8/8 [==============================] - 57s 7s/step - loss: 0.1839 - acc: 0.9531 - val_loss: 0.1391 - val_acc: 0.9375
Epoch 13/20
8/8 [==============================] - 59s 7s/step - loss: 0.2578 - acc: 0.9219 - val_loss: 0.1365 - val_acc: 1.0000
Epoch 14/20
8/8 [==============================] - 58s 7s/step - loss: 0.1662 - acc: 0.9688 - val_loss: 0.1350 - val_acc: 0.9688
Epoch 15/20
8/8 [==============================] - 59s 7s/step - loss: 0.2153 - acc: 0.9375 - val_loss: 0.1471 - val_acc: 0.9688
Epoch 16/20
8/8 [==============================] - 59s 7s/step - loss: 0.2383 - acc: 0.9141 - val_loss: 0.0780 - val_acc: 0.9688
Epoch 17/20
8/8 [==============================] - 58s 7s/step - loss: 0.1322 - acc: 0.9531 - val_loss: 0.1455 - val_acc: 0.9375
Epoch 18/20
8/8 [==============================] - 60s 8s/step - loss: 0.1727 - acc: 0.9375 - val_loss: 0.0793 - val_acc: 1.0000
Epoch 19/20
8/8 [==============================] - 58s 7s/step - loss: 0.1720 - acc: 0.9453 - val_loss: 0.1234 - val_acc: 0.9688
Epoch 20/20
8/8 [==============================] - 58s 7s/step - loss: 0.1476 - acc: 0.9531 - val_loss: 0.0589 - val_acc: 0.9688
<keras.callbacks.History at 0x7f6a1e994f10>
```

```
path = "/content/drive/MyDrive/Dataset/Val/Covid/4e43e48d52c9e2d4c6c1
fb9bc1544f_jumbo.jpeg"
imag = image.load_imag(path, target_size=(256,256))
 imag = image.imag_to_array(imag)/255 imag =
nmp.array([imag]) imag.shape
```

Output:

```
(1, 256, 256, 3)
```

```
modl.predict(imag)
```

Output:

```
array([[0.00307155]], dtype=float32)
```

```
if modl.predict(imag)<0.5:
 print("Covid") else:    print("Non-
covid")
```

Output:

```
Covid
```

Result:

This project is based on binary classification for the detection of Covid-19 using X-ray images.
Convolutional Neural Network or CNN Model is used here for classification of X-ray images into Covid or Non-Covid Cases.
For evaluation of the model, accuracy and precision has been used here.

Precision is the number of true positives per positive predictions and an accurate model is the one which is the best at finding patterns and relationships among variables based on input in a dataset.

Accuracy = (TP + TN)/(TP + FP + FN + TN),

where TP, TN, FP, and FN are True Positive, True Negative, False Positive and False Negative, respectively. The confusion matrix is used to calculate the value of TP, TN, FP, and FN.

This model gives an accuracy of approximately 98% with a precision of 0.98.

Analysis and Advantages:

1. The model proposed doesn’t require any feature extraction technique to perform the classification.
2. The accuracy of the model is approximately 98% for binary classification.
3. The model could also be deployed on the cloud so that it would be helpful to the radiologists in validating their initial screening so that they can correctly identify whether the patient is COVID positive or not for further treatment.
4. This model is beneficial and effective in COVID diagnosis by experts. This model used chest X-Ray images rather than CT images, which is affordable for poor people as X-ray requires less cost and is also readily available in less developed places

Experimental setup

Experimental setup and Model used

Experimental Setup

In this project, the model was implemented to classify COVID-19 diagnosis using Keras, Numpy and Matplotlib. Some intermediate tasks such as image extraction and image balancing were conducted at Google Colaboratory. Image data generator is used for augmentation of the data.

The hardware configuration of the system on which the experiment is conducted is

- Lenovo Ideapad S145-Ryzen 5
- Installed RAM is 8.00 GB (7.73 GB usable)
- System type is a 64-bit operating system

Challenges

- The images were blurred and flipped; hence pre-processing was challenging.
- We had limited resources for performing the experiments.
- To prevent the model from overfitting.
- One of the challenges is to find a model that can give greater accuracy in binary classification.
- Searching for the correct number of iterations and they need to deactivate some of the neurons to save from overfitting.

Conclusion and Future Work

Conclusion and Future work

The work reported in this project gives an insight into various deep learning models used for COVID-19 detection. In this work, CNN model is used for binary classification with a limited dataset. After performing the experiments, the study concluded that for binary classification, the model Sequential shows an accuracy of 98% and a Precision of 0.98, which is the best among all the models used for the classification.

From the above, we can conclude that the Sequential model can be used for early detection of COVID-19 along with other biomarkers.

The study will be helpful for researchers to carry out the research for multi-class classification in the future. This suggests that the models used in our work can also be used to identify other chest-related conditions, such as tuberculosis. The study used a restricted number of COVID-19 X-Ray pictures is a restriction.

To make the models more reliable and precise, we intend to use additional photos from the hospitals in our area. The models can also be trained using chest CT scan pictures.

References

[1] Kaggle https://www.kaggle.com/datasets/fu sicfenta/chest-xray-for-covid19https://www.kaggle.com/datasets/fusicfenta/chest-xray-forhttps://www.kaggle.com/datasets/fusicfenta/chest-xray-for-covid19-detectioncovid19-detectiondetection

[2] COVID-19 Infection Detection Using Machine Learning – IEEE Explore

https://ieeexplore.ieee.org/Xplore/home.jsp

[3] GreeksForGreeks

https://www.geeksforgeeks.org/detecting-covid-19-from-chest-xhttps://www.geeksforgeeks.org/detecting-covid-19-from-chest-x-ray-images-using-cnn/rayhttps://www.geeksforgeeks.org/detecting-covid-19-from-chest-x-ray-images-usinghttps://www.geeksforgeeks.org/detecting-covid-19-from-chest-x-ray-images-using-cnn/cnn/images-using-cnn/

[4] GitHub

https://github.com/arghyadeep99/Covid-19-Detection-from-X-Rayshttps://github.com/arghyadeep99/Covid-19-Detection-from-X-Rays-using-

CNNusing-CNN

[5] Ezz El-Din Hemdan, Marwa A Shouman, and Mohamed Esmail Karar. Covidxnet: A framework of deep learning classifiers to diagnose covid-19 in x-ray images. *arXiv preprint arXiv:2003.11055*, 2020.

[6] M.; Chen B.; et al. Howard, A. G.; Zhu. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arxiv. 2015, arxiv:17040486. 5 2020.

[7] Rachna Jain, Meenu Gupta, Soham Taneja, and D Jude Hemanth. Deep learning based detection and analysis of covid-19 on chest xray images. *Applied Intelligence*, 51(3):1690–1700, 2021.

[8] Iosif Mporas and Prasitthichai Naronglerdrit. Covid-19 identification from chest xrays. In *2020 International Conference on Biomedical Innovations and Applications (BIA)*, pages 69–72. IEEE, 2020.

[9] Ali Narin, Ceren Kaya, and Ziynet Pamuk. Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *Pattern Analysis and Applications*, 24(3):1207–1220, 2021.

[10] Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim, and U Rajendra Acharya. Automated

detection of covid-19 cases using deep neural networks with x-ray images.
Computers in biology and medicine, 121:103792, 2020.

[11] Prabira Kumar Sethy and Santi Kumari Behera. Detection of coronavirus disease (covid-19) based on deep features. 2020.

[12] Shrinjal Singh, Piyush Sapra, Aman Garg, and Dinesh Kumar Vishwakarma. Cnn based covid-aid: Covid 19 detection using chest xray. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pages 1791–1797. IEEE, 2021.

[13] Ying Song, Shuangjia Zheng, Liang Li, Xiang Zhang, Xiaodong Zhang, Ziwang Huang, Jianwen Chen, Ruixuan Wang, Huiying Zhao, Yutian Chong, et al. Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images. *IEEE/ACM transactions on computational biology and bioinformatics*, 18(6):2775– 2780, 2021.

[14] Linda Wang, Zhong Qiu Lin, and Alexander Wong. Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest xray images. *Scientific Reports*, 10(1):1–12, 2020.