Early prediction and analysis of corona pandemic outbreak using deep learning technique

Veerraju Gampala
CSE, Koneru Lakshmaiah Education Foundation, Guntur, India
Praful Vijay Nandankar
Government College of Engineering, Nagpur, India
M. Kathiravan
Department of Computing Science, Hindustan Institute of Technology and Science, Chennai, India
S. Karunakaran
Vardhaman College of Engineering, Hyderabad, India, and
Arun Reddy Nalla and Ranjith Reddy Gaddam
Chaitanya Bharathi Institute of Technology, Hyderabad, India

Abstract

Purpose – The purpose of this paper is to analyze and build a deep learning model that can furnish statistics of COVID-19 and is able to forecast pandemic outbreak using Kaggle open research COVID-19 data set. As 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 to forecast the outbreak of COVID-19 with better accuracy, and an empirical data exploration with advanced picturing has been made to comprehend the outbreak of coronavirus.

Design/methodology/approach – This research work presents a fine-tuned LSTM deep learning model using three hidden layers, 200 LSTM unit cells, one activation function ReLu, Adam optimizer, loss function is mean square error, 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 data set.

Originality/value – The fine-tuned LSTM model is developed and tested for the first time on COVID-19 data set to forecast outbreak of pandemic according to the authors' knowledge.

Keywords Data analysis, Recurrent neural network, Deep learning, Coronavirus, Pandemic, COVID-19, Pandemic, LSTM

Paper type Research paper

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 (Pastor-Satorras and Vespignani, 2001). Old age people and the people with causal medical backgrounds such as chronic respiratory syndrome, cardiovascular syndrome, cancer and diabetes 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 (Watts and Strogatz, 1998), so it is 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 COVID-19. However, several clinical trials are ongoing to evaluate potential treatments. So, the best way to prevent this coronavirus and slow down transmission is to be well-informed about COVID-19, the disease and its cause and the way it spreads. The only way to protect ourselves and others from the virus is by frequently cleaning hands with soap or alcohol based-rub, use sanitizer, maintain social distance (1 m), practice respiratory hygiene and do 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 (Han *et al.*, 2009) have recognized a 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–

The current issue and full text archive of this journal is available on Emerald Insight at: **https://www.emerald.com/insight/1708-5284.htm**

World Journal of Engineering
 [ISSN 1708-5284]
[DOI 10.1108/WJE-03-2021-0145]

Received 11 March 2021
Revised 1 April 2021
Accepted 16 April 2021

2003 (Kwok *et al.*, 2019). COVID-19 has spread across the globe around 209 countries (Keras Documentation, 2021). The virus is rapidly raising infection in the USA, Spain, Italy, France, Germany and the UK. As per statistics as on June 12, 2020, this virus is infecting 7.27 million people and has 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 USA with 2.06 million and deaths are 115,000. When deaths are concern, Italy is suffering with 34,167 deaths. The spread of COVID-19 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 COVID-19are fever, cough, sneeze, the shortage of sniff and breathing complications (Watts and Strogatz, 1998; Pastor-Satorras and Vespignani, 2001). In most of the serious cases, these viruses can cause 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 COVID-19-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; for example, let us say Ravi got infected with COVID-19 yesterday, but he will not know about the same for the subsequent 14 days; Ravi felt he is healthy, but he can infect averagely 10 persons a day. Now, these 10 persons who felt they are completely healthy can infect 100 persons per day. These 100 persons felt they are completely alright, but they can infect 1,000 persons and so on. Here, no one knows who is healthy and who can infect who. So, the only prevention is stay at home or in quarantine.

However, monitoring and anticipating the pandemic of infection turns out to be incredibly critical for the decision-making against the general well-being calamity. Measurable scientific models have gained more consideration in clinical science and the study of disease transmission (Anderson, 1999; Koopman, 2004; Gomathi *et al.*, 2020). Some dynamic models are likewise intended for a particular epidemic (Nair *et al.*, 2021). For instance, the dynamic models are intended to simulate and forecast the outbreak of human immunodeficiency virus and SARS (Kwok *et al.*, 2019; Soni *et al.*, 2020a; Mukesh and Gomathi, 2020; Levin *et al.*, 1997). As the enlargement of different models, machine learning and the complex network are joined with the fragmented model and generate a likely route of the widespread forecast (Ng *et al.*, 2003; Dye, 2003).

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 the USA/Brazil/ Russia/ the UK/Italy/South Korea/China;
- to explore and analyze complete worldwide data; and
- to propose a recurrent neural network (RNN) prototypical model to predict and forecast COVID-19 in India.

The organization of remaining paper is as follows: the methodology of the proposed model is presented in Section 2, the data sets are explored in Section 3; experimental outcomes and analysis are discussed in Section 4, and finally, final remarks are presented in Section 5.

2. Methodology

This article proposes a solution based on standard long short-term memory (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 RNN. 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 Figure 1. Those are input layer, LSTM layer and fully connected layer (dense).

2.1 Input layer

The COVID-19 data sets (Sudalai, 2021a; Sudalai, 2021b) are taken in the input layer, and it contains several comma-separated values (CSV) files, which contains information about coronavirus. Next, data preprocessing is applied to acquire required data for the model.

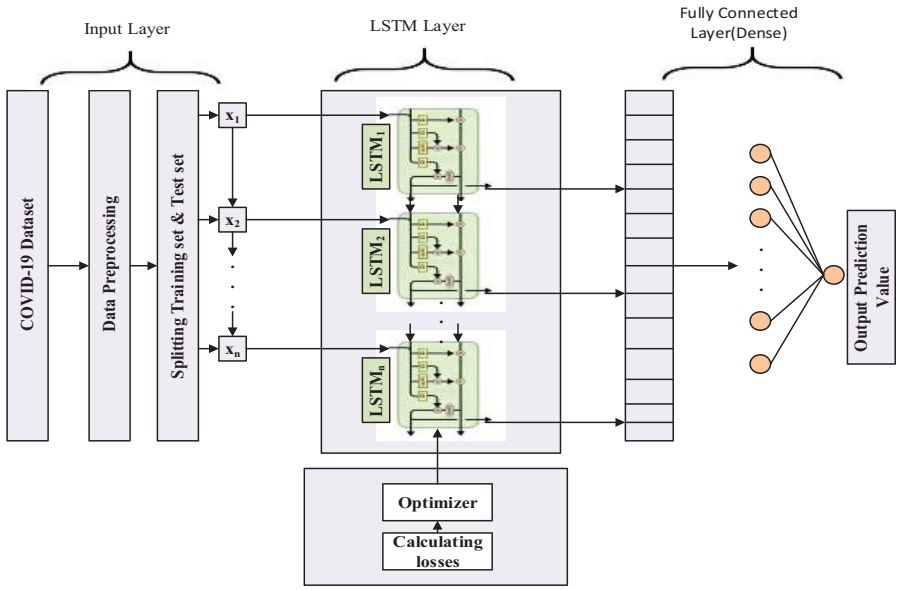
The first step in data preprocessing is importing libraries such as Numpy, Matplotlib and Pandas. Next, CSV files are loaded to pandas, which are used to do some simple scrubbing operations such as eliminating info you do not want that will causes data processing dawdling. For instance, eliminating the 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 (Abadi *et al.*, 2016; WHO, 2020). 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 (Drosten *et al.*, 2003; Chen *et al.*, 2020). Finally, the data is passed to the next layer, i.e. LSTM layer.

2.2 Long short-term memory layer

The LSTM layer contains several LSTM units. The initial phase in the LSTM unit is to select what data is going to pass from the cell state. A "forget gate layer" that 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:

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

Figure 1 Proposed LSTM architecture



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" is termed sigmoid layer. It chooses what values are to be update, and it is shown in equation (2). Further, a tanh layer generates a vector of new candidate values, $\widetilde{C}_t$ 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:

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

$$\tilde{C}_t = \tanh\left(W_C.\left[h_{t-1,}x_t\right] + b_{C)}\right. \tag{3}$$

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 $i_t * \widetilde{C}_t$. The generated values are candidate values, scrabbled by how much obvious to reform every state cell:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}\mathrm{t} \tag{4}$$

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:

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

$$h_t = o_t * tanh(C_t) \tag{6}$$

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 (Gorbalenya *et al.*, 2020). Especially, it associates the yield of LSTM layer to a preferred yield size. The sigmoid function is used in the dense layer 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. Data set description

In this article, several data sets are studied for COVID-19 analysis, visualization and prediction. Finally, two different sources of data set are used. Those are "COVID-19 in India" data set (Sudalai, 2021a) and "novel coronavirus 2019" data set (Sudalai, 2021b). COVID-19 India data set 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 data set contains a spreadsheet consists of worldwide confirmed, deaths and recovered persons. Table 1 gives knowledge on every data set 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 used CSV files.

4. Experimental analysis and discussion

This section describes the main aspects of experimental work carried out in this study. Concretely, new coronavirus data sets (Sudalai, 2021a; Sudalai, 2021b) are used to analyze the COVID-19 outbreak in various countries and predict future spread. The proposed LSTM model is implemented using Keras (Dye, 2003), which is a deep learning package of Tensorflow (Lipsitch, 2003). 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 (Abadi *et al.*, 2016). An advanced

Table 1 Different data set sources of COVID-19

Data set	Description	Columns
COVID-19 in India data set (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 data set (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 data sets from different sources

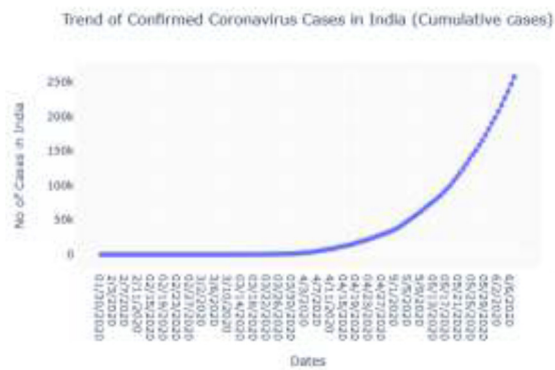
Column	Description
S No	Serial number
Observation Date	Observation date in mm/dd/yyyy
state	Province or state of the observation
Country	Country or region of the observation
Last Update	Last update date time in UTC
Confirmed	Cumulative quantity of confirmed cases
Deaths	Cumulative quantity of deaths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

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 Figure 2, which is a scatter plot. This graph shows the sum of

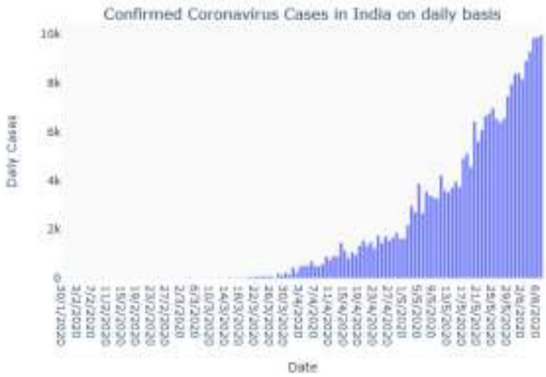
Figure 2 Visualization of confirmed corona cases in India



cumulative confirmed corona cases in India day wise. The date format used in the graphs is "MM/DD/YYYY." The first corona case was found in India on January 30, 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 7 June is 257,486.

The day-to-day confirmed corona cases in India are visualized in Figure 3. From the plot, it is observed that very little cases in February, i.e. two 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 1,397, which is presented in the Figure 2. The graph depicts that the corona cases are exponentially increasing. From 13 April onwards, confirmed cases are rising in thousands every day. on 13 April, 1,243 cases are recorded. Finally, the highest number of cases, i.e. 9,971 was noted on 7 June up to now. The daily cases are visualized up to 7 June, but still it is continuing.

Figure 3 Confirmed corona cases in India on a daily basis



The trend of coronavirus cases in India are compared in Figure 4. The confirmed cases are exponentially growing. The first death happened on 11 March. The deaths are growing almost sequentially. The first three coronavirus patients recovered on 16 February. The recovered cases are successively growing. As per observation in the Figure 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 June 7, 2020, is 7,207, whereas recovered cases are 123,848, and total confirmed cases are 257,486. The total recovered cases are 48.09% of confirmed cases.

4.2 Worldwide COVID-19 outbreak analysis

The comparison among most affected five countries with COVID-19 including China, Italy and South Korea is shown in Figure 5. The trend of coronavirus in eight countries such as the USA, Brazil, Russia, the 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 the USA. In the

Figure 4 Comparison of COVID-19 cases in India

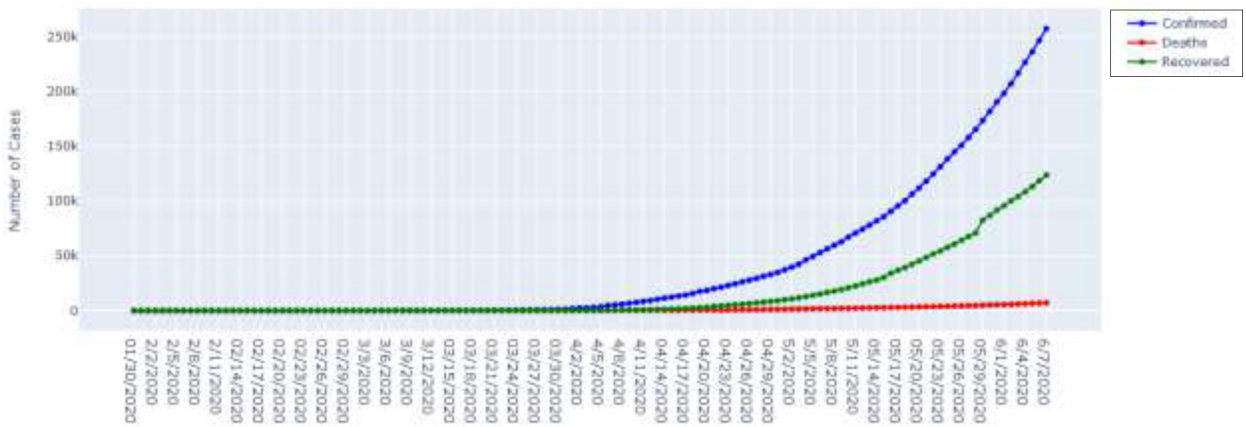
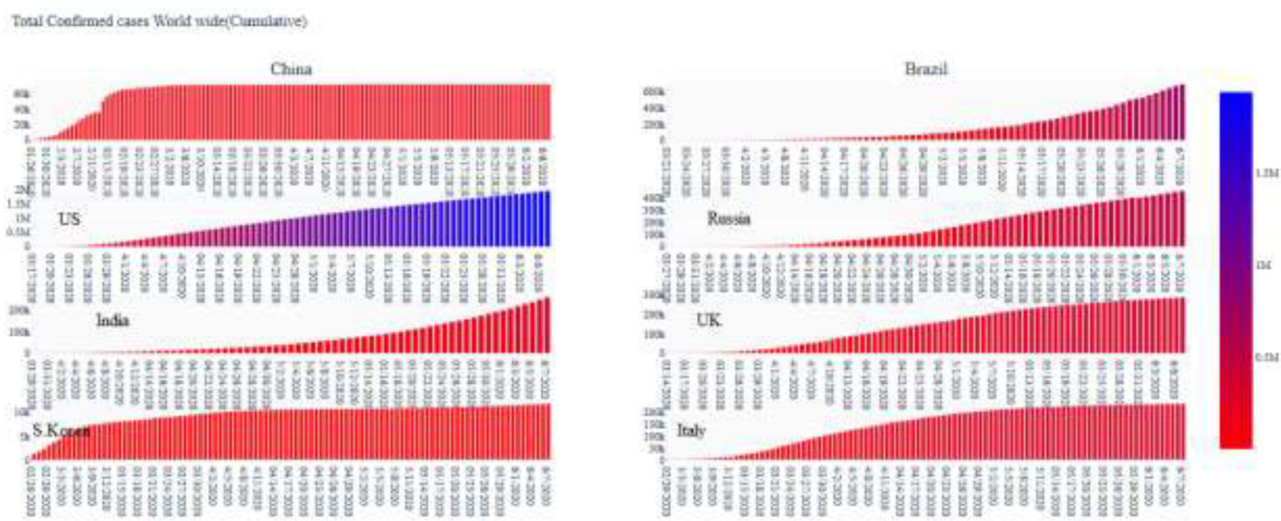


Figure 5 Comparison of coronavirus cases in various countries



USA, the highest total number of cases was recorded on June 7, 2020, i.e. 1.95 million. The US graph indicates that the corona growth in the USA is a completely exponential curve, and a greater number of cases were recording in the USA because the then US president did not consider COVID-19 as a serious pandemic. Next highest cases werenoted in Brazil, i.e. 691,758. The Brazil plot shows that the growth is exponential. Brazil took too much time to know the serious about the COVID-19 pandemic.

Russia graph is also an exponential curve, which shows a logistic growth in certain parameters. The UK is started a completely sigmoid curve since the 15 May because an extensive testing is done. Next, if you observe the India plot, it is an exponential graph, especially from 28 March, but the number of cases are very less because the Prime Minister of India announced the first phase of lockdown from 23 March to 14 April, the second phase up to 3 May, and the third phase up to May 17, 2020. The pattern between the USA and India is almost same, but the number of cases in India are very less compared to the USA, but after the third phase, lockdown cases are increasing rapidly. The total quantity of corona cases in India up to June 7, 2020 are 257,486. The South Korea graph is a sigmoid curve, which shows logistic growth in certain parameters. The South Korea is started completely sigmoid curve since 7 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 12 and 13 February. Later, it goes on rising; again, this is also following a sigmoid pattern successfully from 3 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

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

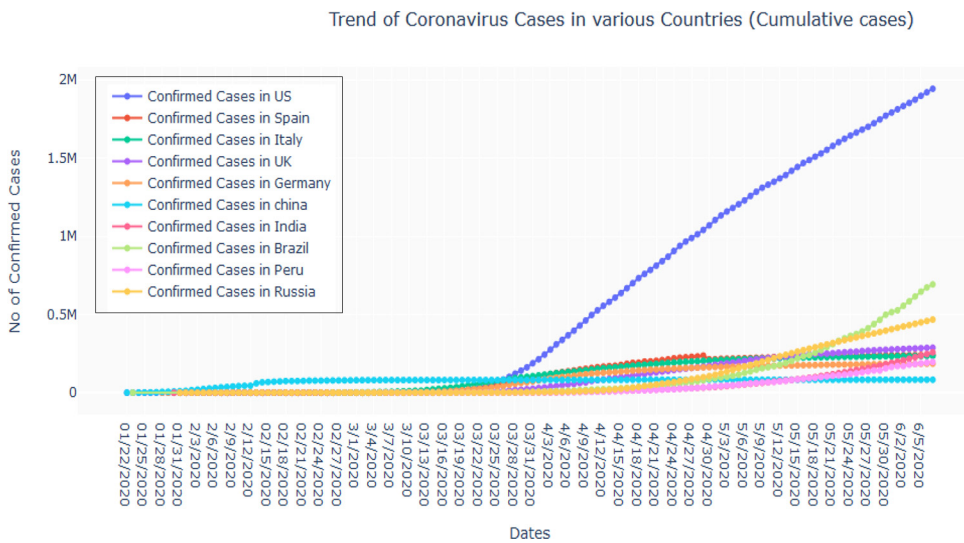


Figure 7 Trend after crossing 1,000 confirmed cases in various countries

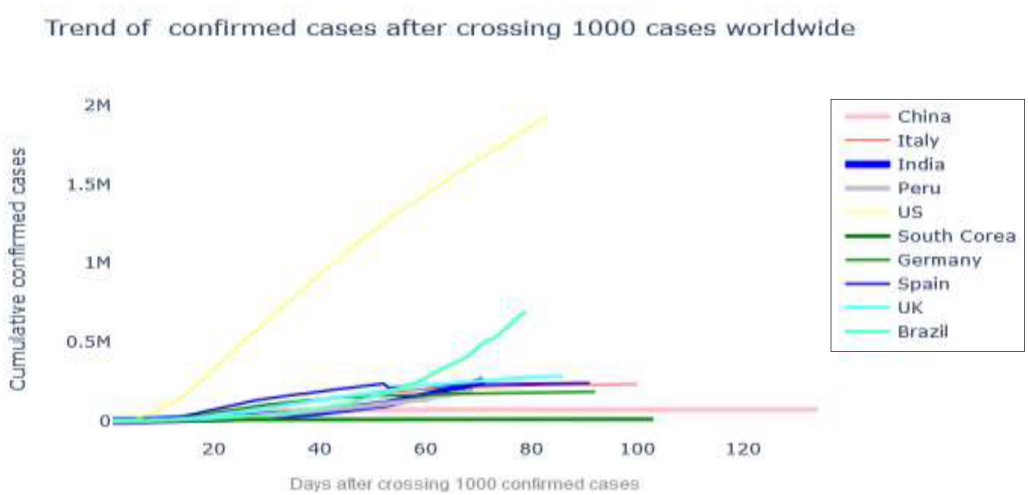


Figure 8 Comparison of worldwide COVID-19 cases

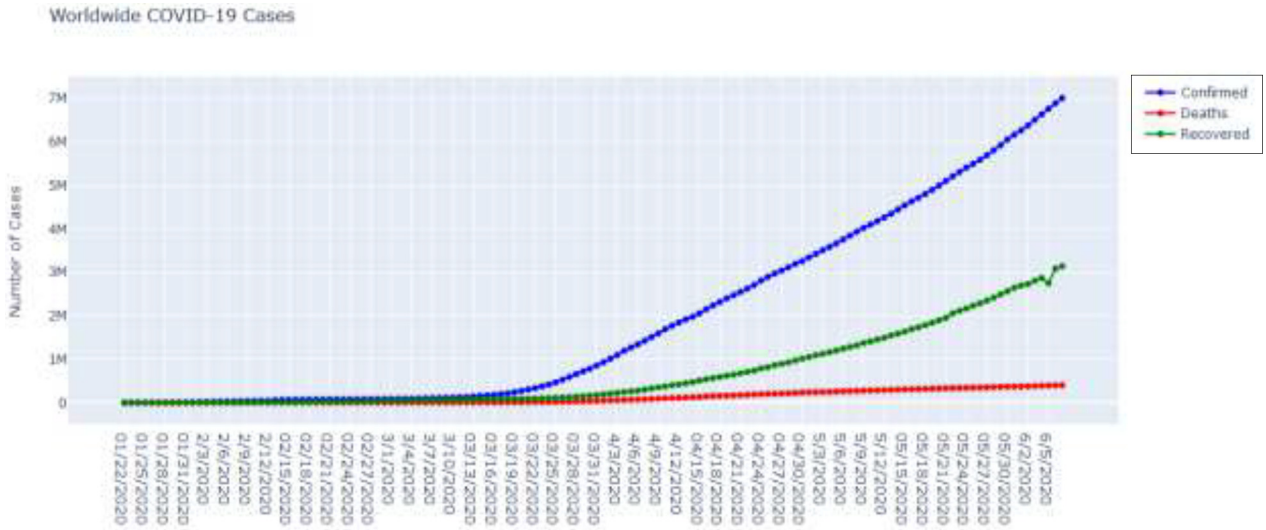


Figure 9 Gender wise confirmed cases analysis

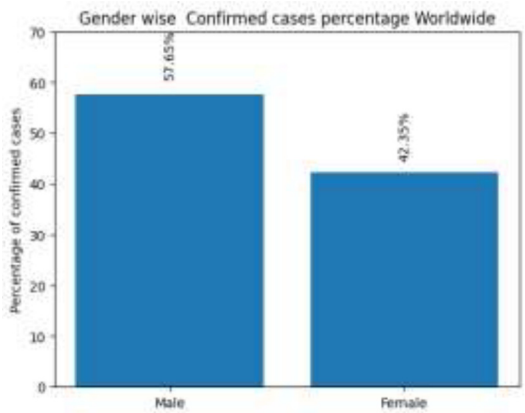
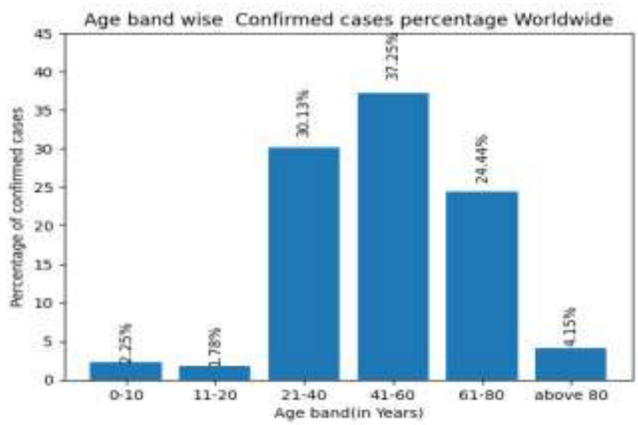


Figure 10 Age wise confirmed cases analysis



and sends them directly into testing. The Italy plot shows that the growth is exponential up to 22 March; later it becomes flatter. Italy took too much time to know the serious about the COVID-19 pandemic.

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

The trend of countries after crossing 1,000 cases has been compared in Figure 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 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 1,000; South Korea has the minimum number of cases out of all other countries, i.e. 11,814 in 103 days, which is better when compared to spread of this pandemic in other countries in the world. The USA has 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. 691,758 cases in 79 days.

The impact of pandemic on the world with three separate spread sheets for confirmed, deaths and recovered cases is plotted in the Figure 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 June 7, 2020, the total confirmed cases worldwide are 7 million. The recovered cases are not exponential, which is shown in the green color in the Figure 8, and still now the recovered cases are

Figure 11 Comparison of COVID-19 actual cases

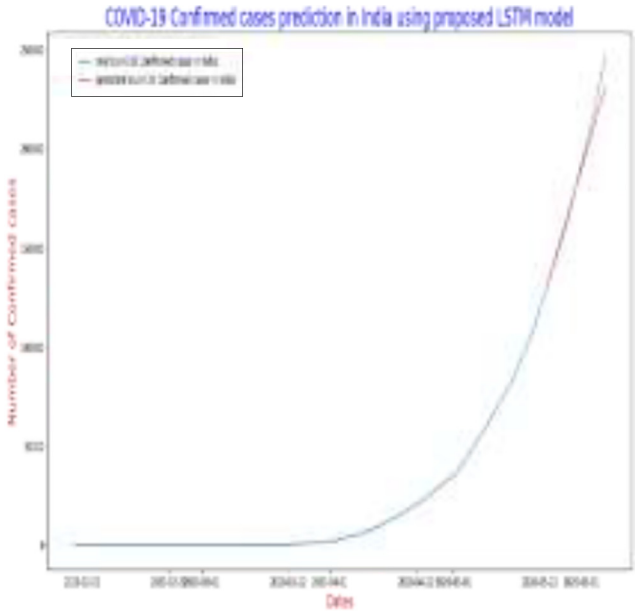
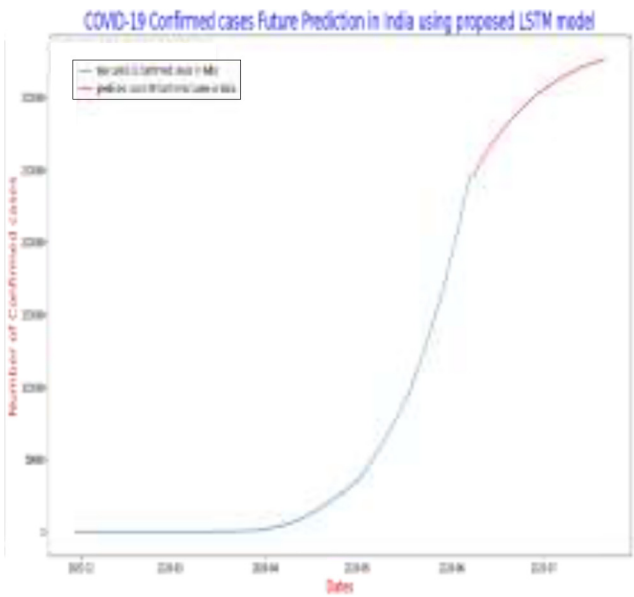


Figure 12 Future prediction of COVID-19 and predicted confirmed cases in India



3.14 million, and it is 44.81% of confirmed cases. The deaths are very less compared to confirmed cases. The number of deaths happened still now is 402,730. The deaths had happened 5.74% out of the confirmed cases.

When the worldwide COVID-19 data is analyzed, then it is noticed that the coronavirus is infected for different category people with different ratio such as shown in Figure 9 and 10. The gender wise infected corona cases are compared in Figure 9. The male-infected worldwide corona cases are 57.65%, whereas the female-infected cases are 42.35%, which

Figure 13 Loss of confirmed cases in India

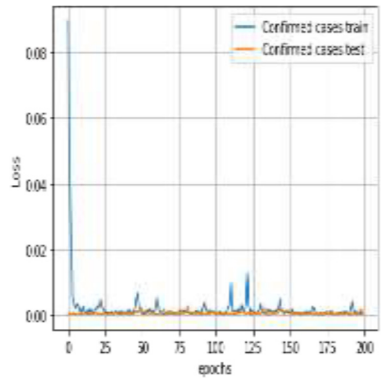
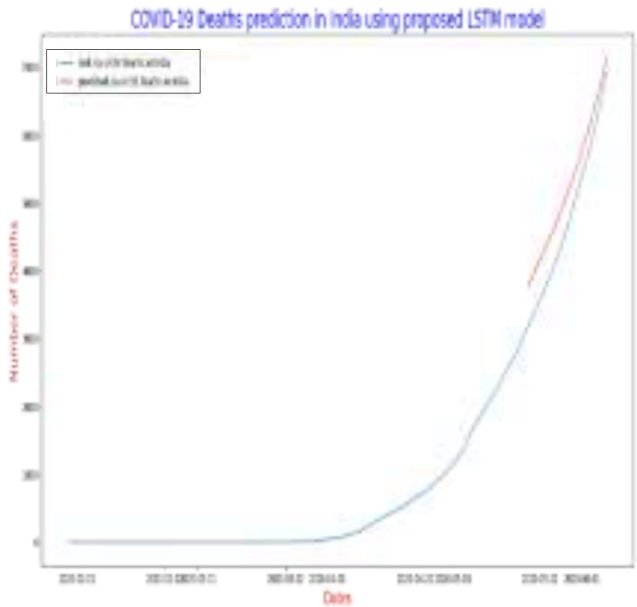


Figure 14 Comparison of COVID-19 actual deaths in India



are shown in Figure 9. The graph indicates that the male are more affected when compared to female. Age wise corona affected cases are compared in Figure 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 home-quarantined 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: three 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

Figure 15 Future prediction of COVID-19 and predicted deaths in India

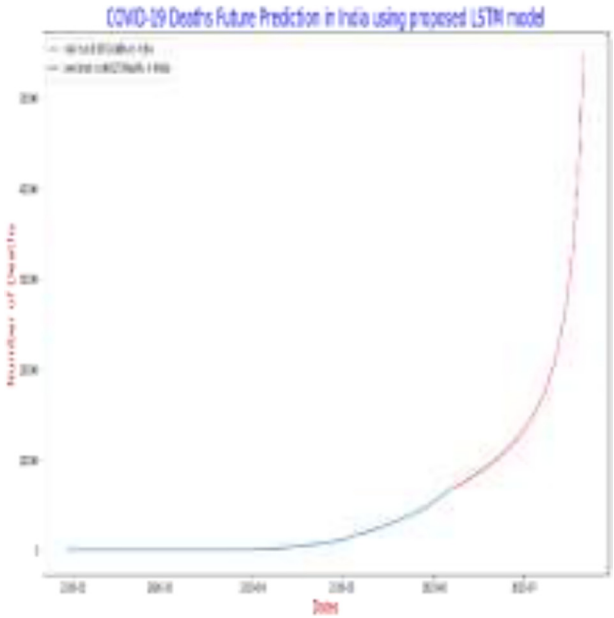
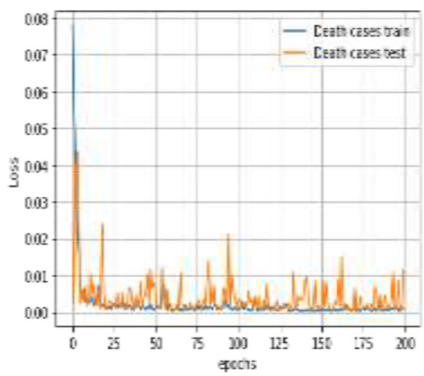


Figure 16 Loss of death cases

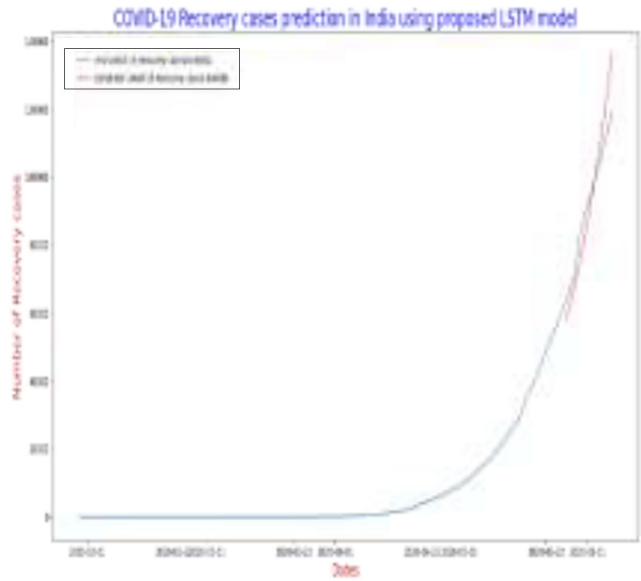


optimizer, loss function is mean square error (MSE), the number of epochs 200, and finally one dense layer to predict one value each time.

4.3.1 Confirmed

The whole data set 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 Figure 11. From the graph, it is observed that the loss is very less. The confirmed cases' data is considered from February 1, 2020, to June 7, 2020. The future confirmed cases are predicted for 45 days, i.e. from June 8, 2020, to July 21, 2020, which is shown in Figure 12. The cumulative confirmed cases are reached the peak on July 15, 2020, i.e. 4 lakhs. Later, the confirmed cases' graph is flattened. From the Figure 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 Figure 13.

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



Initially, the loss is more for training when number of epochs 5. Later, it decreased when the number of epochs increases.

4.3.2 Deaths

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 Figure 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 Figure 15. The proposed model is predicted that maximum death cases may reach 50,000 and may happen on July 21, 2020. Later, the curve is expected to flatten. The MSE loss of predicted death cases for training and testing is shown in Figure 16. Initially, the loss is more for training when number of epochs 5. Later, it decreased when the number of epochs increases. The testing loss is fluctuating between 0 and 0.02.

4.3.3 Cured

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 Figure 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 Figure 18. The proposed model is predicted the maximum recovery cases may reach 160,100, and may happen on July 10, 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 Figure 19. Initially, the loss is more for training when number of epochs 5. Later, it is decreased when the number of epochs increases.

Figure 18 Future prediction of COVID-19 recovery cases in India

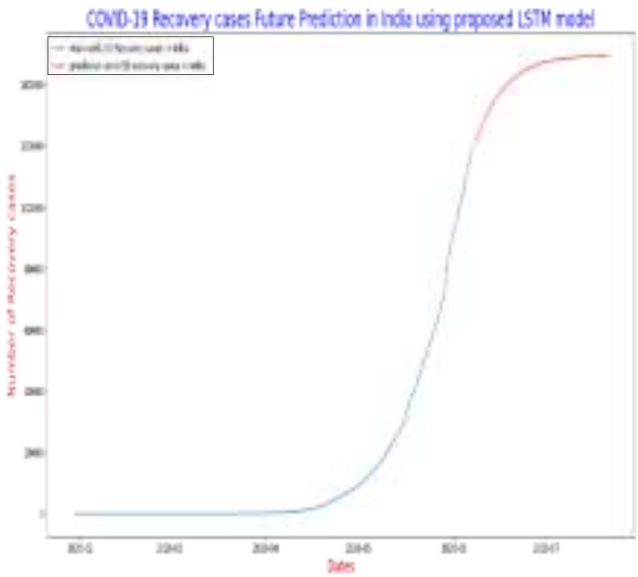
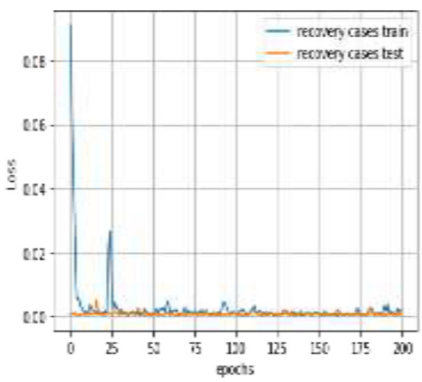


Figure 19 Loss of recovery cases



5. Conclusion

In conclusion, "COVID-19 in India" and "Novel coronavirus 2019" data sets are used to examine the data with visualization and 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 July 15, 2020, deaths of 50,000 on July 21, 2020, and recovered cases of 160,100 on July 10, 2020 in India. Moreover, the curves are flattened after peak cases recorded. It is an uncommon flare-up of COVID-19 in the 20th 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

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M. and Kudlur, M. (2016), "Tensorflow: a system for large-scale machine learning", *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pp. 265-283.

Anderson, R.M. (1999), "The pandemic of antibiotic resistance", *Nature Medicine*, Vol. 5 No. 2, pp. 147-149.

Chen, Y., Liu, Q. and Guo, D. (2020), "Emerging coronaviruses: genome structure, replication, and pathogenesis", *Journal of Medical Virology*, Vol. 92 No. 4, pp. 418-423, doi: 10.1002/jmv.25681.

Colab research framework (2021), "Colab research framework", available at: https://colab.research.google.com/

Drosten, C., Günther, S., Preiser, W., Van Der Werf, S., Brodt, H.R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A. and Berger, A. (2003), "Identification of a novel coronavirus associated with severe acute respiratory syndrome", *New England Journal of Medicine*, Vol. 348 No. 20, pp. 1967-1976.

Dye, C. (2003), "Epidemiology: modeling the SARS epidemic", *Science*, Vol. 300 No. 5627, pp. 1884-1885.

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*, doi: 10.1108/WJE-09-2020-0450.

Gorbalenya, A.E., Baker, S.C., Baric, R., Groot, R.J.D., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A.M., Neuman, B.W. and Penzar, D. (2020), "Acute respiratory syndrome-related coronavirus: the species and its viruses – a statement of the coronavirus study group", *bioRxiv*, doi: 10.1101/2020.02. 07.937862.

Han, X.N., De Vlas, S.J., Fang, L.Q., Feng, D., Cao, W.C. and Habbema, J.D.F. (2009), "Mathematical modelling of SARS and other infectious diseases in China: a review", *Tropical Medicine & International Health*, Vol. 14, pp. 92-100, doi: 10.1111/j.1365-3156.2009.02244.x.

Keras Documentation (2021), "Keras documentation", 609, available at: https://keras.io

Koopman, J. (2004), "Modeling infection transmission", *Annual Review of Public Health*, Vol. 25 No. 1, pp. 303-326.

Kwok, K.O., Tang, A., Wei, V.W., Park, W.H., Yeoh, E.K. and Riley, S. (2019), "Epidemic models of contact tracing: systematic review of transmission studies of severe acute respiratory syndrome and Middle east respiratory syndrome", *Computational and Structural Biotechnology Journal*, Vol. 17, pp. 186-194.

Levin, S.A., Grenfell, B., Hastings, A. and Perelson, A.S. (1997), "Mathematical and computational challenges in population biology and ecosystems science", *Science*, Vol. 275 No. 5298, pp. 334-343.

Lipsitch, M. (2003), "Transmission dynamics and control of severe acute respiratory syndrome", *Science*, Vol. 300 No. 5627, pp. 1966-1970.

Mukesh, S. and Gomathi, S. (2020), "Cotton leaf spot disease detection using multi-class SVM", *International Journal of Research in Engineering and Advanced Technology*, Vol. 8 No. 5, p. 149.

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*, doi: 10.1108/WJE-10-2020-0529.

Ng, T.W., Turinici, G. and Danchin, A. (2003), "A double epidemic model for the SARS propagation", *BMC Infectious Diseases*, Vol. 3 No. 1, p. 19.

Pastor-Satorras, R. and Vespignani, A. (2001), "Epidemic spreading in scale-free networks", *Physical Review Letters*, Vol. 86 No. 14, pp. 3200-3203.

Soni, M., Chauhan, S., Bajpai, B., (2020a), and T. and Puri, "An approach to enhance fall detection using machine learning classifier", *2020 12th International Conference on Computational Intelligence and Communication Networks (CICN)*, *Bhimtal*, pp. 229-233, doi: 10.1109/CICN49253.2020.9242634.

Sudalai, R.K. (2021a), "COVID-19 in India", available at: www.kaggle.com/sudalairajkumar/covid19-in-india

Sudalai, R.K. (2021b), "Novel corona virus 2019", available at: www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset

Watts, D. and Strogatz, S.H. (1998), "Collective dynamics of 'small-world' networks", *Nature*, Vol. 393 No. 6684, pp. 440-442.

WHO (2020), "Novel coronavirus – China 12 January 2020", available at: www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/ (accessed 19 January 2020).

Further reading

Dharmadhikari, S.C., Gampala, V., Rao, C.M., Khasim, S., Jain, S. and Bhaskaran, R. (2021), "A smart grid incorporated with ML and IoT for a secure management system", *Microprocessors and Microsystems*, Vol. 83, doi: 10.1016/j.micpro.2021.103954.

Gampala, V., Anupriya, K. and Balajee, M. (2020), "Enhance the educational outcome in higher educational institutes through deep neural network", *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 9 No. 5, pp. 8315-8324.

Gampala, V., Kumar, M.S., Sushama, C. and Raj, E.F.I. (2021a), "Deep learning based image processing approaches for image deblurring", *Materials Today: Proceedings*, doi: 10.1016/j.matpr.2020.11.076.

Gampala, V., Shilpa, V.J., Kishore, S.K. and Mydhili, S.K. (2021b), "Enhancement of resolution and image reconstruction in digital image correlation", *Materials Today: Proceedings*, doi: 10.1016/j.matpr.2020.11.758.

Soni, M., Gomathi, S. and Bhupendra Kumar Adhyaru, Y. (2020b), "Natural language processing for the job portal enhancement", *2020 7th International Conference on Smart Structures and Systems (ICSSS)*, *Chennai*, pp. 1-4, doi: 10.1109/ICSSS49621.2020.9202046.

Corresponding author

Veerraju Gampala can be contacted at: veerraju5a8@gmail.com