IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

COVID-19 PREDICTION USING CONVOLUTIONAL NEURAL NETWORK APPROACH

[1]Yella Siva Naga Sai Meghana, [2]Sarikonda Vaishnavi, [3]Sai Charan Sahu, [4]Gaddam Uday Kiran Reddy

[1,2,3,4]Student, Department of Computer Science and Engineering,
[1]GITAM University, Visakhapatnam, India

Abstract: The novel coronavirus disease outbreak of 2019 (COVID-19) has impacted the health of many civilians around the world. It has developed a difficult situation by bringing life to a standstill throughout the globe and killing thousands of civilians. Several applications have been developed for analyzing and detecting the Covid-19 disease. Our proposed system consists of two modules. One is prediction of Coronavirus based on basic symptoms using supervised Machine Learning algorithms which are Logistic Regression and Random Forest Classification. Another is detection by applying Convolutional Neural Network on Chest X-ray image Dataset

Key words: Convolutional Neural Network, COVID-19 detection, Python, Logistic Regression, Random Forest Classifier, Machine Learning.

I. INTRODUCTION

1.1 What is COVID-19?

COVID is a huge, extremely infectious virus that can cause a wide range of diseases, starting from the cold to much more significant medical disorders including Acute Respiratory Distress Syndrome (MER Covid) and acute respiratory syndrome (sars syndrome (SARS) (SARS-Covid). Severe acute respiratory syndrome is a highly infectious pathogen which can lead to serious sickness. Coronavirus is pathogenic, which means that may be passed through one animal to another such as human being. As per extensive study, SARS-Covid is transferred from civet cats to humans, while MERS-Covid was spread from dromedary camels to beings. Numerous coronaviruses are proliferating in wildlife which have not reached humans.[1]

1.2 History of COVID-19

The new human coronavirus epidemic 2019 was originally found in Wuhan, China, in 2019, and it also has consequently spread across the globe, being the sixth reported epidemic ever since 1918 influenza epidemic. COVID-19 has caused over 2 billion reported illnesses and 4.6 million deaths globally by September 2021, over 2 years after it had been originally discovered. As per specialists, SARS-Covid-2 is considered to have originated in bats. SARSCoV-2 was discovered in one of Wuhan's open-air "wet markets." Fresh meat and fish, as well as animals killed on the spot, were available for purchase. Whenever a virus encounters major changes, it has the potential to infect and spread among people. Yet that, bats were not sold in the Wuhan market during the epidemic. That's why animals, also known as spiny anteaters, were suspected of being unlawfully traded in Chinese markets at first. SARS-CoV-2 is similar to a number of pangolin-infecting coronaviruses. SARS-CoV-2 infected people that had no personal contact with animals and spread within and without the body, and outside China's borders This indicates that the illness spreads from individual to individual. The coronavirus, that is presently expanding throughout the United States, and around the globe, is being transferred accidentally by people. As a result of the rising international spread, a pandemic has formed.

1.3 Covid-19 statistical Survey

The earliest case of covid-19 was reported on December 31, 2019, and it was formally designated a pandemic in January of 2020. In February, the virus was given the name covid-19, and in March, 1, 00, 000 corona cases were reported in the same month. Since the covid-19 outbreak, Europe has become an epi-center for the virus, and the number of cases registered has increased day by day. In April, there were 1 million cases registered all over the world, and several variants were discovered after the outbreak, the variants are alpha and lambda, and since the total number of cases registered till date is 245 373 039, and the total number of deaths registered is4 979 421, the total number of people who have been cured is 223 962 519.

1.4 Logistic regression algorithm

There in Supervised Learning approach, logistic regression is among very often utilized Algorithms. It is indeed a method to predict a categorized target variable out of a number of distinct factors. The outcome of a continuous attribute is predicted using logistic regression. As a solution, output has to be finite or binary in nature. It can also be 0 or 1,yes or no, true or false, rather than precise numbers such as zero and one, it produces probabilities possibilities mostly in middle. Since it can generate predictions and categorizes extra information using discrete and continuous datasets, logistic regression is indeed a powerful machine learning approach.[2]

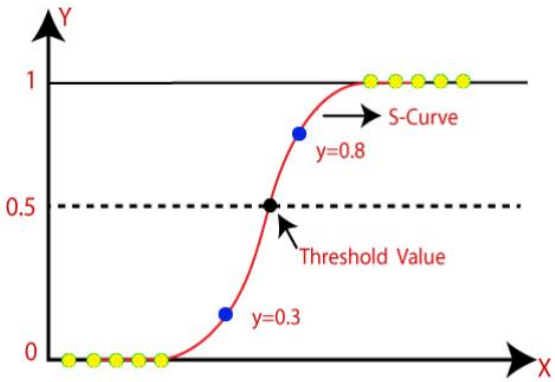

Figure 1: Logistic Regression Algorithm

1.5 Random Forest Classification

Like the name implies, a random forest classifier is made up of a huge number of distinct decision trees that work together like an aggregation. The random forest develops a classification method for each tree, and even the model's prediction is determined by the group by far the most voting. Instead of depending on a decision tree classifier, the random forest gathers recommendations from every tree and guesses the ultimate conclusion dependent on the amount of total voting. "The greater the validity and also the smaller the risk of errors, therefore more trees in the forest".[3]

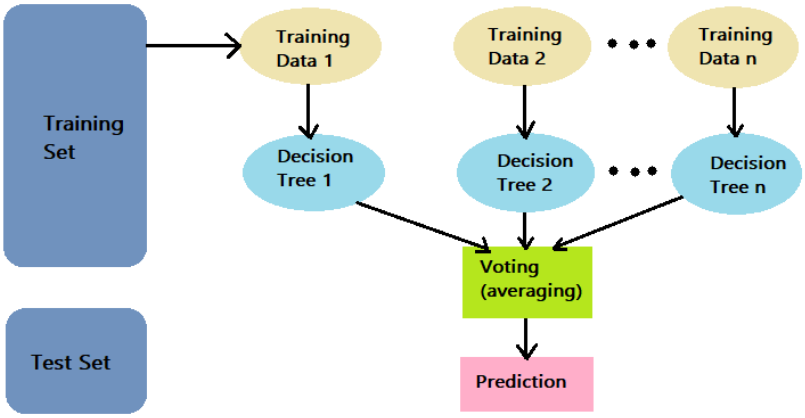

Figure 2: Random Forest Classification

II. Convolutional Neural Network

A well-known approach in computer vision applications is the convolutional neural network, commonly known as convnets or CNN. It's a type of deep neural network that's used to analyze visual data. This architecture is commonly used to detect entities in an image or video. Image or video recognition, neural language processing, and other applications employ it.[4]

Algorithm:

Step 1: Upload the dataset

The first step is to upload the data set. With scikit to learn, you can access the MNIST dataset. Please save it to your Downloads folder. You may use fetch ML-data ('MNIST original') to upload it.

Step 2: Create an input layer
The data gets reshaped in this stage. The square root of the number of pixels determines the form. The form of a photograph of 156 pixels, for example, is 2626. You must indicate whether the image is colored or not. If you answered yes, you had 3 to the shape three for RGB and one if you said no.

Step 3: Add a convolutional layer to your image
The convolutional layers must then be created. To allow the network to learn key features, you apply various filters. You can choose kernel size and the number of filters.

Step 4: a layer for pooling
A pooling layer is added in the third phase. The size of the input is reduced by this layer. It achieves it by taking the sub-greatest matrix's value. If the sub-matrix is [3, 1, 3, 2], for example, the pooling will return the maximum which is 3.

Step 5: Convolutional Layer and Pooling Layer should be added
You can add as many conv levels and pooling layers as you wish in this stage. More than 20 conv layers are used in Google's architecture.

Step 6: a dense layer
Flatten the preceding layers to make them totally linked. You can utilize several activation functions and a dropout effect in this stage.

Step 7: Logit Layer
The prediction is the last step

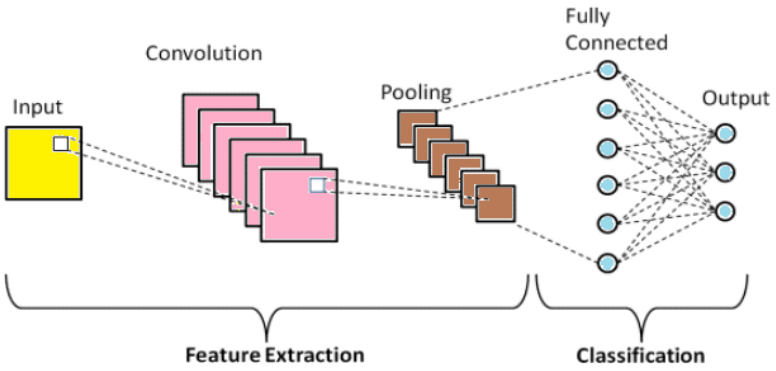

Figure 3: CNN Architecture

III. LITERATURE SURVEY

In the existing system, MOHAMMAD (BEHDAD) JAMSHID explored conceptual structures and platforms in the study field of AI-based approaches that are suited for dealing with COVID-19 concerns. Various algorithms using Corona virus's diagnostic systems, such as RNN, LSTM, GAN, and ELM, have been created. The primary concerns with coronavirus were geographical issues, high-risk persons, and identifying and radiography, all of which have been investigated and described in the author's work.[5]

Nora El-Rashidy S suggested a model that predicts recurrence in both the covid-19 and sears situations, using a hyper feature extraction approach comprising four filters, including the MPEG-7 histogram filter, fuzzy 64, and local binary histogram, to build a hybrid classification technique with 96 percent accuracy. They also used optimization approaches to eliminate false positive situations, and the model was only good for a few days. With such a short dataset, generalization ability can only be guaranteed for 51 photos.[6]

Pathak et al. also utilized the transfer and resnet-50 pre-trained models to create a 2Dclassification for distinguishing sick chest x-rays from normal chest x-rays, and the model performed well. This model has a training accuracy of 96032 percent and a testing accuracy of 93.11 percent; however, it requires a lot of work a significant amount of time to train the data.[7]

A segmentation and classification model for chest x-rays has been presented by Wang et al. There were two essential phases in the process. The first phase is segmentation, which is based on deep learning models such as u-net and 3-net, and so on Second, the categorization is based on a pre-trained model that assessed 732 examples. the system of classification. The accuracy of the model was 99 percent.[8]

Wang et al developed a model which analyzes the changes that occurred in chest images of infected patients and they built a cnn model that utilizes an inception pre-trained model which transfers the learning techniques and will build an effective model for the diagnosis of the patients this model achieved a performance of 88% accuracy and has saved a lot of time the training and testing does not take a lot of time. [9]

Akib khanda have developed a model. they have taken 212 clinical reports and various features are extracted from the clinical reports and 3 machine learning algorithms are used Multinomial naive bayesian, decision tree and stochastic gradient boosting machine learning algorithms are used for classification after the classification of the data result obtained is multinomial naive Bayesian has achieved the accuracy of 93% ,decision tree has achieved the accuracy of 92.5% and stochastic gradient boosting achieved 92% accuracy multinomial has achieved the best accuracy .[10]

Yazeed Zoabi, Shira Deri-Rozov & Noam Shomron established a machine learning approach they have trained a model the model has been tested on 51,831 individuals the model has predicted covid-19 test results with high accuracy using only 8 features like gender,age,fever,body pains, sore throat and the predictions are generated using gradient boosting machine model using decision tree classifiers the accuracy achieved is 90% .[11]

Dandi Yang, Cristhian Martinez, Lara Visuña, Hardev Khandhar, Chintan Bhatt & Jesus Carretero have proposed some framework models which are VGG16,DenseNet121, ResNet50,and ResNet152 the models are designed to find out the best architecture, pre-processing and training parameters for achieving the best accuracy the accuracy achieved is 93% for the diagnosis of covid-19 using chest x-rays.[12]

IV. OUTLINE OF PROPOSED MODEL

1. Prediction of Coronavirus based on basic symptoms using supervised Machine Learning Algorithms which are Logistic Regression and Random Forest Classification.

Algorithm:

Step1: Import the dataset.
Step2: Process the dataset.
Step3: Create and train the model.
Setp4: Perform logistic regression and random forest algorithms.
Step5: Prediction

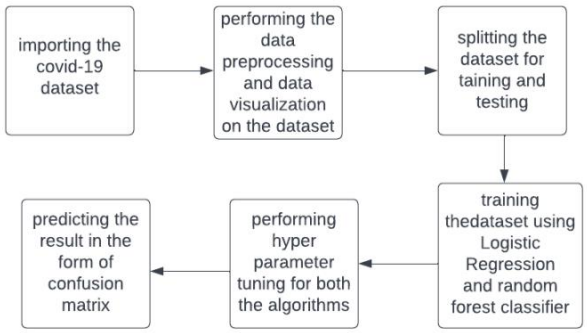

Figure 4: Block diagram of coronavirus prediction using symptoms

2. Detection of Covid-19 by applying Convolutional Neural Network on Chest X-ray image Dataset.

Algorithm:

Step1: Collect and store the data.
Step2: Data pre-processing.
Step3: Building and utilization of CNN model.
Step 4: Make the model to learn from the different input features
Step 5: Predict the results

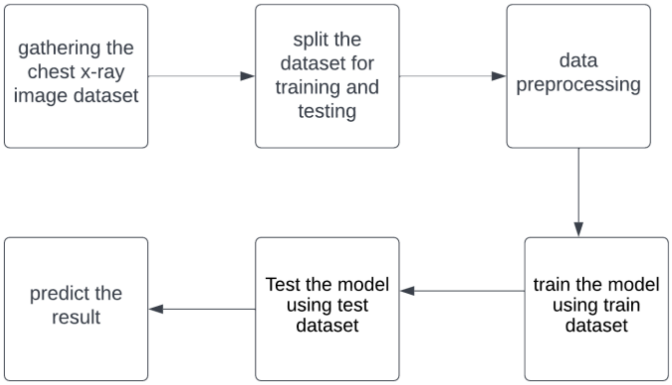

Figure 5: Block diagram of coronavirus detection using Chest x-ray images

V. EXPERIMENTAL SET-UP

Python:
In our proposed system we used Python programming language for implementing our code and importing the required libraries.
Python libraries utilized:
Pandas:
Pandas is a popular open Python module for computer science, analysis techniques, and machine learning activities. pandas is imported to support multi-dimensional arrays.
NumPy:
NumPy is used for converting the 2D images or text in the form of arrays for prediction.
Matplotlib:
Matplotlib is used for data visualization such as plotting the graphs, pie charts etc for the data.
SK-learn:
Sk-learn is very useful and robust package for machine learning in Python. It is used for importing the algorithms such as logistic regression and random forest classification which are used for predicting the symptom based dataset.
Keras:
Keras is a Python-based deep learning API that extend the functionality of TensorFlow, a machine learning environment.
Tensor flow:
TensorFlow is an open-source end-to-end machine learning platform. It is utilized for performing image classification in cnn.
OS:
In Python, the OS package has methods for dealing with the windows OS.

VI. METHODOLOGY

The text-based dataset consists of a total of 2500 entries, and the image-based dataset with 208 images that includes both covid and non-covid.

The following are the steps used to predict covid using both text-based and image-based inputs:

1. **Prediction of Covid-19 based on basic symptoms using supervised Machine Learning algorithms which are Logistic Regression and RandomForest Classification.**

   Step 1: For creating a machine learning model, a textual dataset which consists of the information related to covid patients in terms of country, age, gender, and symptoms like fever, body pain, breathing difficulty, runny nose, nasal congestion, severity, sore throat, contact with covid patient and infected patients must be imported.
   Step 2: After importing the dataset, preprocessing on the dataset is performed to make raw dataset into a clean dataset without any errors.

```
df = pd.read_csv('covidreports.csv')
df.head()
```

	Country	Age	Gender	fever	Bodypain	Runny_nose	Difficulty_in_breathing	Nasal_congestion	Sore_throat	Severity	Contact_with_covid_patient	Infect
0	China	10	Male	102	1	0	0	0	1	Mild	No	
1	Italy	20	Male	103	1	1	0	0	0	Moderate	Not known	
2	Iran	55	Transgender	99	0	0	0	1	1	Severe	No	
3	Republic of Korean	37	Female	100	0	1	1	0	0	Mild	Yes	
4	France	45	Male	101	1	1	1	1	0	Moderate	Yes	

Table 1: Sample text-based dataset

Step 3: After preprocessing of the data, a model is created and trained using the dataset. 70% -30% split is used for training and testing the data.
Step 4: logistic regression and random forest algorithms are used for categorizing the data in terms of infected and disinfected patients. Whenever the accuracy of logistic regression and random forest algorithms are same, hyper parameter tuning can be done for better results.

a) logistic regression being an algorithm that could do binary classification helps in predicting the infected and disinfected covid patients.

b) Random forest being an algorithm that works on decision trees helps in categorizing the common decision trees that helps in predicting the infected and disinfected covid patients.

Step 6: After fitting and checking scores of the model, checking for the precision,recall,f1 score and accuracy of the model is done, and Parameters of the model are changed which best fits the algorithm with best accuracy score.
Step 5: Prediction of the infected and disinfected covid patients using Confusion matrix.

An accuracy of 94.6 for random forest and 93.3 for logistic regression and now predicting the confusion matrix and now it has predicted the and we are displaying it in the form of confusion matrix.

Final Evaluation

- Final evaluation is on the Random Forest Classifier
  - Accuracy = 94.6%
- where as Logistic Regression
  - Accuracy = 93.3%

```
np.random.seed(45)
clf = RandomForestClassifier(max_depth = None, min_samples_leaf = 1,
                             min_samples_split = 2 , n_estimators = 110)
clf.fit(X_train , y_train)
pred = clf.predict(X_test)
```

```
print(confusion_matrix(pred , y_test))
```

```
[[318  20]
 [ 29 383]]
```

```
sns.heatmap(confusion_matrix(pred , y_test) , annot=True , fmt='g' )
plt.xlabel('True Values')
plt.ylabel('Predicted Values');
```

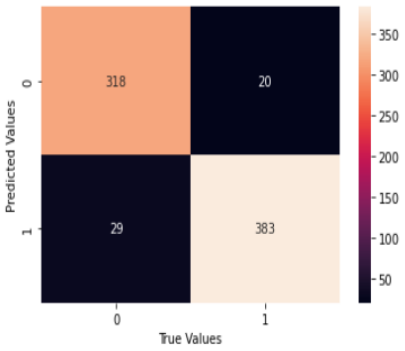

Figure 6: Heat map

Below shown is a classification report of the model which tells the accuracy, precision, f1 score and recall score.

```
print(classification_report(pred , y_test))
```

	precision	recall	f1-score	support
0	0.92	0.94	0.93	338
1	0.95	0.93	0.94	412
accuracy			0.93	750
macro avg	0.93	0.94	0.93	750
weighted avg	0.94	0.93	0.93	750

Figure 7: Classification Report

Testing:

Testcase1:

- Testcase objective: To achieve accuracy for logistic regression model
- Expected output: To predict the coronavirus infected and disinfected persons
- Obtained Output: we have obtained the prediction of coronavirus disinfected and Infected persons
- Test Result: pass

Testcase 2:

- Testcase objective: To achieve accuracy for Random Forest Classifier model.
- Expected Output: To predict the Coronavirus infected and disinfected persons.
- Obtained Output: we have obtained the prediction of Coronavirus disinfected and Infected persons
- Test Result: pass

2. Detection of Coronavirus by applying Convolutional Neural Network on Chest X-ray image Dataset.

Step 1: Collect and store the Data:

Collection of datasets include Non COVID person CXR Scan images, COVID patient CXR images from Kaggle. The number of train images are 104 and the number of test images are 104. The size of images in the image datasets are 256*256.

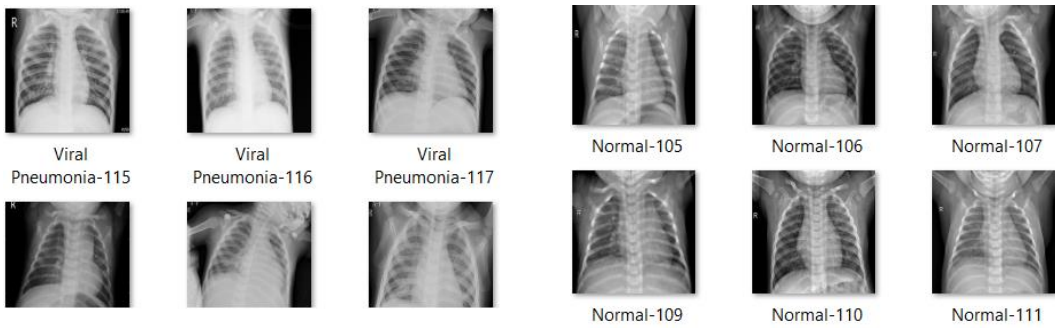

Figure 8: Sample image-based dataset

Step 2: Data pre-processing:

The images were collected from Kaggle so they Comparison of the COVID-19 and normal X-ray images they might need preprocessing of images as they are raw data. First, the images were resized to 64 × 64 pixels to make them ready to be given as input to the system and are checked in the aspects of rotation reshaping and scaling.

Step 3: Building and utilization of CNN model:

A Convolutional Neural Network model has been created and trained using the datasets. CNN extracts the most relevant features from the image datasets and classifies them into coronavirus and normal images.

Step 4: Make the model to learn from the different input features:

This step entails feeding the learning algorithm with all of the training data. Then the learning algorithm discovers the mapping between the inputs and the output class label that reduces the prediction error in the project. The model is fitted to the training dataset and so that it gives the required accuracy a after the training.

Step 5: Predict the results:

The CNN classification results in prediction of covid images to identify whether the patient has covid or not.

Testing:

Testcase 1:

- Testcase objective: To detect an image with Coronavirus from the dataset
- Expected output: The CNN deep learning model is tested with a coronavirus X-ray Chest image and the model should predict the image as Covid.
- Observed output: Covid-19 is successfully predicted.
- Pass/fail: pass

```
[8] import numpy as np
    from keras.preprocessing import image
    test_image = image.load_img('/content/drive/MyDrive/covid/test/covid/Viral Pneumonia-111.png', target_size = (64, 64))
    test_image
```

Figure 9: Testcase 1 image

```
[12] if result[0][0] == 0:
         prediction = 'Covid'
     else:
         prediction = 'None'
     print("Detected is %s"%prediction)

     Detected is Covid
```

Figure 10: Testcase 1 output

Testcase 2:

- Test objective: To classify between chest x-ray images which are having covidand which are not having covid
- Expected output: Image classification of Coronavirus and normal images.
- Observed output: The CNN model is tested with a normal Chest x-ray image the model is able to successfully differentiate between a covid Chest X-ray image and a normal chest x-ray image.
- Test Result: pass

```
[13] import numpy as np
     from keras.preprocessing import image
     test_image = image.load_img('/content/drive/MyDrive/covid/test/none/Normal-101.png', target_size = (64, 64))
     test_image
```

Figure 11: Testcase 2 image

```
if result[0][0] == 0:
    prediction = 'Covid'
else:
    prediction = 'None'
print("Detected is %s"%prediction)

Detected is None
```

Figure 12: Testcase 2 output

VII. RESULT AND DISCUSSION

1. Prediction of Coronavirus based on basic symptoms using supervised Machine Learning Algorithms which are Logistic Regression and Random Forest Classification.

	Logistic Regression	**Random Forest Classification**
Precision	0.962779	0.950372
Recall	0.910796	0.929612
F1 Score	0.936068	0.939877
Accuracy	0.929333	0.934667

Table 2: Prediction for text-based input

Through random forest Classification we are able to obtain better results than LogisticRegression.

2. Detection of Coronavirus by Convolutional Neural Network on Chest X-ray imageDataset.

No. of epochs	**Train Accuracy**	**Test Accuracy**
50	88	94
60	92	95
80	95	92
90	92	96
100	95	94

Table 3: Accuracy and Loss Prediction for image-based input

At epoch=100 we are able to achieve an accuracy of 95% for the trained model and a test accuracy of 94%.

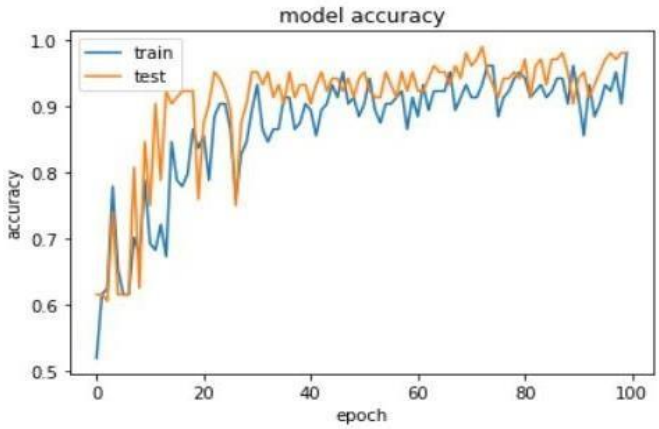

Figure 13: epoch Vs accuracy

VIII. CONCLUSION

In this paper we have presented an approach for COVID-19 image classification.

1. We have taken a textual dataset consisting of covid-19 symptoms like fever, cough, body pain etc. and predicted the covid-19 infected and disinfected persons using logistic regression and random forest classifier. we observed that the random forest classification algorithm provides moreaccurate results in comparison with Logistic Regression.
2. We have taken a chest x-ray image dataset containing of coronavirus and normal images and trained a CNN model that could accurately predict a coronavirus affected chest x-ray.

IX. REFERENCES

[1] Feng He,Yu Deng, Weina Li, "Coronavirus disease 2019: What we know?" .14th March 2020,7pages

[2] Schober, Patrick MD, "Logistic Regression in Medical Research", Anesthesia and analgesiaJournal, February 2021,volume 132 ,Issue 2.

[3] Jehad Ali, Rehanullah Khan, Nasir Ahmad, Imran Maqsood, "Random Forests and DecisionTrees", Research Gate, September 2012 ,Vol. 9, Issue 5, No 3.

[4] Keiron O'Shea, Ryan Nash "An Introduction to Convolutional Neural network" arxiv.org,2 Dec2015,10pages.

[5] Mohammad Jamshidi, "Artificial Intelligence and COVID-19: Deep Learning Approaches for Diagnosis and Treatment", IEEE Access, Vol. 8.

[6] S. A. T. A. 3, *ORCID, E. A. F. A. J.-W. H. andShaker E.-S. Nora El-Rashidy, "Comprehensive Survey of Using Machine Learning in the COVID-19 Pandemic," *MDPI*, 2021.

[7] Y. Y. F. M. X. ZhengWang, "Automatically discriminating and localizing COVID-19 from community-acquired pneumonia on chest X-rays," *ScienceDirect*, vol. 110, 2021.

[8] di Dong; Zhenchao Tang; Shuo Wang; Hui Hui; Lixin Gong; Yao Lu, "The Role of Imaging in the Detection and Management of COVID-19: A Review," *IEEE Xplore logo*, vol. 14.

[9] P. K. S. A. T. S. S. S. S. P. K. S. Y.Pathaka, "Deep Transfer Learning Based Classification Model for COVID-19 Disease," *ScienceDirect*, vol. 43, no. 2, pp. 87–92, Apr. 2022.

[10] S. T. R. Q. R. K. N. R. & M. M. U. D. Akib Mohi Ud Din Khanday, "Machine learning based approaches for detecting COVID-19 using clinical text data," *SpringerLink* , 2020.

[11] S. D.-R. & N. S. Yazeed Zoabi, "Machine learning-based prediction of COVID-19 diagnosis based on symptoms," *npj Digital Medicine*, 2021.

[12] C. M. L. V. H. K. C. B. & J. C. Dandi Yang, "Detection and analysis of COVID-19 in medical images using deep learning techniques," *Scientific Reports*, Oct. 2021.

[13] Vladimir Nasteski, "An overview of the supervised Machine Learning Methods" researchgate,December 2017.