

Spatiotemporal Analysis of COVID-19 Pandemic and Predictive Models based on Artificial Intelligence for different States of India

Paramita Guha¹

Received: 15 August 2020 / Accepted: 15 May 2021 / Published online: 9 June 2021
© The Institution of Engineers (India) 2021

Abstract Geographical and spatial diversities play important roles in dynamics of spread of COVID-19 virus. These phenomena are not properly addressed in the literature yet. In this paper, COVID data of various states of India are collected. The data had been processed and analysed using an open-source software. A framework based on Susceptible, Infectious, Hospitalised, Recovered and Deaths model to determine the effects of geographical diversities of Indian states on COVID-19 pandemic has been developed. The confirmed, cured and death cases due to the virus have been analysed for different state. Reasons behind the differences in number of cases in different states are identified. An improved Long-Short-Term-Memory algorithm has been developed to forecast the virus spread and recovery of patients for the next one month. Numerical results along with discussions are given.

Keywords Spatiotemporal analysis · COVID-19 · SIHRD · Predictive model · LSTM · RNN artificial intelligence

Introduction

The novel coronavirus COVID-19 has created havoc all over the world. It has made an unparalleled health emergency in recent history. The COVID-19 outbreak started in Wuhan Province of China in December 2019 and has travelled all over the world with more than 2 crore people

affected and around 7.5 lakhs deaths. The first confirmed case in India was reported on 30 January 2020. As of now, the number of total affected people in India is more than 22 lakhs with 45 thousand people died so far.

Mathematical modelling along with temporal and spatial dynamics are important to provide the epidemiological characteristics of the disease. The transmission dynamics of infections are discussed in the literature [1]. This work is important to understand the early stages of any new infectious disease outbreak. Other than obtaining insights of epidemiological characteristics, temporal and spatial dynamics can be useful for forecasting of potential future burden of the disease along with the identification of disease hot spots. The researchers [2, 3] have used spatiotemporal analysis methods to identify the drivers of local transmission and population at risk and guide the designing of targeted interventions in resource limited settings. Temporal trajectories of the disease under populations and intervention scenarios are derived and discussed in the literature [1, 4–7]. The investigators [8–10] have assessed spatiotemporal dynamics of the disease. It is observed from the literature that research on disease transmission dynamics has not been progressed much. In the literature [11–15] characteristics of different infectious diseases like respiratory infections with variations geographical locations, uneven distribution of population and spatial diffusions of pathogens have been discussed. The relations between the unequal distribution of capacity of healthcare systems and disease outcomes are discussed in the prior investigations [16]. It is known that regions across India have not been equally affected with this virus. The northern and western states are mostly affected, whereas the North-East states, islands are least affected. The underlying reasons of these differences are not well understood yet. Studies made in other countries show that socio-economic

✉ Paramita Guha
paramita.guha@csio.res.in

¹ CSIR-Central Scientific Instruments Organisation,
New Delhi, India

conditions like, higher populations; more footprints; environmental factors like air pollutions; location of major airports, highways are key factors for geographical diversities of COVID-related issues [17–20]. This type of study is not much done in Indian context. Hence, in this paper, authors have developed a framework based on Susceptible, Infectious, Hospitalised, Recovered and Deaths (SIHRD) model to determine the effects of geographical diversities of Indian states on COVID-19 pandemic. An improved Long-Short-Term-Memory algorithm has been developed to forecast the virus spread and recovery of patients for the next one month.

Data Collection and Problem Formulation

The COVID-19 data of various states of India have been collected from the website www.kaggle.com. The website was accessed for data collection on 7 August 2020. The number of confirmed, cured and deaths of different states was collected and stored in *.csv* file. Also, the number of hospitals, beds available in both urban and rural areas were also collected from the same website. The data were cleansed and pre-processed, and missing points were imputed using Python open-source software. After that they were arranged in ascending order so that clear status of different states can be obtained. A mathematical model to simulate the transmission dynamics of the disease has been developed. The data have been classified based on SIHRD *i.e.* Susceptible (*S*), Infected (*I*), Hospitalized (*H*), Recovered (*R*) and Death (*D*) model. The SIHRD model has been calibrated using the data obtained from the website. Details of the classifier model are given in next section. It has been observed from the literature that airports, hubs, dense populations, main roads are important geospatial attributes to spread of corona diseases, hospitalization and mortality [17, 21]. Based on this idea, in this study, the states of India are classified into two categories, viz. states with high populations, having airports more than 50 thousand passengers travelling per year and major highways passing through them and the other category of states is having less populations, not having major airports and highways. The dynamics of disease transmission and infection progression in each group has been modelled using the epidemiological compartments of SIHRD COVID-19 model.

Modelling

It is well known that at the end of December 2019, Chinese Province Wuhan had first reported of the outbreak of novel coronavirus (COVID-19). Within the span of three months

this virus has travelled all over the world and affected more than 202 lakhs people with death of 7.5 lakhs so far. It has affected very badly India as well with more than 23 lakhs people affected and around 45 thousand people died. The WHO, ICMR and other government organizations of India are scrambling to minimize the spread of this deadly virus. Mathematical modelling of such infectious disease can be an essential part of this effort as a well-designed disease model can be helpful to predict the likely course of the epidemic and reveal the most promising realistic strategies for combating it. In this paper, we have considered two different approaches for modelling of COVID data. Modelled data obtained from SIHRD are passed through an artificial intelligence-based network to forecast the nature of transmission and spread of the virus. The approaches are discussed here.

Susceptible Infectious Hospitalised Recovered Deaths Framework

The most popular method to model and forecast any virus outbreak is compartmental method based on Susceptible (*S*)—Infectious (*I*)—Recovered (*R*) (SIR) framework. Each compartment is related to each other, and the relations are governed by nonlinear coupled equations. Here *Susceptible* compartment indicates group of individuals who have less or almost no immunity to the disease. They are more prone to any virus/disease. People belonging to this group can easily move into the *Infection* compartment if there is any contact with an infected person. It is observed from the literature that around 75% of total population of India belong to this class.

The next compartment of the classifier model is *Infection* or infectious persons. These persons are suffering from the disease and spread to others. Individuals of this compartment can move into the *Recovered* stage after recovering from the illness. It is well known that nowadays COVID-19 has two categories of affected persons, viz. symptomatic and asymptomatic. Here, we have considered only symptomatic persons. Data for them are collected and analysed.

The last compartment of this simple classifier is *Recovered*. Individuals belong to this group have recovered from this infection, and it is assumed that they have developed immunity from the exposure. Nowadays, the virologists have monitored that the persons recovered from COVID-19 virus may not have a long-lasting immunity like other diseases (SARS, CoV-2, chicken pox, etc.). However, for the sake of simplicity in this paper we have assumed that recovered persons have long-lasting immunity power. A simplified block diagram of SIR model is shown in Fig. 1.

In this model, S represents susceptibility rate, I is infections rate, and R is the recovery rate. Here, μ^* is the control input, in terms of individuals becoming susceptible to the disease; μ is the mortality rate; λ is the rate at which the susceptible individuals become infectious; and γ is the rate at which the infectious individuals recover from the disease. The parameter λ is proportional to transmission rate (β). The susceptibility rate can be given as

$$\lambda(I) = \beta \frac{I}{N}$$

with N is the total population of the city or country. The recovery rate at ‘

$$\frac{dS}{dt} = -\lambda(I)S \tag{1}$$

$$\frac{dI}{dt} = -\lambda(I)S - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I \tag{3}$$

After solving these coupled equations, number of persons belonging in each compartment can be obtained. The SIR model is very simple and does not take account of hospitalization and death data rates. Hence, in order to make the model more practical, in this paper two more compartments are added to the previous model to form an improved SIHRD classifier. The block diagram of SIHRD is shown in Fig. 2.

Mathematically the model for different states can be described as,

$$\frac{dS_i}{dt} = -\varepsilon\lambda_i S_i \left(\frac{I_i}{N_i}\right) \tag{4}$$

$$\frac{dS_i}{dt} = -\varepsilon\lambda_i S_i \left(\frac{I_i}{N_i}\right) - I_i(\delta + \eta_i + \psi) \tag{5}$$

$$\frac{dH_i}{dt} = \eta_i I_i - H_i \sigma \tag{6}$$

$$\frac{dM_i}{dt} = \mu_i H_i + \psi_i I_i \tag{7}$$

$$\frac{dR_i}{dt} = \delta_i I_i + \sigma_i H_i \tag{8}$$

Here, the hazard rate of infection for state ‘ i ’ is given by λ_i ; reduction of hazard rate after lockdown was over is

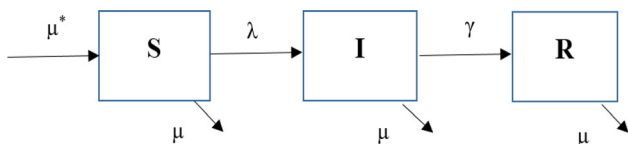


Fig. 1 Block diagram of SIR model

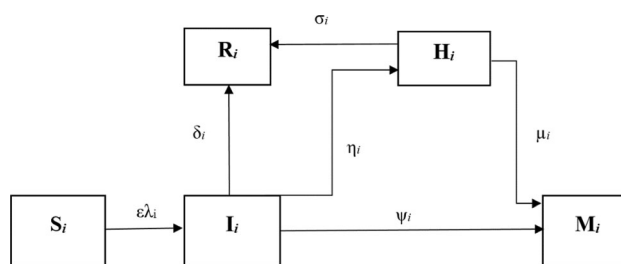


Fig. 2 Block diagram of SIHRD model

given as ε ; natural recovery rate without hospitalization is denoted by $1/\delta$. The hospitalization and discharge rates from hospitals are denoted as η and σ , respectively. Similarly, mortality rates in hospitals and home quarantines are given as μ and ψ , respectively. These are all fitted parameters and obtained through least square estimation of input data.

Artificial Intelligence-based Prediction

After obtaining the modelled data, they are predicted through a Long-Short-Term-Memory (LSTM)-based recurrent neural network (RNN). The LSTM model is the most popular and common forms of RNN. It avoids long-term dependence problems and suitable for processing and predicting time series. It consists of unique set of memory cells which replace the hidden layers of RNN, and its key is the state of memory cells. It filters the information through gate structure to maintain and update the state of memory cells. LSTM consists of input, forgotten and output gates. Each memory cell consists of three sigmoid layers and one tanh layer. The structure of LSTM cell is shown in Fig. 3.

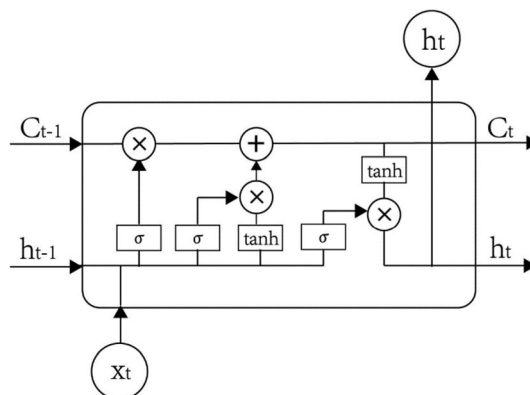


Fig. 3 Structure of a Long-Short-Term Memory RNN model [22]

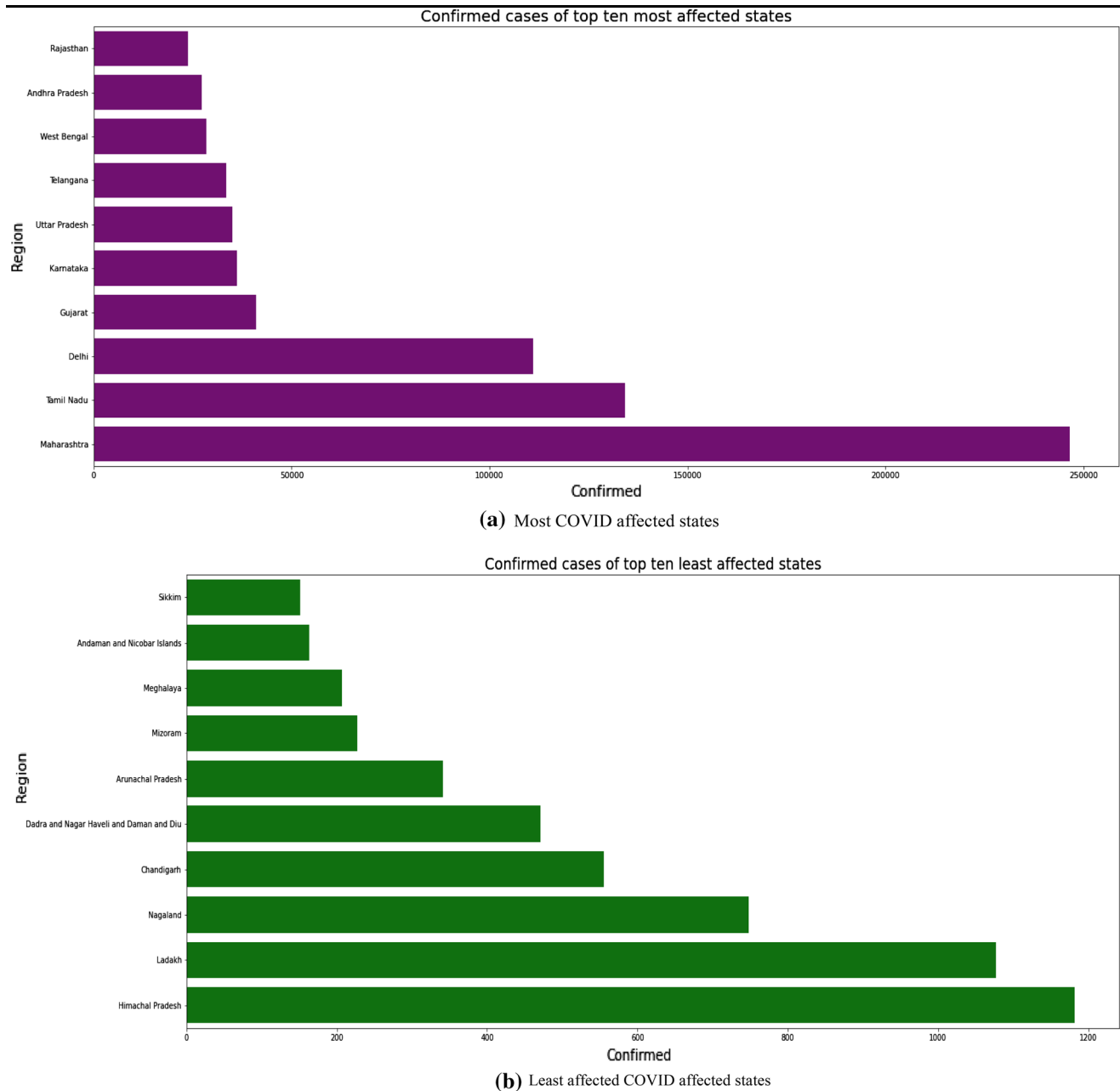


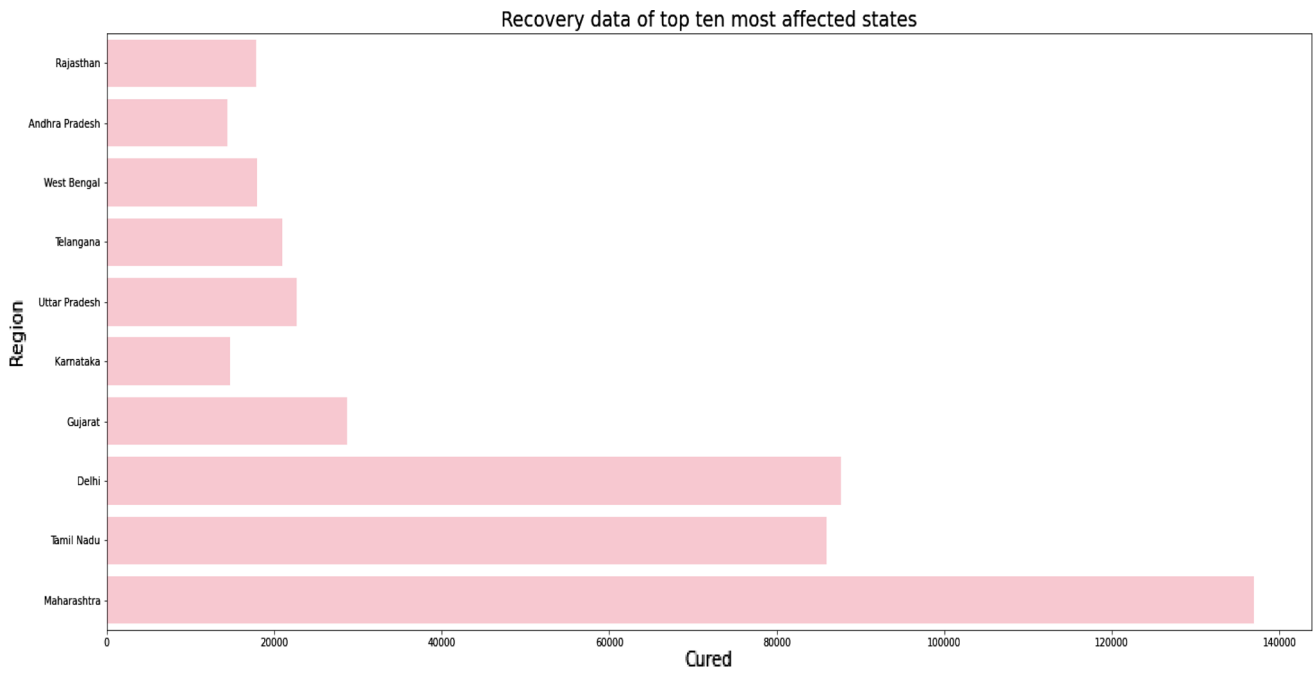
Fig. 4 a Most COVID-affected states. b Least COVID-affected states

Above-discussed algorithms have been applied on the downloaded data, and relations among confirmed, cured and deaths with hospital beds availability are obtained.

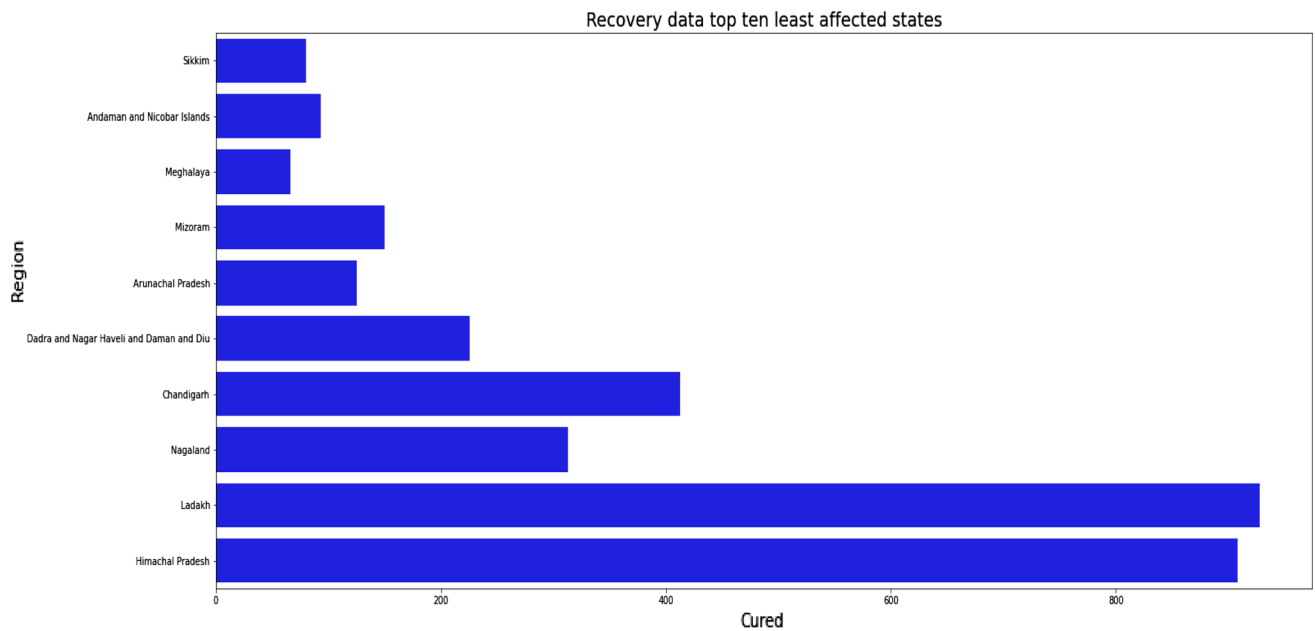
Results and Discussion

As discussed earlier, the data had been obtained from www.kaggle.com website. The data of different states had been processed and cleansed. For this purpose, authors

have used Python open-source software. The data from each state had been modelled through SIHRD framework, and the confirmed, cured and death cases are plotted. It has been observed that Maharashtra, Tamil Nadu, Delhi, Gujarat and Karnataka are the five major states severely affected by the COVID-19, whereas the regions like Sikkim, Andaman and Nicobar, Meghalaya, Mizoram and Andhra Pradesh are least affected with this virus. It is also known that the severely affected states have airports with large number of footprints, high populations and major



(a) Recovery data of most affected COVID affected states



(b) Recovery data of least COVID affected states

Fig. 5 **a** Recovery data of most affected COVID-affected states. **b** Recovery data of least COVID-affected states

highways passing through them compared to the least affected states/UTs. This validates the SIHRD model developed in this paper. The results of COVID cases for most and least affected states are shown in Figs. 4, 5, 6, 7,

8, 9. In this paper, the model has been developed using data from 26 March 2020 to 28 July 2020. Number of footprints in the most affected regions during the considered period is taken as 10 millions, whereas the least affected regions had

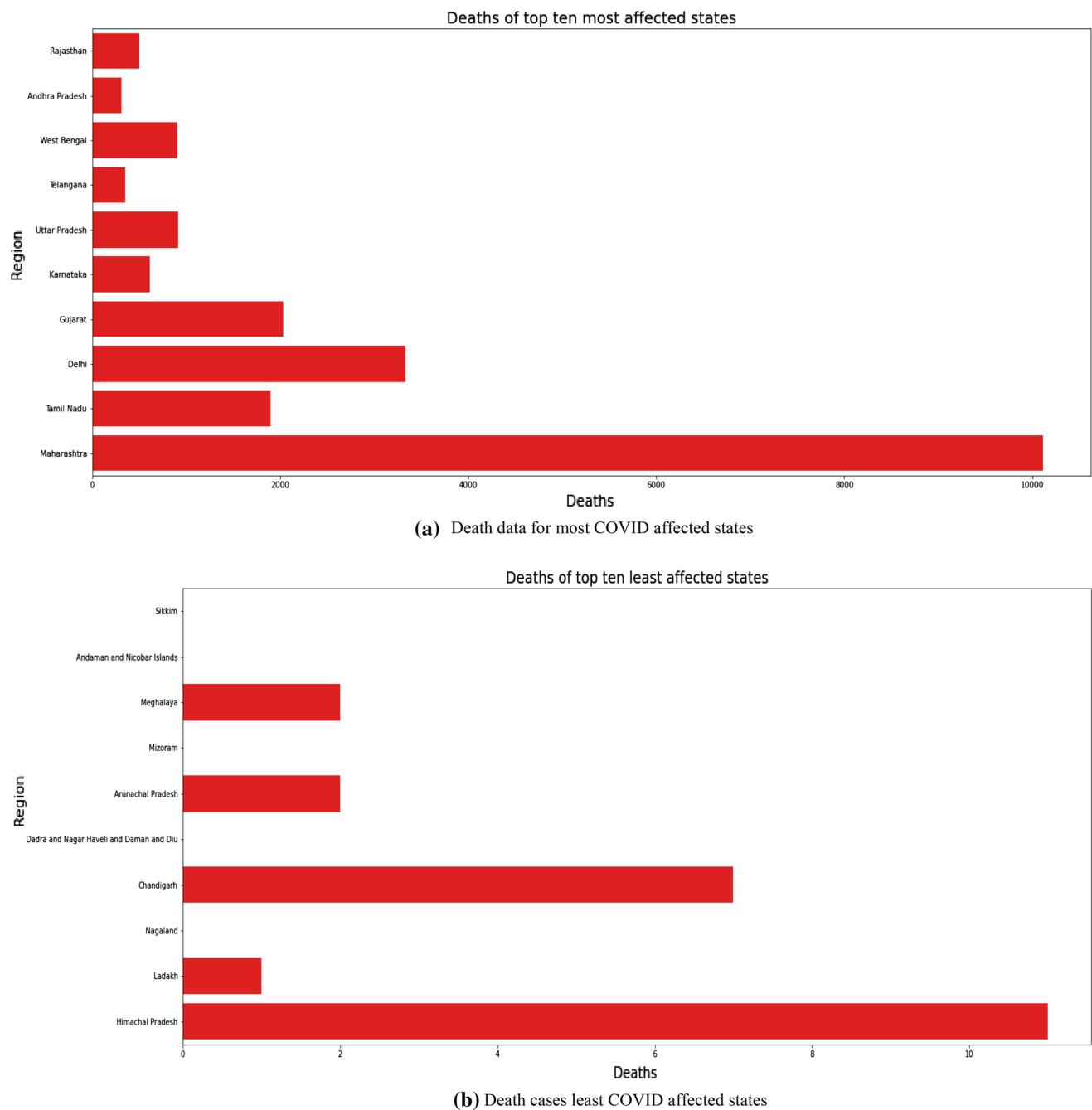


Fig. 6 **a** Death data for most COVID-affected states. **b** Death cases least COVID-affected states

around 1 million footprints. Total populations of the most affected states are more than 77 crores with that of least affected states have less than 2 crore populations.

The fitted parameters for Eqs. (4–8) are obtained as $\lambda_{1,2} = 0.45, 0.16$; $\varepsilon = 0.3$; $\delta = 10$ days; $\eta_{1,2} = 0.05, 0.17$; $\sigma = 0.095$; $\mu_{1,2} = 0.22, 0.25$; $\psi = 0.005$. Here, ‘1’ is the regions having large populations, surrounded with high-ways and big airports, whereas ‘2’ is for other regions.

The model estimated a total number of 22.59 lakhs (with 95% confidence level) COVID cases compared to 20.23 lakhs confirmed cases as the data collected on 7 August 2020. Similarly, the death and cured cases at 95% of confidence level are obtained as 43.671 thousand and 12.33 lakhs, respectively. These data are compared with original data and observed that the overall errors for death and cured cases are 5.23% and 3.95%, respectively. The

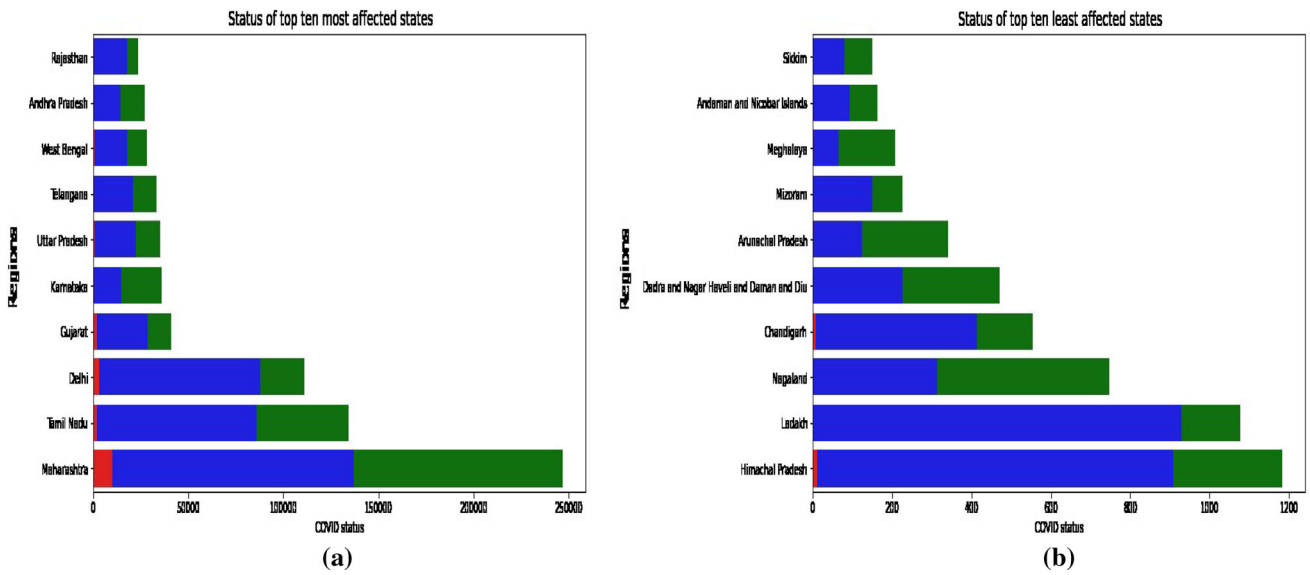


Fig. 7 Overall status of COVID-affected states; green: confirmed cases, blue: recovered; red: death cases. a Most affected; b least affected

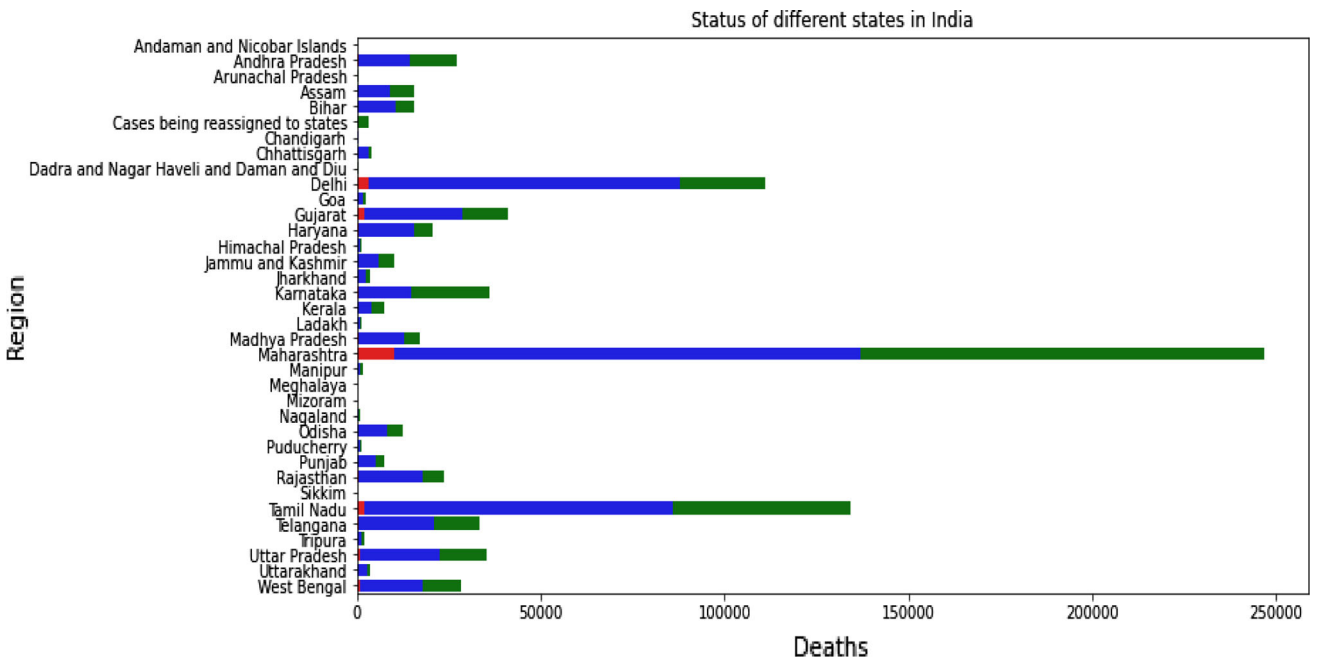


Fig. 8 COVID-affected states data overall; green: confirmed, blue: recovered; red: deaths

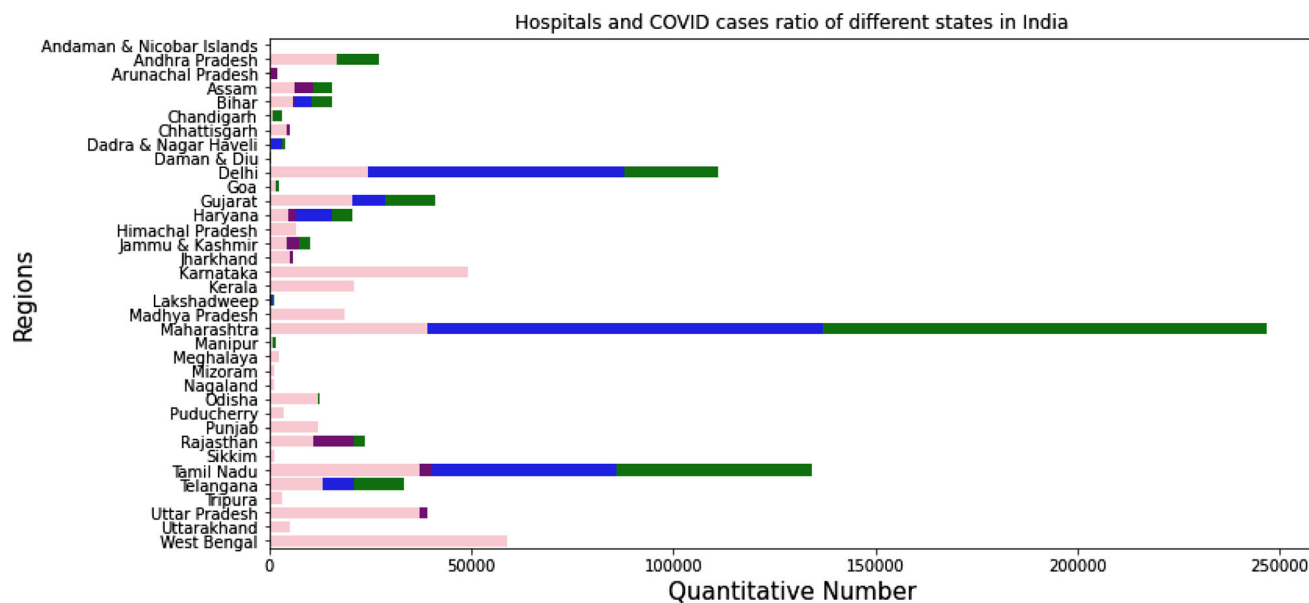


Fig. 9 COVID-affected states and medical facilities available; green: confirmed, blue: recovered; pink: beds available at urban areas, purple: beds available at rural beds; red: deaths

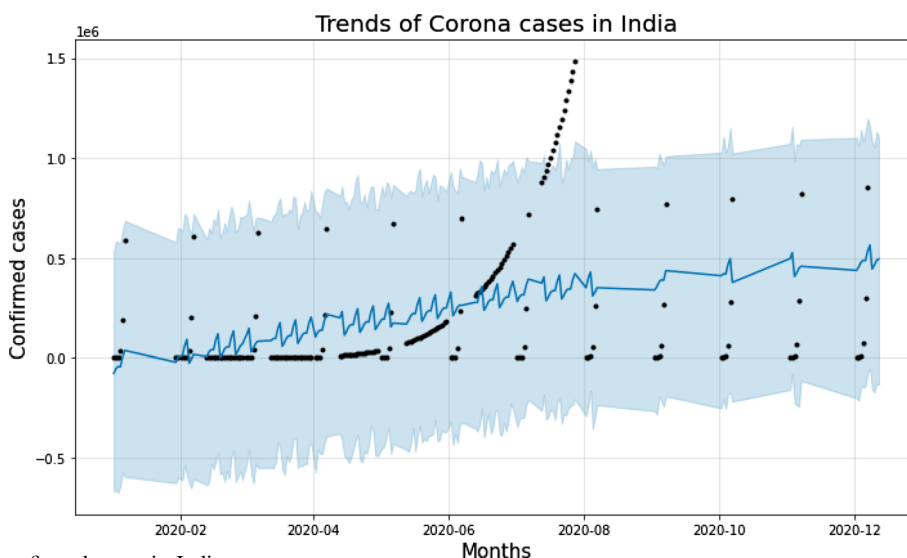


Fig. 10 Forecast of confirmed cases in India

predicted data for confirmed, cured and death cases for next one month are shown in Figs.10, 11, 12.

Conclusion

In this paper, online COVID-19 data of India have been obtained from an open access data source. The data have been processed and cleaned using Python software. A

framework based on SIHRD model has been developed to model the data obtained for different states of India. It has been observed from the analysis that regions across India have not been equally affected with this virus. The northern and western states are mostly affected, whereas the North-East states, islands are least affected. The main reasons behind this are socio-economic conditions like higher populations; more footprints; environmental factors like air pollutions; location of major airports, highways. Also, an

Fig. 11 Forecast of cured cases in India

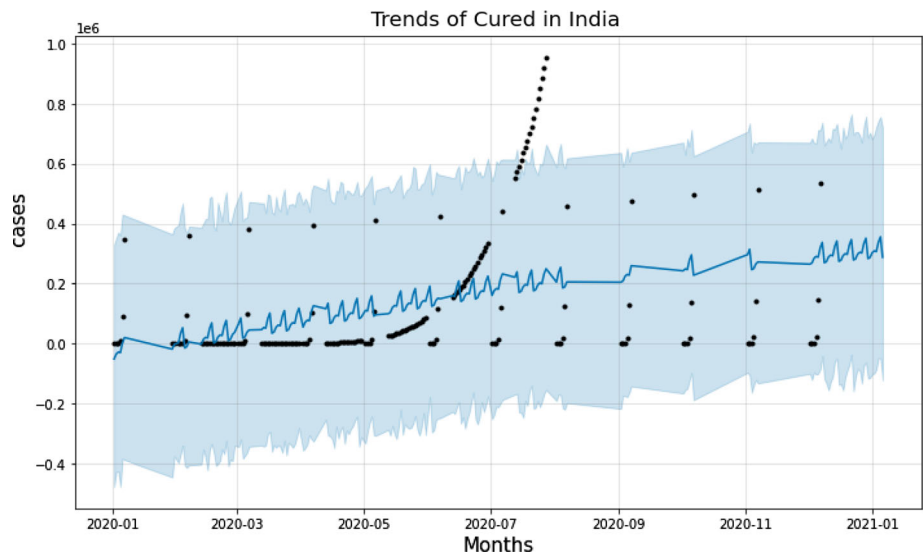
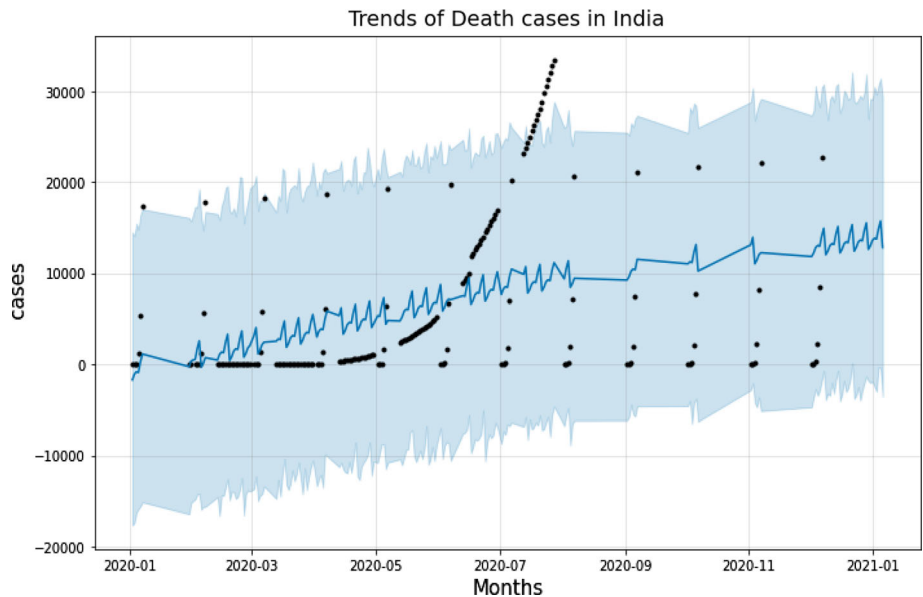


Fig. 12 Forecast of death cases in India



improved Long-Short-Term-Memory algorithm has been developed to forecast the virus spread and recovery of patients for the next one month.

References

1. A.J. Kucharski, T.W. Russell, C. Diamond, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* (2020)
2. G. Meyer-Rath, J.B. McGillen, D.F. Cuadros et al., Targeting the right interventions to the right people and places: the role of geospatial analysis in HIV program planning. *AIDS* **32**(8), 957 (2018)
3. D. Wilson, D.T. Halperin, “Know your epidemic, know your response”: a useful approach, if we get it right. *Lancet* **372**(9637), 423–426 (2008)
4. H.H. Ayoub, H. Chemaitelly, G.R. Mumtaz, S. Seedat, S.F. Awad, M. Makhoul, L.J. Abu- Raddad, Characterizing key attributes of the epidemiology of COVID-19 in China: model-based estimations. (2020a) medRxiv. 2020. <https://doi.org/10.1101/2020.04.08.20058214>
5. J. Hellewell, S. Abbott, A. Gimma et al., Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* (2020). [https://doi.org/10.1016/S2214-109X\(20\)30074-7](https://doi.org/10.1016/S2214-109X(20)30074-7)
6. Z. Mukandavire, F. Nyabadza, N. Malunguza, D. Cuadros, T. Shiri, G. Musuka, Quantifying early COVID-19 outbreak transmission in South Africa and exploring vaccine efficacy scenarios. *PLoS One* (2020) (in press)
7. K. Prem, Y. Liu, T.W. Russell, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Publ. Health.* (2020)

8. M. Gatto, E. Bertuzzo, L. Mari et al., Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *Proc. Natl. Acad. Sci. Unit States Am.* **117**(19), 10484–10491 (2020)
9. R. Huang, M. Liu, Y. Ding, Spatial-temporal distribution of COVID-19 in China and its prediction: a data-driven modeling analysis. *J. Infect. Dev. Ctries.* **14**(246–253), 03 (2020)
10. J.S. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia, N.A. Christakis, Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* (2020). <https://doi.org/10.1038/s41586-020-2284-y>
11. S. Caini, C. El-Guerche Seblain, M.A. Ciblak, J. Paget, Epidemiology of seasonal influenza in the middle east and north africa regions, 2010–2016: circulating influenza A and B viruses and spatial timing of epidemics. *Influenza Other Respir. Vir.* **12**(3), 344–352 (2018)
12. C.A. Green, D. Yeates, A. Goldacre et al., Admission to hospital for bronchiolitis in England: trends over five decades, geographical variation and association with perinatal characteristics and subsequent asthma. *Arch. Dis. Child.* **101**(2), 140–146 (2016)
13. M. Park, A.R. Cook, J.T. Lim, Y. Sun, B.L. Dickens, A systematic review of COVID-19 epidemiology based on current evidence. *J. Clin. Med.* **9**(4), 967 (2020)
14. L.A. Real, R. Biek, Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. *J. R. Soc. Interface* **4**(16), 935–948 (2007)
15. C. Sloan, M.L. Moore, T. Hartert, Impact of pollution, climate, and sociodemographic factors on spatiotemporal dynamics of seasonal respiratory viruses. *Clin. Transl. Sci.* **4**(1), 48–54 (2011)
16. Y. Ji, Z. Ma, M.P. Peppelenbosch, Q. Pan, Potential association between COVID-19 mortality and health-care resource availability. *Lancet Glob. Health* **8**(4), e480 (2020)
17. E. Correa-Agudelo, T. Mersha, A. Hernandez, A.J. Branscum, N.J. MacKinnon, D.F. Cuadros, Identification of vulnerable populations and areas at higher Risk of COVID-19 Related Mortality in the U.S. *medRxiv*, 2020.2007.2011.20151563. (2020)
18. C. COVID, Response Team. Geographic differences in covid-19 cases, deaths, and incidence-United States, February 12-April 7, 2020. *MMWR Morb. Mortal. Wkly. Rep.* **69** (15), 465–471 (2020)
19. A. van Dorn, R.E. Cooney, M.L. Sabin, COVID-19 exacerbating inequalities in the US. *Lancet* **395**(10232), 1243–1244 (2020)
20. X. Wu, R.C. Nethery, B.M. Sabath, D. Braun, F. Dominici, Exposure to air pollution and COVID-19 mortality in the United States. (2020). *medRxiv*
21. D. Cuadros, N. MacKinnon, Y. Xiao, A. Hernandez, E. Correa, H. Kim, Z. Yao, R. Mavi, A. Hincapie, Identification of vulnerable areas of high COVID-19 mortality risk in Ohio (Press release). (University of Cincinnati, 2020)
22. Y. Zifeng, Z. Zhiqi, Ke. Wang, W. Sook-San, L. Wenhua, Z. Mark, L. Peng, C. Xudong, G. Zhongqiang, M. Zhitong, L. Jingyi, L. Xiaoqing, L. Shiyue, L. Yimin, Y. Feng, G. Weijie, Y. Yifan, L. Fei, L. Shengmei, X. Yuqi, L. Bin, W. Zhoulang, Z. Shaobo, W. Yaonan, Z. Nanshan, H. Jianxing, Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J. Thorac. Dis.* **12**(3), 165–174 (2020). <https://doi.org/10.21037/jtd.2020.02.64>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.