Copyright Transfer Agreement

The transfer of copyright for the Work from author to publisher must be clearly stated to enable the publisher to assure maximum dissemination of the work. Therefore, the following agreement (hereafter known as the "Agreement"), executed and signed by the author, is required with each manuscript submission.

Emerald Publishing Ltd, subsequently referred to as "Emerald", recommends that you keep a copy of this completed form for reference purposes. Emerald is unable to publish your work until a completed and signed copyright transfer agreement has been received.

Journal Title (hereafter known as the "Journal"):	World Journal of Engineering
Article Title (hereafter known as the "Work"):	Early Prediction and Analysis of Corona Pandemic Outbreak using Deep Learning Technique

Title	Mr.	
Name	RANJITH REDDY	GADDAM
Job Title	Student	
Organisation	Chaitanya Bharathi Institute of Technology	
Address	10-93, Rollakal	
Address		
County/State	Rollakal	Telangana
Country	India	508257
Telephone		Fax
Email	grreddy2726@gmail.com	

All author names	gampala, veerraju; Nandankar, Praful Vijay; M, Kathiravan; Karunakaran, S; NALLA, ARUN REDDY; GADDAM, RANJITH REDDY
All author email addresses	veerraju5a8@gmail.com, pppful@gmail.com, kathirrec1983@gmail.com, s.karunakaran@vardhaman.org, arunreddy096@gmail.com, grreddy2726@gmail.com

In consideration of **Emerald** agreeing to consider the above-named previously unpublished original Work for publication (both parties agree that such consideration shall be deemed sufficient), I/We, by signing this form hereby assign worldwide copyright of the Work in all forms and media (whether now known, or hereafter developed), in all languages for the full term of copyright and all extensions and renewals thereof.

I/We understand that **Emerald** will act on my/our behalf to publish, reproduce, distribute and transmit the

Work and will authorise other reputable third parties (such as document delivery services) to do the same, ensuring access to and maximum dissemination of the Work.

Licence to Author: **Emerald** grants to Author a non-exclusive licence to use and reproduce in printed form all or part of the Work (after first publication by the Journal): as photocopies for an Author's use for classroom teaching to be distributed to students free of charge, and in any literary work written or edited by the Author. This licence is granted providing that all such copies include full attribution to the Journal and the appropriate copyright line. For further information about additional Author rights, please see Emerald's Author Charter.

Authors submitting articles to **Emerald** warrant the following:

- I/We have the full power and authority to enter into and execute this Agreement and to convey the rights granted herein.
- The Work is an original work which I/We have created independently. It has not been published before in its current or a substantially similar form. Please refer to Emerald's Originality Guidelines
- The Work is not currently being considered for publication by any other journal or publication and will not be submitted for such review while under review by the Journal.
- Subject to the use of any third party rights where consents have been obtained in accordance with the paragraph below, I/We own all intellectual property rights vesting in the Work.
- If third party material has been used in the Work, I/We have obtained the necessary permission from the copyright holder/s to reproduce in the Work, in all media in all countries, and transmit via all reputable third parties, any such materials including tables, figures and photographs not owned by me/us (Please upload any permissions documents.).
- The Work does not contain any unlawful statements, does not infringe any existing copyright or violate any proprietary rights, rights of privacy or publicity, or any other rights of any third party. "Proof of consent" has been obtained for studies of named organisations and people (Please upload any evidence).
- All authors have received a final version of the Work, take responsibility for the content, agree to its publication and the order of authors listed on the paper.
- Anyone who has made a significant contribution to the research and the Work has been listed as an author. Minor contributors have been noted in the Acknowledgements section.
- I/We have declared any potential conflict of interest in the research. Any support from a third party has been noted in the Acknowledgements.
- I/We have read and adhered to the Journal author guidelines.
- I/We will not permit others to electronically gather or harvest and save to a separate server my/our Work.

I/We assert my/our moral rights to be identified as the author/s of the Work, in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988. I/We acknowledge that **Emerald** will ensure fair and faithful representation of my/our Work in all media and will take the necessary steps to protect the Work from unlawful copying.

I/We indemnify and shall keep **Emerald** Group Publishing indemnified against any loss, expense, injury or damage (including any legal costs and disbursements paid by them to compromise or settle any claim) howsoever caused incurred by **Emerald** directly or indirectly as a result of a breach of the above warranties.

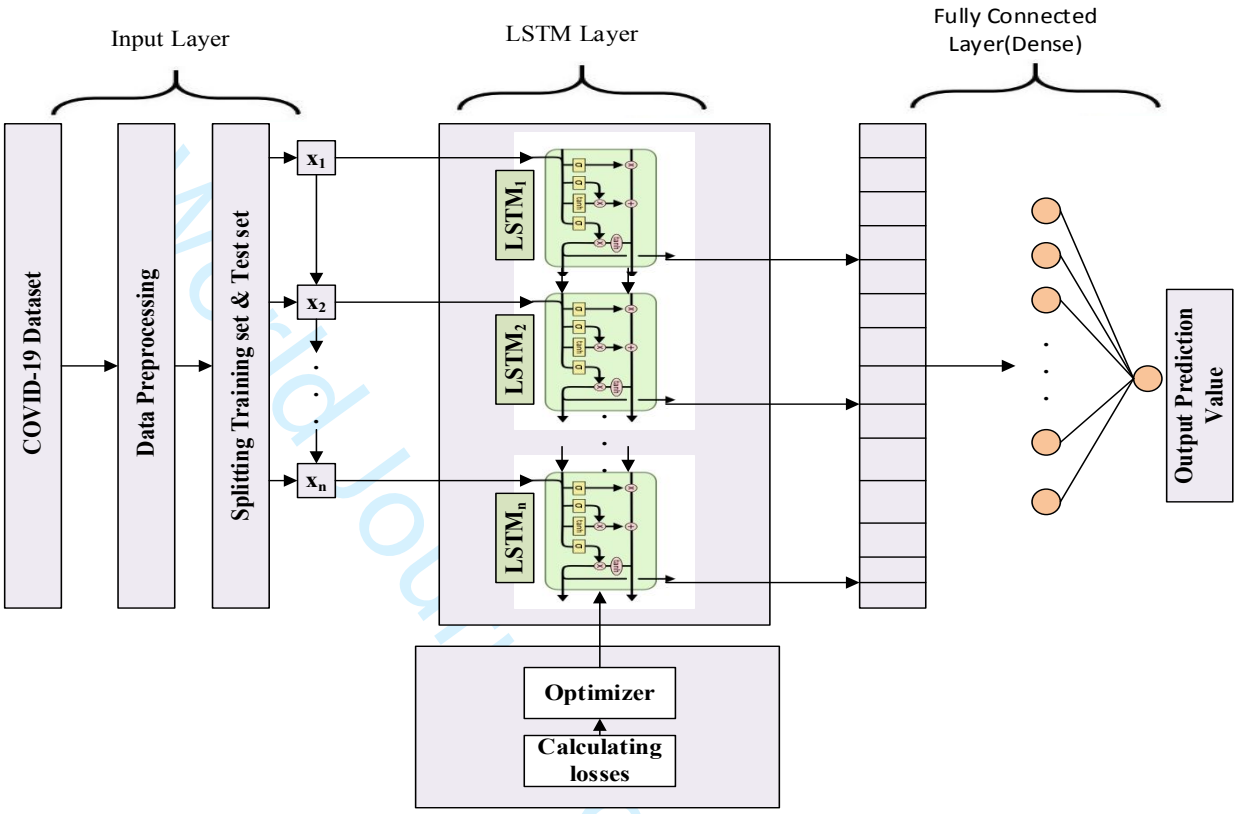
✓ req By clicking here you agree to the terms and conditions detailed above

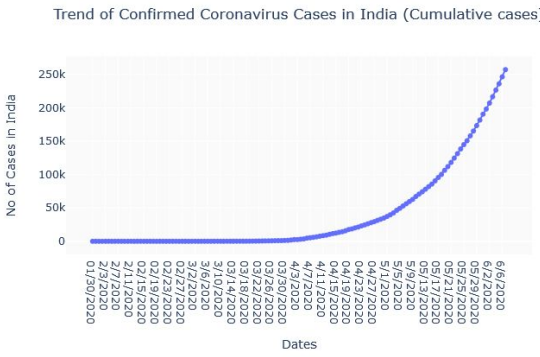
req

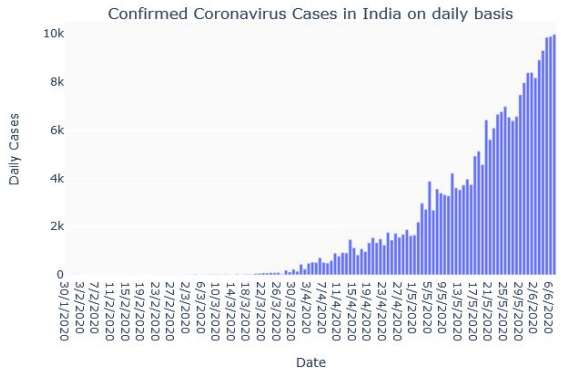
Date 16-Apr-2021

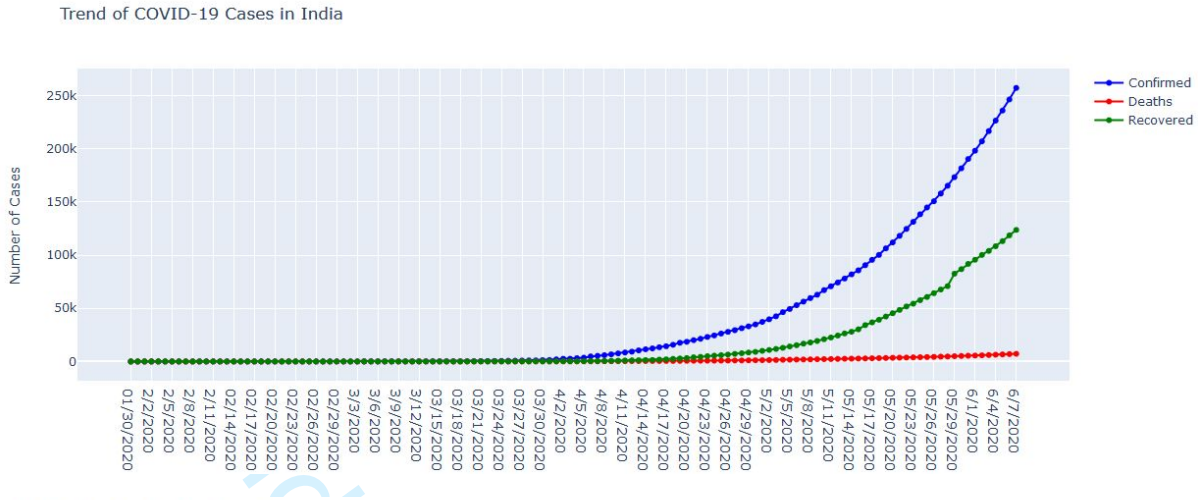
Early Prediction and Analysis of Corona Pandemic Outbreak using Deep Learning Technique

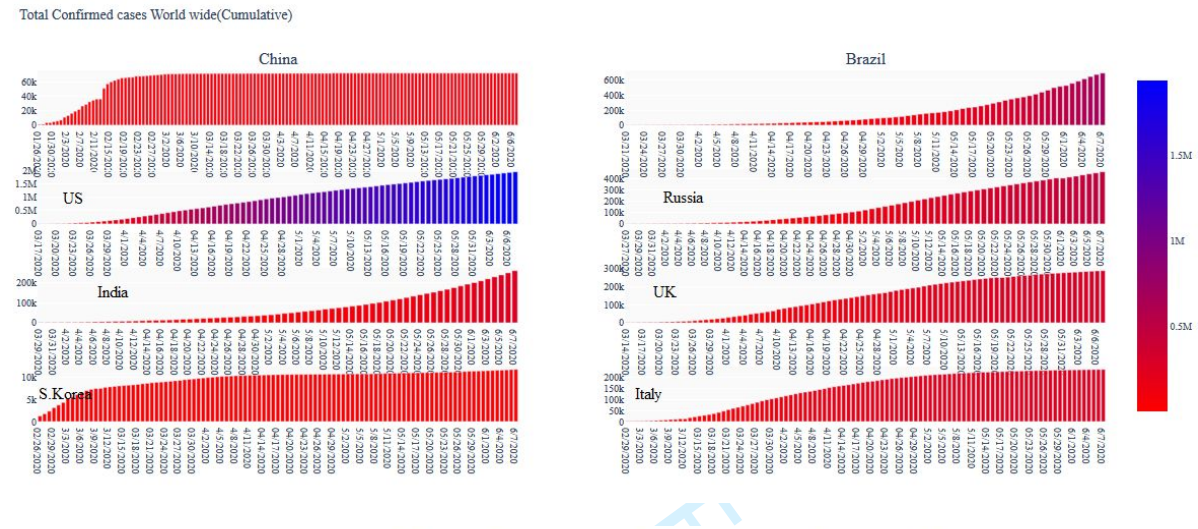
Journal:	*World Journal of Engineering*
Manuscript ID	WJE-03-2021-0145.R1
Manuscript Type:	Research Paper
Keywords:	COVID-19, Coronavirus, Pandemic, Deep Learning, Recurrent Neural Network, Data analysis

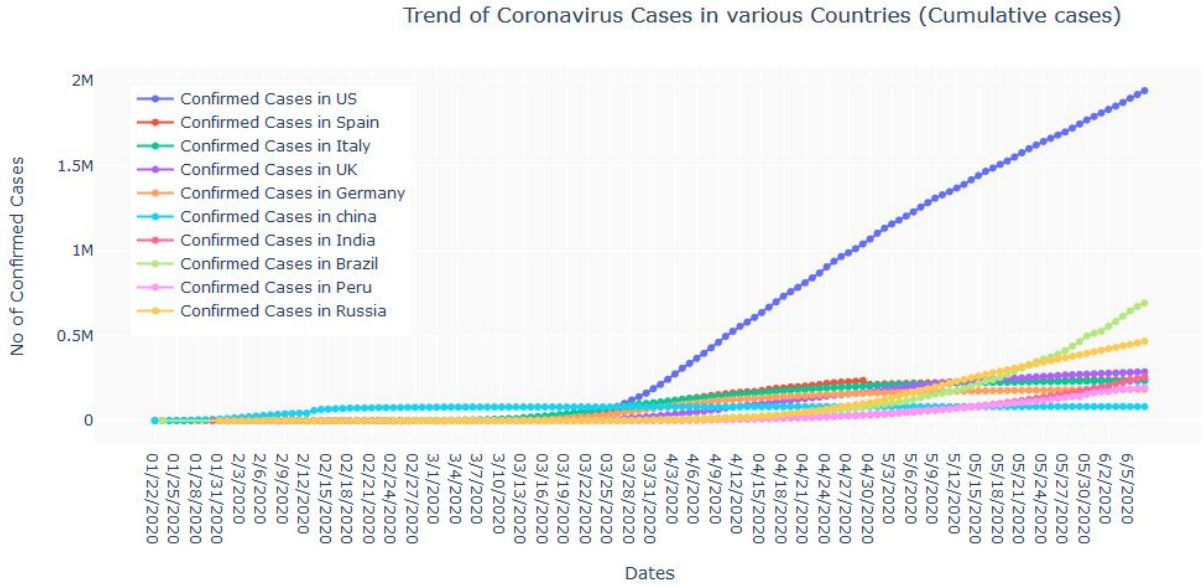


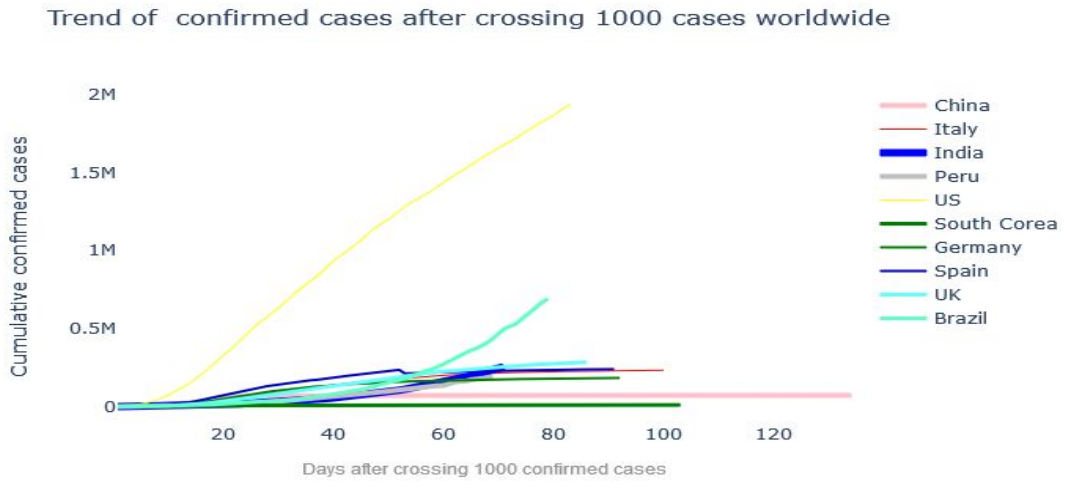


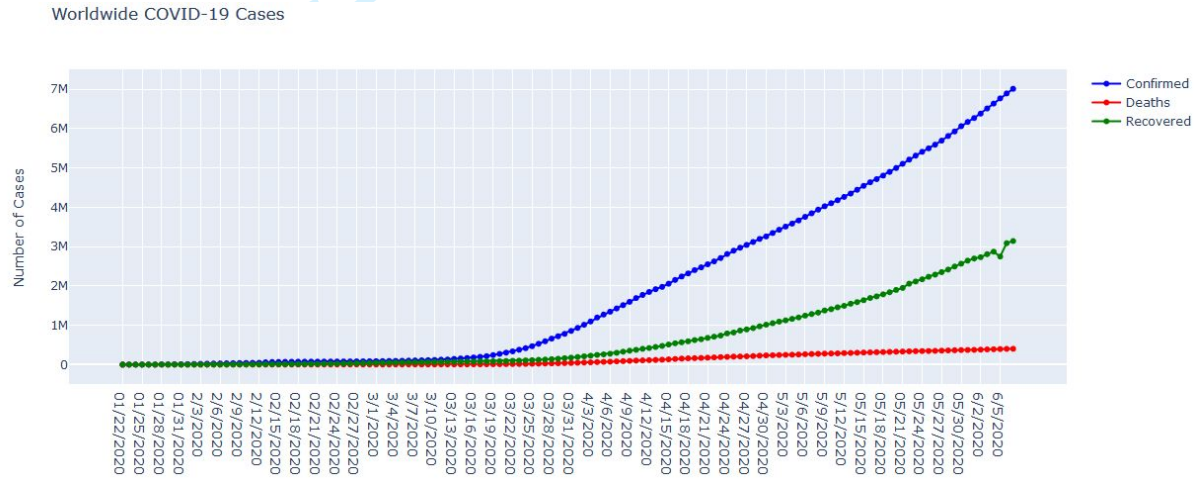


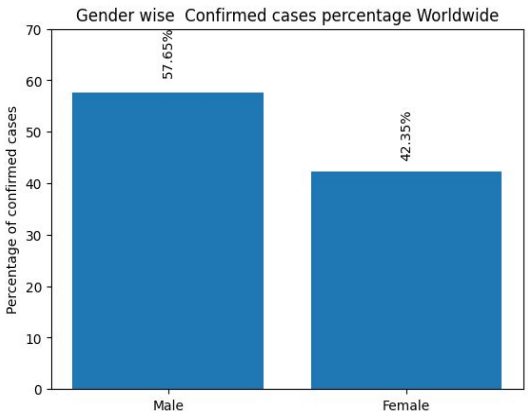

Trend of COVID-19 Cases in India
Number of Cases
Confirmed
Deaths
Recovered

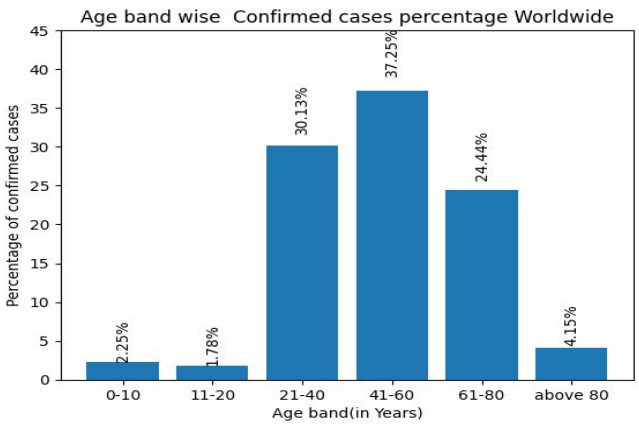

Total Confirmed cases World wide(Cumulative)
China
US
India
S.Korea
Brazil
Russia
UK
Italy

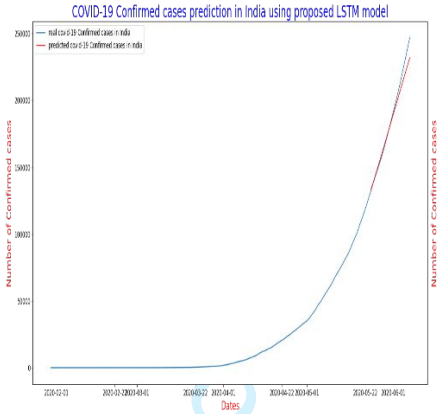

Trend of Coronavirus Cases in various Countries (Cumulative cases)
No of Confirmed Cases
Confirmed Cases in US
Confirmed Cases in Spain
Confirmed Cases in Italy
Confirmed Cases in UK
Confirmed Cases in Germany
Confirmed Cases in china
Confirmed Cases in India
Confirmed Cases in Brazil
Confirmed Cases in Peru
Confirmed Cases in Russia
Dates

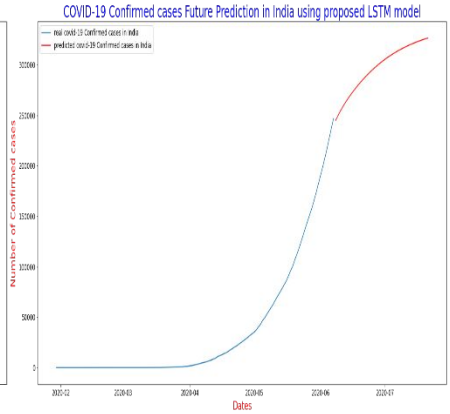

Trend of confirmed cases after crossing 1000 cases worldwide
Cumulative confirmed cases
2M
1.5M
1M
0.5M
0
20
40
60
80
100
120
Days after crossing 1000 confirmed cases
China
Italy
India
Peru
US
South Corea
Germany
Spain
UK
Brazil

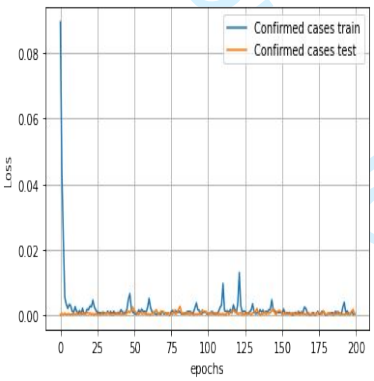

Worldwide COVID-19 Cases
Number of Cases
7M
6M
5M
4M
3M
2M
1M
0
Confirmed
Deaths
Recovered

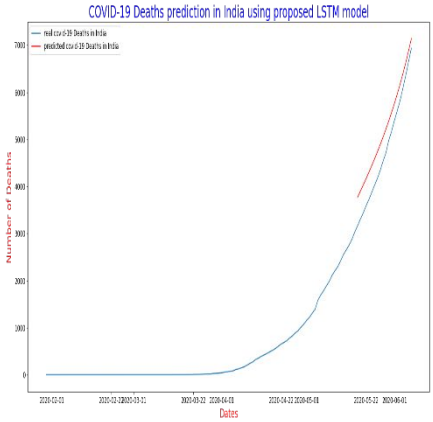

Gender wise Confirmed cases percentage Worldwide
Percentage of confirmed cases
57.65%
42.35%
Male
Female

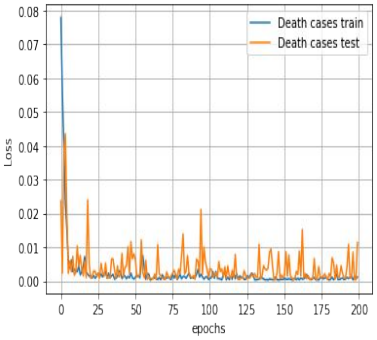

Age band wise Confirmed cases percentage Worldwide
Percentage of confirmed cases
2.25%
1.78%
30.13%
37.25%
24.44%
4.15%
0-10
11-20
21-40
41-60
61-80
above 80
Age band(in Years)

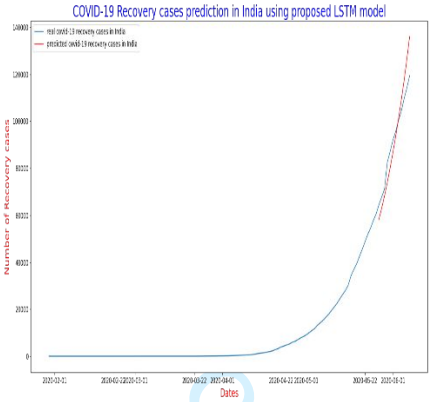


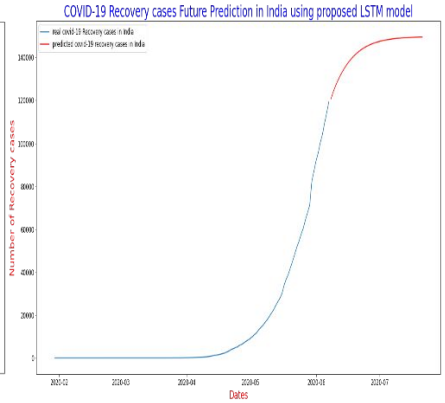


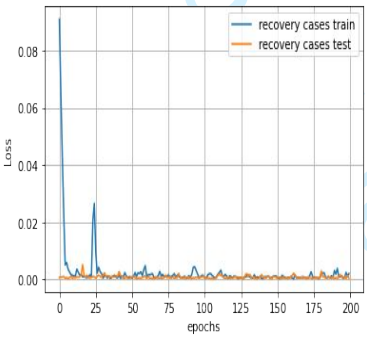


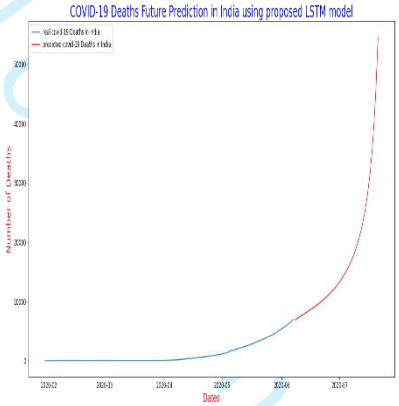












Early Prediction and Analysis of Corona Pandemic Outbreak using Deep Learning Technique

Abstract:

Purpose:

The purpose of this paper is to analyze and build a deep learning model that can furnish statistics of COVID-19 and able to forecast pandemic outbreak using Kaggle open research COVID-19 dataset. Since COVID-19 has an up to date data collection from the government, deep learning techniques can be used to predict future outbreak of coronavirus. The existing Long Short-Term Memory (LSTM) model is fine-tuned in order to forecast outbreak of COVID-19 with better accuracy as well as an empirical data exploration with advanced picturing has been made to comprehend the outbreak of coronavirus.

Methodology:

This research work presents a fine-tuned LSTM deep learning model using 3 hidden layers, 200 LSTM unit cells, one activation function ReLu, Adam optimizer, loss function is Mean Square Error (MSE), the number of epochs 200, and finally one dense layer to predict one value each time.

Findings:

LSTM is found to be more effective in forecasting future predictions. Hence, fine-tuned LSTM model predicts accurate results when applied to COVID-19 dataset.

Originality:

The fine-tuned LSTM model is developed and tested for the first time on COVID-19 dataset in order to forecast outbreak of pandemic.

Keywords: COVID-19, Coronavirus, Pandemic, Deep Learning, Recurrent Neural Network, LSTM, Data analysis

1. Introduction

The newly discovered novel coronavirus is an infectious disease called COVID-19. The majority of the individuals contaminated with coronavirus will encounter minor to moderate breathing sickness and recoup without extraordinary treatment [17]. Old age people and the people with causal medical backgrounds such as chronic respiratory syndrome, cardiovascular syndrome, cancer, diabetes, and are further likely to progress stern illness. COVID-19 outbreaks predominantly through droplets of saliva or expulsion from the nose when an infected person sneezes or coughs [16], so it's essential that you also practice breathing etiquette, for example, by coughing or sneezing into a flexed elbow. At present, there is no vaccine or medicine or treatment for this coronavirus. However, several clinical trials to evaluate potential treatments are ongoing. So, the best way to prevent this coronavirus and slow down transmission is, well informed about coronavirus, the disease and its cause, and the way it spreads. The only way to protect ourselves and others from the virus by cleaning hands with soap or alcohol based rub frequently, use sanitizer, maintain social distance (1 meter), practice respiratory hygiene, and not touch eyes, mouth, and nose with hands.

The coronavirus is first noted in Wuhan, China in December 2019. China government and the World Health Organization (WHO) [15] have recognized a coronavirus (COVID-19) as the causative virus that belongs to the Severe Acute Respiratory Syndrome (SARS) virus category, which out broke in South China in 2002-2003 [3]. The coronavirus has spread across the globe around 209 countries [18]. The virus is rapidly raising infection in the United States, Spain, Italy, France, Germany, and the United Kingdom. As per statistics as on 12th June 2020, this virus is infecting 7.27 million people and killed 4.13 lakhs so far worldwide. In India, the COVID-19 affected cases are nearly 3 lakhs and deaths 8,498. The most infected country in the world is the United States with 2.06 million and deaths are 1,15,000. When deaths are concern, Italy is suffering with 34,167 deaths. The spread of coronavirus is developing a major international and national crisis. It influences the daily life of the people. It has a massive impact on the global market and economy such as travel, and press and media.

The most common signs of coronavirus are fever, cough, sneeze, the shortage of sniff, and breathing complications [16]-[17]. In most of the serious cases, these viruses can basis for pneumonia, severe critical respiratory disease, and equally it causes some people to death. The symptoms appear within 2-14 days period. This virus outbreaks primarily through contact with the coronavirus infected person when they cough and sneeze. It also spreads when the infected persons touch an object or a surface and a healthy person touches the same, and then he touches his eye, nose or mouth. The spread of this virus can be explained in a simple way, i.e. for example, let's say Ravi got diseased yesterday, but, he won't recognize still subsequent 14 days and Ravi felt he is healthy, but he can infect averagely 10 persons a day. Now, these 10 persons felt completely healthy can infect 100 persons per day. These 100 persons felt they are completely alright, but they can infect 1000 persons and so on. Here, no one knows who is healthy and who can infect you. So, the only prevention is stay at home or quarantine.

However, monitoring and anticipating the pandemic of infection turns out to be incredibly critical for the decision making against the general wellbeing calamity. Measurable scientific models have gained more consideration in clinical science and the study of disease transmission [4]-[6]. Some dynamic models are likewise intended for a particular epidemic [7]. For instance, the dynamic models are intended to simulate and forecast the outbreak of HIV and SARS [3], [9]-[11]. As the enlargement of different models, machine learning and the complex network are joined with the fragmented model and generated a likely route of the widespread forecast [12], [13].

So, a mathematical model is required to predict the COVID-19 spread across several countries and states. The main objectives of this study are as follows:

- To analyze the present condition in India through different visualizing techniques.
- To check the trend of India is as similar as United States/Brazil/Russia/UK/Italy/S. Korea/ China.
- To explore and analyze complete world wide data.
- To propose a Recurrent Neural Network (RNN) prototypical model in order to predict and forecast COVID-19 in India.

The organization of remaining paper is the methodology of the proposed model is presented in section 2, the datasets are explored in section3, experimental outcomes and analysis are discussed in section 4, and final remarks are presented in section 5.

2. Methodology

This article proposes a solution based on standard LSTM algorithm to develop a prototypical model to forecast the COVID-19 pandemic outbreak. Generally, the LSTM model is used for predicting time series data. An LSTM control flow is as similar as the recurrent neural network. LSTM are proficient of long-term dependencies. All RNN networks have the chain of repeating units of neural network. LSTM also contains chain configuration, but the reiterating unit has a dissimilar configuration.

In this paper, LSTM architecture is proposed that aims to find COVID-19 predictions for confirmed, death, and recovered cases in India. Concretely, the proposed model is divided into three layers as shown in Fig. 1. Those are i) input layer, ii) LSTM layer, and iii) fully connected layer (dense).

2.1. Input Layer

The COVID-19 datasets [1]-[2] is taken in the input layer, and it contains several CSV files, which contains information about coronavirus. Next, data preprocessing is applied to acquire required data for the model.

The first step in data preprocessing is importing libraries such as Numpy, Matplotlib, and Pandas. Next, CSV files are loaded to pandas, which are used to do some simple scrubbing operations such as eliminating info you don't want that will causes data processing dawdling. For instance, eliminating the first line, it comprises unessential content rather than the column titles. Now, data should be explored to understand what features each column represents. This will evade errors in the data study and the modeling procedure [19-20]. From the start you ought to choose on the suitable column to use as a target column for modeling. Finally, features need to be prepared for the proposed model. The input data is split into two: train and test set, and passed as input to the LSTM layer. The machine learning models

uses numerical values. To keep all features in the same scale, you can use standard scalar [21-22]. Finally, the data is passed to the next layer, i.e. LSTM layer.

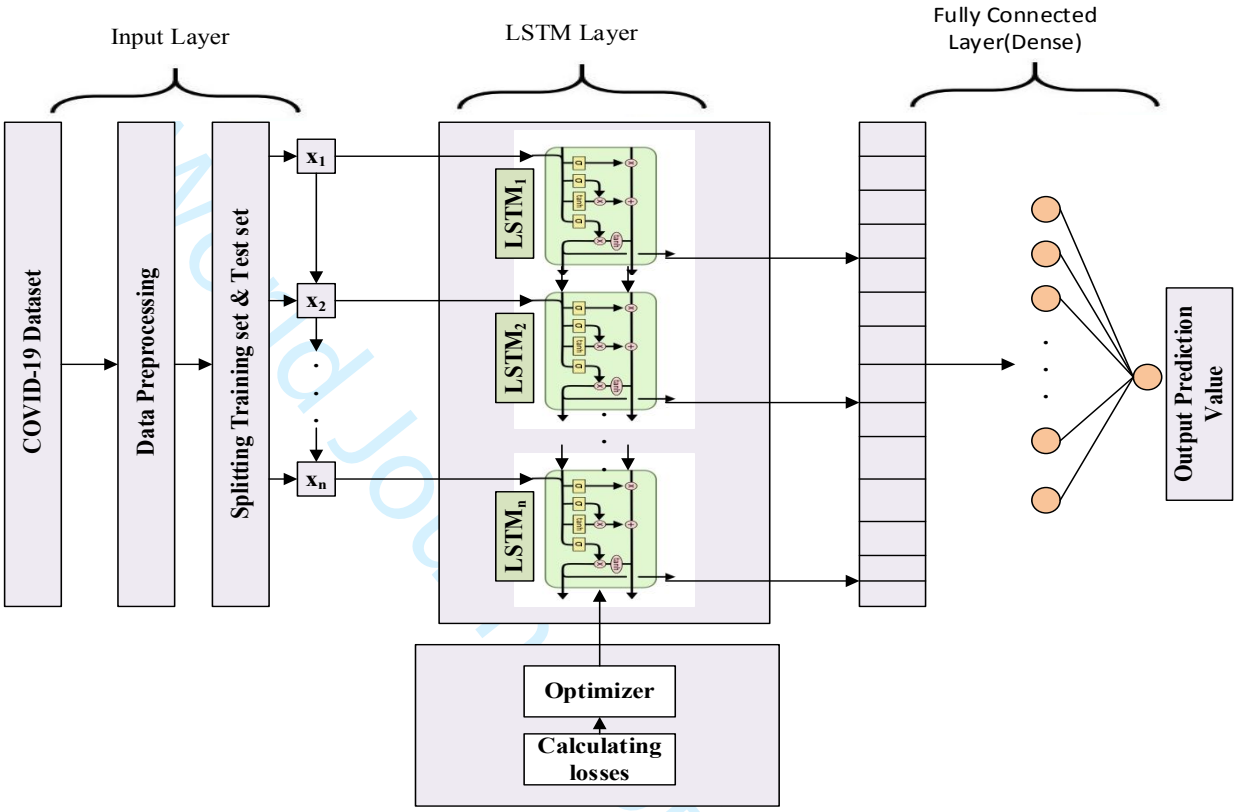


Fig. 1. The proposed LSTM architecture.

2.2. LSTM Layer

The LSTM layer contains several LSTM units. The initial phase in the LSTM unit is to select what data going to pass from the cell state. A "forget gate layer" makes a decision is called the sigmoid layer. It takes two parameters h_{t-1} and x_t output a numeral among 0 and 1for each digit in the cell state C_{t-1}. Here, a 0 denotes "completely get rid of this" and a 1 denotes "completely keep this". The formula for sigmoid function is given below:

$$F_t = \sigma (W_f . [h_{t-1}, x_t] + b_f) \quad (1)$$

The second phase is to choose what new data is to be supply in the cell unit. It includes two sections. First, the "input gate layer" termed sigmoid layer. It chooses what values to be update, and it is shown in equation 2. Further, a tanh layer generates a vector of new candidate values, $\tilde{C}_t$ that could be supplied to the cell. The formula for tanh function is shown in equation 3. In the subsequent phase, these two are joined to update the cell state.

$$i_t = \sigma (W_i . [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh (W_C . [h_{t-1}, x_t] + b_C) \quad (3)$$

Currently, the previous cell state is to be update, C_{t-1} into the new cell state C_t. Next, the previous state is multiplied by f_t. Then sum $i_t * \tilde{C}_t$. The generated values are candidate values, scrabbled by how much obvious to reform every state cell.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (4)$$

At last, the yield will be founded on the cell. Initially, sigmoid layer is going to run that chooses what portions of the cell state-run going to yield. At this moment the cell is put over tanh and multiplies it by the harvest of the sigmoid gate.

$$o_t = \sigma (W_o . [h_{t-1,} x_t] + b_o) \quad (5)$$

$$h_t = o_t * tanh(C_t) \quad (6)$$

The LSTM layer result is passed to next layer i.e. fully connected layer.

2.3. Fully Connected Layer

The dense layer (fully connected layer) forages all output from the preceding layer to all its neurons, every neuron gives one output to the subsequent layer [23]. Especially, it associates the yield of LSTM layer to a preferred yield size. The sigmoid function is used in the dense layer in order to seizure all yield value from the fully connected layer into a value among 0 and 1. The final output is predicted with two steps: the output is reshaped into original shape, and the output is obtained after last the last time sequence.

3. Dataset Description

In this article, several datasets are studied for COVID-19 analysis, visualization, and prediction. Finally, two different sources of dataset are used. Those are "COVID-19 in India" dataset [1] and "novel coronavirus 2019" dataset [2]. COVID-19 India dataset contains a CSV file, which consists of day wise Confirmed Indian National, Confirmed Foreign National, Cured, Deaths, and Confirmed cases, and it also gives the information state wise daily cases. Novel coronavirus 2019 dataset contains a spreadsheet consists of worldwide confirmed, deaths, and recovered persons. Table 1 gives knowledge on every dataset and their respective CSV files with the field description. Similarly, Table 2 is additionally created for giving the data analysis using exhaustive information of each field for utilized CSV files.

Table 1. Different dataset sources of COVID-19.

Dataset	Description	Columns
COVID-19 in India dataset (January-April 2020) [1] covid_19_india.csv	This file contains total of covid-19 cases in India at everyday level	SNo, Date, Time, State, ConfirmedIndianNational, ConfirmedForeignNational, Cured, Deaths, Confirmed
IndividualDetails.csv	Individual case level details in India	Id, government_id, diagnosed_date, age, gender, detected_city, detected_district, detected_state, nationality, current_status, status_change_date, notes
Novel corona virus 2019 dataset (January-April 2020) [2] Covid_19_data.csv	Day-to-day information on the number of COVID 2019 pretentious persons across the world	SNo, ObservationDate State, Country, Previous Update, Confirmed, Deaths, Recovered
time_series_covid_19_confirmed.csv	Time series records on numeral of confirmed cases	State, Country, Lat, Long 1/22/20 to 4/30/20
time_series_covid_19_deaths.csv	Time series records on numeral of deaths	State, Country, Lat, Long 1/22/20 to 4/30/20
time_series_covid_19_recovered.csv	Time series records on numeral of recovered cases	State, Country, Lat, Long 1/22/20 to 4/30/20

Table 2. Column depiction of COVID-19 datasets from different sources

Column	Description
SNo	Serial number
ObservationDate	Observation date in mm/dd/yyyy
state	Province or state of the observation
Country	Country or region of the observation
Last Update	Last update date time in UTC
Confirmed	Cumulative quantity of confirmed cases
Deaths	Cumulative quantity of deaths cases
Recovered	Cumulative quantity of recovered cases
Lat	Latitude
Long	Longitude
1/22/20	No. of confirmed till this day, No. of deaths till this day, and No. of recovered till this day
Date	Date of observation
Time	Time of observation
State	Name of the State / Union territory
ConfirmedIndianNational	Cumulative number of confirmed Indian nationals
ConfirmedForeignNational	Cumulative number of confirmed foreign nationals
diagnosed_date	Date of diagnosis
age	Age
gender	Gender

4. Experimental Analysis and Discussion

This section describes the main aspects of experimental work carried out in this study. Concretely, new coronavirus datasets [1]-[2] are employed in order to analyze the COVID-19 outbreak in various countries and predict future spread. The proposed LSTM model is implemented using Keras [13], which is a deep learning package of Tensorflow [14]. The whole data analysis and model development are coded in Python 3 and executed in the Colab research framework with 12.72 GB RAM and Google Compute Engine backend [19]. An advanced visualization library plotly is used to represent COVID-19 data graphically. The plotly is not just a static graph, but it is also a very interactive graph.

4.1. COVID-19 Outbreak Analysis in India

The COVID-19 pandemic rise in India is presented in Fig. 2, which is a scatter plot. This graph shows the sum of cumulative confirmed corona cases in India day wise. The date format used in the graphs is 'MM/DD/YYYY'. The first corona case is found in India on 30th January 2020. If you observe the data, there is a quick switch between March 3 and March 6 i.e. 9 and 28 cases respectively. The total sum of overall confirmed corona cases recorded up to 7th June is 2,57,486.

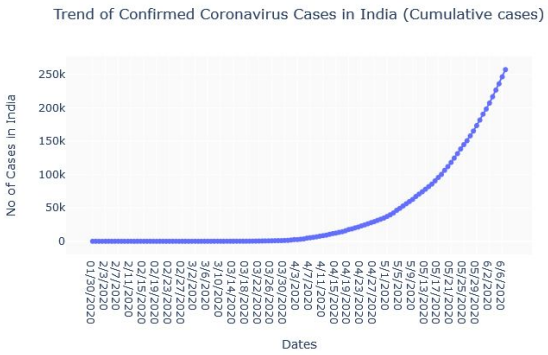


Fig. 2. Visualization of Confirmed corona cases in India

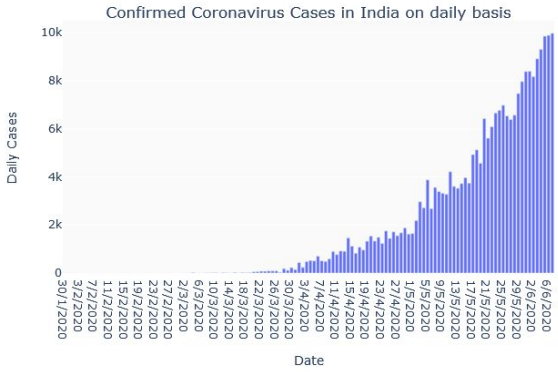


Fig. 3. Confirmed corona cases in India on a daily basis.

The day-to-day confirmed corona cases in India are visualized in Fig. 3. From the plot, it is observed that very little cases in February, i.e. 2 cases and then no cases up to March. On March 23, 103 cases are confirmed and on the ending day of March 309 cases are recorded. Moreover, on the same day overall confirmed cases are 1397, which is presented in the Fig. 2. The graph depicts that the corona cases are exponentially increasing. From 13th April onwards, confirmed cases are rising in thousands every day. 1243 cases are recorded on 13th April. Finally, the highest number of cases i.e. 9971 noted on 7th June up to now. The daily cases are visualized up to 7th June, but still it is continuing.

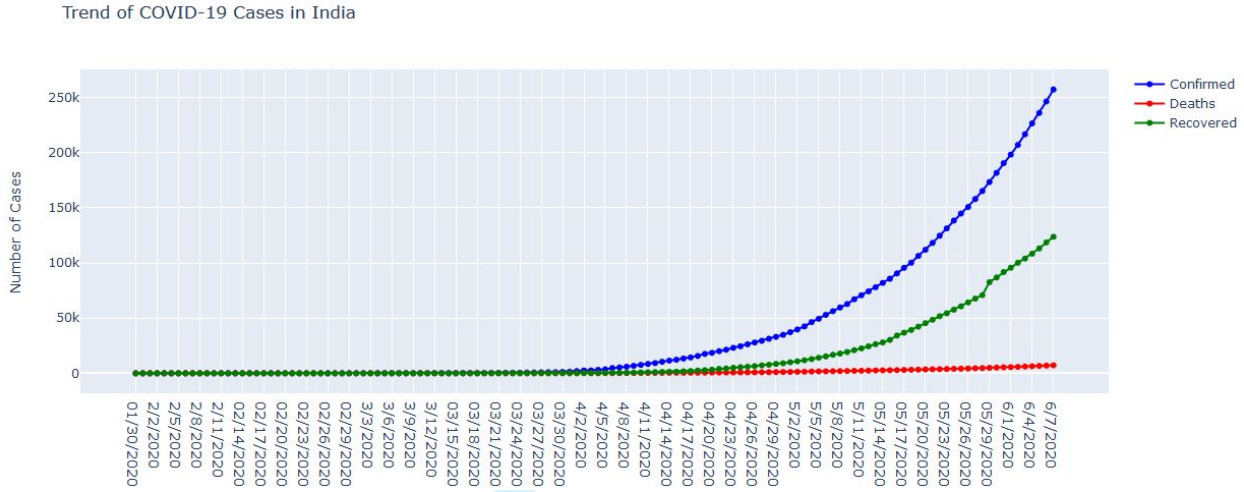


Fig. 4. Comparison of COVID-19 cases in India

The trend of coronavirus cases in India are compared in Fig. 4. The confirmed cases are exponentially growing. The first death happened on 11th March. The deaths are growing almost sequentially. The first 3 coronavirus patients are recovered on 16th February. The recovered cases are successively growing. As per observation in the Fig. 4, the deaths are very low, and it is 2.79% of conformed cases. The recovered cases are little more when compared to death cases. The total deaths are recorded as on 7th June 2020, is 7207, whereas recovered cases are 1,23,848, and total confirmed cases are 2,57,486. The total recovered cases are 48.09% of confirmed cases.

4.2. Worldwide COVID-19 Outbreak Analysis

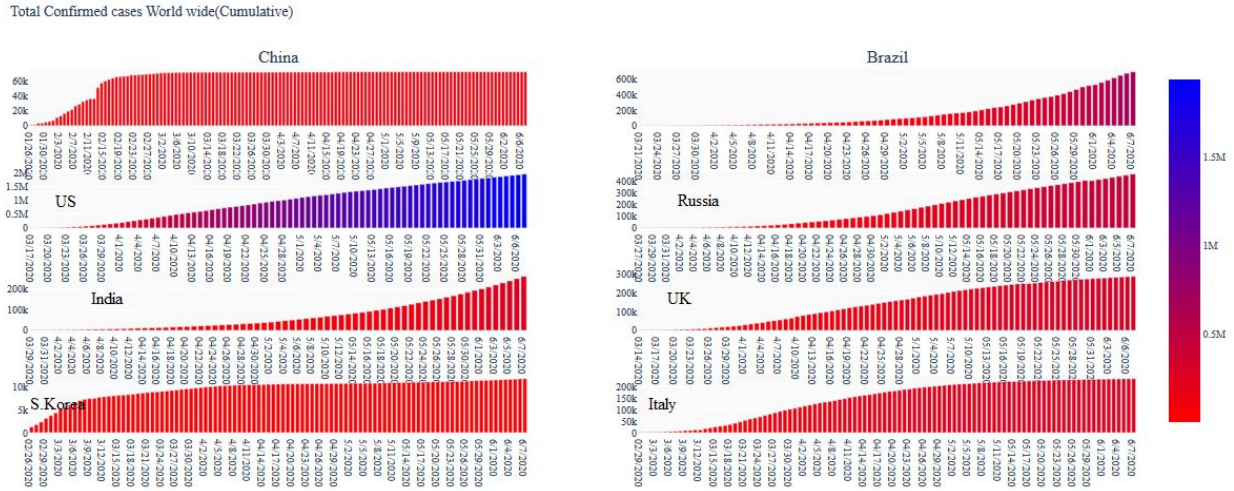


Fig. 5. Comparison of coronavirus cases in various countries.

The comparison between most affected five countries with COVID-19 including China, Italy, and South Korea is shown in Fig. 5. The trend of coronavirus in eight countries such as the US, Brazil, Russia, UK, India, China, Italy, and South Korea are compared in one common plot. The legend of the graph is shown at the right side. The complete blue color indicates a more number of cumulative daily cases, and pure red color represents a less number of cases. It observed that the blue color is recorded from 1 million to 2 million cases, and the red color is from 0 to 0.5 million cases. Among the eight countries, the blue color is recorded only for US. In the US, the highest total number of cases

recorded on 7th June, 2020, i.e. 1.95 million. The US graph indicates that the corona growth in the US is a completely exponential curve, and still more number of cases are recording in the US because the US president not considered coronavirus as serious pandemic. Next highest cases noted in Brazil, i.e. 6,91,758. The Brazil plot shows that the growth is exponential. Brazil took too much time to know the serious about pandemic coronavirus.

Russia graph is also an exponential curve, which shows a logistic growth in certain parameters. UK is started a completely sigmoid curve since the 15th May because an extensive testing is done. Next, if you observe the India plot, it is an exponential graph, especially from 28th of March, but the number of cases are very less because the prime minister of India announced the first phase of lockdown from 23rd March to 14th April, the second phase up to 3rd May, and the third phase up to 17th May 2020. The pattern between the US and India is almost same, but the number of cases in India are very less compared to US, but after 3rd phase lockdown cases are increasing rapidly. The total quantity of corona cases in India up to 7th June 2020 is 2,57,486. The South Korea graph is a sigmoid curve, which shows logistic growth in certain parameters. The South Korea is started completely sigmoid curve since 7th March because an extensive testing is done anybody with mildest infection has been quarantined in this country since March, which is why they have successfully flattened the curve almost. Next, if you observe the China plot, there is a steep rise between 12th of February to 13th Feb. Later, it goes on rising, again this is also following a sigmoid pattern successfully from 3rd of March because China has started to adopt some artificial intelligence models, which will help them to diagnose people with flu on a very extensive scale. The Baidu Company comes up with a heat mapping on human bodies in China. This is a machine learning technique, which picks out the people with mild temperatures and straight out of the roads and sends them directly into testing. The Italy plot shows that the growth is exponential up to 22nd March, later it becomes flatter. Italy took too much time to know the serious about pandemic coronavirus.

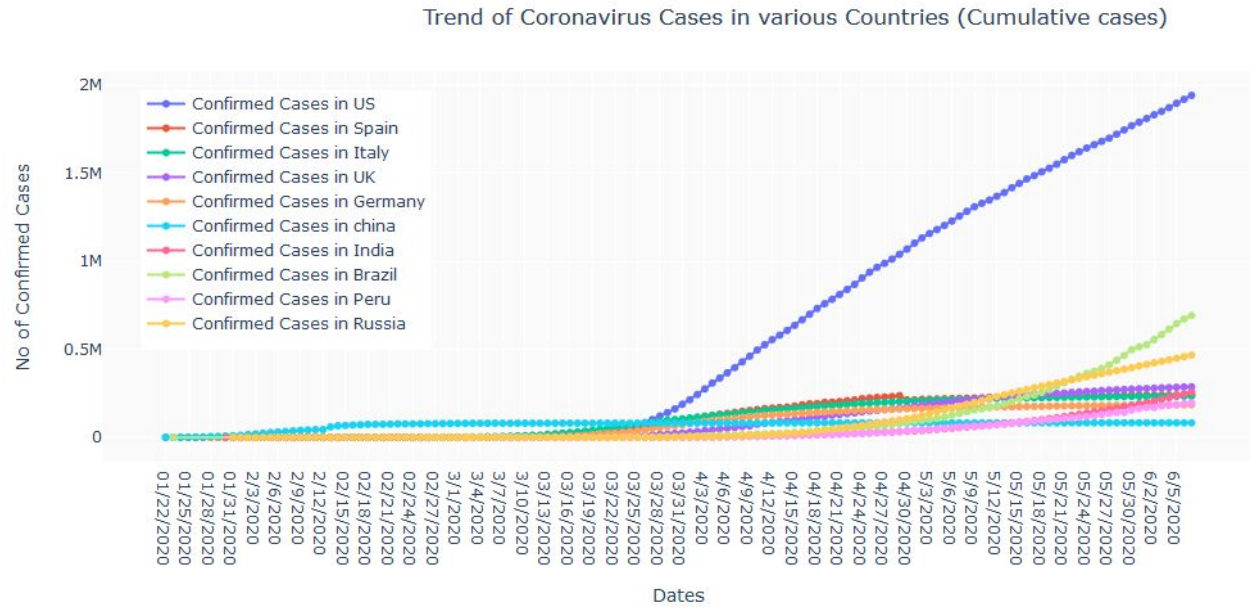


Fig. 6. Trend of cumulative COVID-19 confirmed cases in various countries.

The trend of confirmed cumulative corona cases in various countries is compared in Fig. 6. The confirmed cases in US are very high from 30th March, 2020, when compared to other countries. Even after 30th March also, the confirmed cases are growing exponentially in the US. China curve is flattened after 13th February. India has recorded less number of cases when compared to other countries still now due to the lockdown announced an early stage of the COVID-19 outbreak, whereas US announced very lately. But after lockdown, the cases are increased rapidly in India. The remaining countries' curve is growing exponentially in between US and China.

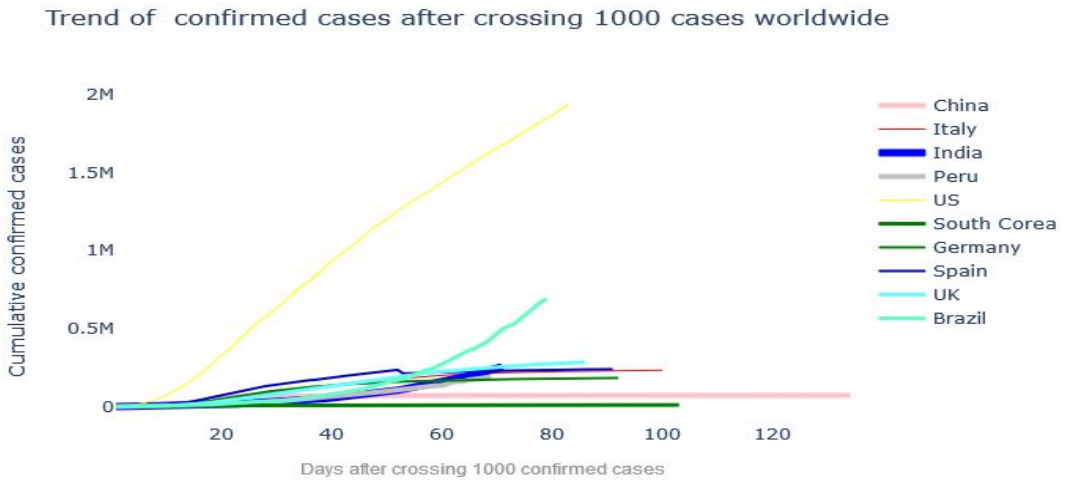


Fig. 7. Trend after crossing 1000 confirmed cases in various countries.

The trend of countries after crossing a thousand cases has compared in Fig. 7. Initial days, it is observed that the progression happened in China really quickly without having a systematic method of recoding the progression. Later, China growth is exponential. Finally, the growth becomes a sigmoid curve. South Korea is recorded less cases when compared to other countries, because it is doing well through testing, tracing, and quarantined. This graph depicts that in the sum of days afterward the COVID-19 cases have crossed a thousand, South Korea has the minimum number of cases out of all other countries i.e. 11,814 in 103 days, which is better when it compared to spread of this pandemic other countries in the world. US have crossed the mark of 1.94 million cases in the next 83 days, which is the highest among all other countries in the world. Next highest is Brazil, i.e. 6,91,758 cases in 79 days.

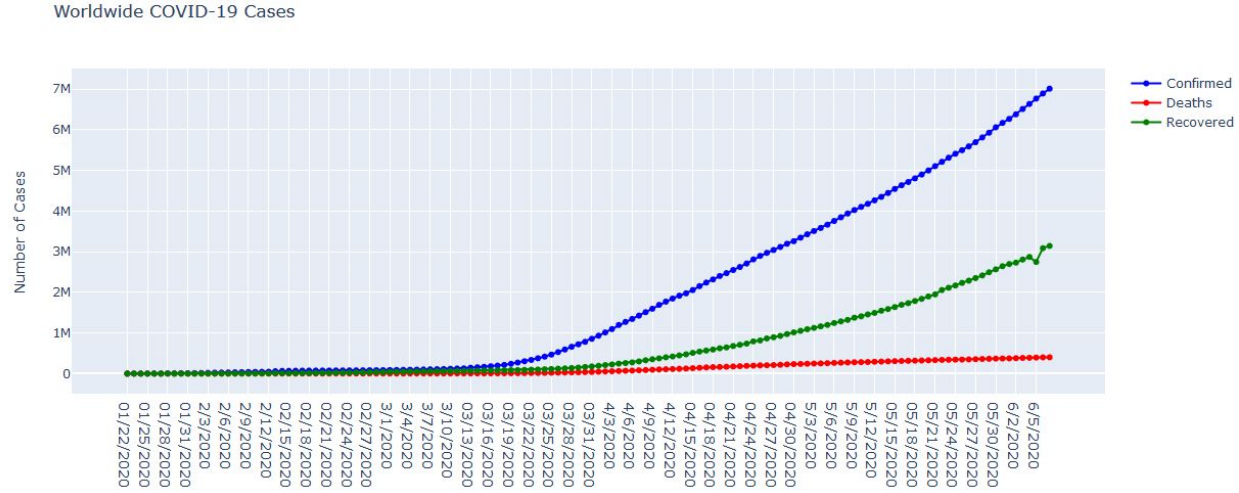
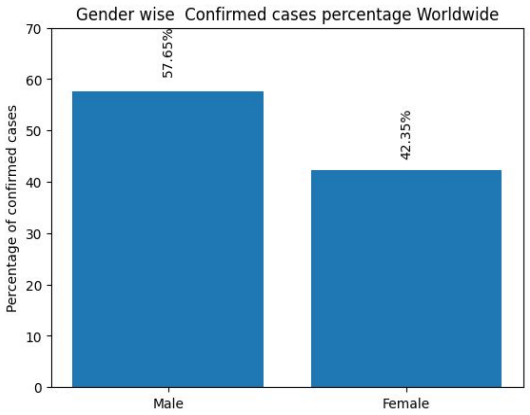


Fig. 8. Comparison of worldwide COVID-19 cases.

The impact of pandemic on the world with three separate spread sheets for confirmed, deaths, and recovered cases is plotted in the Fig. 8. The graph shows the day-wise cumulative cases, in which the blue line represents the confirmed cases. It is a perfect exponential curve. As on 7th June 2020, the total confirmed cases worldwide is 7 million. The recovered cases are not exponential, which is shown in the green color in the Fig. 8, and still now the recovered cases are 3.14 million, and it is 44.81% of confirmed cases. The deaths are very less when compared to confirmed cases. The number of deaths happened still now is 4, 02, 730. The deaths had happened 5.74% out of the confirmed cases.



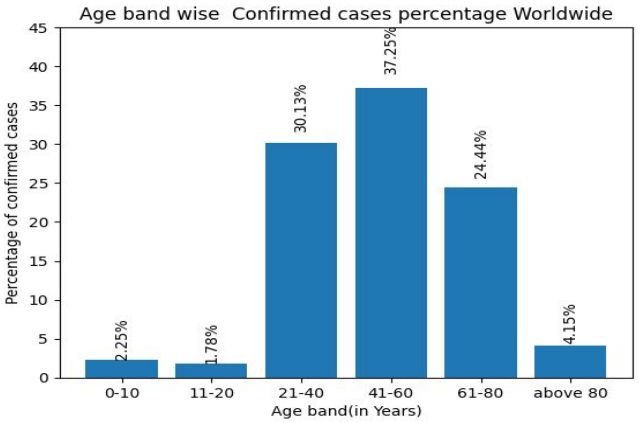


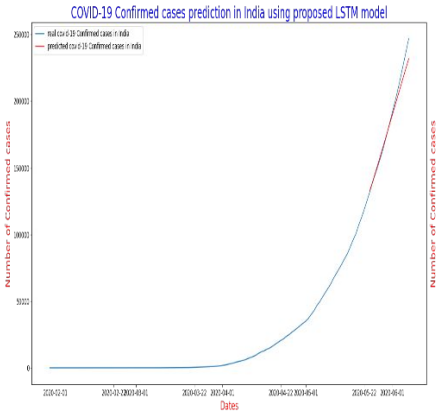
Fig. 9. Gender wise confirmed cases analysis. **Fig. 10.** Age wise confirmed cases analysis.

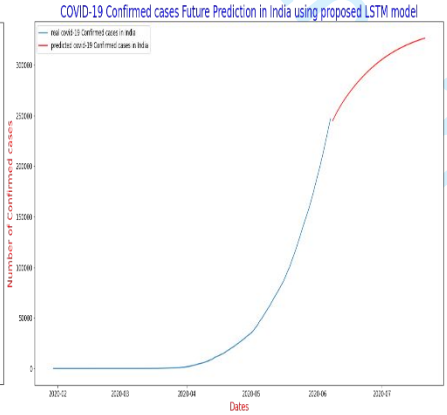
When the worldwide COVID-19 data is analyzed, then it is noticed that the coronavirus is infected for different category people with different ratio like shown in Fig. 9 & 10. The gender wise infected corona cases are compared in Fig. 9. The male infected worldwide corona cases are 57.65%, whereas the female infected cases is 42.35%, which is shown in Fig. 9. The graph indicates that the male are more affected when compared to female. Age wise corona affected cases are compared in Fig. 10. This graph shows that the more corona affected age band is 41-60, and it is 37.25%. The next infected age band is 21-40 and it is 30.13%. The least infected age band is 11-20, and it is 1.78%. From this observation, age bands 41-60 and 21-40 people have to take more precautions. Hence, these people should be a house quarantined in order to reduce the infection.

4.3. Prediction of COVID-19 outbreak in India

The coronavirus outbreak in India is predicted using the proposed LSTM model in three cases: confirmed, deaths, and recovery cases. The proposed model is designed using following hyper parameters: 3 hidden layers, 200 LSTM unit cells, the number of inputs 15,the number of features 1, batch size 6, dropout 0.15, activation function ReLu, Adam optimizer, loss function is Mean Square Error (MSE), the number of epochs 200, and finally one dense layer to predict one value each time.

4.3.1. Confirmed





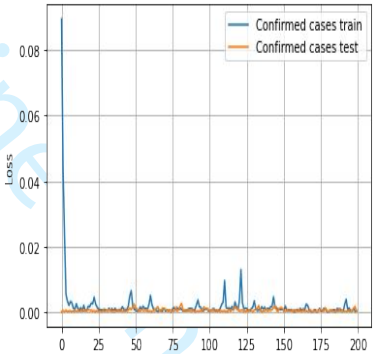


Fig. 11. Comparison of COVID-19 actual cases **Fig. 12.** Future prediction of COVID-19 and predicted confirmed cases in India. **Fig. 13.** Loss of Confirmed confirmed cases in India.

The whole dataset is split into two sets: 80% for training and 20% for testing. To observe the accuracy of the suggested model, the confirmed cases are predicted for the existing last 15 days' data and compared with the actual confirmed cases, which is shown in Fig. 11. From the graph, it is observed that the loss is very less. The confirmed cases' data is considered from 1st February, 2020 to 7th June 2020. The future confirmed cases are predicted for 45 days, i.e. from 8th June 2020 to 21th July 2020, which is shown in Fig. 12. The cumulative confirmed cases are reached the peak on 15th July 2020, i.e. 4 lakhs. Later, the confirmed cases' graph is flattened. From the Fig. 12, it is perceived that the

progress of the confirmed cases forms sigmoid curve. The MSE loss of predicted confirmed cases for training and testing is shown in Fig. 13. Initially, the loss is more for training when number of epochs 5. Later, it is decreased when the number of epochs increases.

4.3.2. Deaths

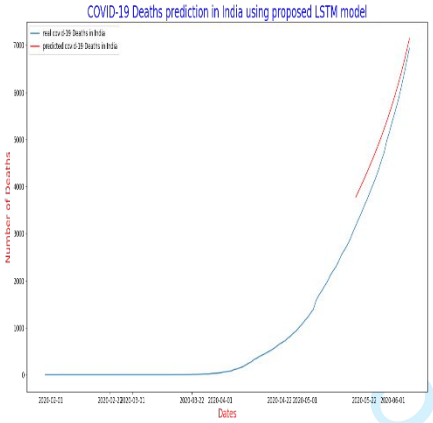


Fig. 14. Comparison of COVID-19 actual and predicted deaths in India.

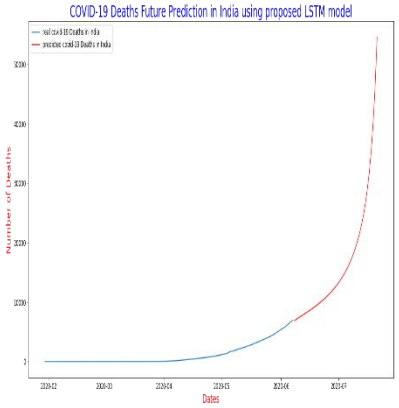


Fig. 15. Future prediction of COVID-19 deaths in India.

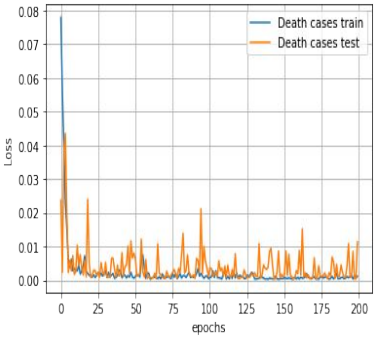


Fig. 16. Loss of Death cases.

To compare the accuracy of coronavirus death cases' prediction in India, the proposed LSTM model is applied to predict the last 15 days' deaths, and it is compared in Fig.14. From the graph, it is observed that actual COVID-19 deaths' curve, and the predicted deaths' curve are very close to each other. Hence, the prediction loss is less and accuracy is more. Using the proposed LSTM model, death cases are predicted for future 45 days, and depicted in Fig. 15. The proposed model is predicted that maximum death cases may reach 50,000, and may happen on 21st July 2020. Later, the curve is expected to flatten. The MSE loss of predicted death cases for training and testing is shown in Fig. 16. Initially, the loss is more for training when number of epochs 5. Later, it is decreased when the number of epochs increases. The testing loss is fluctuating between 0 and 0.02.

4.3.3. Cured

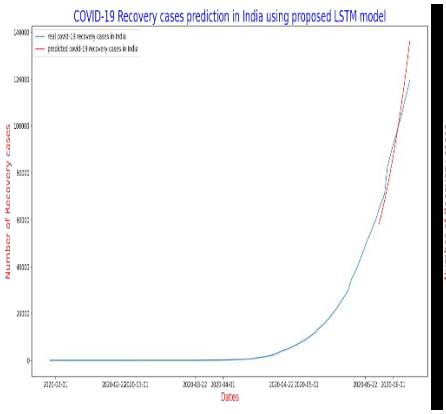


Fig. 17. Comparison of COVID-19 actual and predicted recovery cases in India.

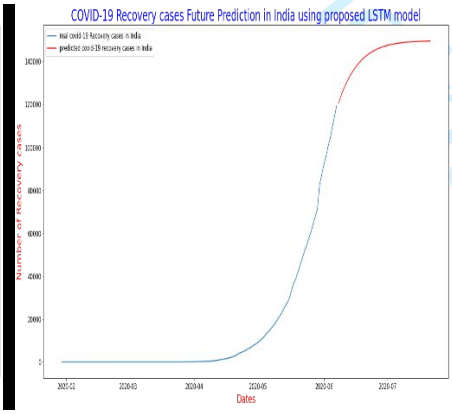


Fig. 18. Future prediction of COVID-19 recovery cases in India.

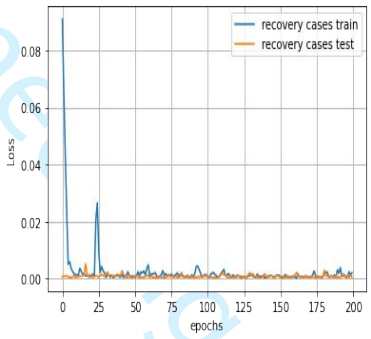


Fig. 19. Loss of recovery cases.

To compare the accuracy of coronavirus recovery cases' prediction in India, the proposed LSTM model is applied to predict the last 15 days' recovery cases, and it is compared in Fig.17. From the graph, it is observed that the actual COVID-19 recovery cases' curve and the predicted recovery cases' curve are very close to each other. Hence, the prediction loss is less and accuracy is more. Using the proposed LSTM model, recovery cases are predicted for future 45 days and depicted in Fig. 18. The proposed model is predicted the maximum recovery cases may reach 1, 60, 100,

and may happen on10th July 2020. Later, the curve is flattened. If you observe from the graph, the curve growth follows sigmoid function. The MSE loss of predicted recovery cases for training and testing is shown in Fig. 19. Initially, the loss is more for training when number of epochs 5. Later, it is decreased when the number of epochs increases.

5. Conclusion

In conclusion, "COVID-19 in India" and "Novel coronavirus 2019" datasets are used to examine the data with visualization and in order to predict the future outbreak of coronavirus. Such visualization can assist us with generating and proliferate total data to the scientific civic researchers, particularly in the beginning periods of pandemic, when there is a little data available, taking into consideration autonomous appraisals of key parameters that impact interventions. The COVID-19 future prediction (confirmed, deaths, and recovered) is made with the proposed LSTM model, which acquires better accuracy. The proposed model predicts the crowning confirmed cases of 4 lakhs on 15th July 2020, deaths of 50000 on 21th July 2020, and recovered cases of 1,60,100 on10th July 2020 in India. Moreover, the curves are flattened after peak cases recorded. It is an uncommon flare-up of COVID-19 in the twentieth era that is not actually the widespread of only a nation, progressively an open apprehension of the entire globe. Hence, further worldwide participation is essential to battle the COVID-19.

References

[1] Sudalai raj kumar, "COVID-19 in India" dataset, URL: https://www.kaggle.com/sudalairajkumar/covid19-in-india

[2] Sudalai raj kumar, "Novel Corona Virus 2019" dataset, URL: https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset

[3] K. O. Kwok, A. Tang, V. W. Wei, W. H. Park, E. K. Yeoh, and S. Riley, ``Epidemic models of contact tracing: Systematic review of transmission studies of severe acute respiratory syndrome and middle east respiratory syndrome," Comput. Struct. Biotechnol. J., vol. 17, pp. 186-194, Jan. 2019.

[4] R. M. Anderson, ``The pandemic of antibiotic resistance," Nature Med., vol. 5, pp. 147-149, Feb. 1999.

[5] J. Koopman, ``Modeling infection transmission," Annu. Rev. Public Health, vol. 25, pp. 303-326, Apr. 2004.

[6] Gomathi, S., Kohli, R., Soni, M., Dhiman, G. and Nair, R. (2020), "Pattern analysis: predicting COVID-19 pandemic in India using AutoML", World Journal of Engineering, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/WJE-09-2020-0450.

[7] Nair, R., Vishwakarma, S., Soni, M., Patel, T. and Joshi, S. (2021), "Detection of COVID-19 cases through X-ray images using hybrid deep neural network", World Journal of Engineering, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/WJE-10-2020-0529.

[8] M. Soni, S. Gomathi and Y. Bhupendra Kumar Adhyaru, "Natural Language Processing for the Job Portal Enhancement," 2020 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2020, pp. 1-4, DOI: 10.1109/ICSSS49621.2020.9202046.

[9] M. Soni, S. Chauhan, B. Bajpai, and T. Puri, "An Approach To Enhance Fall Detection Using Machine Learning Classifier," 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), Bhimtal, India, 2020, pp. 229-233, DOI: 10.1109/CICN49253.2020.9242634.

[10] Mukesh Soni, S. Gomathi, Cotton Leaf Spot Disease Detection using Multi-Class SVM, International Journal of Research in Engineering and Advanced Technology, Volume-8, Issue-5 page-149, 2020.

[11] S. A. Levin, B. Grenfell, A. Hastings, and A. S. Perelson, ``Mathematical and computational challenges in population biology and ecosystems science," Science, vol. 275, pp. 334-343, Jan. 1997.

[12] T. W. Ng, G. Turinici, and A. Danchin, ``A double epidemic model for the SARS propagation," BMC Infectious Diseases, vol. 3, p. 19, Sep. 2003.

[13] Dye., C. Epidemiology: Modeling the SARS Epidemic. Science, 2003, vol. 300, no. 5627, pp. 1884-1885.

[14] M. Lipsitch, ``Transmission dynamics and control of severe acute respiratory syndrome," Science, vol. 300, no. 5627, pp. 1966-1970, 2003.

[15] X. N. Han, S. J. De Vlas, L. Q. Fang, D. Feng, W. C. Cao, and J. D. F. Habbema, ``Mathematical modelling of SARS and other infectious diseases in China: A review," vol. 14, no. s1, pp. 92-100, 2009, doi: 10.1111/j.1365-3156.2009.02244.x.

[16] D. Watts and S. H. Strogatz, ``Collective dynamics of `small-world' networks," Nature, vol. 393, no. 6684 pp. 440-442, 1998.

[17] R. Pastor-Satorras and A. Vespignani, ``Epidemic spreading in scale-free networks," Phys. Rev. Lett., vol. 86, pp. 3200-3203, Apr. 2001.
[18] "Keras Documentation". 609 https://keras.io.
[19] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M., 2016. Tensorflow: A system for large-scale machine learning. In 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), 265-283.
[20] WHO. Novel Coronavirus—China January 12, 2020. http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/. Accessed 19 January 2020.
[21] Drosten C, Günther S, Preiser W, et al. Identification of a novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med. 2003;348:1967-1976.
[22] Chen Y, Liu Q, Guo D. Emerging coronaviruses: genome structure, replication, and pathogenesis. J Med Virol. 2020;92:418-423. https://doi.org/10.1002/jmv.25681
[23] Gorbalenya AE, Baker SC, Baric RS, et al. Acute respiratory syndrome-related coronavirus: the species and its viruses—a statement of the Coronavirus Study Group [published online ahead of print February 11, 2020]. bioRxiv. 2020. https://doi.org/10.1101/2020.02. 07.937862
[24] Colab research framework, URL: https://colab.research.google.com/
[25] V. Gampala, M. Sunil Kumar, C. Sushama et al., Deep learning based image processing approaches for image deblurring, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.11.076
[26] V. Gampala, V. Jean Shilpa, Sajja Krishna Kishore et al., Enhancement of resolution and image reconstruction in digital image correlation, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2020.11.758
[27] V. Gampala, Anupriya Koneru, Balajee Maram, Enhance the Educational Outcome in Higher Educational Institutes through Deep Neural Network, International Journal of Advanced Trends in Computer Science and Engineering, Volume 9, No.5, September - October 2020, pp. 8315- 8324.
[28] S.C. Dharmadhikari, Veerraju Gampala, Ch. Mallikarjuna Rao, et al., A smart grid incorporated with ML and IoT for a secure management system, Microprocessors and Microsystems, Volume 83, 2021, https://doi.org/10.1016/j.micpro.2021.103954.