

COVID-19 DETECTION TECHNIQUES WITH X RAY IMAGES USING MACHINE LEARNING

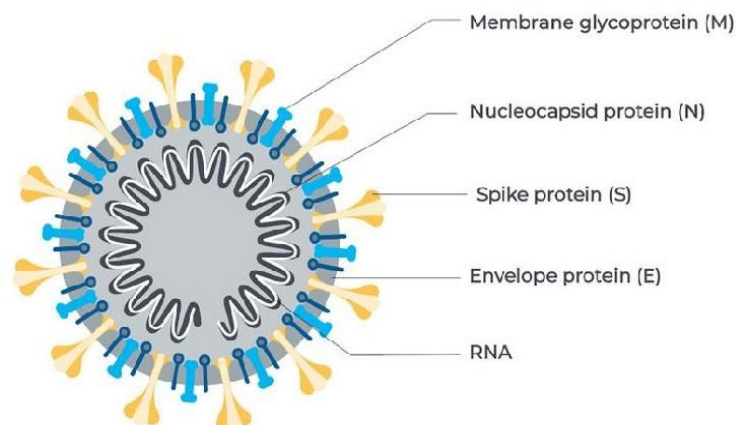
R. UMAMHESWARI, G. YAMINI, B. SWARNA CHANDANA, Y. NAGARJUNA REDDY
Dr. V. S. R. KUMARI, Professor, PRINCIPAL

SAI SPURTHI INSTITUTE OF TECHNOLOGY, B. Gangaram, Sathupalli mandal,
Khammam District, Telangana.

Abstract: COVID-19 has been declared as a pandemic in over 200 countries of the world. COVID-19 is an infectious disease that is primarily caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). According to the latest figures by the world health organization, the number of confirmed cases for the COVID-19 pandemic worldwide is more than 20 million worldwide and the number of fatalities reported is over 700,000. It has been found from several studies that medical imaging coupled with machine learning methods holds great promise in the detection and follow-up of the COVID-19 disease due to the enhanced accuracy in results of the experiments performed by the researchers. Machine Learning (ML)-based solutions can be used to simultaneously analyse multiple input computed tomography (CT) images of chest and lungs. A large number of papers have been published that show the application of machine learning methods in successful detection of the COVID-19 disease. Such applications demonstrate the suitability of feature prediction, identification of involved risks and therefore managing and intercepting the outbreak of such diseases. This paper describes some of the techniques in machine learning that can be used detection of COVID-19 disease.

Keywords: SARS-CoV-2, COVID-19, CT scan, image processing, Machine Learning, Deep Learning.

I. INTRODUCTION



A coronavirus contains four structural proteins, including spike (S), envelope (E), membrane (M) and nucleocapsid (N) proteins

Source: J Peiris, Y Guan & K Yuen, Severe acute respiratory syndrome, Nature Medicine Supplement 2004, 10 (12)

Fig1 Coronavirus Structure

COVID-19 is an infectious disease that is primarily caused by the severe acute respiratory syndrome coronavirus 2((SARS-CoV-2) of the genus Beta coronavirus [1].In comparison with other strains of coronavirus such as the severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), the SAR-CoV-2 is highly contagious with a rapid transmission capability leading to the onset of respiratory distress symptoms and even death[2,3].TheCOVID-19 has been declared as a pandemic in 213 countries owing to the steep increase in the number of confirmed cases as is evident from the papers published by different research studies[1, 4]. According to the latest figures by the world health organization(WHO), the number of confirmed cases for the pandemic is more than 20 million worldwide and the number of fatalities reported is over 700,000[5]. The COVID-19 disease is a pandemic and a worldwide catastrophe as suggested by the United Nations which is directly affecting the lives of billions of people worldwide in terms of economic, environmental & social developments. The pandemic is spreading through newer means and mechanics is yet to be sufficiently understood [11].Thestructure of coronavirus can be understood with the help of the figure-1 below

It has been found that the SARS-CoV-2 is mainly transmitted through minute droplets exhaled by an infected person through sneezing, coughing, or close interaction with non-infected persons. These droplets can then be inhaled, or they can land on surfaces that others may come into touch with, who can then get infected when they contact their eyes, mouth, or nose.

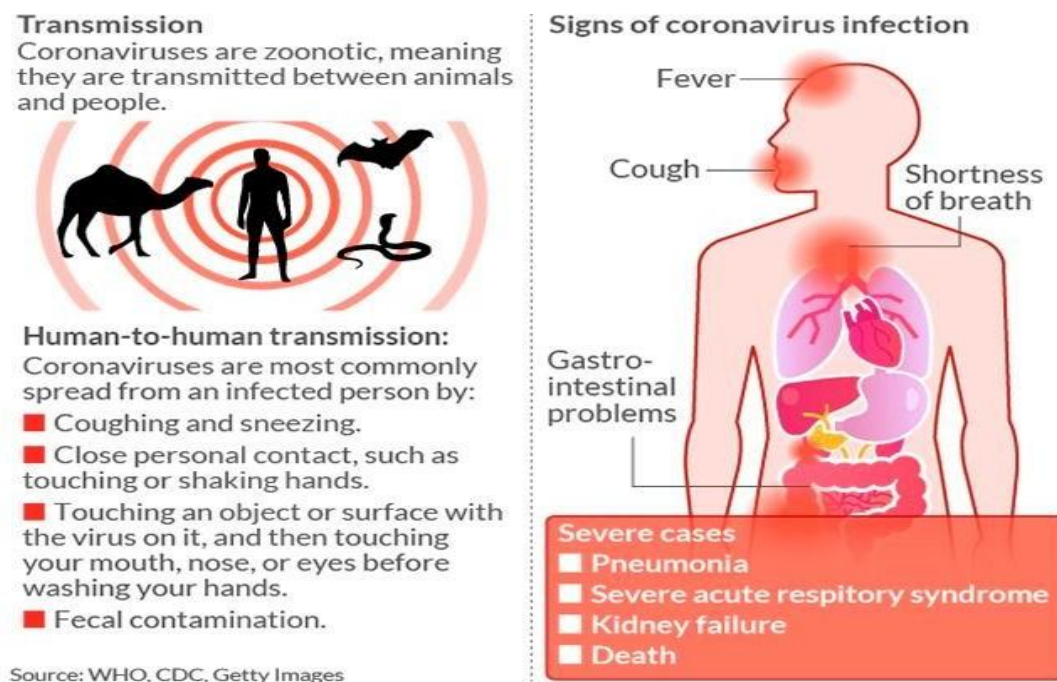


Fig-2 Transmission of coronavirus

There are several stages involved in the spread of the virus which is described below-

Stage 1- Imported Cases: These are those who have travelled to virus-hit foreign countries and have come back to India.

Stage 2- Local Transmission: These are those cases that have come in contact with patients who have a travel history.

Stage 3- Community Transmission: Community transmission is when a patient not exposed to any infected person or one who has traveled to any of the affected countries tests positive. Large areas get affected when community transmission takes place.

Stage 4- Epidemic: This is the last and the worst stage where the disease takes the shape of an epidemic with no clear endpoint like it did in China.

Machine learning (ML) is a branch of artificial intelligence that deals with algorithms that can continuously evolve and improve through experience [14]. ML algorithms typically build a numerical model based on available data, known as "training data", for purpose of prediction without being

Explicitly programmed to do so. Generally, ML algorithms are used for applications where it becomes cumbersome or infeasible to employ or build models using conventional programming algorithms. Today however we find that the ML algorithms are being used effectively for a large number of applications involving robotics, computer vision, data mining & forecasting.

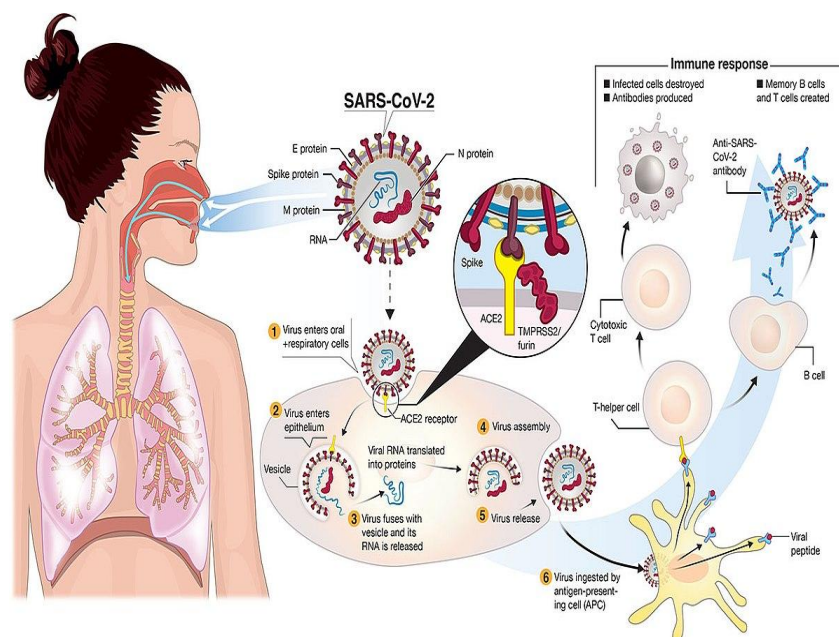


Fig 3 Covid transmission in body

Deep learning is a subset of the machine learning paradigm which deals with artificial neural network algorithms that follow the structure and function of the human brain. Deep learning is concerned with deep neural networks that employ a large number of network layers for operation. Due to the usage of a large number of layers, the performance continuously increases as we train the network layers with more data. An

important characteristic of deep learning is called feature learning which allows for automatic feature extraction from raw data. Deep learning methods aim at learning feature hierarchies with features from higher levels of the hierarchy formed by the composition of lower-level features. The hierarchy of concepts allows the computer to learn complicated concepts by building them out of simpler ones.

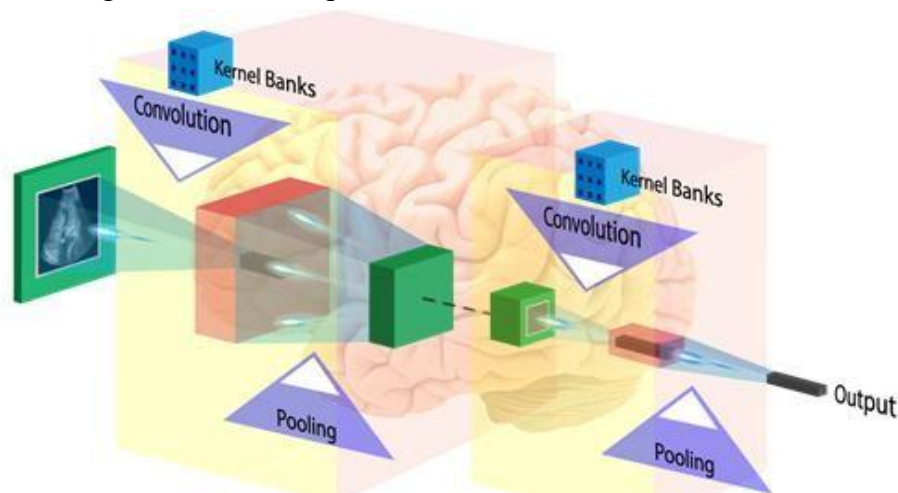
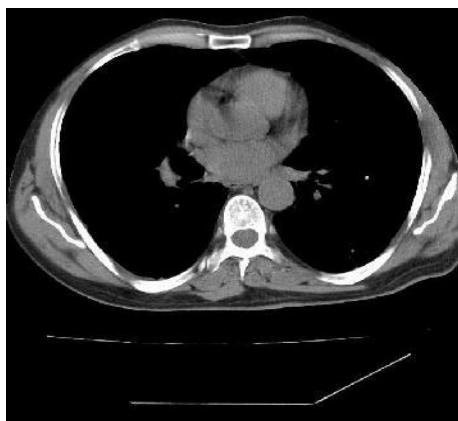


Fig 4 A basic convolution neural network model

Medical imaging can be used to create a visual representation of the interior portion of the human body so that anomalies can be detected and further intervention can be carried out. Medical imaging modalities such as thoracic, chest, and lung CT scans are widely being used in the diagnosis and follow-up of the COVID-19 disease. These imaging techniques coupled with ML solutions are quite effective in the early detection of disease. The major advantage of such methods is better prediction capability, lower miss diagnosis rates, and lowers the workload of radiologists [1]. The imaging techniques along with ML can also be employed for classification and positive confirmation of COVID-19 from also a large number of cases analyzed in a short period [4].

Techniques such as deep learning (DL) enable the effective management of COVID-19 patients and can help avoid transmission [6, 7, 8]. The imaging techniques [1, 4, 6] commonly used for positive confirmation and detection of the COVID-19 are low resolution computed tomography (LR CT), High resolution computed tomography (HR-CT), Non-contrast thoracic CT, etc as shown in Figure 5. Convolutional neural network-based approaches have been used for screening and detection of COVID-19 from X-ray image datasets [11, 12]. This review is dedicated to the study of some ML techniques for



the detection o

Fig 5 Lung CT image infected by COVID-19 pneumonia

COVID-19. The paper is divided into nine sections: section 1 gives the introduction. Section 2 describes some specialized CNN models used for COVID-19 detection and 3,4, 5, 6, 7, presents 5 working ML models for the classification of lung CT images, classification of X-ray images and detection of COVID-19 disease. Finally, the paper concludes in section 8.

II. LITERATURE SURVEY

A. Deep Learning Based COVID-19 Detection

In this paper they've taken the PA view of chest x-ray scans for covid-19 affected patients as well as healthy patients. After cleaning up the images and applying data augmentation, they've used deep learning-based CNN models and compared their performance.

B. COVID-19 Detection Using Convolutional Neural Network

In this paper they've applied three different models (InceptionV3, Xception, and ResNeXt). The analysis of this collected data is done with the help of CNN. This work mainly focuses on the use of CNN models for classifying chest X-ray images for coronavirus infected patients.

C. COVID-19 Prediction using SEIR model

This paper presents a comparative analysis of machine learning and soft computing models to predict the COVID-19 outbreak as an alternative to susceptible-infected-recovered (SIR) and susceptible-exposed-infectious-removed (SEIR) models.

Among a wide range of machine learning models investigated, two models showed promising results (i.e., multi-layered perceptron, MLP; and adaptive network-based fuzzy inference system, ANFIS). Based on the results reported here, and due to the highly complex nature of the COVID-19 outbreak and variation in its behaviour across nations, this study suggests machine learning as an effective tool to model the outbreak. This paper provides an initial benchmarking to demonstrate the potential of machine learning for future research. This paper further suggests that a genuine novelty in outbreak prediction can be realised by integrating machine learning and SEIR models.

D. COVID-19 Prediction using ARIMA model

The objective of the paper is to formulate a simple average aggregated machine learning method to predict the number, size, and length of COVID-19 cases extent and wind-up period crosswise India. This study examined the datasets via the Autoregressive Integrated Moving Average Model (ARIMA).

The study also built a simple mean aggregated method established on the performance of 3 regression techniques such as Support Vector Regression (SVR, NN, and LR), Neural Network, and Linear Regression. The results showed that COVID-19 disease can correctly be predicted. The result of the prediction shows that COVID-19 ailment could be conveyed

through water and air ecological variables and so preventives measures such as social distancing, wearing of mask and hand gloves, staying at home can help to avert the circulation of the sickness thereby resulting in reduced active cases and even mortality.

E. Specialized CNN models for covid-19 detection

COVID-19 Detection Neural Network (COVNet) architecture was introduced by [17]. This is a 3D deep learning architecture to detect COVID-19. This architecture can extract both 2D local and 3D global illustrative features. The Covent architecture is made with a ResNet architecture as the base model. A max-pooling operation was used for the extracted features from all slices. The finishing feature map is connected with a fully connected layer and they used a soft ax activation function for

the probability score for every type (COVID-19, Community-Acquired Pneumonia (CAP), and non- pneumonia).

COVID-Net architecture is specially adapted for COVID-19 case detection from chest X-Ray images. So obviously it has high architectural diversity and selective long-range connectivity. The massive use of a projection-expansion-projection design pattern in the COVID-Net [18] architecture is also observed. COVID-Net network architecture is incorporated into a heterogeneous mix of convolution layers. The proposed COVID-Net was pre-trained on the ImageNet dataset and then applied to the COVIDx dataset. Applying this architecture, they got accuracy about 93.3% on the COVIDx dataset.

ChexNet is originally a DenseNet-121 type of deep network which is trained on Chest X-ray images introduced by the paper [19]. So, this architecture has been specially designed to diagnose COVID- 19.1024-D feature vectors are extracted for the compact classifiers in ChexNet. They used the Softmax activation function to classify COVID-19, Normal, Viral Pneumonia and Bacterial Pneumonia. The number of trainable parameters in this model is 6,955,906.

COVID-CAPS is a capsule-based network architecture invented by [22]. This model has 4 convolutional layers and 3 capsule layers. 3-dimensional chest X-Ray images are the input of this architecture. The primary layer is a convolutional layer, and then batch-normalization is attached. The second layer is also a convolutional layer, followed by a pooling layer. Correspondingly, the third and fourth layers are convolutional, and the fourth layer is reshaped as the first Capsule layer. Three Capsule layers are embedded in the COVID-CAPS to perform the routing. The last Capsule layer contains the classification parameters of the two classes of positive and negative COVID-19. The trainable parameter is 295,488 for this model. Pre-trained COVID-CAPS gave 98.3% accuracy.

Detail-Oriented Capsule Networks (DECAPS) architecture was introduced by the paper [23]. It uses a ResNet with three residual blocks because the base network which

outputs 1024 feature maps, followed by a 1×1 convolutional layer with 512 channels and a ReLU non-linear layer. This architecture is trained in CT images. This model obtained an area under the curve (AUC) of 98%.

Besides these, some papers used different types of approaches Like Details Relation Extraction neural network (DRE-Net) [26] which is ResNet-50 on Feature Pyramid Network [FPN] for extracting top K details from each image and an attention module combined to learn the importance of every detail. In the training stage,[21] and [25] employed the least absolute shrinkage and selection operator (LASSO) to traverse the optimal subset of clinical-radiological features to classify. GLCM, HOG, and LBP were used by [24]. Moreover, [20] used commercial off-the-shelf software that detects nodules and small opacities within a 3D lung volume and subsystem.

F. UNET dl model for chest CT scans

The paper of Menget al. [6] discusses a standard framework for the discharge of COVID-19 patients from the hospital .research was carried out using a combination of a novel deep learning algorithm and chest CT of COVID-19 patients with respiratory pathogen nucleic acid test results. Pharyngeal swabs were used to gather Respiratory secretions for testing. Subsequently, low dose CT scans of all patients were also collected. A U-Net DL model was employed in the study due to the advantages of lesion detection with a large size variation and pneumonia volume was an important clinical index to evaluate the stage of illness and prognostic. Furthermore, precise pixel-level segmentation was required and U-Net provided a natural choice for its segmentation capability. The average intersection over union is 0.7894. Fifty-seven (21.1%) and 213 (78.9%) patients exhibited normal lung findings and pneumonia, respectively. 54.0% (115/213) involved mild interstitial fibrosis. 18.8% (40/213) had a total volume ratio of lesions to the lung of more than and equal to 50% according to our severity scale and were monitored continuously in hospital, and three cases of which had a positive follow-up nucleic acid test during hospital observation.

G. CNN for lung CT classification

Sally et al.[9]paper discuss an Artificial Intelligence-inspired Model for COVID-19 diagnosis and prediction for patient response to treatment AIMDP.The functionalities of the system are performed using two modules namely the Diagnosis Module (DM) and Prediction Module (PM). The study used the TensorFlow framework which is a popular open-source library for DL.The diagnosis module uses a convolutional neural network (CNN) shown in Figure 2and processes CT scans which can be usedfor the identification of positive COVID-19 patients. The prediction module on the other side employs a whale optimization algorithm [10] and can predict the treatment response-ability of the patient based on different factors such as the stage of infection, organ failure, and procedure of treatment. The major phases in the proposed model are pre-processing, segmentation, feature selection; classification, prediction, and diagnosis recommendation phase. The diagnosis and prediction modules performed well over a dataset having a large number

of CT images and data. The results obtained from the AIMDP coupled with the implementation of FSWOA with CNN has been compared with the performance of popular diagnostic models.

H. CNN based transfer learning for automatic detection of covid-19 from x-ray images.

The paper by Ioannis D. Apostolopoulos et al[11]. uses the technique called transfer learning for automatic detection of the COVID-19 disease. The advantages of transfer learning is that it provides better results for detection of abnormalities in medical image datasets They have acquired datasets of X-ray images having a manifestation of bacterial pneumonia, confirmed COVID-19, and normalcases. They have made an effort to study the performance of popular convolutional neural network architectures for medical image classification. A total of two datasets were utilized for the study purpose.

The convolutional neural network architectures used for the study in this paper were VGG19, MobileNet v2, Inception, Xception, Inception ResNet v2. A parameter called layer cut-off was defined after several experiments which are defined as the number of un-trainable layers beginning from the bottom of CNN. Remaining upper-level layers closer to the output were made trainable and thus were able to extract meaningful information being supplied from the other convolutional layers of the network

Another parameter called the neural network which can be explained by the total number of hidden layers and the number of nodes was used as a classifier at top of the CNN for classification of extracted features. Also, all convolutional layers were activated by the Rectified Linear Unit (ReLU) function. The CNNs were then compiled using a well-known optimization method called Adam. The training was conducted for ten epochs, with a batch size of 64.

The results from the study conclude that t Deep Learning with X-ray imaging can successfully find out the significant biomarkers about the Covid-19 disease, while the best accuracy, sensitivity, and specificity obtained was 96.78%, 98.66%, and 96.46% respectively.

I. Dark Covid Net model for automatic COVID-19 detection using X-ray images.

Tulin et al. [12] have performed a study for automatic COVID-19 detection using X-ray images which uses a proposed model for binary classification and multi-class classification that can accurately detect whether an X-ray image belongs to a non-COVID-19, COVID-19 or a pneumonia patient. They have suggested that their model can be used by radiologists for validating the initial screening of patients for COVID-19 and pneumonia diseases.

They have used a DarkNet-19 model called DarkCovidNet model as a classifier that can be used as the basis for a real-time object detection system named YOLO (You only look

once).in the proposed approach fewer layers and filters were used in comparison to the originalDarkNet architectures. They have used 80% of X-ray images for training and 20% for validation.The model proposed in the study uses 17 convolution layers with filters being applied to each layer.

Every layer in the DarkNet uses a convolution layer.Two operations are then carried out. Batch normalization is used to enhance the stability of the model and also reduces the training time.LeakyReLU is then subsequently applied to prevent dying of neurons. It is a variant of the ReLU operation used in CNN models.The DarkCovidNet model was trained to categorize the input X-ray images into one of the following COVID-1, No-Findings, Pneumonia. Again, the model was trained to detect two classes: COVID-19 and No-Findings. The researchers obtained a classification accuracy of 98.08% for binary classes and 87.02% for multi-class cases.

III. SYSTEM ARCHITECTURE

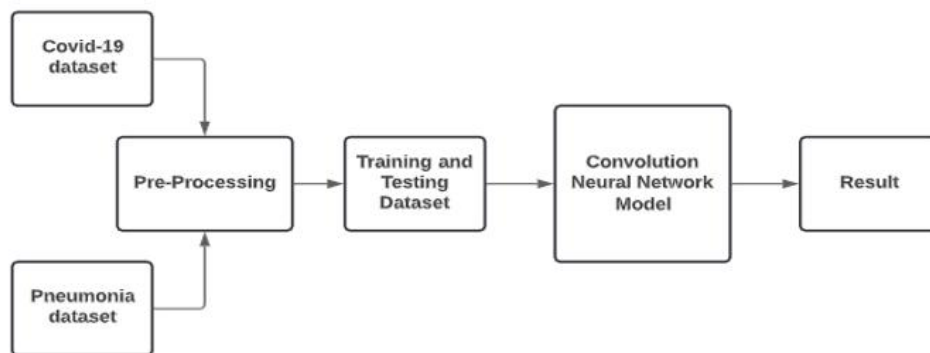


Fig 6 Complete System architecture

A. COVID-19 Dataset Description

First dataset involves collecting important datasets from github. COVID-19 dataset consists of X-Ray images of patients infected with COVID-19. These images are used to identify COVID-19 positive samples.

B. Pneumonia Dataset Description

The second dataset consists of Pneumonia negative patients. COVID-19 shares some similar chest conditions as Pneumonia so this dataset is used to identify COVID-19 negative samples.

C. Pre-Processing Description

Image Pre-processing here involves two phases namely image sorting, which includes merging COVID-19 dataset with pneumonia dataset and digital image processing which includes running of algorithms to perform image processing on digital images. It involves applying a wider range of algorithms to the input data to improve image data by suppressing unwanted distortions so that the CNN model may benefit from this and become more optimised and precise.

D. Training and Testing Data Description

Training dataset consists of thoroughly cleaned data, the success rate of a model depends upon the dataset. Cleaning is one of the major processes in model based projects. The training set is the material through which the computer learns how to process information. The CNN model applies its algorithms on this piece of dataset to generate its main intelligence. It is this dataset which trains the CNN model to differentiate between COVID-19 positive and negative samples when presented with a chest X-Ray image. Testing dataset evaluates the model's intelligence gained by the training data. It lists important parameters like Accuracy and Loss rate.

E. CNN block Description

This block represents the base model of image detection which is Convolutional Neural Network Model. It makes use of iterative training on each of the images to eventually be able to recognize features, shapes. Multiple filtering techniques are used to scan images using convnets. Repetitive filtering scans are carried out to expel any errors and provide the most accurate results for this sensitive test of COVID-19.

F. Result Block Description

Final result is displayed on the screen about whether the scanned Chest X-Ray of the patient in doubt is coronavirus positive or negative

G. Dataset Description

First phase of this prediction model is to collect COVID-19 daily case numerical data. The data needs to be updated regularly for precise predictions.

H. Preparation Description

This phase of the project involves modifying the dataset by using python libraries. Pandas is used to transform the dataset into data frames, numpy is used in crunching large datasets, it is one of the most used libraries by a data scientist, Matplotlib is used to convert data into corresponding x and y coordinates. Using matplotlib a graphical representation of the dataset is generated.

I. Training/Regression Description

In the training phase, the Machine learning model learns using sklearn library and the data which is collected and prepared. A Regression Model is used which is useful when predicting number based problems like probability of an event.

J. Prediction Count

The results are evaluated based on predicted trend and actual trend which provides strong evidence regarding the accuracy and operability of the Model. Model is then tuned up according to the results.

K. Dataset Description

Dataset consists of three main components namely number of active cases, deaths and recovered cases. It contains records of all the countries making it possible to analyse the COVID-19 situation globally. Dataset is extracted from John Hopkins repository which is updated daily.

L. Pre-processing Description

In this phase the dataset is exploited to extract specific data and group up this data into similar arrays. Arrays are then processed to do a comparative analysis according to the desired output by the developer

M. Analysis Description

Arrays of data created in the pre-processing phase are used in this phase to calculate certain country specific values and time specific values and learn the pattern of outbreak of Coronavirus. Analysis provides a deep knowledge of the trend of the virus in different regions and the growth rate of different countries.

N. Training Description

In this phase python libraries such as numpy, pandas, scikit-learn are used to train the model to provide desirable outputs. Model specific algorithm is the base on which the model is trained.

O. Competitive Modelling Description

This phase focuses on developing alternate predictive models to calculate possible future values. Accuracy and loss are the parameters observed in this phase which decide the success rate of the models developed.

IV. CONCLUSION

ML techniques are used to accurately predict the manifestation of COVID-19 disease and can prove to be useful for radiologists for validating the results of clinical diagnosis. Medical imaging coupled DL holds great promise in the detection and follow-up of the COVID-19 disease due to the enhanced accuracy in the results of the experiments performed by the researchers. The imaging techniques alongwith AI can also be employed for classification and positive confirmation of COVID-19 from also a large number of cases analyzed in a short period. Techniques such as DL have enabled the effective management of COVID-19 patients. AI-based solutions can be used to simultaneously analyze multiple input Computed Tomography (CT) images of the chest may lead to early detection of the disease efficiently. Convolutional neural networks can also be used accurately for COVID-19 classification from X-ray images.

References

[1] Shuyi Yang, Longquan Jiang, Zhuoqun Cao, et al. Deep learning for detecting coronavirus disease 2022 (COVID-19) on high-resolution computed tomography: a pilot

- study, *Ann Transl Med* 2020;8(7):450 | <http://dx.doi.org/10.21037/atm.2020.03.132>.
- [2] Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2020. *N Engl J Med*. 2020 Feb 20; 382 (8):727-733.
- [3] Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020 Feb 15;395 (10223):497-506.
- [4] OphirGozes , Heather Greenspan 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, 2020, ResearchGate.
- [5] Available Online. <https://covid19.who.int>
- [6] QingchengMeng, Wentao Liu , Xinmin Dou, Jiaqi Zhang, et al, Role of novel deep-learning-based CT used in management and discharge of COVID-19 patients at a “square cabin” hospital in China, 2020, Research Square.
- [7] Ye H, Gao F, Yin Y, et al. Precise diagnosis of intracranial hemorrhage and subtypes using a three dimensional joint convolutional and recurrent neural network. *European Radiology*, 2019, 29: 6191–6201.
- [8] Lin Li, Lixin Qin, ZeguoXu et al. Artificial Intelligence Distinguishes COVID-19 from Community Acquired Pneumonia on Chest CT. *Radiology* 2020.
- [9] Sally Elghamrawy, Aboul Ella Hassanien, Diagnosis and Prediction Model for COVID-19 Patient’s Response to Treatment based on Convolutional Neural Networks and Whale Optimization Algorithm Using CT Images. medRxiv preprint doi:<https://doi.org/10.1101/2020.04.16.20063990>.
- [10] Mirjalili, S. and Lewis, A., 2016. The whale optimization algorithm. *Advances in engineering software*, 95, pp.51-67.
- [11] Ioannis D. Apostolopoulos, Tzani A. Mpesiana, Covid-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine* (2020) 43:635–640.
- [12] <https://doi.org/10.1007/s13246-020-00865-4>
- [13] Tulin Ozturk, Muhammed Talo, Eylul Azra Yildirim, Ulas Baran Baloglu, Ozal Yildirim
- [14] U. Rajendra Acharya, Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine* 121 (2020) 103792. <https://doi.org/10.1016/j.combiomed.2020.103792>
- [15] Chuansheng Zheng et al., Deep Learning-based Detection for COVID-19 from Chest CT using Weak Label. medRxiv preprint doi: <https://doi.org/10.1101/2020.03.12.20027185>
- [16] <http://www.cs.cmu.edu/~tom/mlbook.html>
- [17] Singh, S., Norving, P., & Cohn, D. (1996). *Introduction to reinforcement learning*. Harlequin Inc.

