How to Cite:
Kumar, C., Singh, P., Hans, S., Sehgal, L., Verma, S., & Sharma, V. K. (2022). Machine learning application: Detecting COVID-19 using X-Ray images. *International Journal of Health Sciences*, *6*(S1), 8406–8417. https://doi.org/10.53730/ijhs.v6nS1.6809

Machine learning application: Detecting COVID-19 using X-Ray images

Chaman Kumar
Department of IT, IIMT college of Engineering, Greater Noida, CSE, NIET, Greater Noida, CSE, Satyug Darshan Institute of Technology Faridabad, CSE, IT, IIMT college of Engineering, Greater Noida, CSE, PSIT Kanpur, MIET Meerut
Email: chamankumar31@gmail.com

Priyanka Singh
Department of IT, IIMT college of Engineering, Greater Noida, CSE, NIET, Greater Noida, CSE, Satyug Darshan Institute of Technology Faridabad, CSE, IT, IIMT college of Engineering, Greater Noida, CSE, PSIT Kanpur, MIET Meerut
Email: priyanka.singh@niet.co.in

Swati Hans
Department of IT, IIMT college of Engineering, Greater Noida, CSE, NIET, Greater Noida, CSE, Satyug Darshan Institute of Technology Faridabad, CSE, IT, IIMT college of Engineering, Greater Noida, CSE, PSIT Kanpur, MIET Meerut
Email: swati.hans@satyug.edu.in

Laveena Sehgal
Department of IT, IIMT college of Engineering, Greater Noida, CSE, NIET, Greater Noida, CSE, Satyug Darshan Institute of Technology Faridabad, CSE, IT, IIMT college of Engineering, Greater Noida, CSE, PSIT Kanpur, MIET Meerut
Email: laveenasehgal30@gmail.com

Shanu Verma
Department of IT, IIMT college of Engineering, Greater Noida, CSE, NIET, Greater Noida, CSE, Satyug Darshan Institute of Technology Faridabad, CSE, IT, IIMT college of Engineering, Greater Noida, CSE, PSIT Kanpur, MIET Meerut
Email: shanu.verma56@gmail.com

Vijay Kumar Sharma
Department of IT, IIMT college of Engineering, Greater Noida, CSE, NIET, Greater Noida, CSE, Satyug Darshan Institute of Technology Faridabad, CSE, IT, IIMT college of Engineering, Greater Noida, CSE, PSIT Kanpur, MIET Meerut
Email: vijay.sharma@miet.ac.in

***Abstract*---**The Coronavirus which is scientifically named COVID-19. Its strain was found in Wuhan, a city of China, at the end of 2019.

After that the case of coronavirus started spreading quickly around the world and has turned it into a huge global pandemic. Now coronavirus has made a huge impact on human lives since the last several years where people are losing their lives, people are losing their jobs. It has a devastating effect on human life already. Since this virus has come as a complete surprise to everyone in 2019 there were not so many detection or screening methods or trained healthcare workers for this medical challenge and the virus being airborne was spreading really very rapidly. It has been found that COVID-19 affects the epithelial cells which are present in the respiratory tract of our body, so we can use X-ray images and various artificial intelligence techniques to detect the virus. We have built a Deep Learning model, and trained over 200 COVID-19 positive X-ray images and 202 Normal X-ray images of lungs of people. The trained model extracted features from the COVID positive and normal X-ray images and base on that predicts whether the person is COVID positive or negative.

***Keywords*---**X-ray images, Coronavirus, Deep Learning, CNN model, Flask.

Introduction

The global pandemic, COVID-19, started from Wuhan, a city of China in 2019, has hada huge impact on human lives since last year. It has a devastating effect on human life already. There are frontline healthcare community workers. They are very busy in identifying various solutions to minimize this impact and are working feverishly day in and day out to prevent or to minimize this impact. On the other hand there is a different set of people who are also busy in the form of the data science community. So they are looking at various innovative solutions, various technology related solutions basically to assist the frontline healthcare community with their solutions. As a responsible citizen and also as a CS student, we would like to find an innovative way, or we wanted to offer some technological help to the healthcare community with Artificial Intelligence. Also it's going to offer a huge productivity benefit in terms of assisting personnel who are busyin screening the covid patients at the same time it is also going to help them in terms of detecting Covid with much more accuracy.

In our assignment, we created a Fully Convolutional Network-based Deep Learning Approach. We tested our model on X-Ray images of a patient ’s chest to see if they had Covid-19. The lungs of individuals with Covid-19 may exhibit an expansion and some patch on X-Ray, but the distinction among a healthy lung and a Covid-19 afflicted lung X-Ray is sometimes so small that it takes a skilled eye to determine the difference. We've automated this technique in our study to detect even the tiniest diseased lungs. Because our data is so unbalanced, we used data augmentation techniques to alter current training set imagery and produce new ones. These data augmentation approaches aid in the development of data-intensive deep learning models. The suggested system's major goal is to create forecasting models that may be used to determine if a person is Covid. The suggested system's major goal is to create a prediction model that can be used to

determine whether or not a person is Covid impacted. The suggested framework will be important because Covid 19 is spreading rapidly and vaccinations are still in short supply. Lockdown is not the answer; instead, rapid testing with greater precision is necessary. Because Covid 19 is a virus that affects the lungs, X-ray pictures of the lungs can be utilized to detect the virus.

Related Work

CoronaVirus disease caused by SARS-CoV2 has spread in 214 countries and infected about 53 million people and has killed over a million. The first patient originated in Wuhan and upon a genetic sequence with another beta coronavirus, it was found to match ~96% with a bat coronavirus and about 80% with SARS-CoV that it was named as SARS-CoV2. Since then, multiple methodologies have been created as a diagnostic test for SARS-CoV2. They are: -

Nucleic Acid Test:
It is a primary way to detect Covid-19. Many RT-PCR or Reverse transcript polymerase sequence reaction test kits have been developed. Samples from the upper respiratory tract through swabs. The test result depends on the viral load in the samples collected. A negative result doesn't rule out the infection. The Centre for disease control (CDC) uses a variant of this test called RT PCR. It also gives the amount of viral load collected so the further conclusion can be taken from that data.

Computed Tomography:
Due to a shortage of kits and many false-negative results of the RT-PCR test CT scans were used in China for diagnosis. CT scan uses multiple x-rays that are taken at dissimilar angles to generate a cross-sectional image of the lungs. This image is then examined by virologists to find abnormal features that are common in the lungs of covid-19 patients. These symptoms start showing within 2 -3 days and peak at about 10 days after exposure. The most common feature is ground-glass opacity in lungs i.e there are areas of hazy opacity. Other features also develop like wavy stone patterns but this pattern also overlaps with other pneumonia-like diseases thus this method only works when done with RT-PCR test so that they both can complement each other's shortcomings.

Isothermal Nucleic Acid Test:
This Method is still under development as a diagnostic test for SARS-CoV2. Unlike PCR this test is done at isothermal temperature and doesn't require a laboratory to perform. Thus, this can be integrated with every hospital.

Protein Testing:
When a body gets infected from a virus protein antigen is produced by the body as a countering this virus. In a similar manner, protein antigens are formed during covid-19 infection. Testing for this antigen will give the result if the person is infected or not. Similar to CT-scan x-rays also take pictures of the lungs and can help in observing abnormal features in them due to infection. Some of these features may be invisible to thehuman eye but through Deep learning models, this can be overcome.

Methods and Materials Required:

X-ray Image Dataset Collection:
Our dataset consists of 400 images classified into two types, Normal and Covid-19 affected. We have divided our dataset into three parts: train, validation and test set. The train and test sets can be used for training and validation, respectively, whilst the validation set is utilised to estimate how successfully your model has been trained as well as model attributes. We gathered X-ray images using two or more sources for our research:

Dataset of X-ray images for Covid-19:
It was created by Joseph Paul Cohen, who used photos from a variety of public domain sources. The repository is updated regularly with photographs given by researchers from around the world. There are currently roughly 200 X-ray images in the collection that have been classified with COVID-19.

Dataset of X-Ray images for common citizens:
The kaggle provided the X-Ray images of typical persons. Because it comprised a large number of X-ray images, 202 normal x-ray photos were picked at random from the collection to preserve the proportion of the both datasets.

Proposed Methodology

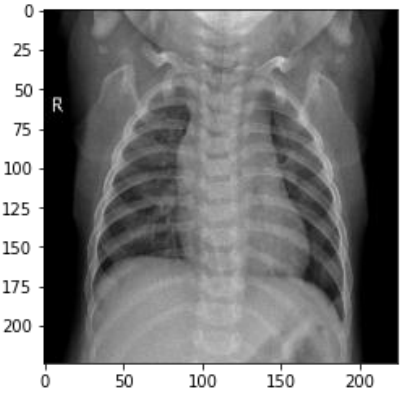

Figure 1: X-Ray image of lungs

In today's world, Machine learning and Deep learning have become essential in every Industry, be it business analytics or agriculture. One such industry is the healthcare industry, which has a huge scope for ML and DL. These techniques could be used for faster detection of fatal diseases which can be treated if they're detected in earlier stages. In our project we're trying to detect Covid-19 using X-ray images. X-ray helps us to determine the extent of virus spread, but this method needs a trained eye to even tell the difference between a normal and affected X-Ray, sometimes the difference is so minimal between a normal Chest X-Ray and Covid-19 affected Chest X-Ray that it goes undetected, hence if we can build a deep learning model, which can detect Covid-19 easily and quickly, with high precision and recall (since a model with low recall and precision can cause

hazards to the patient, if the disease goes undetected) we can help reduce the test detection time and help reduce the spread of virus.

4(a) Deep Learning

Deep learning is a type of intelligent machines that mimics the human brain's processing of information and creation of correlations for judgment. Deep learning is a branch of machine learning and AI that employs neural network models to understand unsupervised from unstructured or disorganized data.

4(b) Neural Networks

A neural structure is a collection of approaches that tries to find hidden patterns in a set of data by simulating how the human brain functions.

4(c) The Proposed CNN Model:

CNN stands for "Convolutional Neural Network," and it is most commonly used for image analysis. CNN can be thought of as an Artificial Neural Network which has been trained to recognize and detect patterns. CNN's ability to discover patterns is what makes it so valuable for image analysis. Convolution layer are hidden layers in CNN that receive input, convert it in a fashion that the machine can comprehend, filter particular patterns in the dataset, and output the changed input to next layer. The quantity that each convolution operation should have must be specified. Actually the filters in CNN detect the patterns. Filters are very vast; there are filters which can detect the edges, filters which can detect a particular feature like maybe the eye or even the retina of the eye from the image. The simplified features are discovered at the beginning of our network; as our network grows deeper, these filtration become more advanced, so that in later layers, instead of corners or basic shapes, the filter can help detect patterns/features, and in even deeper levels, the filters can identify much more advanced objects. The way in which CNN works in image analysis is that it runsa filter over the pixel matrix of the image, now it iterates through each sub-matrix of (filter x filter) size of the image. These filters pick up the pattern present in the image and make a feature map to map features these filters have detected. Now the process is followed on each image of the dataset. Actually iterating through each image and making the feature map is a very computationally expensive process so to optimize the process concept of pooling was introduced. In pooling, the size of the feature map is reduced such that computation time can be reduced. In the proposed system, we have used a sequential model, which means it has layers. Each layer has certain filters, layer-1 has 32 filters, layer-2 has 64 filters, and then we have a max-pooling layer, in the max-pooling layer the most prominent value is chosen from the feature map thus the size of the feature map is reduced. After max-pooling we have a dropout layer, the main function of the dropout layer is to set some input units to zero, this is done to prevent overfitting. After the dropout layer there is a dense layer which basically performs the function of relating the output from the previous layer with input of the next layer. As we input the output of the previous layer into the current layer, this makes the neural network a densely Connected network as well. And then we have a flatten layer, which basically converts the output of the previous matrix into a 1dimensional array thus making it a vector.

Layer (type)	Output Shape	Param #
conv2d_4 (Conv2D)	(None, 222, 222, 32)	896
conv2d_5 (Conv2D)	(None, 220, 220, 64)	18496
max_pooling2d_3 (MaxPooling2	(None, 110, 110, 64)	0
dropout_4 (Dropout)	(None, 110, 110, 64)	0
conv2d_6 (Conv2D)	(None, 108, 108, 64)	36928
max_pooling2d_4 (MaxPooling2	(None, 54, 54, 64)	0
dropout_5 (Dropout)	(None, 54, 54, 64)	0
conv2d_7 (Conv2D)	(None, 52, 52, 128)	73856
max_pooling2d_5 (MaxPooling2	(None, 26, 26, 128)	0
dropout_6 (Dropout)	(None, 26, 26, 128)	0
flatten_1 (Flatten)	(None, 86528)	0
dense_2 (Dense)	(None, 64)	5537856
dropout_7 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 1)	65

Total params: 5,668,097
Trainable params: 5,668,097
Non-trainable params: 0

Figure 2: CNN Model Workflow

Design and Implementation of Proposed Methodology

In this, proposed methodology we are representing the system that how will design the model internally and implemented the proposed system in this application which is easily understandable for users. These applications are recognizing the covid-19 effected people in very simple manner using X-ray. Here we are taking the both X-ray one is covid-19 effected people and another one normal people X-ray.

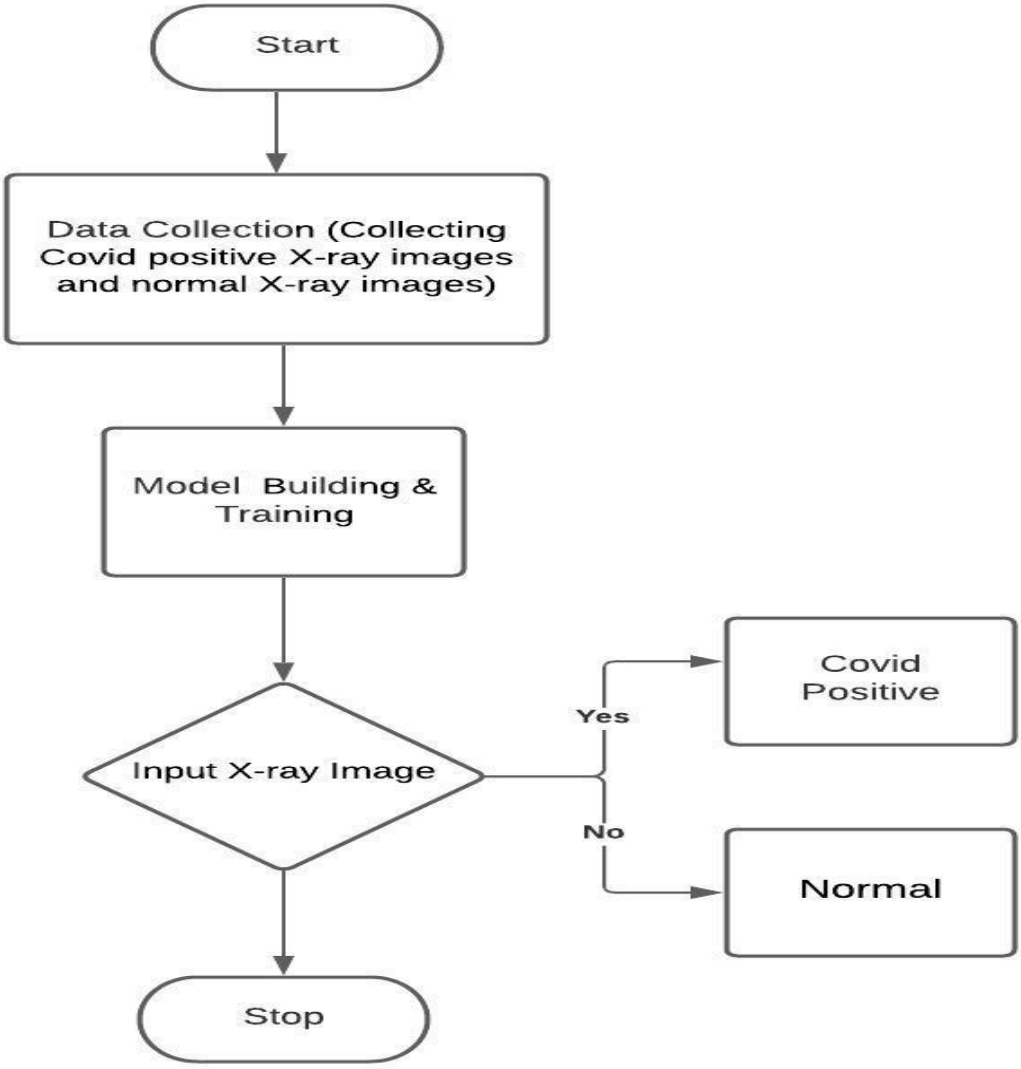

Figure 3: Design of proposed methodology

For the proposed system, the entire CNN model is being trained on the Google Colab. The datasets for Covid positive X-rays and Covid negative X-rays were collected from Chest X-ray dataset by J. Cohen and Kaggle respectively. The collected datasets were preprocessed using Keras preprocessing library. In preprocessing the data, various techniques like Data Augmentation where images were flipped, rotated to match the complete batch. The whole purpose of Data Augmentation is to make the images uniform so that they can be feeded to the model and the filters present in the layers of the model could detect the patterns in the images.

Models steps:

- The first and most crucial step is to collect the data, here we have collected datafrom J. Cohen's X-ray chest dataset and Kaggle.
- The data collected was found to be non-uniform and could not be fed to the CNN model. So various data preprocessing techniques were performed like data augmentation, reshaping the images to the same size each, collecting only grayscale images.
- The operating model has been constructed. Lower layers discovered features in a relatively tiny section of the image; hence the number of filters was initially minimal (about 32). As we progress, the variety of filters we use grows rapidly. It's known as the Open to listening Field Concept.
- The training data is in supervised form. In this step we fed the training data which was previously prepared by us to the CNN Model.
- After training just on training sample, the CNN model has been validated on a testing set.
- Once trained we save the model on the drive and then perform tests. The tests were performed on the model as it was supposed to predict from the unlabelleddata whether it is Covid positive or not.

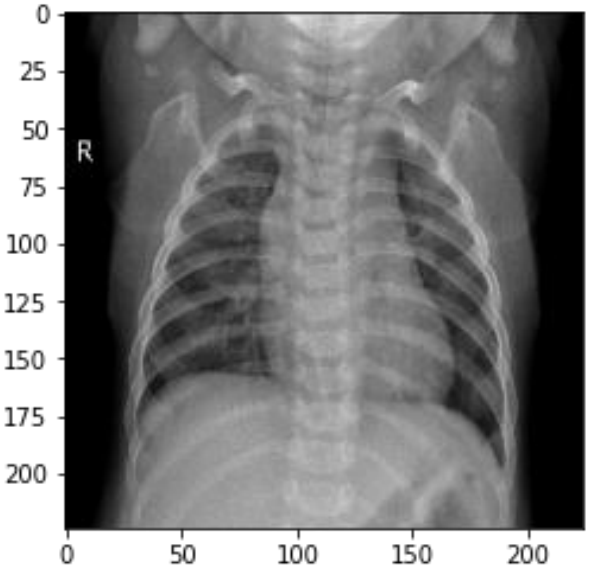

Figure 4(a): X-Ray image of Covid Positive Lungs

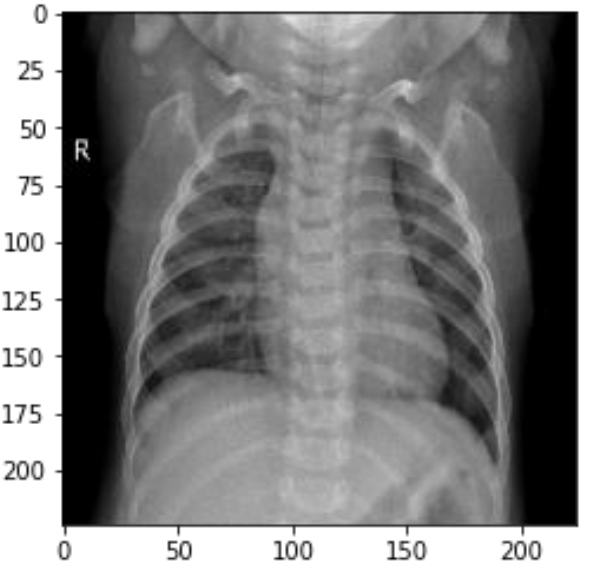

Figure 4(b): X-Ray image of Covid Negative Lungs

Result Analysis

Figure 5(a) shows the UI of the flask app. On the navigation panel of the website various helpful links are provided to ease the navigation between pages of the site. To run the app python 3.8, tensorflow 2.2.0 version must be available on the system.Then install all the libraries and packages using anaconda distribution.

Once all requirements are met, open the terminal into the directory which contains *app.py file* and run the command ***python app.py***
Then open the ***localhost:*** ***http://127.0.0.1:5000/*** in the browser

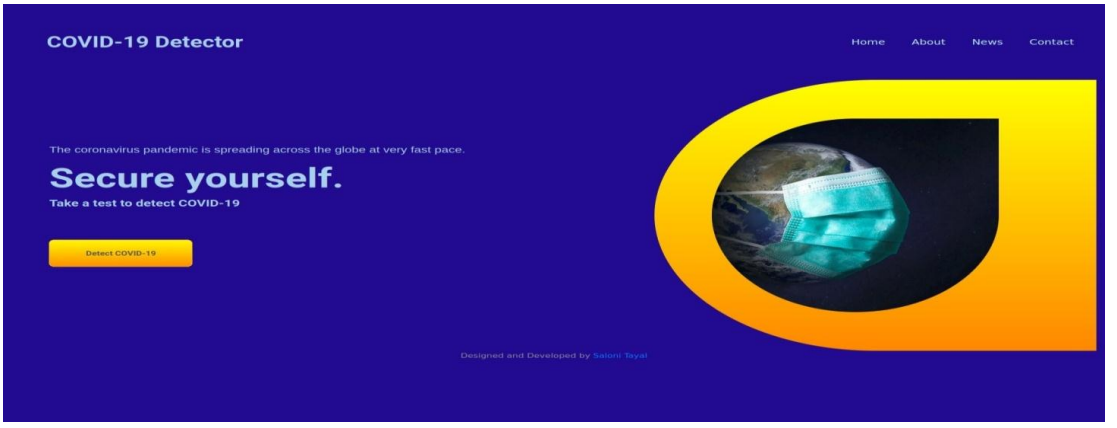

Figure 5(a): Applications Layouts

On clicking the Detect Covid-19 button user can easily reach to the webpage (as shownin Figure 5(b)) of the app where X-ray image is uploaded and prediction is given.

Figure 5(b): Inputs Application layout

Figure 5(c): Covid-19 Infected X-ray

The figure 5(d) below specifies on the application about the proposed system.

Figure 5(d): Access the output of given input

Conclusion

We have created a deep model to detect if an individual is Covid positive or not in the suggested work. Since we all aware, X-ray equipment is utilized in healthcare to scan various human organs at a minimal price and with faster results. An expert radiologist usually performs the manual interpretation of various X-ray pictures. However, we used Deep Learning to train those collected photos, which will be a big help to clinical personnel in detecting COVID-19 patients. This algorithm essentially trains from the information it is fed, then finds the features in X-ray images and categorizes them independently. This will help developing countries where there is X-ray equipment but a specialist is still a silly idea.

The future work that can be done to expand this project includes the development for an app. It also includes prediction of COVID-19 cases by observing different patterns from different areas from the past days and predicting how many more positive cases will come in future. It will help hospital, government to take necessary steps towards it. The app would be helpful to the users to verify their X-ray report, connect and manage their accounts through different platforms.

Reference

[1] Mohammad-Rahimi, Hossein, Mohadeseh Nadimi, Azadeh Ghalyanchi-Langeroudi, Mohammad Taheri, and Soudeh Ghafouri-Fard. "Application of machine learning in diagnosis of COVID-19 through X-ray and CT images: a scoping review." *Frontiers in cardiovascular medicine* 8 (2021): 185.

[2] Fusco, Roberta, Roberta Grassi, Vincenza Granata, Sergio Venanzio Setola, Francesca Grassi, Diletta Cozzi, Biagio Pecori, Francesco Izzo, and Antonella Petrillo. "Artificial Intelligence and COVID-19 Using Chest CT Scan and Chest X-ray Images: Machine Learning and Deep Learning Approaches for Diagnosis and Treatment." *Journal of Personalized Medicine* 11, no. 10 (2021): 993.

[3] Rasheed, Jawad, Alaa Ali Hameed, Chawki Djeddi, Akhtar Jamil, and Fadi Al-Turjman. "A machine learning-based framework for diagnosis of COVID-19 from chest X-ray images." *Interdisciplinary Sciences: Computational Life Sciences* 13, no. 1 (2021): 103-117.

[4] Saygılı, Ahmet. "A new approach for computer-aided detection of coronavirus (COVID-19) from CT and X-ray images using machine learning methods." *Applied Soft Computing* 105 (2021): 107323.

[5] Apostolopoulos, Ioannis D., and Tzani A. Mpesiana. "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks." *Physical and Engineering Sciences in Medicine* 43, no. 2 (2020): 635-640.

[6] Mijwil, Maad M. "Implementation of Machine Learning Techniques for the Classification of Lung X-Ray Images Used to Detect COVID-19 in Humans." *Iraqi Journal of Science* (2021): 2099-2109.

[7] Wang, Dingding, Jiaqing Mo, Gang Zhou, Liang Xu, and Yajun Liu. "An efficient mixture of deep and machine learning models for COVID-19 diagnosis in chest X-ray images." *PloS one* 15, no. 11 (2020): e0242535.

[8] Sekeroglu, Boran, and Ilker Ozsahin. "<? covid19?> Detection of COVID-19 from Chest X-Ray Images Using Convolutional Neural Networks." *SLAS*

TECHNOLOGY: Translating Life Sciences Innovation 25, no. 6 (2020): 553-565.

[9] Abbas, Asmaa, Mohammed M. Abdelsamea, and Mohamed Medhat Gaber. "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network." *Applied Intelligence* 51, no. 2 (2021): 854-864.

[10] Nayak, Soumya Ranjan, Deepak Ranjan Nayak, Utkarsh Sinha, Vaibhav Arora, and Ram Bilas Pachori. "Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study." *Biomedical Signal Processing and Control* 64 (2021): 102365.

[11] Aradhya, VN Manjunath, Mufti Mahmud, D. S. Guru, Basant Agarwal, and M. Shamim Kaiser. "One-shot cluster-based approach for the detection of COVID–19 from chest X–ray images." *Cognitive Computation* (2021): 1-9.

[12] Saha, Prottoy, Muhammad Sheikh Sadi, and Md Milon Islam. "EMCNet: Automated COVID-19 diagnosis from X-ray images using convolutional neural network and ensemble of machine learning classifiers." *Informatics in medicine unlocked* 22 (2021): 100505.

[13] Ohata, Elene Firmeza, Gabriel Maia Bezerra, João Victor Souza das Chagas, Aloísio Vieira Lira Neto, Adriano Bessa Albuquerque, Victor Hugo C. de Albuquerque, and Pedro Pedrosa Reboucas Filho. "Automatic detection of COVID-19 infection using chest X-ray images through transfer learning." *IEEE/CAA Journal of Automatica Sinica* 8, no. 1 (2020): 239-248.

[14] Rahimzadeh, Mohammad, and Abolfazl Attar. "A new modified deep convolutional neural network for detecting COVID-19 from X-ray images." *arXiv e-prints* (2020): arXiv-2004.

[15] Apostolopoulos, Ioannis D., Sokratis I. Aznaouridis, and Mpesiana A. Tzani. "Extracting possibly representative COVID-19 biomarkers from X-ray images with deep learning approach and image data related to pulmonary diseases." *Journal of Medical and Biological Engineering* (2020): 1.

[16] Johri, Shikhar, Mehendi Goyal, Sahil Jain, Manoj Baranwal, Vinay Kumar, and Rahul Upadhyay. "A novel machine learning-based analytical framework for automatic detection of COVID-19 using chest X-ray images." *International Journal of Imaging Systems and Technology* 31, no. 3 (2021): 1105-1119.

[17] Asif, Sohaib, Yi Wenhui, Hou Jin, and Si Jinhai. "Classification of COVID-19 from Chest X-ray images using Deep Convolutional Neural Network." In *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pp. 426-433. IEEE, 2020.

[18] Erdaw, Yabsera, and Erdaw Tachbele. "Machine Learning Model Applied on Chest X-Ray Images Enables Automatic Detection of COVID-19 Cases with High Accuracy." *International Journal of General Medicine* 14 (2021): 4923.

[19] Ozturka, Tulin, Muhammed Talob, Eylul Azra Yildirimc, Ulas Baran Baloglud, Ozal Yildirime, and U. R. Acharyafgh. "Automated detection of COVID-19 cases using deep neural networks with X-ray images." *Comput Biol Med* 121, no. 103792 (2020): 10-1016.

[20] Alqudah, Ali Mohammad, Shoroq Qazan, Hiam Alquran, Isam Abu Qasmieh, and Amin Alqudah. "Covid-19 detection from x-ray images using different artificial intelligence hybrid models." *Jordan Journal of Electrical Engineering* 6, no. 2 (2020): 168-178.

[21] Hussain, Lal, Tony Nguyen, Haifang Li, Adeel A. Abbasi, Kashif J. Lone, Zirun Zhao, Mahnoor Zaib, Anne Chen, and Tim Q. Duong. "Machine-learning classification of texture features of portable chest X-ray accurately classifies COVID-19 lung infection." *BioMedical Engineering OnLine* 19, no. 1 (2020): 1-18.
[22] https://europepmc.org/article/pmc/pmc7254021#bib001
[23] Keras : https://www.keras.io/
[24] Coronaviruses. https://www.niaid.nih.gov/diseases-conditions/coronaviruses
[25] Kaggle Chest X-Ray images: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia
[26] Public COVID-19 dataset of X-ray and CT scans: https://josephpcohen.com/w/public-covid19-dataset/
[27] Flask python framework: https://flask.palletsprojects.com/en/1.1.x/