A simple mathematical model to predict and validate the spread of Covid-19 in India

Harish Kumar [a,*], Pawan K. Arora [b], Meena Pant [a], Anil Kumar [c], Shahroz Akhtar Khan [b]

[a] National Institute of Technology Delhi, India
[b] Galgotias College of Engineering and Technology, Gautambudh Nagar, U.P, India
[c] H.R.Group of Institutions, Ghaziabad, U.P, India

ARTICLE INFO

ABSTRACT

The new outbreak of the corona virus (Covid-19) is expanding rapidly worldwide, disrupting millions and prompting authorities to take swift measures to avoid the disease. National lockdown imposed by the Indian government since 25 March 2020, the early lockdown action shows as compared to many other Countries/states can benefit from limiting the final size of the epidemic. A report on the issue of spreading the Covid-19 modeling in India is under review. This study analyzes Covid-19 infections by 20Dec 2021 and presents a mathematical approach for forecasting new cases or cumulative cases in practical situations. This forecast is much needed to schedule/continue medical set-ups for possible action to tackle the Covid-19 outbreak. It is important to mention here that the number of authors has proposed different models for predicting the expansion of Covid-19 to India and other countries; almost no model has yet to be demonstrated viable. With this mathematical model, it is simple to forecast the transfer of Covid-19. It is clear from the data that lockdown has played a significant role in controlling the transmission of the disease. A close match between the predicted empirical results and the available results proves the derived model similarity.

1. Introduction

As reports showed, a pandemic of unknown pneumonia started in Wuhan, China, in December 2019, but there are reports that it began by the end of November 2019 [1,2]. A new variant of the corona virus COVID-19 similar to SARS-CoV, for which health experts have approved no valid medicine or vaccine. This virus is transmitted by the respiratory tract or by contact with infected particles or pathogenic organisms over an incubation period of 2 to 14 days [3,4]. Until then, the world could be protected by various preventive measures; COVID-19 pandemic has taken over the entire globe. Over an early duration of 4 months, many as five million cases are publicly revealed, with much more than 300,000 causalities in over 150 countries [5,6]. COVID-19 is not explicitly treated, and the mortality rate varies across different countries and ranges from 2 to 15 percent [7,8]. The epidemiological data verify that the spread rate is very high, and the WHO confirmed that SARS-CoV is

about 10 times fatal [9]. Fig. 1 illustrates the countries reporting Covid-19 cases in the globe. It can see that Covid-19 cases are registered in each country except in two nations around the world [10]. Fig. 2 shows a map of India and various states of India, with an estimated number of Covid-19 infections reported to date [11].

1.1. Present scenario

It is essential to determine the needed potential facilities, which may be necessary for terms of hospitals, beds, ventilators, medicines, etc. In regards to health personnel including doctors, nurses, and support personnel, a variety of researchers, including physicists, biologists work for thorough evaluation and further consequences of their developed model in the study of infected data, recovered data, and death rates. Infection, recovery, and mortality rates depend not only on the form of Covid-19 but also on the country's reaction. In this scenario, the researcher would like to examine data on infected, recovered, and dead people in India directly compared with other countries and the worldwide mean. The specific modeling was designed and evaluated by a few
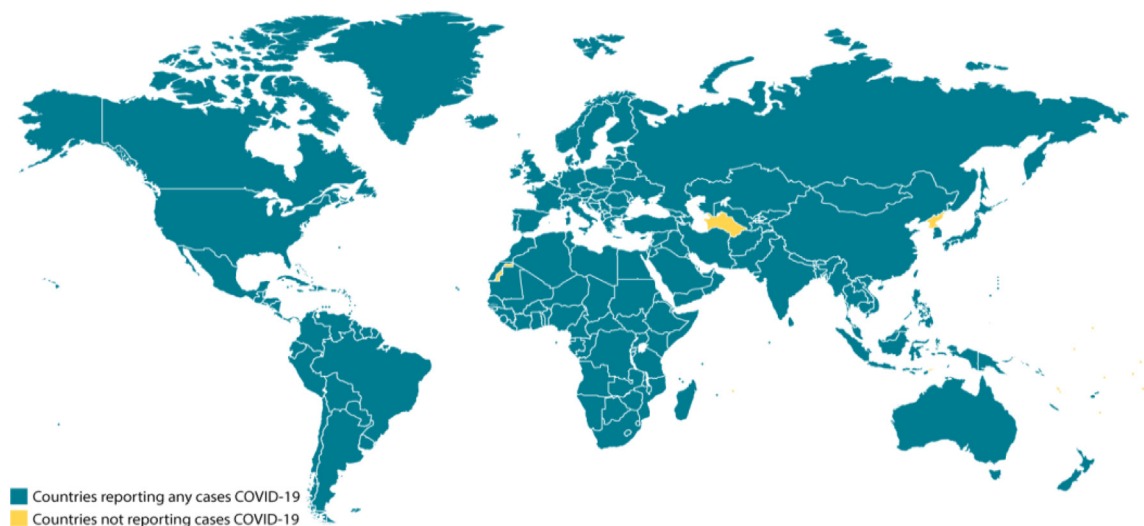
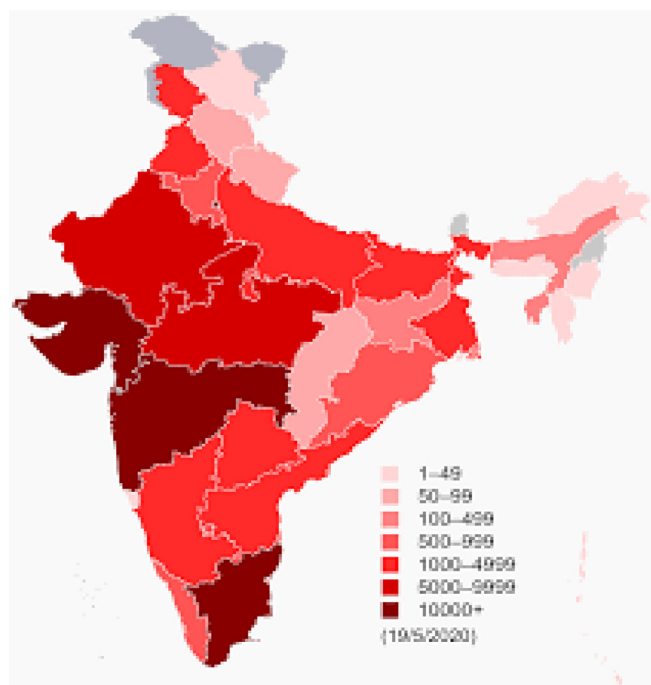**Fig. 1.** Countries reported Covid-19 infections so far [10].

Fig. 2. States of India showing Covid-19 infections so far [11].

researchers, and their findings were presented [12–15]. Some of the models are discussed below, along with their brief concept and shortcomings.

Julia R Gog (2020) explained the classic epidemic susceptible-infectious-recovered (SIR) model, a very simple model proposed in the 1920 s. It is called the SIR model, and S is the number of people who are susceptible to disease but have not been infected so far, I form the number of people who are ill, and R are the people who are either dead or healed from infection [16,17]. SanglierContrearas (2020) suggested the prediction of COVID-19′s epidemiological progress through modeling, analysis, and interpretation of data, and recommending various quadratic equations and appears to fit for critical data evaluation. The multiple factors studied and evaluated are the number of people infected, the death rate, the number of people recovered, the duration of treatment, etc. [18].

Rashad Eletrey (2020) investigates the impact of evolutionary adaptations, referring COVID-19 on the transmission of processes in complex networks [19].

GianlucaMalato (2020) evaluated the pattern of infected persons using a mathematical model based on a logistic model (equation 1) and an exponential model (equation 2) for publicly available data through the portal of the Government of Italy. The data was compared to the actual data available for far, and a critical consistency is obtained with significant deviation. X is the time in the equation, a, b, c are parameters/variables [20].

Tain-Mu-Chen, et. Al. (2020) provided a mathematical model is used to simulate the transmissibility process of the novel coronavirus in China. He used Berkeley Madonna 8.3.18 software (developed by R Macey and G Oster at Berkeley University of California) to match the data collected from the competent Chinese authority, which is in the public domain [21].

In a recent work researchers have analyzed the global impact of Covid-19 with emphasis given on daily cumulative infections and cumulative deaths. The data for Europe and Americawere analyzed and impact was judged in relation to the data for the worldfor Covid-19 outbreak [22].

However, due to limitations, only specific mathematical models are listed here, but at the same time, researchers and practitioners are seeking to create/examine current statistical/mathematical methods in times of hardship for the human race. On further observations, authors have found that the models proposed by various researchers are focused on the constraints depending on a specific city o country and are not accurate for each area. Different constants assumed/evaluated in different models, as discussed above, are very specific, limiting the use of models as general-purpose models. Addressing the limitations, authors present simple mathematical models based on extensive data available in India (the first case of Covid-19 reported in India on 30 January 2020, the confirmed cases then reached to 100,000 on 19 May 2020 andcrossed11046000 cases up to now), which do not include different particular constants / computed factors [11]. The relevant data and subsequent evaluation of the mathematical model are described in the following section.

2. Methodology

Fig. 3 gives a brief glimpse of Covid-19 spread in India. One can find the confirmed cases of Covid-19 in India as on 20 Dec 2020.
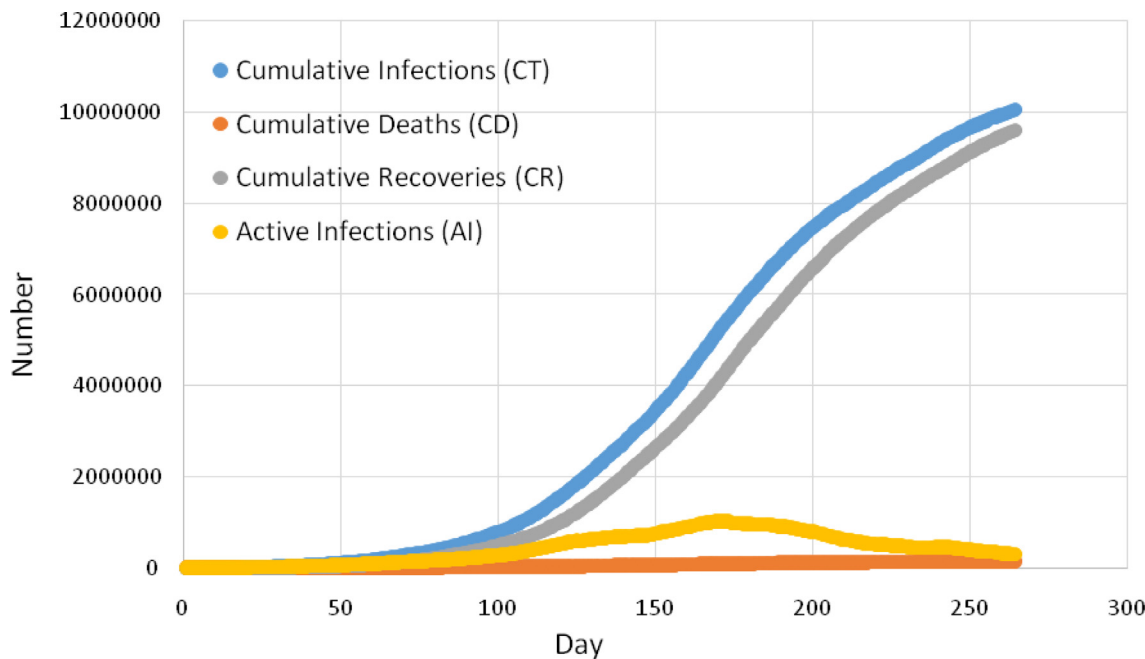
Fig. 3. Impact of Covid-19 in India since reporting of first case until Dec20, 2020.

The data for each state has been shown in tabulated form in table 1. The same information of cumulative sum of different parameters since inception for India are shown in Fig. 3. Out of the total infected people so far, the cured, deaths and infected are represented in form of pie-chart through Fig. 4.

Authors have attempted a number of different generalised mathematical models to predict some function for cumulative infections, cumulative recoveries, cumulative deaths and cumulative active infections, few of them areQuartic / bi-quadratic regression,exponential function, sum of sine function, Gaussian model,

Table 1
Statistics of the Covid-19 information as on Dec20, 2020 [11].

| S. No.	State / Union Territory	Total cases	Deaths	Recoveries	Active cases
1	Andaman and Nicobar Islands	4881	61	4732	88
2	Andhra Pradesh	878,723	7076	867,445	4202
3	Arunachal Pradesh	16,630	55	16,350	225
4	Assam	215,409	1017	210,869	3523
5	Bihar	247,244	1352	240,915	4977
6	Chandigarh	19,073	308	18,328	437
7	Chhattisgarh	267,219	3181	247,480	16,558
8	Dadra and Nagar Haveli and Daman and Diu	1640	1	1632	1
9	Delhi	617,005	10,277	596,580	10,148
10	Goa	50,064	721	48,371	972
11	Gujarat	235,299	4234	219,225	11,840
12	Haryana	257,644	2821	248,935	5888
13	Himachal Pradesh	52,329	864	46,221	5196
14	Jammu and Kashmir	118,263	1841	112,568	3854
15	Jharkhand	113,025	1010	110,307	1708
16	Karnataka	909,469	12,009	882,944	14,497
17	Kerala	705,870	2817	641,285	61,600
18	Ladakh	9279	124	8800	355
19	Lakshadweep	0	0	0	0
20	Madhya Pradesh	231,284	3481	216,485	11,318
21	Maharashtra	1,896,518	48,746	1,783,905	62,743
22	Manipur	27,648	337	25,646	1665
23	Meghalaya	13,253	134	12,620	499
24	Mizoram	4122	7	3963	152
25	Nagaland	18	0	9	0
26	Odisha	8901	29	8765	107
27	Puducherry	37,748	626	36,777	345
28	Punjab	163,042	5201	152,223	5618
29	Rajasthan	298,996	2617	283,957	12,422
30	Sikkim	5573	124	4981	374
31	Tamil Nadu	806,891	11,983	785,315	9593
32	Telangana	281,414	1513	273,013	6888
33	Tripura	33,164	375	32,509	257
34	Uttarakhand	2961	28	2443	490
35	Uttar Pradesh	2170	20	2068	82
36	West Bengal	536,828	9360	509,697	17,771
Total		**9,069,597**	**134,350**	**8,657,363**	**276,393**
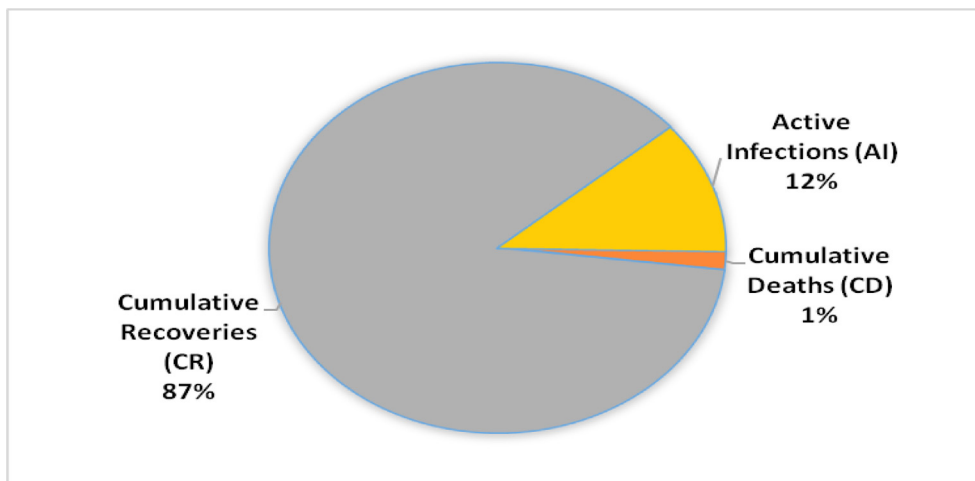
Fig. 4. Distribution of Covid-19 infections in India as on Dec 20, 2020.

Fourier series and sum of exponential function series with confidence level of 95%. But, there was not a single polynomialorder or function, which could be helpful in present scenario. Authors are hereby proposing the use of generalized polynomialregression of order 6 and discussing the function equations for cumulative infections, cumulative recoveries, cumulative deaths and cumulative active infections respectively as a function of time.Data for different functions as mentioned earlier are considered for first 115 days of spread of Covid-19 in India, i.e., during April 01 – Jul24, 2020. Table 2 summarizes the equation for each function with the R-squared value for better understanding. The model equations (as a function of time in terms of days) are used to plot the existing data from April 01 – Jul24, 2020and projection has

been madeto 20 days forward on the basis of these model equations. The plot of the same has been shown in Fig. 5 while Fig. 6 is supplied to compare the results of model equations with the actual plotted data for Covid-19 infections for first 115 days.

If we carefully analyse and compare the Fig. 5 with Fig. 3, we will see that on Day 130, (Aug13, 2020), the line showing number of cumulative active infected people is changing slope in having a small peak, thereafter the number will keep of reducing which indicates that the recovering people are more than the freshly infected people. It has been found that the projected and the actual data are in the good agreement exhibiting the similar pattern.

3. Results and discussions

This paper has attempted to investigate the spread of Covid-19 in India on the basis of information available through authenticated sources. India reported its first Covid-19 infection on January 30, 2020 and for about five weeks, there were very little or negligible infections reported across the country. During this period, lot of careful considerations were made by Government of India like thermal screening at airports, awareness among the citizens regarding Covid-19 and its spread, sanitization of probable or confirmed zones of Covid-19 infection etc. The Covid-19 infections started surging (per day reporting of fresh infections in two digits) in March 2020 and the Government of India imposed nation-wide lock-down with minimal / emergent connectivity to the rest of

Table 2
Equation for different functions.

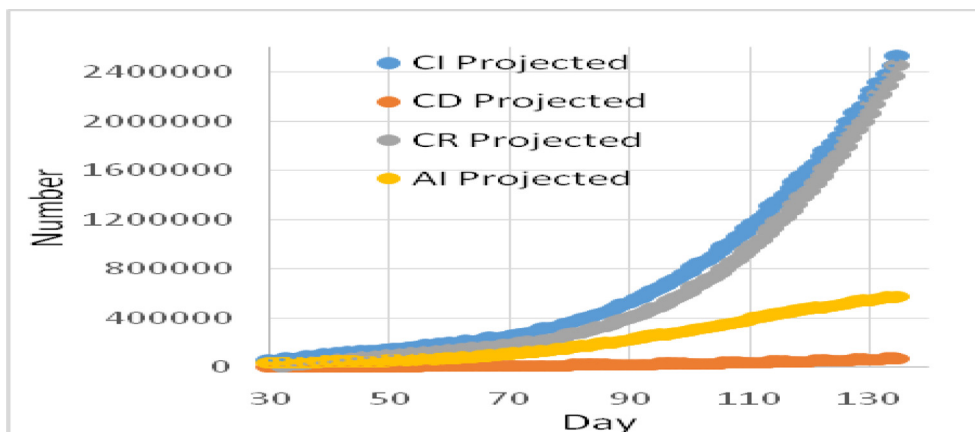| S. No.	Function for	Model Equation	R - Squared
1.	Cumulative infections (CI)	$1E\text{-}06 * x^6 - 0.0007 * x^5 + 0.1751 * x^4 - 17.607 * x^3 + 791.72 * x^2 - 10431 * x$	0.9998
2.	Cumulative recoveries(CR)	$8E\text{-}07 * x^6 - 0.0006 * x^5 + 0.1726 * x^4 - 18.97 * x^3 + 912.71 * x^2 - 14113 * x$	0.9999
3.	Cumulative deaths (CD)	$9E\text{-}09 * x^6 - 6E\text{-}06 * x^5 + 0.0015 * x^4 - 0.1376 * x^3 + 7.4558 * x^2 - 97.284 * x$	0.9998
4.	Cumulative active infections (AI)	$1E\text{-}07 * x^6 - 5E\text{-}05 * x^5 + 0.0011 * x^4 + 1.4914 * x^3 - 128 * x^2 + 3772.5 * x$	0.9795

Fig. 5. Projection of Function on the basis of model equations Vs span of time (days)upto Aug 13, 2020 (day 135).
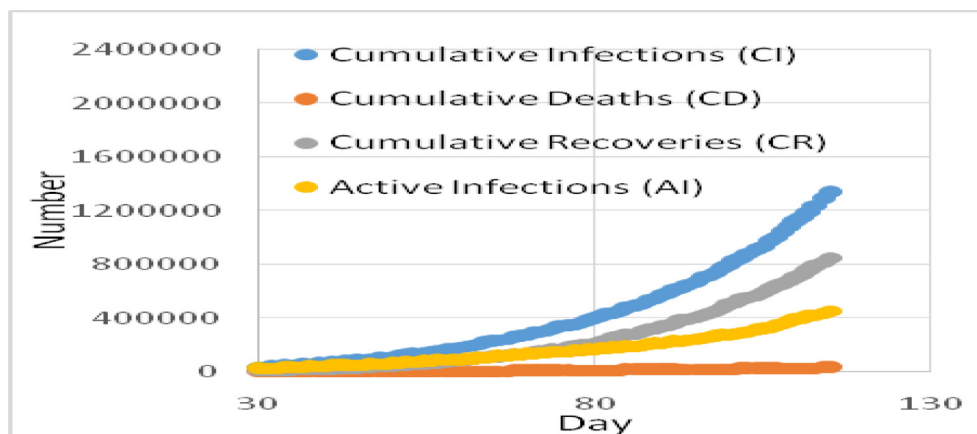
Fig. 6. Actual data of Covid-19 infectionVs span of time (days)upto Jul 24, 2020 (day 115).

world and within country, there was no transport permitted with complete stop to industry, institutions, retails etc. excluding health, law and order sector.

Initially, as evident from the international databases, there Covid-19 infections grew in small number, but because of its propagating nature, soon the infections keep on rising and the country extended the lock-down on April 15, May 04 and May 18, 2020 respectively. As of now, about 972,121,246 Covid-19 infections reported so far (Day 264) with a propagation rate of approximately 5% per day. The databases have been used to investigate the pattern of spread of Covid-19 infections and the following parameters are investigated:

(a) Cumulative infections, CI (derived from infections reported each day)
(b) Cumulative recoveries, CR (derived from recoveries reported each day).
(c) Cumulative deaths, CD (derived from deaths reported each day)
(d) Cumulative active infections, AI (derived from active CI, CR and CD respectively)

Authors attempted to investigate different models reported and applied across the globe for prediction of spread of Covid-19 and felt that the models like SIR, SIRD and SIER, though have been applied through the globe, but they are dependent over various factors depending upon the circumstances over that location and may not be applied unilaterally to any situation. As of now, there has been enormous information available regarding Covid-19 infection, its spread; it is now possible to employ simple mathematical models to predict the same. Even in India, such models had been used by various researchers, but either their projections were very much conservative or too extrapolated, and hence did not able to present a practically viable model.

Authors have implied generalized regression model for critical investigation of Covid-19 infection information available regarding India and has used the information of 115 days to predict the future spread of Covid-19.Fig. 5 can be employed to make estimates/projections for Covid-19 infections on days in advance.

4. Conclusions

The following conclusions have been drawn from the investigation presented herewith:

(a) A detailed assessment of Covid-19 infection and related aspects regarding India has been presented.

(b) Brief discussion regarding the recent work reported by various researchers through the globe is presented along with the mathematical models developed / implied by them.
(c) A simple mathematical model for assessment of Covid-19 infections in India has been presented using generalized polynomial regression. The model equations for Cumulative infections (CI), Cumulative recoveries (CR), Cumulative deaths (CD), and Cumulative active infections (AI) are derived and the projected values are compared with information available regarding the same.
(d) The Covid-19 infections are seen surging in the country from day 130 and a local peak of infections is expectedwith projections of decay in infection rate thereafter. The same can be validated through actual data plotted in Fig. 4 near to the day 130.
(e) The predicted data for CD is found in close agreement with the actual CD value and its pattern. Predicted CD pattern has a little rise in slopebeyond the day 130,showing increase in cumulative deaths in the country per day. The similar pattern can be seen in the actual data for CD plotted in Fig. 4.

CRediT authorship contribution statement

Harish Kumar: Investigation. **Pawan K. Arora:** Methodology. **Meena Pant:** Writing - original draft. **Anil Kumar:** Writing - review & editing. **Shahroz Akhtar Khan:** Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Cui, F. Li, Z.-L. Shi, Origin and evolution of pathogenic coronaviruses, Nat. Rev. Microbiol. 17 (3) (2019) 181–192.
[2] Muhammad Adnan Shereen, Suliman Khan, Nadia Bashir and Rabeea Siddique, COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses, J. Adv. Res. 24 (7) (2020) 91–98.
[3] N. Zhong, B. Zheng, Y. Li, L. Poon, Z. Xie, K. Chan, et al., Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003, Lancet 362 (9393) (2003) 1353–1358.
[4] N. Wang, X. Shi, L. Jiang, S. Zhang, D. Wang, P. Tong, et al., Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4, Cell Res. 23 (8) (2013) 986.
[5] https://www.worldometers.info/coronavirus/ (accessed on May 20, 2020).
[6] https://ourworldindata.org/mortality-risk-covid (accessed on May 20, 2020).
[7] https://www.healthline.com/health/coronavirus-covid-19 (accessed on May 20, 2020).

[8] Ivan Fan-Ngai Hung, Kwok-Cheung Lung, Eugene Yuk-Keung Tso et.al., Triple combination of interferon beta-1b, lopinavir–ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial, The Lancet, 10.1016/ S0140-6736(20) 31101-6.

[9] http://www.ucihealth.org/blog/2020/04/why-is-covid19-so-dangerous (assessed on May 20, 2020).

[10] https://www.cdc.gov/coronavirus/2019-ncov/global-covid-19/world-map. html (assessed on May 20, 2020)

[11] https://en.wikipedia.org/wiki/COVID-19_pandemic_in_India (assessed on May 20, 2020).

[12] Kaustuv Chatterjee, Kaushik Chatterjee, Arun Kumar and Subramanian Shankar, Healthcare impact of COVID-19 epidemic in India: A stochastic mathematical model, Med J Armed Forces India, 10.1016/j.mjafi.2020.03.022.

[13] Tarun Bhatnagar, Sandip Mandal, Nimalan Arinaminpathy, et al., Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India: A mathematical model-based approach, Indian J. Med. Res. 151 (2) (2020) 190–199.

[14] Jyoti Bhola, Vandana Revathi Venkateswaran and Monika Koul, Corona Epidemic in Indian context: Predictive Mathematical Modelling, 10.1101/ 2020.04.03.20047175.

[15] Jaspreet Singh, P. K. Ahluwalia and Ashok Kumar, Mathematical Model Based COVID-19 Prediction in India and its Different States, 10.1101/ 2020.05.16.20104232

[16] J.R. Gog, How you can help with COVID-19 modelling, Nat Rev Phys (2020), https://doi.org/10.1038/s42254-020-0175-7.

[17] https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model (assessed on May 20, 2020).

[18] G. Sanglier Contreras, M. Robas Mora, P. Jimenez Gómez, Estimation of the Epidemiological Evolution Througha Modelling Analysis of the Covid-19, Outbreak, Acta Sci. Mic. 3 (4) (2020) 152–158.

[19] Rashad Eletreby, Yong Zhuang, Kathleen M Carley, Osman Yağan, and H Vincent Poor, The effects of evolutionary adaptations on spreading processes in complex networks, PNAS 117 (1) (2020) 5664–5670.

[20] https://towardsdatascience.com/covid-19-infection-in-italy-mathematical-models-and-predictions-7784b4d7dd8d (retrieved on April 14, 2020)

[21] T. Chen, J. Rui, Q. Wang, et al., A mathematical model for simulating the phase-based transmissibility of a novel coronavirus, Infect Dis Poverty 9 (2020) 24.

[22] Ayush Kumar Agrawal, P.K. Arora, MusarratNafees, Shahroz Akhtar Khan, Harish Kumar,Assessment of health infrastructure in tackling COVID-19: A case study of European and American scenario,Materials Today: Proceedings,2021,ISSN 2214-7853,10.1016/j.matpr.2021.01.916.