International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-11 Issue-2, December 2021

Machine Learning Approach using CNN for COVID-19 Pandemic Detection

Vikas S, Thimmaraju S N

Abstract: Data science and machine learning are domain names in which data generation can assist with inside the fight towards the disease. Early caution systems which can are expecting how much a disease might effect society and permit the authorities to take suitable measures without disrupting the economy are extremely important. In the confrontation towards COVID-19 methods for forecasting the future cases primarily based totally on present data are extremely beneficial. The preceding are three strategies of machine learning which are discussed: Two for predicting the wide variety of positive cases in the coming ten days, and one for identifying COVID-19 infection via way of means of analyzing the patient's chest x-ray image. Various algorithms had been tested, and the only that produced the maximum accurate consequences become selected for use on this take a look at to forecast confirmed cases in India. Various government entities can leverage the findings to take corrective action. Now that methods for forecasting infectious disease are available, COVID-19 can be less complicated to combat.

Keywords*: *COVID-19 Infection Via Way of Means of Analyzing the Patient's Chest X-Ray Image.

I. INTRODUCTION

Coronavirus disease 2019 (COVID-19) is a new disease as a result of the Severe Acute Respiratory Syndrome Coronavirus 2 virus, which become recently discovered (SARS-CoV-2). On March 11th 2020, the World Health Organization proclaimed the specific disease, which first seemed Wuhan, China in December 2019, to be a pandemic [1]. COVID- 19 outbreaks now no longer best endangered people's lives, however additionally they have a significant monetary effect at the country. On January 30, 2020, the World Health Organization declared it a worldwide health emergency. (WHO) By July 11, 2021, the virus had infected more than 30 million Indians[2]. SARS-CoV-2-related illness was labelled COVID-19 by the WHO, which stands for "coronavirus disease 2019." The name was chosen to avoid glamorising the virus's origins in terms of demography, geography, or animal associations [3]. (By July 11, 2021, the virus had infected greater than 30 million Indians[2]. SARS-CoV-2-associated contamination turned into labelled COVID-19 by the WHO, which stands for "coronavirus disease 2019." The name was selected to keep away from glamorising the virus's origins in phrases of demography, geography, or animal associations [3].

Here we're presently in a completely unique situation wherein new information concerning the virus and infection is being published on a day by day basis through the scientific community. It is important to file and describe cases inflamed with SARS-CoV-2 and the direction in their sicknesses so one can achieve important information on the sickness, its characteristics in numerous groups of patients, and the natural history of COVID-19. It is essential to capture context-specific statistics in patients with tuberculosis or malnutrition as a comorbidity as it will influence medical decision-making and policy updates in India [5].

Data science and machine learning are domain names in which information generation can assist with inside the fight against the disease. Early warning systems that could are expecting how much a disease could impact society and permit the government to take appropriate measures without disrupting the financial system are extremely important. In the war of words against COVID-19, methods for forecasting the future cases based on current facts are extraordinarily beneficial[8]. The preceding are three strategies of machine learning which are discussed: Two for predicting the number of fantastic instances withinside the coming ten days, and one for figuring out COVID-19 infection through analyzing the patient's chest x-ray image. Various algorithms had been tested, and the only that produced the maximum accurate consequences become selected for use on this take a look at to forecast confirmed cases in India. Various government entities can leverage the findings to take corrective action. Now that methods for forecasting infectious disease are available, COVID-19 can be less complicated to combat[13].

A.Scope of the Work

• The primary purpose is to forecasting the number of positive cases in the coming ten days • Detecting COVID-19 infection through inspecting the patient's chest x-ray image through the use of CNN algorithm and the main objectives are 1) Gathering the covid-19 cases dataset and the covid-19 patients x-ray images. 2) Processing the accumulated data for the as requirement desires. 3) Train the pre-processed data through the use of machine learning models. 4) Examine the obtained outcomes for its accuracy.

Manuscript received on December 09, 2021.
Revised Manuscript received on December 13, 2021.
Manuscript published on December 30, 2021.
* Correspondence Author
Dr. Vikas S*, Assistant Professor, Department of Computer Science and Engineering, Visvesvaraya Technological University, Mysuru (Karnataka), India. E- mail. vikas.smg@gmail.com
Dr. Thimmaraju S N, Assistant Professor, Department of Computer Science and Engineering, Visvesvaraya Technological University, Mysuru (Karnataka), India. E- mail. thimmaraju_sn@yahoo.com

II. LITERATURE SURVEY

A country wide lockdown become applied in addition to the prohibition of all modes of locomotion which include airports, railways, neighborhood transportation, marketplaces, malls, theatres and cinemas, to preserve COVID-19 from spreading to parts of the country that are not affected. Patients who're suspected of being threatening also are remoted or quarantined. Despite the reality that the entire federal bureaucracy is absolutely engaged in preventing its spread withinside the communities, COVID-19 positive cases are growing each day[15]. India has more or less 31 million confirmed cases with 418,987 deaths and 30,429,339 recoveries on the of this writing and the numbers are usually attempting to climb. For the cause of to forecast the growth of COVID-19 cases withinside the country, diverse statistical and mathematical simulations and studies are being undertaken and Technologies were proposed to estimate the future situation, that is known as N-days in forecasting. Many machine learning strategies can be effective in building a reliable forecasting version because of the nonlinear nature of COVID-19 data. According to studies, the cornerstone to design a prediction version is the employment of computational intelligence technologies. The neural network is efficient for forecasting data due to its capability to study from data. The pandemic inflamed hundreds of thousands of people across the world, in line with information uploaded and updated through health organisations and the governments across the world[8]. The maximum risky infection prompted through COVID-19 is pneumonia which destroys the lungs. Fever, runny nose and cough are only some of the signs and symptoms. A chest X-ray imaging look at for abnormalities is the maximum common method for diagnosing those conditions [8]. X-ray imaging can be a better opportunity for a mass, easy and rapid pandemic diagnosis tool, given the prevailing worldwide healthcare infrastructure issue.X-ray imaging has a range of advantages over traditional testing procedures as a COVID-19 alternative diagnosis procedure. Only many of the advantages are Low cost, X-ray equipment is typically available, reduced time consumption and Equipment affordability[17].

III. MATERIALS AND METHODS

A basic dataset of 228 X-ray pictures was used as a base dataset in this study's studies. 114 X-ray photos belonged to proven COVID-19 patients, whereas the remaining 114 Picture belonged to healthy people or illnesses of other People's such as pneumonia. The dataset utilised in this study is available on GitHub [13]. There are two COVID-19 classes in the core dataset, one with 114 samples and the other with 114. As a result, the dataset was uneven, necessitating preprocessing in order to obtain encouraging results. In a previous attempt, here we trained the CNN on the provided original information dataset and achieved an accuracy of 74.3 percent, which was insufficient for the current application area.

1) Dataset Preprocessing

To balance the dataset and in order to improve the performance of the CNN Technique in detecting COVID-19 patients, 1114 normal chest X-ray Pictures were used. These concatenated additional X-ray Pictures are downloaded using Kaggle [21]. After balancing the obtained dataset and training the methodology again on the created dataset, the accuracy of the given CNN Technique was enhanced by upto 94.53 percent

2) Data Augmentation

Data augmentation is a technique for dramatically expanding the quantity of data instances in a dataset so that it may be utilised to train a machine learning model [23]. The technique enhances photo collections with Flipping, rotating, cropping, and padding are examples of basic image processing techniques. These modified photographs from the original image set are then added to the dataset, making it larger for neural network training [21]. The data augmentation approach was used in this study to address the issue of a limited dataset, which hampered the performance of the proposed CNN. This method increased the dataset size while also supplying more learning characteristics to the learning model. This study used two image processing techniques for data augmentation: flipping and rotation. In the first round of information augmentation, the 90 X-ray Pictures were inverted to produce an additional 90 Pictures. The generated dataset was enlarged to 180 images using this method. The original 90 shots were rotated by 90 degrees to produce 90 new images, then 180 degrees to produce 90 more images, and finally 270 degrees to produce even more 90 images in the second step.

3) Convolutional Neural Networks (CNNs)

The visual system of the human brain inspired CNNs. CNNs aim to give computers the ability to observe the world in the same way that humans do. CNNs can assist with picture recognition and analysis, image classification, and natural language processing in this fashion [20]. CNNs have convolutional, max pooling, and nonlinear activation layers, and are a type of deep neural network. The "convolution" operation that gives the network its name is performed by the convolutional layer, which is considered a CNN's core layer. The convolutional layer's kernels are applied to the layer inputs. All of the convolutional layers' outputs are combined into a feature map. The Rectified Linear Unit (ReLU) was employed in this work with a convolutional layer in the activation function, which helps to enhance the nonlinearity in the input image, which is significant because images are nonlinear in nature. As a result, in the current setting, CNN with ReLU is easier and faster. For all negative inputs, the ReLU is zero, hence it can be defined as

$$z = \max(0, i). \quad \text{Eq (1)}$$

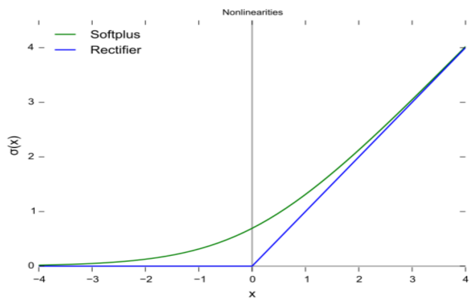


Fig 1: Relu activation graph

Retrieval Number: 100.1/ijeat.B33021211221
DOI: 10.35940/ijeat.B3302.1211221
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The function suggests that for all negative values, the output z is zero, whereas the positive value remains constant illustrated in the Figure 1.

Max pooling was utilised in this experiment since others may not be able to identify the acute traits as quickly as others. Because this study involves the training of a very deep neural network, the batch normalisation layer was also utilised. To normalise the input layer and speed up the learning procedure between hidden units, the approach alters the scale and activation. By removing neurons at random during training, the dropout layer, which has a 20% dropout rate, has also been used to minimise the overfitting problem. There is a flatting layer near the end of the CNN utilised in the study that converts the output of convolutional layers into a single-dimensional feature vector. To put it another way, the flattening layer takes all of the pixel data from the convolutional layers and mixes it into a single vector. The vector input is then passed into the CNN's next levels, known as fully connected layers or dense layers, after flattening. Every neuron in the layer above it is connected to every neuron in the layer before it. Dense layers' principal role is to take the flattened output results of the convolution and pooling layers as input and categorise the image using a given class label. The likelihood of a feature belonging to a certain class is represented by each integer in the flattened feature set. As a result, the fully connected network with deep layers eventually drives the classification decision.

IV. CNN'S PROPOSED ARCHITECTURE

The suggested CNN model has 38 layers, including 6 convolutional, 6 max pooling, 6 dropout, 8 activation function, 8 batch normalisation, 1 flatten layer, and 3 fully connected layers; the input picture shape of the CNN model is a 150-by-150 RGB image. The filter size increased after every two Con2D layers. 64 filters were used to learn from input in the first and second levels of Con2D, 128 filters in the third and fourth layers, and 256 filters in the fifth and sixth layers. The output of the final Con2D layer's 256 output neurons is followed by max pooling, batch normalisation, activation, and the dropout layer. Because the last pooling and convolutional layer generates a three-dimensional matrix, a flattening layer was employed to turn the matrix into a vector that could be used as input for three dense layers[18].

The CNN was employed with varied configurations in terms of the number of convolution layers in the model throughout the early trials. An incremental method was used to determine the number of convolution layers in the model. First, only one convolutional layer was used to test the CNN, and the results were analysed[14]. After that, a CNN with two layers was developed, and the findings were analysed, and so on. The method had been used until the model's output was accurate and effective demonstrated in Figure 2. Six convolution layers made up the final model, which was extremely viable based on the results.

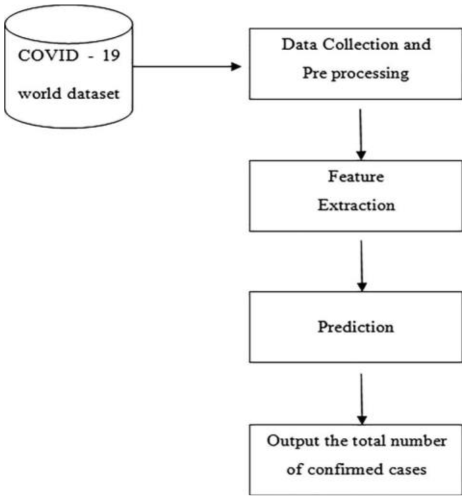


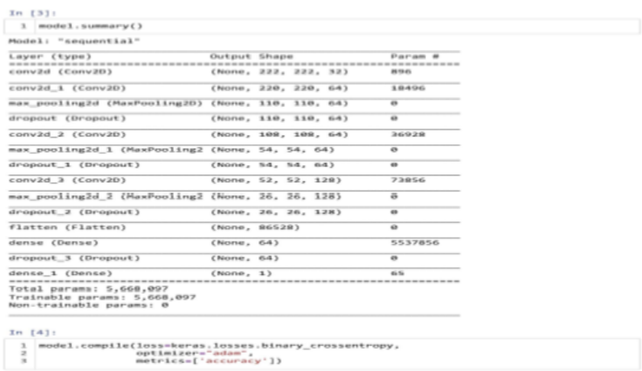
Fig 2: DFD for Covid-19 prediction

V. RESULT AND DISCUSSION

The accuracy measurements of Convolutional Neural Networks are discussed in this section (CNNs). We successfully trained and tested the model illustrated in Figure 3 as stated in the prior section.

```
In [5]:
#Data preprocessing

train_datagen=image.ImageDataGenerator(rescale=1./255,
        horizontal_flip=True, zoom_range=0.2,shear_range=0.2)

train_data=train_datagen.flow_from_directory(
    directory="C:\\Users\\hp\\Desktop\\Project\\CovidDataset\\Train",
    target_size=(224,224),batch_size=16,class_mode='binary')

train_data.class_indices

Found 224 images belonging to 2 classes.
Out[5]:
{'Covid': 0, 'Normal': 1}

In [6]:
test_datagen=image.ImageDataGenerator(rescale=1./255)
test_data=test_datagen.flow_from_directory(
    directory="C:\\Users\\hp\\Desktop\\Project\\CovidDataset\\Val",
        target_size=(224,224),batch_size=16,class_mode='binary')

test_data.class_indices

Found 60 images belonging to 2 classes.
Out[6]:
{'Covid': 0, 'Normal': 1}
```

Fig 3: Pre-processing the collected X-Ray image dataset according to the needs for CNN Model



```
In [3]:
model.summary()
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 222, 222, 32)      896
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 220, 220, 64)      18496
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 110, 110, 64)      0
_________________________________________________________________
dropout (Dropout)            (None, 110, 110, 64)      0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 108, 108, 64)      36928
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 54, 54, 64)        0
_________________________________________________________________
dropout_1 (Dropout)          (None, 54, 54, 64)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 52, 52, 128)       73856
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 26, 26, 128)       0
_________________________________________________________________
dropout_2 (Dropout)          (None, 26, 26, 128)       0
_________________________________________________________________
flatten (Flatten)            (None, 86528)             0
_________________________________________________________________
dense (Dense)                (None, 64)                5537856
_________________________________________________________________
dropout_3 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 65
=================================================================
Total params: 5,668,097
Trainable params: 5,668,097
Non-trainable params: 0
_________________________________________________________________

In [4]:
model.compile(loss=keras.losses.binary_crossentropy,
              optimizer="adam",
              metrics=['accuracy'])
```

Fig 4: Training the CNN Model with the pre-processed X-Ray dataset

Retrieval Number: 100.1/ijeat.B33021211221
DOI: 10.35940/ijeat.B3302.1211221
Journal Website: www.ijeat.org

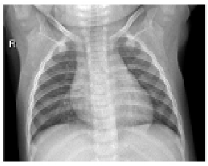
Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

We achieved a validation accuracy of 94.53 percent after validation/testing, which enhanced the validation accuracy from the beginning to the completion of the training shown in Figure 4. The percentage of people who lost weight dropped over time demonstrated in Fig 5 and 6.

In [11]:

```
#Manually testing working of CNN model by giving Normal x-ray image
path="C:\\Users\\hp\\Desktop\\Project\\CovidDataset\\Train\\Normal\\IM-0182-0001.jpeg"
img=image.load_img(path,target_size=(224,224))
img
```

Out[11]:



In [12]:

```
img=image.img_to_array(img)/255
img=np.array([img])
model.predict_classes(img)
```

Out[12]:

```
array([[1]])
```

Fig 5: Manually testing the working of CNN Model by giving COVID-19 negative X-Ray dataset.

In [14]:

```
#Manually testing working of CNN model by giving Covid x-ray image
path2="C:\\Users\\hp\\Downloads\\covid_cxnet_dataset\\Dataset\\Covid\\290.jpg"
img=image.load_img(path2,target_size=(224,224))
img
```

Out[14]:

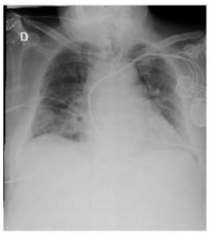


Fig 6: Manually testing the working of CNN Model by giving COVID-19 positive X-Ray dataset.

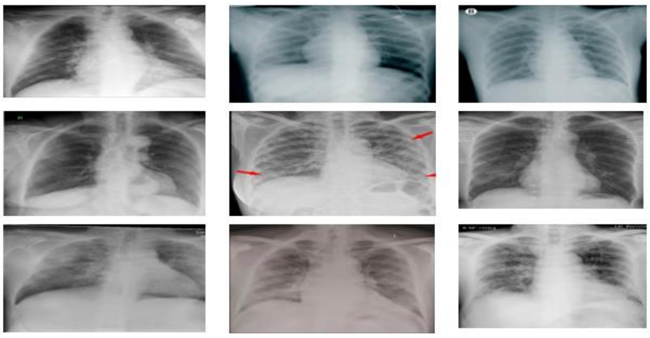


Fig 7: COVID X-Rays

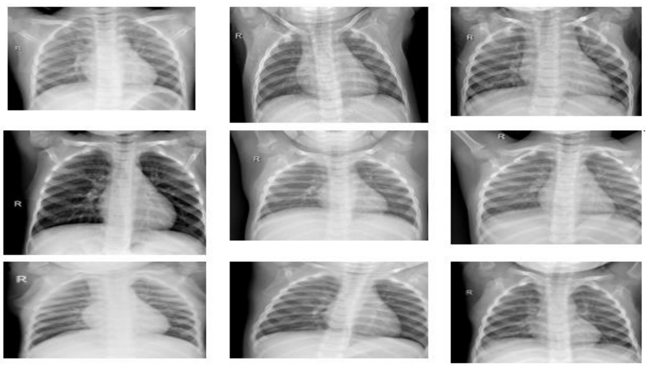


Fig 8: Normal X-Rays

The training yields two indices: 'COVID':0 and 'NORMAL':1. Here we test the model for accuracy metrics after it has been trained. We constructed a function validation generator for testing that has the same input image size and batch size as the train generator. For testing, we used stepsperepoch=12 result is presented in Fig 7 and 8(the number of epochs in each iteration is referred to as steps perepoch), epoch=15, and validation steps=3.

VI. CONCLUSION

Using 114 COVID-19 positive and 114 COVID-19 negative images for training and 30 COVID-19 positive and 30 COVID-19 negative images for testing, we achieved a 94.53% curacy rate. CNN approach may be utilised in India to deal with a radiology shortage. Other chest-related disorders, together with tuberculosis and pneumonia, also can be identified the usage of such models. The study's deployment of a limited range of COVID-19 X-ray scans is a limitation. To enhance the firmness and reliability of our model, we purpose to incorporate additional similar images from local hospitals. The accuracy we had been able get is about 94.53%.

REFERENCES

1. https://www.icmr.gov.in/tab1ar1.html
2. BMJBest Practice, 2020. Overview of Coronaviruses. https://bestpractice.bmj.com/topics/en-us/3000165
3. Bae S, Kim MC, Kim JY, Cha HH, Lim JS, Jung J, et al. Effectiveness of Surgical and Cotton Masks in Blocking SARS-CoV-2: A Controlled Comparison in 4 Patients. Ann Intern Med. 2020 Apr 6. [Medline].
4. Deepak Painuli, Divya Mishra, Suyash Bhardwaj, and Mayank Aggarwal, "Forecast and prediction of COVID-19 using machine learning" https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8138040/#bib4
5. World Health Organization. Coronavirus disease 2019 (COVID-19) Situation Report –World Health Organization. Available at https://www.who.int/docs/defaultsource/ coronaviruse/situation-reports/20200308-sitrep-48-covid-19.pdf?sfvrsn=16f7ccef_4. March 8, 2020; Accessed: March 9, 2020
6. Aijaz Ahmad Reshi,Furqan Rustam,Arif Mehmood, "An Efficient CNN Model for COVID-19 Disease Detection Based on X-Ray Image Classification" https://www.hindawi.com/journals/complexity/2021/6621607/
7. I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," Physical and Engineering Sciences in Medicine, vol. 43, pp. 635–640, 2020.
8. L. Wang and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," 2020, http://arxiv.org/abs/ 2003.09871.
9. A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network," 2020, http://arxiv.org/abs/ 2003.13815.
10. D. J. Cennimo, "Coronavirus disease 2019 (COVID-19) clinical presentation," vol. 8, pp. 101489–101499, 2020, https:// emedicine.medscape.com/article/2500114-clinical#b2, 2020.
11. Z. H. Chen, "Mask-RCNN detection of COVID-19 pneumonia symptoms by employing stacked autoencoders in deep unsupervised learning on low-dose high resolution CT," IEEE Dataport, 2020.
12. M. Ahmad, "Ground truth labeling and samples selection for hyperspectral image classification," Optik, vol. 230, Article ID 166267, 2021.
13. E. B. Fatima, B. Omar, E. M. Abdelmajid, F. Rustam, A. Mehmood, and G. S. Choi, "Minimizing the overlapping degree to improve class-imbalanced learning under sparse feature selection: application to fraud detection," IEEE Access , vol. 9, pp. 28101–28110, 2021.

Retrieval Number: 100.1/ijeat.B33021211221
DOI: 10.35940/ijeat.B3302.1211221
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

14. CDC. 2019 Novel Coronavirus, Wuhan, China: 2019 Novel Coronavirus (2019-nCoV) in the U.S. Centers for Disease Control and Prevention (CDC). Available at https://www.cdc.gov/coronavirus/2019-ncov/cases-in-us.html. March 18, 2020; Accessed: March 19, 2020.
15. Feuer W, Higgins-Dunn N, Lovelace B. US now has more coronavirus cases than either China or Italy. CNBC. Available at https://www.cnbc.com/2020/03/26/usa-now-hasmore-coronavirus-cases-than-either-china-or-italy.html. March 26, 2020; Accessed: March 27, 2020.
16. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72□314 Cases From the Chinese Center for Disease Control and Prevention. JAMA. 2020 Feb 24.
17. Dong Y, Mo X, Hu Y, Qi X, Jiang F, Jiang Z, et al. Epidemiological Characteristics of 2143 Pediatric Patients With 2019 Coronavirus Disease in China. Pediatrics. 2020 Mar 16.
18. CDC. 2019 Novel Coronavirus, Wuhan, China: Interim Healthcare Infection Prevention and Control Recommendations for Patients Under Investigation for 2019 Novel Coronavirus. CDC. Available at https://www.cdc.gov/coronavirus/2019-ncov/infectioncontrol. html. January 18, 2020; Accessed: January 27, 2020.
19. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors formortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020 Mar 11.
20. Funk S, Camacho A, Kucharski AJ, Eggo RM, Edmunds WJ. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. Epidemics 2018; 22:56–61.
21. Yamana, T.K., Kandula, S. & Shaman, J. (2017). Individual versus super ensemble forecasts of seasonal influenza outbreaks in the United States. PLoS Comput Biol 13: e1005801.
22. Funk S, Camacho A, Kucharski AJ, Eggo RM, Edmunds WJ. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. Epidemics 2018; 22:56–61.
23. Tuite AR, Fisman DN. Reporting, epidemic growth, and reproduction numbers for the 2019 novel coronavirus (2019-nCoV) epidemic. Ann Intern Med. 2020 Feb 5. doi: 10.7326/M20- 0358.

AUTHORS PROFILE

Dr.Vikas S. received Ph.D. degree in Computer Science in the year 2021 from Visvesvaraya Technological University and Master of computer Applicaions (MCA) in the year 2007 from Visvesvaraya Technological University and Bachelors Degree in Computer Science in the year 2004 from kuvempu University. He is currently working as Assistant Professor in the Department of MCA, Visvesvaraya Technological University, PG Center, Mysore, Karnataka, where he is involved in research and teaching activities. He is having 14 years of teaching experience and 02 years of Industrial experience. He is a Life member of India Society for Technical Education (LMISTE), Computer Society of India (CSI) and Doing Research in the Area Big data Analytics.

Dr. Thimmaraju S N, he is presently a professor and heading the Department of Computer Science and Engg, and Regional Director for Mysuru centre, Visvesvaraya Technological University, Mysore, Karnataka, he has received his Ph.D degree from Visvesvaraya Technological University(VTU), Belgaum in the year 2010, M.E., degree in Computer Science and Engineering from University Visvesvaraya College of Engineering (UVCE), Bangalore in 2002 and Bachelors Degree in Computer Science and Engineering from PESCE, Mandya in the year 1999. He is involved in research and teaching activities. His major areas of research are Computer Networks, WSN's and Graph theory. He is having 19 years of teaching experience. He has published around 18 research papers which include International Journals, International Conferences and Notional Conferences and Two Research Scholars are Awarded under his guidance and still more has have to be Awarded.

Retrieval Number: 100.1/ijeat.B33021211221
DOI: 10.35940/ijeat.B3302.1211221
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)