A deep learning approach for prediction of SARS-CoV-2 cases using the weather factors in India

Kantha Rao Bhimala [a], Gopal Krishan Patra [a], Rajashekar Mopuri [b], Srinivasa Rao Mutheneni [b]*

[a] CSIR Fourth Paradigm Institute (CSIR-4PI), Wind Tunnel Road, Bangalore-560037, Karnataka, India.

[b] ENVIS Resource Partner on Climate Change and Public Health, Applied Biology Division, CSIR-Indian Institute of Chemical Technology (CSIR-IICT), Tarnaka, Hyderabad-500007, Telangana, India.

Running Title: Role of climate on COVID-19 transmission and Prediction

*** Corresponding Author:**

Dr M Srinivasa Rao
Senior Scientist
Coordinator ENVIS Resource Partner of Climate Change & Public Health
Applied BiologyDivision
CSIR-Indian Institute of Chemical Technology
Tarnaka, Hyderabad-500007
Telangana, India.

Abstract:

Advanced and accurate forecasting of COVID-19 cases play a crucial role in management of hospital facility, policy decision, logistic support, and economy of the country. Artificial Intelligence (AI) techniques have proved its capability in time series forecasting of the non-linear problems. The present study assessed the relationship between weather parameters and COVID-19 cases and found the specific humidity have strong positive association, maximum temperature have negative and minimum temperature have positive association in most of the states in India. Further, we have developed a weather integrated LSTM (long short term memory) models for advanced (1-14 days) forecasting of the COVID-19 cases over different states in India. To achieve the goal we have utilized the humidity and temperature time series data along with the COVID-19 confirmed cases data (1^{st} April-30^{th} June 2020) to optimise the LSTM model in univariate and multivariate modes. The optimised models are utilized to forecast the COVID-19 cases for the period 1^{st} July, 2020 to 31^{st} July 2020 with 1 to 14days lead time. The results shows that the univariate LSTM model (past COVID-19 input) have reasonably good skill (Relative Error < 20%) in short range forecast (1day lead) for most of the selected states, whereas the skill is degraded with the medium and long range forecast. The major finding of the current study is that the medium range (1-7days) forecasting skill is enhanced in some of the states with the weather integrated multivariate LSTM models. The states (Maharashtra, Gujarat, Rajasthan, Madhya Pradesh, Haryana, and Punjab) located in West and North West India region, humidity play a key role in enhancement of medium range forecasting skill of the LSTM model. It is also observed that the states located in high humid regions (Kerala, Tamil Nadu, and West Bengal) temperature plays a key role in model enhancement.

Keywords: SARS-CoV-2, COVID-19, Humidity, Temperature, Prediction, LSTM, India

Introduction:

The corona virus disease 2019 (COVID-2019) epidemic was caused by the severe acute respiratory syndrome coronavirus 2 (SARS COV-2) began in Wuhan city, Hubei, China in early December, 2019 [Li et al., 2020; Shen et al., 2020]. On 7th January 2020, this novel strain of SARS COV-2 was isolated and confirming the circulation in populace and causes coronavirus disease (COVID-2019). Since then the disease has quickly spread globally (216 countries/areas or territories) and established local epidemics in many countries including USA, Europe and Asia. The rapid spread of the COVID-19 cases may be due to lack of information at the early stage of the epidemic [Zhong et al., 2020]. On 30th January 2020, WHO (World Health Organisation) declared the COVID-19 outbreak as a public health emergency of international concern [WHO 2020a]. As of 13th October 2020, more than 37 million of COVID-19 confirmed positive cases and morethan one million deaths reported in the world [WHO 2020b]. Human to human contact or through respiratory droplets (produced when infected individual sneezes or coughs), transmission through touch of a surface or object contain the virus and aerosol transmission are the main routes of the transmission of coronavirus [Huang et al., 2020; Xu et al., 2020]. The typical clinical symptoms of COVID-19 consist of fever, dry cough, myalgia, pneumonia and may cause alveolar damage leads to respiratory failure and death occurs [Huang et al., 2020]. This novel coronavirus has spread faster than its two ancestor viruses severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East Respiratory Syndrome (MERS) may be due to high transmission rates produced by asymptomatic carries [Vellingiri et al., 2020; Bai et al., 2020]. Similarly, the SARS-COV 2 is closely related to bat derived viruses bat-SL-CoVZC45 and bat-SL-CoVZXC21 and distinct from SARS-CoV (-79% similarity) and MERS-CoV (-50% similarity) [Lai et al., 2020; Jiang et al., 2020; Liu et al., 2020].

Along with other countries the COVID-19 cases also reported in India. The first case of COVID-19 was reported on 30th January 2020 in Kerala, a student who returned from the Wuhan University in China [Rawat 2020]. The number of corona cases are gradually increasing across the country hence, to flatten the curve, India suspended visas for all international travelers from 13th March 2020 onwards. Followed by travel ban, the Government of India announced a phase wise national wide lockdown (1st lockdown from 25 March to 14 April; 2nd lockdown from 15 April to 3 May; 3rd lockdown 4 May to 17 May; 4th lockdown from 18 May to 31 May) to minimize the human activity in the country [Ministry of Health & Family Welfare, GOI]. From 1st June, 2020, the unlock processes started except for containment zones and the testing capability (~1million/day in the month of September) increased rapidly to identify and isolate the infected people to minimize the spread over most of the states in India. The all India positivity rate (percentage of confirmed among the total tests) is between 8-9%, whereas some of the states located in south India have more positivity rates including Maharashtra (20%), Andhra Pradesh (12.3%), Karnataka (12%), Goa (10.4), and Tamil Nadu (8.6%) (ICMR). Indian Council of Medical Research (ICMR) conducted the COVID-19 tests among the severe acute respiratory illness (SARI) patients at the early stage of the pandemic in India, they found that 1.8% (104 out of 5911) of SARI patients tested positive for COVID-19 from 52 districts located in 20 states/Union Territories, the positivity rate was zero during the period February 15 to March 14, 2020 and increased up to 2.6% during the period March 15 to April 02, 2020. ICMR also conducted the population-based seroepidemiological study to measure the extent of COVID-19 infection in the country and found that 0.73% (6.4 million) of the adults exposed to the corona virus by the early May 2020 (ICMR COVID study group Abraham et al., 2020; Gupta et al., 2020; Murhekar et al., 2020).

Transmission of viruses can be influenced by several factors including contact, droplet, airborne, fomite, mother to child, and animal to human (WHO). In case of airborne transmission mode, climatic conditions (temperature, humidity, and wind speed) play an important role in viability of the virus in the surrounding environment. At the early stage of pandemic, number researchers have studied the relationship between weather parameters and COVID-19 cases and reported that temperature has positive association and humidity have a negative association with the COVID-19 cases over most of the regions in the world (Liu, et al., 2020; Briz-Redón et al., 2020; Chen et al., 2020, Oliveiros et al., 2020, Sahin 2020; Bashir et al., 2020; Ma et al., 2020; Wang et al., 2020). Whereas the detailed experimental study by Dbouk and Drikakis, 2020 shown that the viability of SARS-CoV-19 in a cloud of airborne saliva droplets is significantly influenced by the evaporation rate of the surrounding environment. The study also found a contrary results from the existing knowledge that the virus viability in the atmosphere significantly reduced with the high temperature and low humidity, it is also found that if relative humidity is high the virus viability is high at any temperature.

Globally, various researchers are trying to develop an accurate COVID-19 prediction models based on the statistical, dynamical, and Artificial Intelligence (AI) techniques, and some of the studies suggested that the AI based approaches shown better skill compared to the other traditional mathematical techniques. However, some of the earlier studies shown that the RNN (Recurrent neural networks) based LSTM models (basically univariate nature) have better skill among the AI based techniques for COVID-19 predictions. In the case of India, a very few number of researchers (Tomar and Gupta, 2020; Arora et al., 2020; Shastri et al., 2020) assessed the skill of LSTM models (Deep LSTM, stacked LSTM, convolutional LSTM, Bi-directional LSTM) for COVID-19 case forecasting and their results shown that the model skill was high in short range prediction compared to the medium and long range prediction. Hence, to improve the LSTM model skill in short, medium, and long range forecasting, the present study proposed a modelling framework which is integrating the weather data into LSTM (multivariate LSTM) system for advanced forecasting of COVID-19 cases over most of the effected states in India.

Methods:

Data:

All 28 states and 08 Union Territories of India covering latitude (8°N-38°N) and longitude (68°E-98°E) were considered for the study. Daily counts of laboratory confirmed COVID-19 cases of all the states of India were collected from Ministry of Health and Family Welfare (MoHFW), Govt. India from 1st April to 31st July, 2020. Similarly, the daily meteorological parameters of specified period consist of temperature (minimum, maximum and mean) and specific humidity (SH) extracted from NCEP/NCAR reanalysis data [Kalnay et al.,1996] (https://psl.noaa.gov/).

Correlation Analysis:

To understand the weather impact on COVID-19 cases, the lag (0-14) correlation coefficients are computed between daily meteorological parameters and daily COVID-19 cases for different states in India during the period 1st April 2020 to 31st July 2020. The popular statistical formula of Pearson Correlation Coefficient (r) is utilized to measure the strength and linear relationship between daily COVID-19 cases (X) and surface meteorological parameters (Y), and the values are ranging between -1.0 to 1.0. The correlation coefficient values are computed as

$$r(X,Y)=\frac{\sum_{i=1}^{N}(X_i-\overline{X})(Y-\overline{Y})}{\sqrt{\sum_{i=1}^{N}(X-\overline{X})^2\sum_{i=1}^{N}(Y-\overline{Y})^2}}$$

Long-Short Term Memory (LSTM) Model:

A Long Short-Term Memory (LSTM) network is a kind of Recurrent Neural Network (RNN) that attempts to model time or sequence dependencies (Hochreiter and Schmidhuber, 1997**;** Sagheer and Kotb 2019; Shastri et al., 2020; Arora et al., 2020). LSTM falls under the category of Deep learning and it is performed by feeding back the output of a neural network layer at time *t* to the input of the same network layer at time *t + 1*. The proposed work was carried out using the Keras implementation of a LSTM network (Fig.1). The computations were carried out on a five-node system each with an eight-core Intel i7-9700 CPU working at 3 GHz and 32 GB memory each with keras.

The block diagram of a basic multi-input LSTM network and the memory transformation between each cell of LSTM was presented in Fig.1a and Fig.1b. The LSTM cell consists of three gates: input gate (i_t), forget gate (f_t) and output gate (o_t) with different functionality (Fig.1c). The forget gate is responsible for forgetting information that is not required anymore, while the input gate is used for adding new useful information. The output gate updates the hidden states at every time step. Each gate is a feed forward neural network with a number of hidden units as shown in Fig.1d. The mathematical representation of LSTM is given below in Eqs. (1)-(5) (Hochreiter and Schmidhuber, 1997).

$$i_t=\sigma(w_i x_t+u_i h_{t-1}+b_i) \text{----------------------------(1)}$$

$$f_t=\sigma(w_f x_t+u_f h_{t-1}+b_f)\text{----------------------------(2)}$$

$$o_t = \sigma\left(w_o x_t + u_o h_{t-1} + b_o\right) \text{-----------------------------(3)}$$

$$h_t = o_t \times \tanh\left(i_t \times \tanh\left(w_g x_t + u_g h_{t-1} + b_g\right) + f_t \times s_{t-1}\right) \text{------------------ (4)}$$

Where σ, i, f, o, g represent the sigmoid function, input gate, forget gate, output gate, and un-gated input transformation respectively. The weights (w_i, w_f, w_o, w_g and u_i, u_f, u_o, u_g) are represented in a matrix format, bias (b_i, b_f, b_o, b_g) are represented in vectors, and S_{t-1} represents the cell state of the previous time step.

The present study utilized both univariate and multivariate LSTM models for forecasting the daily cases of a given state. The univariate model (Control Experiment (CTL)) utilizes the confirmed daily COVID-19 cases time series data to forecast the future (next day) COVID-19 cases of the selected state in India. To understand the weather impact on corona virus transmission, four experiments were conducted with multivariate LSTM model (Table1). The time series data (1st April 2020 till 31st July 2020) was divided into two parts and the first three months (April-June) data utilized for training and the last one month (July) data was utilized for testing purpose. The LSTM model was optimized with minimum error by manually considering different hyper-parameters, such as the number of units in the hidden layer, number of hidden layers etc. The univariate and multivariate LSTM model were optimized with different weather parameters separately and utilized for the forecasting purpose. The forecasts are generated with univariate (CTL) and multivariate (CTL_SH, CTL_Tmax, CTL_Tmin, CTL_Tmean) LSTM models and evaluated with observed data (Table1). Further, we have also generated the forecasts with different combination of the weather parameters and evaluated with the observed data of worst effected states for COVID-19 in India.

Experiments	Input data (Time series data)	Output data
CTL	COVID-19 cases	Forecasted COVID-19 cases
CTL_SH	COVID-19 cases & specific humidity	Forecasted COVID-19 cases
CTL_Tmax	COVID-19 cases & maximum temperature	Forecasted COVID-19 cases
CTL_Tmin	COVID-19 cases & minimum temperature	Forecasted COVID-19 cases
CTL_Tmean	COVID-19 cases & mean temperature	Forecasted COVID-19 cases

*CTl: Control experiment, SH: Specific Humidity, TMax: Maximum Temperature, TMin: Minimum Temperature, Tmean: Mean Temperature.

Table1: Description of the LSTM models utilized for the experimental forecast.

Model evaluation:

Relative Error (RE):

The relative error is the ratio between the absolute error and the absolute value of the observation.

$$\Re = \frac{100 * |X_m - X_o|}{|X_o|}$$

Where X_m is the model forecasted and X_o is the observed COVID-19 cases in a single day.

Results:

Figure-2 depicts the spatial distribution of state wide cumulative COVID-19 cases reported between the months of April and July, 2020. In India, the first COVID-19 case was reported from the state of Kerala and the virus rapidly expanded to the other regions during the pre-monsoon season. Before the south west monsoon onset, there was only 182,143 cumulative cases observed in India and the majority of the cases were reported from the states including Maharashtra, Delhi, Tamil Nadu, and Gujarat. After the onset of monsoon, there was a rapid growth in confirmed cases (cumulative cases during June and July > 14 lakhs) and by the end of July more than 16 Lakh cases reported in India. However, the maximum number of cases were reported from the southern part of India (Maharashtra, Andhra Pradesh, Tamil Nadu, and Karnataka), moderate cases were reported from the states located in central, east and west India. The low number of cases were reported from the states located in north and north east India.

Figure-3 dipicts the spatio-temporal variation (current year 2020) of specific humidity (SH), maximum temperature (Tmax), minimum temperature (Tmin), mean temperature (Tmean), at two meters and windspeed at ten meters during the months of April, May, June, and July over India. It was observed that the monthly average SH values was very low (<0.01 kg/kg) over Central India (CI), Northwest India (NWI), and North India (NI), moderate (0.01-0.02 kg/kg) over the states located in east & west coast of India, high (>0.02 kg/kg) over Kerala and some parts of Tamil Nadu during the early stage of COVID-19 in India (April and May). Whereas the SH was slowly increased from south to north during monsoon season (June and July) and the high values were observed in the month of July over Central and East India region. In the case of surface air temperature, we have observed that, a high (low) diurnal temperature range (DTR) during pre-monsoon (monsoon) season over central, North West and north east India regions.

To understand the environmental effects on COVID-19 cases, the lag (0-14 days) correlation coefficients (CC) are computed between daily COVID-19 cases and daily surface meteorological parameters (SH, Tmax, Tmin, Tmean) for the period 1st April 2020 to 31st July 2020. The present study considered 14 days lag correlations due to the symptoms of COVID-19 will appear after the incubation period which is typically ranging between 1 to 14 days. The correlation coefficient values for lag1, lag7, and lag14 over different states of India shown in figure-4. The correlation maps describes that the specific humidity have strong positive association with COVID-19 cases for most of the states in India. Maximum correlation (>0.75) values found in Central and North-West India and moderate correlation (0.5-0.75) values found in East coast & some parts of north India. In majority of the states it was noticed that the lag7 correlations are slightly better than the lag1. In the case of mean temperature and maximum temperature have strong negative association with COVID-19 cases over south India and positive association over foot hills of Himalaya region. Similarly, minimum temperature have strong positive association over north, north west and north east India and weak negative association over south India region.

The COVID-19 transmission rates are mainly depends on the evaporation rate of the contaminated saliva droplets which is released from the infected person to the surrounding environment. The evaporation rate is mainly depends on the humidity, temperature, and wind speed. The contaminated droplets are more resistant to evaporation when the relative humidity is close to the saturation point, which will allow the contaminated droplet cloud move longer distances from the source. The recent study by Roland R. Netz and William A. Eaton 2020, shown that the droplets (released from the infected person while speaking) size larger than 50 μm fall to the ground very fast, whereas the droplet less than this size slowly reduce their radii based on the evaporation rate of surrounding environment and remain airborne for longer duration. Hence, the higher (lower) relative humidity increase (decrease) the airborne virus viability during the calm wind conditions and possible pathway for acceleration in COVID-19 disease outbreak. Our analysis shows that the potential evaporation rates are drastically reduced in India from pre-monsoon season to the monsoon season. The possible reason for significant increase of COVID-19 cases in India during monsoon season is may be the aggravation of airborne transmission (apart from the other modes of transmission) due to less potential evaporation rates in most of the states.

Univariate LSTM model forecast skill:

The present study utilized the three months (01st April 2020 to 30th June 2020) data for training and one month data (01st July to 31st July 2020) for testing the model. The proposed univariate LSTM model was trained and optimised with confirmed case time series data of COVID-19 and fit the model for forecasting mode. The model simulations (COVID-19 cases) are carried out for 31 days in July month for testing the model over selected states in India. The model performance is evaluated with the robust statistical technique of relative error for each forecasted day. Our results shows that the average relative error (31 days) of the univariate LSTM (CTL) is reasonably good (<20%) with lag1 (short term forecast, i.e. 24 hour forecast) for most of the selected states in India. It is also noted that the univariate LSTM model out performed compared to the multivariate LSTM model for the states of Andhra Pradesh, Karnataka, Delhi, Bihar, Odisha, and Uttar Pradesh (Fig.5). The time series data also shows that the univariate LSTM captured the observed trend very well for the selected states (Fig.6). The major weakness of the univariate model is that the model skill is decreasing with long term lead data.

Skill of multivariate LSTM model:

In the case of states (Maharashtra, Madhya Pradesh, Gujarat, Rajasthan, Haryana, and Punjab) located in West, North West India, the forecasting skill of the multivariate LSTM model (CTL_SH; model optimised with the specific humidity time series data along with the COVID data) out performed compared to the univariate model and other multivariate models which is optimised with temperature data. It was also noted that the correlation coefficient between specific humidity and COVID-19 cases was significant in these regions. The major finding of this experiment is that the forecasting skill of the model was improved with the lagged specific humidity data (lag1-lag7) over these regions which is very important for medium range prediction.

In India, the maximum COVID-19 cases are observed from the state of Maharashtra and the developed multivariate LSTM model (CTL_SH) with specific humidity time series data shown better performance (RE<8%) with the lag7 data (Fig.7a). The time series forecasting plot (with one week advance data) for the state Maharashtra shows that the other proposed models (CTL, CTL_Tmax, CTL_Tmin, and CTL_Tmean) were overestimating the daily cases whereas the specific humidity (CTL_SH) followed the observed trend and close to the observed data (Fig.8a). Similarly, the forecast skill was adequate with CTL_SH model for the states of Punjab (lag5), Rajasthan (lag3), Madhya Pradesh (lag3), Gujarat (lag1), and Haryana (lag1) which is presented in Fig.8

In the case of high humid regions (Kerala, Tamil Nadu, and West Bengal) the forecast skill is improved with the multivariate LSTM model which is optimised with the temperature data (Fig.9). The time series forecast plot clearly shows that the temperature based LSTM models close to the observations compare to the humidity based model (Fig.10).

Conclusions:

The COVID-19 cases were started during the winter season (the first case reported on January 30, 2020) and the maximum number of cases were reported over Maharashtra and Kerala to compare to the other states before the national wide lockdown (March 25, 2020) implemented in India. The virus transmission was so rapid after the onset of monsoon and the maximum number of positive cases are reporting form few of the states in India. There is need of accurate and advanced forecasting system for short, medium, and long range prediction of COVID cases for better management of logistics and policy decisions. Based on the earlier studies, the RNN based LSTM models have shown adequate skill in short range (one day lead) forecasting of COVID-19 cases in India. Hence, the present study proposed to develop a weather integrated multivariate LSTM models to improve the prediction skills in short to long-range forecasting of daily cases over different states in India.

Our results suggested that the skill of univariate LSTM model which is optimised with confirmed COVID-19 time series data performed very well for highly effected states like Andhra Pradesh, Karnataka, Uttar Pradesh, Deli, Bihar, and Odisha. It was also noticed that the skill of univariate model is good in short range forecasting (lag1) and the skill is decreasing with increasing lead period. The major findings of the study is that the medium range (1-7 days lead) forecasting skill have shown adequate skill in some of the states in India when the LSTM models are integrated with time series weather data including specific humidity and temperature. The results shows that the developed multivariate LSTM models optimised with specific humidity (CTL_SH) shown adequate skills in medium range forecast of daily COVID cases over the states located in west and North West India region. It was also observed that the developed multivariate LSTM models with temperature time series data performed very well over the states located in high humid regions including Kerala, Tamil Nadu, and west Bengal.

The present study demonstrated the forecasting skill of LSTM model is improved at medium and long range scale due to the integration of weather data in India. The skill may improve with the high resolution weather data, increasing the length of training data and optimization methods in LSTM models.

Acknowledgments:

The authors are grateful to the Directors of the Council of Scientific and Industrial Research-Indian Institute of Chemical Technology, Hyderabad and 4PI, Bangalore for their encouragement and support. The present work is supported by the DST (Department of Science and Technology) under Epidemiology Data Analytics (EDA) of Interdisciplinary cyber physical systems (ICPS) programme (Grant number: DST/ICPS/EDA/2018), Govt. of India. The authors are grateful to the MoHFW (Ministry of Health and Family Welfare), Govt. India for providing the state level COVID-19 data, and NCEP/NCAR for providing meteorological data. Srinivasa Rao Mutheneni acknowledges Ministry of Environment, Forest & Climate Change (MoEF & CC), Government of India for funding the project environmental information system (ENVIS: Resource Partner on Climate Change and Public Health). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. CSIR-IICT communication number of the article is IICT/Pubs./2020/152.

Competing financial interests:

The authors declare no competing financial interests exist.

Ethical Statement:

The authors declare that ethical statement is not applicable because the case information has been gathered.

Data Availability Statement:

The data used in this study are available from the corresponding author upon request.

References:

1. ICMR COVID Study Group Abraham et al. (2020). Laboratory surveillance for SARS-CoV-2 in India: Performance of testing & descriptive epidemiology of detected COVID-19, January 22 - April 30, 2020. The Indian journal of medical research, 151(5), 424–437. https://doi.org/10.4103/ijmr.IJMR_1896_20

2. Arora, P., Kumarc H., Panigrahi, B,K. (2020). Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. Chaos Solitons Fractals, 139, 110017. doi:10.1016/j.chaos.2020.110017

3. Bai, Y., Yao, L., Wei, T., Tian, F., Jin, D. Y., Chen, L., & Wang, M. (2020). Presumed Asymptomatic Carrier Transmission of COVID-19. JAMA, *323*(14), 1406–1407. https://doi.org/10.1001/jama.2020.2565.

4. Bashir, M. F., Ma, B., Komal, B., Bashir, M. A., Tan, D., & Bashir, M. (2020). Correlation between climate indicators and COVID-19 pandemic in New York, USA. Science of The Total Environment, 138835.

5. Briz-Redón, Á.,& Serrano-Aroca, Á. (2020). A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. Science of The Total Environment, 138811.

6. Chen, B., Liang, H., Yuan, X., Hu, Y., Xu, M., Zhao, Y., & Zhu, X. (2020) Roles of meteorological conditions in COVID-19 transmission on a worldwide scale. MedRxiv, 20037168. DOI: 10.1101/2020.03.16.20037168.

7. Gupta, N., Praharaj, I., Bhatnagar, T., Vivian Thangaraj, J. W., Giri, S., Chauhan, H., Kulkarni, S., Murhekar, M., Singh, S., Gangakhedkar, R. R., Bhargava, B., & ICMR COVID Team (2020). Severe acute respiratory illness surveillance for coronavirus disease 2019, India, 2020. The Indian journal of medical research, 151(2 & 3), 236–240. https://doi.org/10.4103/ijmr.IJMR_1035_20

8. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

9. Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet, 395 (10223): 497−506. doi:10.1016/S0140-6736(20)30183-5

10. ICMR: Indian Council of Medical Research https://www.icmr.gov.in/ Accessed on 08.06.2020.

11. Jiang, S., Du, L., Shi, Z. (2020). An emerging coronavirus causing pneumonia outbreak in Wuhan, China: calling for developing therapeutic and prophylactic strategies. Emerg Microbes Infect, 9(1), 275−277. doi:10.1080/22221751.2020.1723441.

12. Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Roy Jenne, Dennis Joseph (1996). The NCEP/NCAR 40-year reanalysis project. Bulletin of the American meteorological Society, 77(3), 437-472.

13. Lai, C.C., Shih, T.P., Ko, W.C., Tang, H.J., Hsueh, P.R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents, 55 (3), 105924. doi:10.1016/j.ijantimicag.2020.105924.

14. Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T.T.Y., Wu, J.T., Gao, G.F., Cowling, B.J., Yang, B., Leung, G.M., Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. N Engl J Med. 382(13): 1199-1207. doi: 10.1056/NEJMoa2001316.

15. Liu, J., Zhou, J., Yao, J., Zhang, X., Li, L., Xu, X., He, X., Wang, B., Fu, S., Niu, T., Yan, J., Shi, Y., Ren, X., Niu, J., Zhu, W., Li, S., Luo, B., & Zhang, K. (2020). Impact of meteorological factors on the COVID-19 transmission: A multi-city study in China. Science of the Total Environment, 138513.

16. Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S.,& Luo, B. (2020). Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. Science of The Total Environment, 138226.

17. Ministry of Health & Family Welfare https://www.mohfw.gov.in/ Accessed on 07.06.2020.

18. Murhekar et al. (2020). Prevalence of SARS-CoV-2 infection in India: Findings from the national serosurvey, 152 (1), 48 Indian Journal of Medical Research.

19. Roland R. Netz, William A. Eaton. (2020). Physics of virus transmission by speaking droplets. Proceedings of the National Academy of Sciences, 202011889; DOI: 10.1073/pnas.2011889117.

20. Oliveiros, B., Caramelo, L., Ferreira, N. C., & Caramelo, F. (2020). Role of temperature and humidity in the modulation of the doubling time of COVID-19 cases. MedRxiv, 20031872. doi: https://doi.org/10.1101/2020.03.05.20031872

21. Rawat, M. Coronavirus in India: Tracking country's first 50 COVID-19 cases; what numbers tell (2020) Retrieved from https://www.indiatoday.in/india/story/coronavirus-in-india-tracking-country-s-first-50-covid-19-cases-what-numbers-tell-1654468-2020-03-12 Accessed on 20 April 2020.

22. Shastri, S., Singh, K., Kumar, S., Kour, P., Mansotra, V. (2020). Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study. Chaos Solitons Fractals, 140, 110227. doi:10.1016/j.chaos.2020.110227

23. Şahin, M. (2020). Impact of weather on COVID-19 pandemic in Turkey. Science of The Total Environment, 138810.

24. Sagheer, A., Mostafa, K. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. Neurocomputing, 323: 203-213.

25. Shen, Q., Guo, W., Guo, T., Li, J., He, W., Ni, S., Ouyang, X., Liu, J., Xie, Y., Tan, X., Zhou, Z., & Peng, H. (2020). Novel coronavirus infection in children outside of Wuhan, China. Pediatr Pulmonol, 55(6):1424−1429. doi:10.1002/ppul.24762

26. Tomar, A., Gupta, N. (2020). Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. Sci Total Environ. 728: 138762. doi:10.1016/j.scitotenv.2020.138762.

27. Vellingiri, B., Jayaramayya, K., Iyer, M., Narayanasamy, A., Govindasamy, V., Giridharan, B., Ganesan, S., Venugopal, A., Venkatesan, D., Ganesan, H., Rajagopalan, K., Rahman, P., Cho, S. G., Kumar, N. S., & Subramaniam, M. D. (2020). COVID-19: A promising cure for the global panic. Sci Total Environ. 725:138277.

28. Wang, J., Tang, K., Feng, K., & Lv, W. (2020). High temperature and high humidity reduce the transmission of COVID-19. (March 9, 2020). Available at SSRN 3551767.

Available at SSRN: https://ssrn.com/abstract=3551767 or http://dx.doi.org/10.2139/ssrn.3551767

29. World Health Organisation (WHO: 2020a) https://www.who.int/emergencies/diseases/novel-coronavirus2019?gclid=EAIaIQobChMI3dD35r3v6QIVVw4rCh0XqwAdEAAYASAAEgLWBvD_BwE accessed on 07.06.2020.

30. World Health Organisation (WHO: 2020b): https://covid19.who.int/ accessed on 13.10.2020.

31. Xu, H., Yan, C., Fu, Q., Xiao, K., Yu, Y., Han, D., Wang, W., & Cheng, J. (2020). Possible environmental effects on the spread of COVID-19 in China. Sci Total Environ. 731:139211. doi:10.1016/j.scitotenv.2020.139211.

32. Zhong, L., Mu, L., Li, J., Wang, J., Yin, Z., Liu, D. (2020). Early Prediction of the 2019 Novel Coronavirus Outbreak in the Mainland China Based on Simple Mathematical Model. IEEE Access, 8:51761−51769. doi:10.1109/ACCESS.2020.2979599

Figure Legends

Figure-1: Keras implementation of multi-parameter LSTM (a) The basic LSTM structure (b) Unrolled representation of LSTM (c) Architecture of an LSTM cell (d) Internal structure of a cell gate.

Figure-2: Spatial maps of monthly cumulated COVID-19 cases over different states in India during pre-monsoon (April and May) and monsoon season (June and July) of the year 2020.

Figure-3: Spatial-temporal variation of surface meteorological parameters (2m-specific humidity, 2m-mean temperature, 2m-maximum temperature, and 2m-minimum temperature) during the pre-monsoon and monsoon season over India.

Figure-4: Correlation between confirmed COVID-19 cases and meteorological parameters (2m-specific humidity, 2m-mean temperature, 2m-maximum temperature, and 2m-minimum temperature) during the period 01st April, 2020 to 31st July, 2020.

Figure-5: Skill (Average relative error) of univariate (CTL) and multivariate (CTL_SH, CTL_Tmax, CTL_Tmin, CTL_Tmean) LSTM models during the test period (1st July 2020 to 31st July 2020) for the states of Andhra Pradesh, Karnataka, Delhi, Bihar, Odisha, and Uttar Pradesh. Where L1 to L14 represent the 1to 14 days of lag data utilized for forecasting of the next day COVID-19 cases.

Figure-6: Time series data of COVID-19 cases forecasted by univariate (CTL) and multivariate (CTL_SH, CTL_Tmax, CTL_Tmin, CTL_Tmean) LSTM models during the test period (1st July 2020 to 31st July 2020) for the states of Andhra Pradesh, Karnataka, Delhi, Bihar, Odisha, and Uttar Pradesh.

Figure-7: Skill (Average relative error) of univariate (CTL) and multivariate (CTL_SH, CTL_Tmax, CTL_Tmin, CTL_Tmean) LSTM models during the test period (1st July 2020 to 31st July 2020) for the states of Maharashtra, Gujarat, Madhya Pradesh, Rajasthan, Haryana, and Punjab. Where L1 to L14 represent the 1to 14 days of lag data utilized for forecasting of the next day COVID-19 cases.

Figure-8: Time series data of COVID-19 cases forecasted by univariate (CTL) and multivariate (CTL_SH, CTL_Tmax, CTL_Tmin, and CTL_Tmean) LSTM models during the test period (1st July 2020 to 31st July 2020) for the states of Maharashtra, Gujarat, Madhya Pradesh, Rajasthan, Haryana, and Punjab.

Figure-9: Skill (Average relative error) of univariate (CTL) and multivariate (CTL_SH, CTL_Tmax, CTL_Tmin, CTL_Tmean) LSTM models during the test period (1st July 2020 to 31st July 2020) for the states of Tamil Nadu, West Bengal, and Kerala. Where L1 to L14 represent the 1to 14 days of lag data utilized for forecasting of the next day COVID-19 cases.

Figure-10: Time series data of COVID-19 cases forecasted by univariate (CTL) and multivariate (CTL_SH, CTL_Tmax, CTL_Tmin, and CTL_Tmean) LSTM models during the test period (1st July 2020 to 31st July 2020) for the states of Tamil Nadu, West Bengal, and Kerala.