Coronavirus disease (novel COVID-19) detection in Chest X-Ray images using CNN model

1Shweta Achyut Adurkar, 2Siddhi Satish Kate, 3Mayuri Sunil Solankure, 4Ankita Ananda Gend, 5Utkarsha Rajvardhan Desai
1Student, 2Student, 3Student, 4Student, 5Student
Sanjay Ghodawat Institute of Engineering & Technology

Abstract - **Science and technology have improved our quality of life, but some industries' rapid development has given up people's future living environment and harms survival. Chest X-Ray (CXR) plays an essential role in the detection. Yet, the less availability of expert radiologists to interpret the CXR images and the subtle appearance of disease radiographic responses remains the major issue in manual diagnosis. Manual diagnosis is very complex andtime-consuming. Automatic COVID (coronavirus) screening (ACoS) system uses radiomics texture descriptors extracted from CXR images to detect the normal, suspected, and COVID-19 infected patients. But this system is also time-consuming. Hence we propose a System for COVID-19 detection. The diagnosis of COVID-19 is typically associated with both the symptoms of pneumonia and Chest X-ray tests. CXR is the first imaging method that plays a vital role in the diagnosis of COVID-19 disease. In the existing system, we find some disadvantages; to overcome this, we will use X-ray data of normal and COVID-19 positive patients and train a model to differentiate between them. We present COVID-19 AI Detector using a deep convolutional neural network model (CNN) to triage patients for appropriate testing.**

keywords - **Deep learning, Covid-19, CNN, Algorithm**

1. Introduction

Science and technology have enhanced people's quality of life, but the rapid change in development may cause an inverse impact on people's life. Now in the pandemic of COVID 19, the sudden increase in the number of patients with the COVID-19 has put a load on healthcare systems worldwide. There are limited kits of diagnosis, limited hospitals, limited beds for patients, limited ventilators. The accessibility of personal protective equipment kits (PPE) and their regular, proper use by healthcare providers and public health professionals is a critical factor in combating any infectious disease in a crisis. The requirement of PPE has exponentially increased [3]. So the patients who suffer from severe acute respiratory illness (SARI) and those suffering from COVID-19 are the challenge to the healthcare to utilize the resources. The diagnosis process of COVID-19 is done with either the symptoms of pneumonia and a chest X-ray test. The chest X- ray image is the first imaging method that plays a role in detecting the COVID-19. It is also affordable to everyone and much more widespread. Several classical machine learning algorithm is used for automatic classification of chest images. Here in the ACoS system, they have used the Support Vector Machine classifier to classify the images into different categories.

A grey-level co-occurrence matrix method was used with the Backpropagation Network to classify images from being normal or cancerous. With enough images, deep learning approaches have revealed their pre-eminence over the classical machine learning approaches.

In our work, we propose the use of an X-ray image to detect COVID-19 infection. Using our tool, we classify the X-ray image in one of four classes: regular, bacterial pneumonia, viral pneumonia, and COVID-19 pneumonia.

The project's primary goal is to propose a novel deep neural network-based to detect the COVID-19 from the X-ray image with high accuracy. Currently, the detection of COVID-19 from X-ray images is done with the radiologist. But it takes few times to diagnosis the COVID-19. Further, many radiologists are unaware of the infection knowledge and may be lacking in highly accurate diagnostic.

With the virus spreading very quickly, the World Health Organization characterizes the COVID-19 as a pandemic to classify the aesthetic of COVID -19 classifications in their natural habitats. This study developed a method using a convolution neural network (CNN). This extract information from Covid-19 classification images captured previously or in real-time by identifying local features. CNN architecture is one of the most popular deep learning models in the medical imaging domain. It is efficient because of its ability to learn features automatically from domain-specific images.

2. Literature Survey

Novel coronavirus disease (nCOVID-19) is the most challenging problem for the entire world. The disease is caused due to severe acute respiratory syndrome coronavirus-2 (SARS COV-2), leading to high morbidity and mortality worldwide.

There is no particular vaccine or medication available to cure the disease and prevent further spread.

Existing System

The study tells that infected patients to exhibit unique radiographic visual characteristics, fever, dry cough, fatigue, etc. **Chest X-Ray images (CXR)** are critical clinical adjuncts that play a vital role in detecting such visual characteristics associated with SARS-COV-2 infection [1].

A) Manual diagnosis:

However, the lack of expert radiologists to predict the Chest X- ray images and the subtle appearance of disease radiographic responses remains the biggest problem in manual diagnosis. The standard confirmatory clinical test – reverse transcription-polymerase chain reaction (RT-PCR) test for detecting nCOVID-19 is manual, complex, and time- consuming [1].

The limited availability of test-kits and domain experts in the hospitals and the rapid increase in the number of infected patients necessitates an automatic screening system. It can act as a second opinion for expert physicians to quickly identify the infected patient's immediate isolation and further clinical confirmation.

However, the **manual interpretation of these subtle visual characteristics on CXR images is challenging and requires a domain expert**. The exponential increase in the number of infected patients makes it difficult for the radiologist to complete the diagnosis in time, leading to high morbidity and mortality [1].

Disadvantages:

- Manual diagnostic takes more time to predict the result that's its time consuming
- There are limited number of test kits so due to lack of kits there is difficulty to predict it fast.
- And the less number of radiologist who predict the result from xray, also for prediction many of radiologist are unaware about the proper knowledge of the infection. So it's not accurate and takes much time too [1].

B) ACoS System:

Prototype of the ACoS System :

The automatic COVID screening (ACoS) system uses radiomics texture descriptors extracted from CXR images to identify the normal, suspected, and novel COVID-19 affected patients. The proposed method uses 2 phase classification approach (normal versus abnormal and nCOVID-19 versus viral pneumonia) using a majority vote- based classifier ensemble which uses five benchmark supervised classification algorithms [1]. The training, testing, and validation of the ACoS system are done using 2088 (696 normal, 696 viral pneumonia, and 696 coronaviruses) and 258 (86 images of each category) CXR images respectively. The obtained validation results for phase I (accuracy (ACC) is 98.062%, area under the curve (AUC) is 0.956) and phase-II (ACC = 91.329% and AUC = 0.831) show the promising performance of the proposed system.

However, the limited availability of annotated CXR images for nCOVID19 cases to train the data-hungry DL models became the biggest bottleneck. After the retrospective analysis of the above literature, we found that the existing studies had been performed using a limited number of CXR or CT images, leading to the under-fitting of the data-hungry DL models [1].

Feature extraction methods :

Features are the data extracted from images in terms of numerical values that are difficult to understand and analyze by a human being. Let us consider the image as data. The information extracted from the data is called features. Usually, features taken from an image are of greater size than the initially available image. The reduction in size decreases the overheads of processing a lot of images.

There are two types of features extracted from the images based on the application. They are local and global features. Features are sometimes referred to as descriptors. Global descriptors are primarily used in image retrieving, object detection, and object classification, while local descriptors help object recognition or identification. There is a difference between detection and identification. Detection is finding an object's existence (Finding whether an object exists in image/video), whereas recognition is finding the identity (Recognizing a person/object) of an object. Global features describe the image; in general, they generalize the entire object, whereas the local features represent the image patches (key points in the image) of an object[2].

Global features include contour representations, shape descriptors, texture features, and local features representing the texture in an image patch. Shape Matrices, Invariant Moments (Hu, Zernike), Histogram Oriented Gradients (HOG), and Co- HOG are examples of global descriptors. SIFT, SURF, LBP, BRISK, MSER, and FREAK are examples of local descriptors [2].

1) GLCM- Grey Level Co-occurrence Matrix
2) FOSF- First Order Statistical Features
3) HOG- Histogram of Oriented Gradients

The study uses eight first-order statistical features (FOSF) (Srinivasan & Shobha, 2008), 88 grey level co-occurrence matrix (GLCM) (Gomez,' Pereira, & Infantosi, 2012; Haralick, Shanmugam, & Dinstein, 1973) features (in four different orientations) and 8100 histograms of oriented gradients (HOG) (Dalal & Triggs, 2005; Santosh & Antani, 2018) features. The FOSF describes the complete image at a glance by using the mean, variance, roughness, smoothness, kurtosis, energy, and entropy, etc. It can easily quantify the global texture patterns; however, it does not contemplate the local neighborhood information. The GLCM and HOG feature descriptor is used to perform the in-depth texture analysis to overcome this shortcoming. The GLCM feature describes the spatial correlation among the pixel intensities in radiographic texture patterns and four distinct directions (i.e.,0◦ , 45◦ , 90◦ , 135◦), whereas the HOG feature encodes the local shape/texture information.

Use of BGWO feature extraction method: In this study, a total of 8196 features (8 FOSF, 88 GLCM, 8100 HOG) are extracted from each CXR image (described in Appendix-A). However, not all the extracted features are relevant for accurate characterization of visual indicators associated with nCOVID-19. Thus, to select the most informative features, they used a recently developed metaheuristic approach called—binary grey wolf optimization (BGWO) (Mirjalili, Mirjalili, & Lewis, 2014 Abdullah, Mohd Saad, Mohd Ali, & Tee, 2018). The method imitates the leadership, encircling, and hunting strategy of grey wolves. Unlike the other evolutionary algorithms, the process does not get trapped in local minima, which motivated us to use it in a study [1].

3. Proposed System

The novel coronavirus (COVID-19) is quickly spreading throughout the world, but hospitals' facilities are limited. Therefore, diagnostic tests are required to timely identify COVID-19 infected patients and thus reduce the spread of COVID-19.

Covid-19 is a respiratory disease that infects our lungs and can cause significant damage to the lungs that causes difficulty in breathing and, in some cases, pneumonia and respiratory failure. Our proposal System for the COVID-19 pandemic is one of the biggest challenges for the healthcare system.

For that purpose, we will use X-ray data of lungs normal and COVID positive patients and train a model to differentiate between them. We present CovidAID: COVID- 19 AI Detector using a Deep Convolutional Neural Network Model (CNN) to triage patients for appropriate testing.

System Architecture:

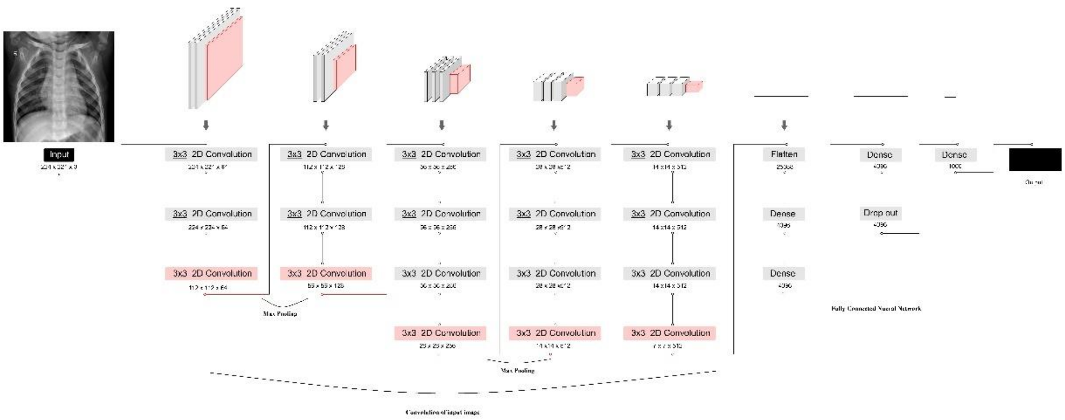

Fig3.1: System Architecture

Methodology:

❖ **Data Processing:**

➢ **Finding Data:**

The dataset is collected from official websites of government and the dataset is limited to country India. For training of model the dataset from Bharat Biotech and ICMR will be used. And to validate the built model, the dataset from WHO will be used.

➢ **Cleaning and Preparing the data:**

▪ **Data Consolidation:**

The received data are from different websites and is not consolidated. The data is gathered according to its type for further ease of operation (Here data is divided into three categories namely COVID-19, Viral Pneumonia and normal).

▪ **Data Wrangling and cleaning:**

The collected information is in raw format and it can't be analysed easily. For ease of analysis and to analyse more complex data more quickly, so accurate results, and because of this better decisions can be made using data wrangling. The missing data and inaccurate data is also handled in data cleaning.

- **Exploratory Data Analysis:**

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations.

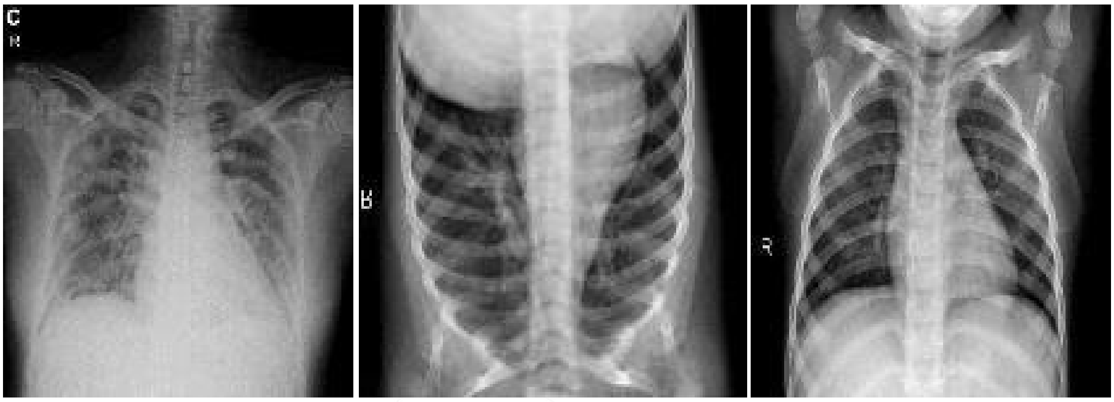

Fig32: Analysis of Images

- **Machine Learning Algorithm:**

As mentioned earlier, we have used the pixel values to classify the given input image into one of three categories-Covid, Viral Pneumonia and Normal. So here classification machine learning algorithm comes into play. We have used advanced version of Artificial Neural Network(ANN) –Convolutional Neural Network i.e. CNN.We have used CNN over ANN as ANN fails to capture a pattern. CNN has slider called 'Stride' who scans bunch of pixels value groups to detect pattern. ANN remembers the common resultant points, like fixed results. And when given a new image, sometimes it fails to detect. But this overfitting problem not occurs in CNN.

And more importantly, as we have huge number of points in an image, and when we flattened them into 1D array, there will be huge number of points. Hence due to this cursive dimensionality of dataset, CNN is best choice as machine learning algorithm.

- **Data Transformation:**

We prepared the data and configured it according to need. But the images available cannot be used as it is. As each image will have different size and shapes and they will be multi-scaled with different rotations also. So before using images for further use it is convoluted and max pooling and padding is used to handle the different sizes of images.

- **Convolution:**

Convolution is done with images to reduce the size of image and produce a feature map. Data augmentation is a done to significantly increase the diversity of data available for training models, without actually collecting new data.

- **Data Augmentation:**
  - Random flipping
  - Rotation
  - Translation
  - Zero phase component analysis whitening
  - Gaussian filtering

- **Pooling:**

All the images provided to the model, will not be of same size. To standardize all images to one size, and to reduce dimensions of the feature maps, we used pooling.We used max pooling as it is useful when the background of the image is dark and we are interested in only the lighter pixels of the image as data in our dataset.

Max Pooling is a convolution process where the Kernel extracts the maximum value of the area it convolves. Max Pooling simply says to the Convolutional Neural Network that we will carry forward only that information, if that is the largest information available amplitude wise.

Max pooling is added after a convolutional layer. This is the output from the convolution operation and is the input to the max pooling operation.The output after max-pooling layer would be a feature map containing the most prominent features of the previous feature map.

Max-pooling on a 4*4 channel using 2*2 kernel and a stride of 2: As we are convolving with a 2*2 Kernel. If we observe the first 2*2 set on which the kernel is focusing the channel have four values 8,3,4,7 in the figure below. Max-Pooling picks the maximum value from that set which is "8".

On the convolutional output, and we take the first 2 x 2 region and calculate the max value from each value in the 2 x 2 block. This value is stored in the output channel, which makes up the full output from this max pooling operation.

We move over by the number of pixels that we defined our stride size to be. We're using 2 here, so we just slide over by 2, then do the same thing. We calculate the max value in the next 2 x 2 block, store it in the output, and then, go on our way sliding over by 2 again.
Once we reach the edge over on the far right, we then move down by 2 (because that's our stride size), and then we do the same exact thing of calculating the max value for the 2 x 2 blocks in this row.

We can think of these 2 x 2 blocks as *pools* of numbers, and since we're taking the max value from each pool, we can see where the name *max pooling* came from.

This process is carried out for the entire image, and when we're finished, we get the new representation of the image, the output channel.

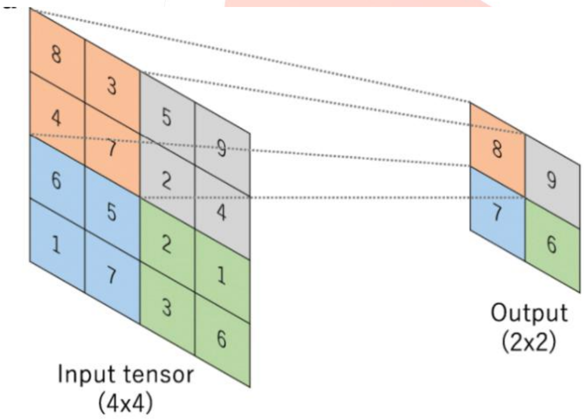

Fig3.3: Max Pooling

- **Feature map Extraction:**

The feature maps of a CNN capture the result of applying the filters to an input image. i.e at each layer, the feature map is the output of that layer. The fully connected network consists of these convolution layers which contains neurons. The input to the fully connected layer is the output from the final Pooling or Convolutional Layer, which is flattened and then fed into the fully connected layer. The inputs from one layer are connected to every activation unit of the next layer.

The feature maps of a CNN capture the result of applying the filters to an input image. i.e at each layer, the feature map is the output of that layer. The reason for visualising a feature map for a specific input image is to try to gain some understanding of what features our CNN detects.

Features are the information extracted from images in terms of numerical values that are difficult to understand and correlate by human. Generally, features extracted from an image are of much more lower dimension than the original image. The reduction in dimensionality reduces the overheads of processing the bunch of images.

Different Conv2D filters are created for each of the three channels for a colour image.Filters for each layer are randomly initialized based on either Normal or Gaussian distribution.Initial layers of a convolutional network extract high-level features from the image, so use fewer filters. As we build further deeper layers, we increase the number of filters to twice or thrice the size of the filter of the previous layer.Filters of the deeper layers learn more features but are computationally very intensive.

CNN uses learned filters to convolve the feature maps from the previous layer. Filters are two- dimensional weights and these weights have a spatial relationship with each other.

The process followed to visualize the feature maps is as follows:

1. Define a new model,that will take an image as the input. The output of the model will be feature maps, which are an intermediate representation for all layers after the first layer. This is based on the model we have used for training.
2. Load the input image for which we want to view the Feature map to understand which features were prominent to classify the image.
3. Convert the image to NumPy array
4. Normalize the array by rescaling it
5. Run the input image through the visualization model to obtain all intermediate representations for the input image.
6. Create the plot for all of the convolutional layers and the max pool layers but not for the fully connected layer. For plotting the Feature maps, retrieve the layer name for each of the layers in the model.

The network consists of an input layer, followed by three convolutional and average pooling layers and followed by a soft max fully connected output layer to extract features. After extracting features, 2 layer hidden neural-network is used for classification.

➢ **Skip Connections:**

We have used skip connections at some places to improve the performance and the convergence of deep neural networks, which is believed to relieve the difficulty in optimization due to non-linearity by propagating a linear component through the neural network layers.Skip connections in deep architectures, skip some layer in the neural network and feeds the output of one layer as the input to the next layers (instead of only the next one).

In general, there are two fundamental ways that one could use skip connections through different non-sequential layers:
a) addition as in residual architectures,
b) concatenation as in densely connected architectures.

In our model, we have used addition.The core idea is to back propogate through the identity function, by just using a vector addition. Then the gradient would simply be multiplied by one and its value will be maintained in the earlier layers.

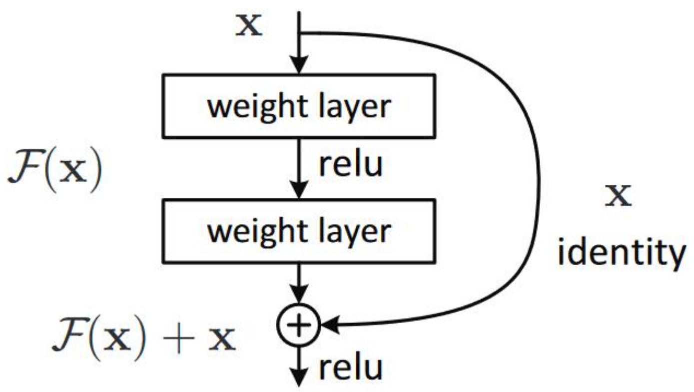

Fig3.4:Skip Connections

➢ **Sampling:**

In total, there are 21165 samples divided into four main classes:

- Covid-19
- Lung Opacity
- Normal
- Viral Pneumonia

All the images are in Portable Network Graphics (PNG) file format and the resolution are 299x299 pixels. On this current update, the database currently holds 3,616 COVID-19 positive cases, 10,192 Normal, 6,012 Lung Opacity (Non-COVID lung infection), and 1,345 Viral Pneumonia images.

Using random sampling, we split our data set into training and testing sets in the ratio of 70:30 respectively. Then we divided the training set such that approximately 60 % of the records were unlabelled and the remaining were labelled. Following this, we used subsets of increasing sizes from the labelled data to train the CNN model. The results of these are discussed in the Experiment and Results' section.

➢ **Plan:**

1. Organise the data into a dataframe to group the path of all images and their respective target. By doing so it is easier to handle data transformations later on.
2. Healthy and Lung Opacity samples compose 80% of the dataset. For this application, the main goal is to recognise Covid-19 patients. It will be interesting to see if the model will have greater difficulty in identifying Pneumonia or Covid samples.
3. Similar to other health conditions prediction problems or unbalanced datasets, it is necessary to prioritise Precision or Recall, since Accuracy can be misleading. The F1-Score is also a reasonable option.

Basic considerations regarding the CNN model used:

- Use ImageDataGenerator for Data Augmentation and organise the files into training and validation set
- train and val_datagen have different settings. Ideally, we have augment the validation set.
- val_datagenhyperparameter **shuffle=False** makes sure the training and validation data do not overlap
- CNN architecture consistently provides reasonable results as a starting point
- The Model predicts the Four types of X-Ray Images
- Confusion Matrix, Accuracy, Precision, Recall and F-Score are analysed for final remarks

4. The validation performance oscillates heavily in the initial epochs, i.e. from 0.00 to 90% in the following epoch. As the Learning Rate was already low, reducing it was not helpful and neither callbacks improved this behaviour.
5. Perhaps, when using 'SGD' as an optimiser,we can play with Learning Rate Scheduling. 'Adam' already has that functionality built-in.
6. Even though we did not use any optimised hyperparameters tuning, the batch_size with a lower learning rate of 0,001 provided satisfactory results.
7. The results are analysed in terms of F1-Score, as Precision and Recall are both relevant metrics for this application.
8. To provide a general overview of the Model performance, the confusion matrix and results for the F1-Score, Precision, Recall and overall Accuracy is also presented in section 'Result and Analysis'.

4. Result & Analysis:

Covid class presents ~ 80% Precision and 60% Recall. While the model is not capable of classifying all the Covid-19 samples correctly (low Recall - higher FN) it is usually correct when it does so (higher precision - low FP).

Lung Opacity and Normal classes have a similar value for Precision and Recall, meaning the model is as good at recognising the samples and properly classifying them.

Viral Pneumonia presents the opposite result we saw in Covid-19. It presents a higher Recall than Precision. This means that the model produces FN for this class. However, it needs to be improved to properly classify the samples, i.e. reducing the FP for the Viral Pneumonia class.

The F-Score is the balance between Precision and Recall. As expected, Lung Opacity and Normal classes have higher scores as their Precision and Recall metrics are similar. Lower results are found for Covid and Viral Pneumonia, as their Precision and Recall metrics differed more intensely.

Overall, it is a good outcome that all metric scores are above 75%. The recall is higher as it was monitored, alongside the Model Loss, to stop the training process.

5. Future Scope:

As of now, the method for detection of Covid-19 is primarily manual. And due to the lack of expert radiologists, it's overhead for doctors and patients also for disease detection. So this system which we are building in our project, will help reduce the time for getting results, and it will also be beneficial for doctors. After completing the system with a user-friendly UI, anyone can detect the effects from a chest x-ray image.

6. Conclusion:

In this study, we have presented a CNN model for the preliminary diagnosis of nCOVID-19 infected patients. Proper precautionary measures (like isolation and RT-PCR test) can prevent the further outbreak of the infection.

Detection of COVID-19 from chest X-ray images is vital for both doctors and patients to decrease the diagnostic time and reduce financial costs.

7. References

[1] **ELSEVIER:** Coronavirus disease (COVID-19) detection in Chest X-Ray images using majority voting based classifier ensemble. Tej Bahadur Chandraa,*, Kesari Vermaa, Bikesh Kumar Singhb, Deepak Jainc, Satyabhuwan

Singh Netamc.
[2] Hari Kishan, (research in Image Processing and Computer Vision)
[3] Personal Protective Equipment: Challenges and Strategies to Combat COVID-19 in India: A Narrative Review
[4] COVID-19 Image Data Collection: Prospective Predictions Are the Future
[5] Joseph Paul Cohen, Paul Morrison, Lan Dao, Karsten Roth, Tim Q Duong, Marzyeh Ghassemi
[6] Nandi, Ritika, and Manjunath Mulimani. "Detection of COVID-19 from X-rays using Hybrid Deep Learning Models." (2021).