



Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art

Gitanjali R. Shinde¹ · Asmita B. Kalamkar¹ · Parikshit N. Mahalle^{1,2} · Nilanjan Dey³ · Jyotismita Chaki⁴ · Aboul Ella Hassanien⁵

Received: 1 April 2020 / Accepted: 28 May 2020 / Published online: 11 June 2020
© Springer Nature Singapore Pte Ltd 2020

Abstract

COVID-19 is a pandemic that has affected over 170 countries around the world. The number of infected and deceased patients has been increasing at an alarming rate in almost all the affected nations. Forecasting techniques can be inculcated thereby assisting in designing better strategies and in taking productive decisions. These techniques assess the situations of the past thereby enabling better predictions about the situation to occur in the future. These predictions might help to prepare against possible threats and consequences. Forecasting techniques play a very important role in yielding accurate predictions. This study categorizes forecasting techniques into two types, namely, stochastic theory mathematical models and data science/machine learning techniques. Data collected from various platforms also play a vital role in forecasting. In this study, two categories of datasets have been discussed, i.e., big data accessed from World Health Organization/National databases and data from a social media communication. Forecasting of a pandemic can be done based on various parameters such as the impact of environmental factors, incubation period, the impact of quarantine, age, gender and many more. These techniques and parameters used for forecasting are extensively studied in this work. However, forecasting techniques come with their own set of challenges (technical and generic). This study discusses these challenges and also provides a set of recommendations for the people who are currently fighting the global COVID-19 pandemic.

Keywords COVID-19 · Forecasting models · Machine learning method · Prediction · Big data · Epidemic · Pandemic

Introduction

The world has been facing threats in the form of pandemics periodically over the centuries. The aftermath of these pandemics have always had a huge impact on the world and have also turned the tables over. COVID-19, the current devastating pandemic is also running its course currently in

the world. Not only economies are crashing but the overall strengths and morals of the heavily impacted nations are being compromised.

In order to do accurate predictions understanding of natural progression of disease is very important. A disease generally progresses because of the exposure to the infection. Because of this exposure to infection hosts are formed. Hosts

✉ Gitanjali R. Shinde
gr83gita@gmail.com

Asmita B. Kalamkar
asmitakalamkar@gmail.com

Parikshit N. Mahalle
aalborg.pnm@gmail.com

Nilanjan Dey
neelanjan.dey@gmail.com

Jyotismita Chaki
jyotismita.c@gmail.com

Aboul Ella Hassanien
aboitcairo@gmail.com

¹ Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

² Department of Communication, Media and Information Technologies, Aalborg University, Copenhagen, Denmark

³ Department of Information Technology, Techno International New Town, Kolkata, India

⁴ School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

⁵ Faculty of Computers and Information, Information Technology Department, Cairo University, Giza, Egypt

refer to the group of people who are more susceptible to get affected. When an infected host comes in contact with more people then disease starts to spread. Figure 1 depicts the host formation and progression [1].

The diseases like COVID-19, SARC, PLAGUE, etc., are acquired diseases. It means diseases spread through pathogenic agents (virus or bacteria or any microorganism).

A traditional model for the cause of the infectious disease is defined. It is called as an Epidemiologic Triad. It is depicted in Fig. 2.

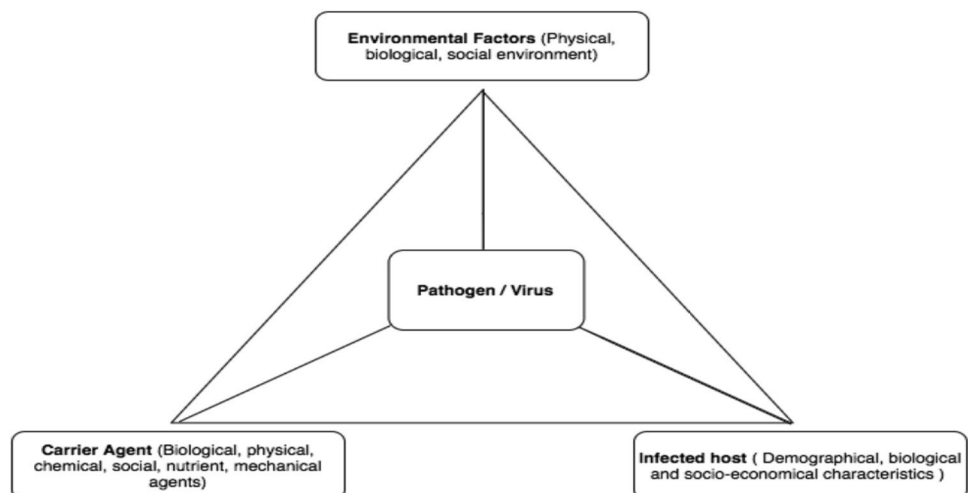
The four important factors involved in the epidemiologic triad are environmental factors, carrier agent, infected hosts and the pathogens. The agent is usually the carrier of the infection. The infection is transmitted to the host when an agent comes in contact with the host under a certain environment. A pathogen is also known as a vector. A vector is an organism that transmits the infection via virus or bacteria from one host to another [2]. Pandemics are often referred to as outbreaks because of their spread pattern. The type of the outbreak determines the mortality rate of the disease. Over the last few years, it has been seen that because of the change in lifestyle, increased global travel and urbanization, infectious diseases quickly escalate into a pandemic. To prevent these epidemics, strong policies need to be administered. Otherwise, the situation can take a drastic turn rapidly. Since the beginning, mankind has faced epidemics and pandemics. The first epidemic faced by mankind was in the early 1300's called black death. It was one of the worst pandemics

seen by humankind. This epidemic took millions of lives. It has been observed that this disease targeted most of the elderly people and people who are exposed to psychological stressors [3, 4]. The next pandemic faced by people was in the early 1500's called smallpox where 50% of the mortality rate was observed [5]. After which mankind had to face one of the deadliest pandemics called the fifth cholera pandemic which took more 1.5 million lives [6]. Following this, in 1918 one of the devastating Spanish flu influenza pandemics was observed. This pandemic took 20–110 million lives. In 1957, the Asian flu influenza pandemic was occurred which took nearly 0.7–1.5 million lives [6, 7]. In 1981, the world witnessed a new pandemic: HIV/AIDS. It was observed that more than 70 million patients were infected with the virus. According to WHO, Global health observatory data 36.7 million deaths occurred due to this pandemic [8, 9]. After the HIV/AIDS pandemic, the world witnessed a new wave of different pandemics starting with SARS in 2003. This pandemic affected 4 continents and 37 countries across the globe [10, 11]. In 2009 swine flu pandemic took place in which about 151,700–575,500 deaths were reported [12, 13]. SARS pandemic was followed by the MERS pandemic in 2012. It affected 22 countries across the globe [14]. Two pandemics then followed the MERS. First was the Ebola pandemic in 2013 followed by the zika pandemic in 2015. Both the pandemics reported deaths in thousands [15, 16]. Currently, the whole world is witnessing the COVID-19 pandemic. More than 100 plus countries till date are majorly affected by COVID-19. This count is increasing as each passing day. Throughout the history of these epidemics, one thing was observed, that is, with the progress in time, these epidemics escalated into pandemics or many times referred to as the outbreak of the virus/disease. An epidemic escalates into a pandemic when the situation gets out of control at the local source where the outbreak was first observed to spread. The novelty of the



Fig. 1 Host formation and progression

Fig. 2 Epidemiologic Triad



disease and the uncertainty that prevails regarding the disease has led to a lot of rumors regarding its whereabouts. People are unclear about the preclinical symptoms and the ways to handle it. Yet another important factor to consider is that lots of people who have preclinical symptoms do not reach the hospitals on time due to negligence or fear of testing positive for the disease. If somebody has the symptoms they have to act on it as soon as possible. This can help to save a lot of lives. If an early outbreak in any nation is successfully controlled then the situation can be prevented from escalating into a pandemic. Whenever these pandemic occur, world economies are majorly hit. Billions of dollars need to be invested in controlling an outbreak as well as in the development of a vaccine for the new disease [17]. While studying the outbreak or spread of any disease it is imminent to take all related factors into the account. Gaudart et al. [18] have taken extensions of the classical Ross-McKendrick-Mac Donald approaches. These approaches are combined with demographic and spatial dependencies of the virus on the host as well as the spread of disease. This research discusses the retro prediction model to study the spread of the COVID-19. To predict the spread of the HIV/AIDS pandemic Kaplan's model was used in [19]. But the prediction focussed on drug addicts using injector/syringe. Hence the study was focused on the spread pattern pertaining to the specific group of people. MERS was another pandemic faced by the world. In order to analyze the transmission route of the MERS, decision tree and apriori algorithms were used in [20]. In [21] a maximum likelihood method was used to assess the spread of the SARS epidemic using the construction of phylogenetic tree. In [22] SVM was used to address the same issue. The neural forecasting model was used in [23] for obtaining a forecast for swine flu.

COVID-19 is a novel disease that has evolved into a pandemic. This novel disease has been reported by the WHO on December 31, 2019, in Wuhan, China. Soon after the outbreak in China lots of countries were in the grasp

of COVID-19. According to WHO globally 634,835 confirmed cases have been registered, 29,891 deaths have been recorded till date. The region-wise statistics are shown in Fig. 3. The regions are as follows: The Western Pacific region, European region, South-East Asia region, Eastern Mediterranean region, American region and African region. Among these regions along with China, Italy, Spain, France, USA, comes under heavily infected regions.

These statistics have been taken from WHO dated March 29, 2020 [24]. From the graph, it is confirmed that this pandemic is spreading its arms across all regions. There are numerous techniques from the field of statistics, data science, ML and AI that can be used for prediction. The detailed study is presented in section three.

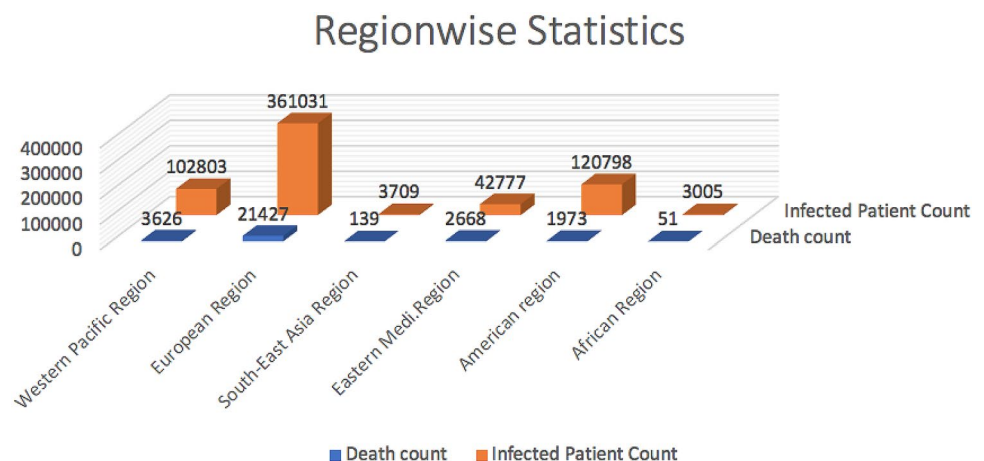
In China, COVID-19 spread took place at an unprecedented rate. Quickly situation escalated into a pandemic. The noble objectives of the researchers were to present study which can be useful for further decision-making process. In decision-making process, past data are analyzed to get perspective. However, data availability in such a short time span is not sufficient to train AI models. Effectively trainable AI models for time-series data is required (insufficient amount of available data during the initial stages of an epidemic spread). The time series data helps in improvement of efficiency of forecasting.

The objectives of the study are as follows:

- To study existing forecasting models.
- To categorize forecasting models based on type of datasets.
- To study of symptomatic and asymptomatic parameters.
- To derive challenges related to forecasting models.
- To formulate recommendations to control the pandemic.

This study is organized into four main sections. The paper starts with the natural course of the disease; categorization of the diseases, along with the global history of pandemics

Fig. 3 Region wise infected patient and death count



where the COVID-19 outbreak is also mentioned. “[Coronavirus Overview](#)” section provides an overview of the COVID-19, the different measures exercised in confining the outbreak. “[Forecasting Techniques](#)” section provides a survey of the multiple forecasting techniques and their categories. “[Discussion](#)” section deals with analysis, policies/recommendations for the control of the outbreak and the challenges that exist in the forecasting models.

Coronavirus Overview

COVID-19 affects the respiratory system of the human body which is caused due to coronavirus-2. This virus is highly contagious. It is spreading through the bodily droplets in the air. Common symptoms include fever, tiredness, and dry cough. Along with these symptoms, a patient also experiences shortness of breath, aches and pains and sore throat. Very few people have experienced diarrhea, nausea or a runny nose. People having high fever, cough or difficulty in breathing should call their doctor and seek medical help immediately. Human to human transmission is exponentially increasing the counts of the infected people. The incubation period of this disease is 1–14 days or even longer [24]. When the COVID-19 started to spread at an unprecedented rate; preventive measures were exercised. These measures included a complete lockdown of the heavily infected areas, ban on international travels, suspending schools and other non-essential daily activities. The main aims of these measures were to limit interpersonal contact, considering the contagious nature of the disease. The curfew was imposed and strictly observed. As the incubation period of the virus is longer than other viruses it is very difficult to analyze the optimal time required to observe a curfew. If the curfew is lifted too soon the situation can become dangerous. The people who get infected fall under three categories. First in the category are the elderly, who are highly susceptible to the virus. Statistics show that because of the weak immune system the elderly succumb to the disease easily. The second category is that of the children. As the immune systems of young children are still under development, the children are at higher risk. The third category is that of the people who have diseases like diabetes, high BP, asthma, cancer, cardiovascular disease, etc. As their immune systems have been compromised already due to a prevailing medical condition, these people become easy targets. Infections experienced by the third category of people can be fatal [17].

Forecasting Techniques

In the literature, forecasting has been done based on various forecasting techniques and different data sources. To understand and improve the forecasting this section categorizes

these techniques into multiple types for better analysis. This categorization is done based on the data sources used, i.e., big data accessed from WHO/National databases and data from social media. However, the main aim of this study is the analysis of forecasting techniques in computing and processing perspectives. In the view of this, data in terms of population statistics is considered for discussion throughout in this paper. The main advantages of using population statistics are there is no need of sampling as the entire population is present in the dataset. Population statistics also help to make reliable prediction and estimates with less computational overhead and there is a lack of bias. In the literature, many studies are also carried out on clinical data. These studies may be useful for physician, doctors, and researchers in the medical domain for investigating better diagnostic methods and for pharmaceutical industries in formulating vaccines, drugs in a short time. Categorization is also done based on techniques that are used for forecasting, i.e., data science/machine learning techniques. However, there are also a few other categories that are used in the literature for forecasting. In nutshell, these categories are broadly divided into the following four sets:

- (a) Big data.
- (b) Social media/other communication media data.
- (c) Stochastic theory/mathematical models.
- (d) Data science/Machine learning techniques.

Various statistical, analytical, mathematical and medical (symptomatic and asymptomatic) parameters are taken into consideration for analysis. However, major significant parameters are listed below:

- (a) Daily death count.
- (b) Number of carriers.
- (c) Incubation period.
- (d) Environmental parameters, i.e., temperature, humidity, wind speed.
- (e) Awareness about COVID-19.
- (f) Medical facilities available.
- (g) Social distancing, quarantine, isolation.
- (h) Transmission rate.
- (i) Mobility.
- (j) Geographical location.
- (k) Age and Gender.
- (l) Highly and least vulnerable population.
- (m) Underlying disease.
- (n) Report time.
- (o) Strategic policies and many more.

Apart from these above-mentioned parameters, there can be many influential factors that need to be further investigated. The following section presents a parametric

evaluation of the state-of-the-art by classifying various studies into four aforementioned categories. Every evaluation is supported by the table where work ref. represents the research work which is referred for study, studied regions indicate the countries which data is taken for study, parameters present the factors on which study is based and remark represents the outcome of the study.

Big Data

Effectiveness of forecasting is based upon the quality of data source used for forecasting. Forecasting results may vary based on the impurities in the data sources. Data mining and big data techniques always play a vital role in healthcare systems [25–28]. In the literature, researchers have done forecasting based upon data sources received from authenticated national and international sources. Here, analysis of big dataset is done by using various techniques like mathematical equations or machine learning techniques. Soumyabrata Bhattacharjee [29] has presented the impact of environmental factors like temperature, wind speed and humidity on the spread rate. This analysis is done based on the data accessed from the WHO and the local weather database. Toda [30] has presented decision-making schemes by analyzing the COVID-19 data of countries like China, Japan, Korea, European countries, and North America obtained from Johns Hopkins University. Caccavo [31], Siwiak et al. [32], Zareie et al. [33], Teles [34] and Russo [35] have analyzed COVID-19 databases accessed from WHO, Italy national data and Johns Hopkins to predict the mortality rate. Liu et al. [36] presented the impact of disease control interventions and traffic restrictions on the spread rate. The analysis has been done on the dataset retrieved from US Centers for Disease Control (CDC). Nadim et al. [37], Pear Hossain et al. [38], Tarcísio et al. [39], Train et al. [40] have presented the importance of quarantine in order to reduce the spread rate of COVID-19. Giordano et al. [41] have presented the data analysis of Italy based on Italy's national data. As per Italy's official release, there are a total of 27,980 infected cases and 2158 deaths of people who were positive of coronavirus. Looking at the effect of the Pandemic in Italy, Giulia Giordano has proposed the SIDARTHE Model that helps in redefining the reproduction number. This epidemic prediction model compares the infected density with the level of symptoms. Wangping [42] has presented a study in which, COVID-19 data from Jan 22, 2020, to Mar 16, 2020, has been used in time series form for analysis. The prediction has been estimated using the Markov Chain Monte Carlo method and results show that the reproductive number in Italy is 4.10 and 3.15 in Hunan. The anticipated endpoint in Italy would be April 25. Details of the literature evaluation are summarized in Table 1.

Social Media Data/Other Communication Media Data

In this digital era, social media communication and internet searches are the most easily accessible platforms that provide more information about COVID-19. The social media and web search correlate with the number of daily COVID-19 cases. Keeping this in mind few researchers have taken datasets from Google, Baidu search engines [43, 44], mobile phones [44, 45], newspapers [50] and various websites [47–48] like Github [49] over a particular duration of time. Analysis of these datasets is done by various techniques as discussed before, i.e., machine learning techniques or mathematical equations/stochastic theory based on the parameters which were discussed earlier. Zhu et al. [45] have presented a spatially pandemic model for predicting the death count. This study aims to build a prediction model that will analyze the growth of the virus for the next month considering the current dynamics of COVID-19. Three different scenarios have been taken into consideration for the study which includes residents, residents with Wuhan travel history and residents affected as a result of a local outbreak. The decay rate has also been introduced in the study to appreciate the efforts of different cities to alleviate the spread of the disease. Phone data has been used to collect the statistics of city-wise residents who had traveled back from Wuhan and the city-based model has been trained using the prevailing statistics and validated against the new cases as on February 11. The same model has been used to predict cases up to March 12, 2020, under the aforementioned three scenarios. The study predicted that the number of infections would be around 72,172, 54,348 and 149,774 by March 12, 2020. The potential outcome of the study is a spatial model and its predictions will certainly help in optimizing the allocation of resources in each city during the next 1 month when the epidemic reaches a serious state of concern. Details of this analysis are summarized in Table 2 as follows.

Stochastic Theory/Mathematical Models

In a few past pandemics, the traditional approach of the mathematical and stochastic theory was used to estimate the loss of human and also to predict the total death count until a particular period or end of the pandemic. This traditional approach is very effective and shows better predictions. Hence in the current pandemic situation of COVID-19 researchers [52–57] have used the same traditional approach for estimating the death count and the spread rate of COVID-19. The approach is also used to predict the total death count till the end of the pandemic. The analysis is done on databases accessed from authorized sources or search engines, mobile phone data and newspaper reports. Sameni [58] has proposed a pattern of the virus with the help of mathematical

Table 1 Evaluation of COVID-19 forecasting on Big Data

Sr. no	Work ref.	Studied regions	Data source	Parameters	Remark
1	[29]	China	WHO, Local Weather Underground	Maximum relative humidity, maximum temperature, and highest wind speed	Impact of environmental factors on spread rate
2	[30]	China, Japan, Korea, European countries, and North America	Johns Hopkins University, GitHub repository	Transmission rate, Infection rate, and recovery rate	Recommendations for decision making
3	[31]	China, Italy	WHO, Johns Hopkins University	The average number of contacts per person per time, kinetic of recovery, death, number of susceptible, infected, recovered and the death	Forecasting numbers of COVID-19 patients
4	[32]	Wuhan, China	Johns Hopkins GitHub repository	Number of susceptible individuals; the number of infected individuals	Forecasting of mortality rate
5	[33]	Iran	WHO	Time-dependent transmission rate, time-dependent recovery rate, and time-dependent mortality rate	Prediction of the number of COVID-19 patients in next month
6	[34]	Italy, Portugal	Italy national data	Number of susceptible, exposed, asymptomatic infected, mild-to-severe infected patients	Forecasting numbers of COVID-19 patients
7	[35]	Italy	Italy national data	Total population, number of confirmed COVID-19 patient, number of unreported infected persons, recovered persons, deaths counts	Forecasting numbers of COVID-19 patients
8	[36]	US	US Centers for Disease Control	Disease control interventions and traffic restrictions	Impact of disease control interventions and traffic restrictions on spread rate
9	[37]	China	WHO	Number of quarantined people, number of isolated people	Impact of management of quarantine and isolation patients on spread rate
10	[38]	China	International Air Transport Association database, Chinese Center for Disease Control and Prevention	Mobility, quarantine	Impact of border control and quarantine on the spread count
11	[39]	Brazil	WHO	Number of susceptible, exposed, infectious and recovered patients	Suggested policy-making for avoiding outbreak in metropolitan cities
12	[40]	Italy	Italy national data	Contact rate, the transmission rate	Impact of quarantine on spread rate
13	[41]	Italy	Italy national data	Transmission rate, probability rate of detection, mortality and recovery	Forecasting numbers of COVID-19 patients
14	[42]	Italy	Johns Hopkins University	Early detection and isolation of individuals with symptoms, traffic restrictions, medical tracking, and entry or exit screening	Impact of policies on spread rate

Table 2 Evaluation of COVID-19 forecasting on social media Databases

Sr. No	Work Ref.	Studied regions	Data source	Parameters	Remark
1	[43]	China	Baidu Search engine and Tab	Number of basic regenerations, the incubation period and the average number of days of cure	Impact of future prediction and backward on the spread of COVID-19
2	[44]	China	Baidu map big data	The incubation period, number of carriers, contact rate	Impact of medical facilities, Social responsibility, administrative responsibility on death count and spread rate
3	[44]	China	Mobile phone data	Domestic and international travel	Impact of domestic and international travel on COVID global spread
4	[45]	China	Mobile phone	Spatially pandemic model, the Decay rate	Prediction of the death count
5	[46]	China and Korea	https://www.worldometers.info/coronavirus/	Sub-population of latently infected individuals, the incubation period	Quarantine is not sufficient and stricter measures are needed
6	[47]	China	Website	Reproduction number	Forecasting of death numbers
7	[48]	China	Internet searches and social media data	Confirmed cases and suspected cases of COVID-19	Prediction of COVID-19 outbreak
8	[49]	US	Github page	Social Distancing (school closure)	Impact of social distancing on the death count
9	[50]	Italy	Websites of the main Italian newspapers	Time and space	Impact of space and time on decision makers to intervene on the local policies

modeling. This study uses a model from the family of the well-known compartmental models known as susceptible infected-recovered (SIR) model. A study has shown that the measures taken by the countries are positively affecting the mortality rate. Along with that, the facilities that are created to house the infected people, has contributed greatly in stopping the spread of the disease. However, this mathematical model has limitations in terms of accuracy because it is developed for the underlined dataset. Yuan et al. [59] presented the Boltzmann's function-based analysis. It has been observed that the prediction accuracy is better and it can also help in assessment of the severity of the situation and take appropriate actions. Dowd et al. [60] proposed the impact of age and gender on the death count using mathematical modeling. It has been observed that this virus is largely affecting the elderly. Now, in this case, the age structure of a particular country plays a vital role. In Italy, 23% of the population is above 65 years of age and hence the threat is maximized for the countries having similar age structure as that of Italy. The same situation can be faced by South Korea. Hence the policies like social distancing and quarantine can help to slow down and stop the spread of the virus. He et al. [61] has presented the impact of pre-symptomatic transmission on the death count using mathematical modeling in this infector-infections and the transmission rate is studied. From the observation, it was inferred that the rate of transmission was at its peak on or before the symptom onset. 44% of transmission can be seen even before the first

symptoms become physically visible. Hence the disease control authorities should take the pre-symptomatic transmission into account while implementing the measure to curb the spread. Giannakeas et al. [62] presented an online tool for healthcare management using stochastic theory. Banerjee et al. [63] presented the impact of underlying conditions like heart disease and diabetes on the death rate. The impact of mobility on the spread rate of COVID19 is presented by Alexander et al. [64]. Chen et al. [65], Ma et al. [66] and Shi et al. [67] presented the impact of environmental factors on death count and spread rate of COVID-19. This analysis is based on the parameters listed as earlier and the details of this analysis are summarized in Table 3.

Data Science/Machine Learning Techniques

Nowadays machine learning techniques are used worldwide for predictions due to its accuracy. However, to use machine learning (ML) techniques, there are a few challenges as very little data is available. For instance, the challenges involved in training a model are the appropriate selection of parameters and the selection of the best ML model for prediction. Researchers have done predictions based on datasets that are available and used the best ML model as per the dataset [17, 69–71]. Kumar and Hembram [72] presented a model based on the Logistic equation, Weibull equation, and the Hill equation to find infection rates in China and Italy. In this research work, data analysis is done to understand the effect

Table 3 Evaluation of COVID -19 forecasting based on the mathematical and stochastic theory

Sr. no	Work ref.	Studied regions	Data source	Parameters	Remark
1	[51]	China, Italy, Iran, Germany, France, USA, South of Korea	WHO	Effectiveness of intervention, public response, and healthcare system	Forecasting numbers of COVID-19 patients
2	[52]	South Korea, Italy, France, and Germany	National Data	Reported and unreported numbers	Forecasting number of COVID-19 cases
3	[53]	Global death count, 129 countries	John Hopkins Hospital, WHO	Exposed and infected population	Prediction of infection rate
4	[54]	China	Centers for Disease Control, China	Stages of spread	Prediction of spread rate
5	[55]	China	WHO	Fitting parameter calculated from the total number of cases and new cases each day	Prediction of spread rate
6	[56]	Brazil	Official Airline Guide Data SUS	Outbreak probability, effective distance, Social vulnerability	Prediction of spread rate
7	[57]	China, Italy, Japan and Germany	Johns Hopkins University	Duration, number of deaths	Presented epidemic spread pattern
8	[58]	Global death count	WHO	Mortality rate, infection rate, re-infection rate, the recovery rate	Disease control measures: Lockdown, social distancing
9	[59]	China	WHO	Number of deaths	Forecasting the cumulative number of COVID-19 deaths
10	[60]	Italy, Nigeria Brazil, USA, UK	National Data	Age and gender	Impact of age and gender on spread rate
11	[61]	China	Data of Hospital, China	The incubation period, pre-symptomatic transmission, post-symptomatic transmission	Impact of incubation period on spread rate
12	[62]	US	American Hospital Association	Type of bed acute care bed, critical care bed, and ventilator	Online tool for no of patients can be admitted in hospital
13	[63]	UK	Electronic health records in England	Impact of underlying conditions like heart disease, diabetic on mortality	Estimation of excess 1- year mortality from COVID-19 with underlying conditions
14	[64]	China	Chinese Center for Disease Control and Prevention	Mobility rate	Impact of mobility on spread rate
15	[65]	China, Italy, Japan	WHO	Air temperature, relative humidity, wind speed, and visibility	Multi-factors can impact on spread rate rather than single factor
16	[66]	China	People's Republic of China	Number of death, temperature, humidity	Environment factors may impact COVID-19 mortality rate
17	[67]	China	China National Health Commission, meteorological authority in mainland China	Impact of temperature and absolute humidity on the COVID-19	Lower and higher temperatures may be positive to decrease the COVID-19, there is no major impact of humidity

of environmental factors on the spread of COVID-19. Data analysis is done on 4 cities in China namely Beijing, Chongqing, Shanghai and Wuhan and 5 cities of Italy namely Bergamo, Cremona, Lodi and Milano. The number of infected people is greater in the above-mentioned cities. Three environmental factors are mainly focused on this study, i.e., maximum environmental temperature, relative humidity, and wind speed. For data analysis, data is collected from a report published by the WHO for China and Italy. Data is taken from the official GitHub repository of the Department of Civil Protection, Italy. The results show that there is a negligible relation between humidity and wind speed with the spread of COVID-19. Similarly, it has been observed that higher/maximum temperatures have a negligible to a moderate impact on the spread of the virus. The result shows that there is no sign of any major effect of temperature on the virus. However, results may vary depending on the dataset. DeCapprio et al. [73] proposed a model using logistic regression, gradient boosted trees, and a hybrid model using Medicare data. The outcome of these models will help to initiate control strategies and to initiate corrective measures in time to control the spread. The details of this analysis are explained in Table 4.

From the literature review, it is evident that all studies have taken data from standardized data sources however, these datasets are not yet standardized by any standardization organization or allied bodies. In these studies, the geo-spatial and statistical anomalies are not considered; however, these may be interesting enablers for better forecasting. In the literature impact of environment factor and mobility on COVID-19 spread is considered [64, 65]. Various stages of COVID-19 outbreak are well-explained in [50], where the understanding of outbreak stages may help to reduce the rate of spread. Various ML models are discussed in the

literature however for better accuracy deep learning models can be used for better predictions [74]. Furthermore, predictions can be more accurate using active learning models in this multitudinal and multimodal data used for predictions instead of single type of data [75].

Tremendous work has been going on the COVID-19 apart from the above discussed work [77–82]. Researchers are working to investigate efficient and accurate models in order to predict the death count. Researchers are also working to provide a list of guidelines that can be followed by the people to reduce the spread rate of the COVID-19.

Discussion

As stated earlier, the literature survey presented above is based on broadly four categories like the size of the dataset, source of the dataset, and techniques applied for forecasting like mathematical/analytical or machine learning/data science. This survey is carried out on various medical and non-medical parameters and it is very clear that the basic purpose of all these studies is to estimate the final size of this COVID -19 pandemic. However, it is very interesting to note that, all the studies have referred to the China epidemic as the basis and all forecasts have been done based on the early statistics which are available from the outbreak in China. Outcomes of these studies are very much useful for multiple purposes like controlling the spread of COVID-19 globally, controlling the spread of COVID-19 for a specific country, deciding its impact, building vulnerability index of COVID-19, establishing a correlation between environmental conditions (metrological conditions) and the spread rate, deciding reproduction number, establishing the correlation between quarantine and isolation with the spread of COVID-19, trend

Table 4 Evaluation of COVID-19 forecasting based on Data Science/Machine Learning Techniques

Sr. no	Work ref.	Studied regions	Data source	Parameters	Remark
1	[17]	China	Small dataset	Corrective feedbacks of model	Forecasting suspected numbers of COVID-19
2	[68]	China	Chinese Center for Disease Control and Prevention	Cost of isolation, cost of treatment, no of suspects, no of confirmed COVID patients	Recommendation for decision making
3	[69]	China	WHO	Daily death count	Forecasting of death count
4	[70]	102 countries	WHO	Degree of intervention and starting intervention time	Impact of a public health intervention on the global-wide spread
5	[71]	China	2003 SARS Data	Death count	Forecasting of death numbers
6	[72]	China and European countries	WHO	Infection rate	Prediction of infection rate
7	[73]	Global Data	International Classification of Diseases	Preexisting medical conditions	Identify individuals who are at the greatest risk

analysis of COVID-19 pandemic and tracking the spread of COVID-19 locally and globally. The COVID-19 pandemic having been in existence for a very short period now, it is very important to analyze the trend of its spread and infected cases. All affected nations are looking toward mitigation plans to control the spread of the disease with the help of some modeling techniques. In the sequel, the outcomes of these forecasts are multi-fold. Every forecast is carried out with some perspective irrespective of which category it may represent. From these studies and the forecasts made, it is very clear that the major outcome is to support healthcare communities to initiate critical action, decisions, control measures and public restrictions in time. Another outcome is to support in establishing mechanisms that provide control measures to be considered internationally for the global control of this pandemic as well as restrictions to the public in terms of quarantine, isolation, contact tracing, the recommendation in terms of metrological conditions (mainly Air, Temperature, relative humidity, wind speed and visibility) and its impact on the spread. However, despite these useful outcomes, there are still many issues and challenges which are still unaddressed. The first and most important issue is whether the modeling and predictions based on China's dataset would suffice to address the issues of all countries. There is a need for reassessment to ensure that the control measures initiated by China to regulate the outbreak are enough to control this global pandemic. Many researchers have presented models for disease predictors to decide the reproductive number, but all of them have relied on similar datasets. It is also a crucial factor to rethink whether the same mathematical or prediction model is also suitable to predict the spread and reproduction number for all the countries across the globe. Literature shows that all the models presented are tested based on the numbers of the China epidemic. It is also equally important to ensure that the same model tested for China's dataset can also be applied to control the outbreak of COVID-19 globally. Another issue is that in the literature very limited details regarding the key characteristics of Coronavirus and the symptoms of COVID-19 are available. In this sequel, the challenge is to identify a vulnerable group of people with these limited details regarding viruses and disease. There is also a need to consider multiple peaks in the model not only for short term prediction but also to predict the outbreak later in the year. Before confirming the forecasting mechanism, there is a need to reconsider these issues and challenges for better accuracy. There are few and important medical and non-medical parameters that still need to be investigated as evident from the literature. A few of which are, the genetic relations pertaining to the geographical locations need to be studied in order to confirm the forecast. The Ethnicity (civilization, society, culture) of the infected people is another important parameter that needs to be reviewed. Correlation between the spread and its impact

on a specific patient considering the underlined preexisting medical complications is also another important parameter to be considered for more effective and accurate forecasting. It should be noted that not a single study or model available in the literature has considered the existing treatment options and has assumed that no vaccination option will be available for the next 1 year [83]. However, these are also some important parameters which need attention to fine-tune the model further.

Challenges of Forecasting Models

Forecasting plays an important role in every domain [85–86] due to its benefits to save resources or to improve the economy. However, it comes with its challenges. In the case of COVID-19, there are also many challenges for forecasting the death count and spread rate as COVID 19 incubation period is very much longer and very fewer datasets are available for this purpose. Few such challenges of forecasting models are listed as follows:

- (1) *Tracking of the people* The tracking of infected personnel and other people who came in contact with them is truly one of the difficult tasks.
- (2) *Longer incubation period* As COVID-19 has an incubation period of 14 days, it is impossible to identify patients beforehand. During the defined incubation period patients can infect all the people who come in contact with him/her.
- (3) *Lack of proper data* Sometimes data are available in unstructured format. Hence is very necessary to maintain quality and quantity of the data before it goes in training stage. Data accuracy is an important factor in achieving effective forecasting methods.
- (4) *Overfitting of the data* If overfitting of the data occurs then it is possible that model in question will not perform well on new data.
- (5) *Overly clean data* It is important to have clean data for the analysis purpose but too clean data sometimes loses its integrity.
- (6) *Abundance of data* Data are available in abundance but feeding all these data to model will not improve the accuracy.
- (7) *Wrong algorithm and attribute selection* If wrong algorithm is selected then the result can be misleading. Same is true in case of wrong of wrong attribute selection.
- (8) *Model complexity* If model is too complex it can affect the overall performance if the model.

Along with these challenges, some more challenges are important to make a note of:

- *Proper lockdown* It is very difficult for any country to implement a lockdown. To decide the proper conditions of a lockdown is a very complicated task.
- *The optimal period for lockdown* The optimal period for lockdown is not only crucial but also a critical task.
- *Aware but do not cause panic* It is important to educate people but in the process, it is important to remember not to create panic.
- *Essential services identification and delivery* It is imminent for any country to identify essential services before lockdown. Even among lockdown lack of these services can cause a massive panic.

Recommendations

Predictive models are used to generate predictions depending on the underlying model. The accuracy of model depends on underlying algorithm. Until now statistical model has been used for predicting or determining the relationship between the attributes or variables. In concern with COVID-19, the situation is changing drastically with each passing day. As previously seen the data is been gathered in many different formats, also size of data is huge. In this scenario deep learning-based models can come in handy. Deep learning algorithms can provide better visualization and optimization. If time-series data are available in future AI algorithms can be helpful. On the same line the collaborative filtering algorithms can be used for detection of certain pattern. The detailed analysis of existing system can be helpful in determining the next course of action.

If an epidemic is controlled properly in the initial stage and if proper measures are taken to prevent it from crossing the geographical boundaries then it could save a lot of lives with lesser impact. Forecasting and proper study of the pattern of disease spread could be very helpful in the planning of control strategies. At this stage, a complete lockdown imposed in the affected area (already implemented by many countries) is a good solution to prevent and hopefully stop the spread (local transmission). But when the epidemic turns into pandemic it covers a larger geographical area. This means tremendous growth in the number of infected people. This stage greatly impacts the nations that have been infected with the virus. Awareness about the pandemic among the people is a very crucial part. If people know what the symptoms are and how to act if they have them, then it can help doctors in the process. If people are afraid then they might not come forward and it can lead to a very disastrous situation. Furnishing proper information about the symptoms and treatments to the public might be helpful. Also keeping tabs on the rumors that prevail around the disease is a necessity. One rumor can turn the entire situation into chaos. As previously mentioned, the incubation period of this virus is

too long as compared to other viruses. It is very important to keep track of the infected victims. Along with that, the people who came in contact with an infected person must be moved into the quarantine facilities. These quarantine facilities should be equipped with isolation units. The people who came in contact with the infected person can be put into isolation until the incubation period is over. Following the incubation period testing should be performed on these patients. Additionally, checking the travel history and daily routine of the patient before he/she encountered the infection is vital. This will help them in identifying the patient zero or the carrier agent. All the people who came into immediate contact with the infected patient needs to quarantine himself/herself. Home quarantine is also a good option if an isolated room with different latrine facilities is available. Whenever in-home quarantine care should be taken that the person remains within the confined limits of access until the test results turn out to be negative. If home quarantine facilities are not available, such facilities should be made available for suspected patients. An early forecast of the upcoming situation may help to construct a feasible solution. If accurate predictions about the growth in several patients are done, then situation can be handled in more efficient manner. COVID-19 is highly contagious. Doctors, nursing staff and supporting staff should wear masks when treating the patients. Alcohol-based sanitizers should be used for sanitization. Also, if possible hazmat suites can be used by the doctors to treat patients who are showing severe symptoms. Provision of masks and all necessary stock to the hospitals is primary need so that they can work effectively. Social distancing is one of the measures that can be implemented. It means people should maintain at least two meters distance between themselves. This can potentially stop the spreading of the disease.

In the face of this pandemic, most of the nations are in complete lockdown state. But in this lockdown searching for alternative measures to deliver food, medications and other essential services to the people is vital. A detailed area wise timetable for delivery as per request from people can potentially stop people from rushing for supplies and creating havoc. Also until the cities are in complete lockdown complete sanitization of the cities can be done. Sanitization process can start with public places and then other parts of the cities can be covered. It has been observed that the people who are having diseases like high BP, diabetes, asthma are more susceptible to the infection. Also, children and elderly people are at a higher risk. Identification of such people and keeping track of their count will be beneficial. Finally, people should take really good care of personal hygiene. Frequent washing of hands, avoiding touching of face and eyes frequently, covering the mouth whenever sneezing or coughing, avoiding physical contact and drinking at least 3 liters

of water daily are a few activities that can help maintain personal hygiene. People should strictly follow the lockdown conditions imposed by the country or city. People should avoid stepping out of the house until it is very necessary. Avoiding air conditioners is a good practice as a controlled temperature can affect the health very easily. Lockdown potentially alters the lifestyle and routine of people. A complete lockdown can cause a massive panic. Entertainment via TV or other mediums like Netflix, Amazon prime, HotStar, etc. can provide a little relief. The complete lockdown severely affects the economy of the world. Work from the home policy can come in handy in unforeseen situations like these. Universities can provide students with online classes. Hence the academic loss can be contained. Also, online assessments can help in the process. Following are a few recommendations to stop the spread of the disease as soon as possible:

- Strict action should be taken against the people or organizations that violate the lockdown conditions without any convincing reason. All public transport services can be suspended except for the transportation of essential services and goods.
- All the places of worship can be closed for prayers. No religious congregations should take place in this period. All types of social gatherings (political, academic, cultural, etc.) should be banned.
- Identification of vacant facilities that can be turned into quarantine facilities is essential. People with recent travel history to any foreign land/infected countries should be kept in strict isolation for the 14 days.
- Identification of the services that come under the essential category is very important. The personnel who are responsible for making essential services available to the public should be provided with passes for the ease of transportation. The export and import of goods should be restricted.
- Strict action against people who black-market essential commodities and services. State-wide borders can be sealed off until the situation gets under control.
- The offices that come under essential services should instruct employees to follow social distancing while working. Also, appropriate sanitization facilities like hand wash, sanitizers, etc. should be made available for employees.
- In the case of death, there should be a restriction on the number of people (not more than 12–15) allowed to attend the funeral.

These are really difficult times but better preparedness can help avoid a lot of panics.

Conclusion

The COVID-19 pandemic is spreading its wings across the globe at a surprisingly faster rate and has already resulted in thousands of deaths across countries. Unfortunately, this number is sure to grow within a short period and healthcare organizations would soon face scarcity of resources. In this sequel, it is important to analyze various forecasting models for COVID-19 to empower allied organizations with more appropriate information possible. An overall comprehensive study on analysis of COVID-19, its forecasting, impacts, and control measures are presented in this study. The major contribution of this study is the analysis of several forecasting models available in the literature and their classification, challenges of these models and recommendations to control this pandemic. Based on the available forecasting methods, we studied various statistical, analytical, mathematical and medical (symptomatic and asymptomatic) parameters. Also, common yet significant parameters have been taken into consideration which includes death count, metrological parameters, quarantine period, medical resources, mobility, etc. In this study, we have done the categorization of various forecasting methods into four major sets which include big datasets accessed from WHO/National data sources, social media/other communication media data, stochastic theory/mathematical models and data science/Machine learning techniques. This classification will surely help researchers to consolidate the forecasting methods more crisply and concisely as presented in this study.

Our study indicates that there is a need to reassess control measures initiated by China and other countries. Prediction of the spread and reproduction number should be analyzed on varied datasets. The models presented in the literature should be tested globally for more accurate global forecasting. On similar grounds, there is also a need to consider multiple peaks in the model not only for short term prediction but also to predict the outbreak later in the year. This study also indicates the challenges of various forecasting models and useful recommendations for the control of this pandemic.

We hope that by providing analysis of various forecasting models of COVID-19 will be more helpful for adapting better intervention policies and explicitly, it will also help to alleviate the alarming effect of this pandemic. We agree that many of the papers referred to in this study for analysis are pre-print, i.e., they do not peer review formally. However, due to the rapid growth of COVID-19 globally, there is a strong need for such a comprehensive survey as a contribution toward the society.

Compliance with Ethical Standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

- Dicker RC, Coronado F, Koo D, Parrish RG. Principles of epidemiology in public health practice; an introduction to applied epidemiology and biostatistics. 2006
- Gordis L. Epidemiology. 4th ed. Philadelphia, PA: Elsevier Saunders; 2014.
- Platt C. King death: the black death and its aftermath in late-medieval England. Oxon, U.K.: Routledge; 2014.
- DeWitte SN. Mortality risk and survival in the aftermath of the medieval black death. *PLoS ONE*. 2014;9(5):e96513.
- Diamond J. Guns, germs, and steel: the fates of human societies. New York: Norton; 2009.
- Frieden NM. The Russian cholera epidemic, 1892–93, and medical professionalization. *J Soc History*. 1977;10(4):538.
- McKibbin WJ, Sidorenko AA. Global macroeconomic consequences of pandemic influenza. analysis. Sydney, Australia: Lowy Institute for International Policy; 2006.
- Dixon S, McDonald S, Roberts J. AIDS and economic growth in Africa: a panel data analysis. *J Int Develop*. 2001;13(4):411–26.
- Flahault A, Valleron AJ. HIV and travel, no rationale for restrictions. *Lancet*. 1990;336(8724):1197–8.
- Keogh-Brown MR, Smith RD. The economic impact of SARS: how does the reality match the predictions? *Health Policy*. 2008;88(1):110–20.
- Achonu C, Laporte A, Gardam MA. The financial impact of controlling a respiratory virus outbreak in a teaching hospital: lessons learned from SARS. *Canad J Public Health*. 2005;96(1):52–4.
- Tizzoni M, Bajardi P, Poletto C, Ramasco JJ, Balcan D, et al. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Med*. 2012;10(1):165.
- Jain S, Kamimoto L, Bramley AM, Schmitz AM, Benoit SR, et al. Hospitalized patients with 2009 H1N1 influenza in the United States, April–June 2015. *New England J Med*. 2009;361(20):1935–44.
- Park J, Kim J. Hong Kong sets ‘serious’ response to South Korea’s MERS outbreak.” *Reuters*, June 8, 2015
- Pandemic risk. Background paper for world development report 2014: Risk and opportunity; managing risk for development, World Bank, Washington.
- UNDP (United Nations Development Programme). A socio-economic impact assessment of the Zika virus in Latin America and the Caribbean: with a focus on Brazil, Colombia, and Suriname. UNDP, New York: Synthesis report; 2017.
- Fong SJ, Li G, Dey N, Crespo RG Herrera-Viedma E. Finding an accurate early forecasting model from small dataset: a case of 2019-ncov novel coronavirus outbreak. *arXiv preprint arXiv:2003.10776*. 2020
- Gaudart J, Ghassani M, Mintsu J, Waku J, Rachdi M, Doumbo OK, Demongeot J (2010) Demographic and spatial factors as causes of an epidemic spread, the copule approach: application to the retro-prediction of the black death epidemic of 1346. In: 2010 IEEE 24th International conference on advanced information networking and applications workshops (pp 751–758). IEEE.
- Greenhalgh D, HAY G. Mathematical modelling of the spread of HIV/AIDS amongst injecting drug users. *Math Med Biol J IMA*. 1997;14(1):11–38.
- Kim D, Hong S, Choi S, Yoon T. Analysis of transmission route of MERS coronavirus using decision tree and Apriori algorithm. In: 2016 18th International conference on advanced communication technology (ICACT). 2016. (pp 559–565). IEEE.
- Amiroch S, Pradana MS, Irawan MI, Mukhlash I. Maximum likelihood method on the construction of phylogenetic tree for identification the spreading of SARS epidemic. In: 2018 International symposium on advanced intelligent informatics (SAIN) 2018. (pp 137–141). IEEE.
- Hu B, Gong J. Support vector machine based classification analysis of SARS spatial distribution. In: 2010 Sixth international conference on natural computation. 2010 (vol. 2, pp. 924–927). IEEE.
- Sultana N, Sharma N. Statistical models for predicting swine flu incidences in India. In: 2018 First international conference on secure cyber computing and communication (ICSCCC). 2018 (pp. 134–138). IEEE.
- World Health Organization online available on <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/>.
- Bhatt C, Dey N, Ashour AS (eds). (2017). Internet of things and big data technologies for next generation healthcare.
- Hassanien AE, Dey N, Borra S (eds) (2018). Medical big data and internet of medical things: advances, challenges and applications. CRC Press, Boca Raton.
- Lan K, Wang DT, Fong S, Liu LS, Wong KK, Dey N. A survey of data mining and deep learning in bioinformatics. *J Med Syst*. 2018;42(8):139.
- Jain A, Bhatnagar V. Concoction of ambient intelligence and big data for better patient ministration services. *Int J Ambient Comput Intell (IJACI)*. 2017;8(4):19–30.
- Bhattacharjee S. Statistical investigation of relationship between spread of coronavirus disease (COVID-19) and environmental factors based on study of four mostly affected places of China and five mostly affected places of Italy. *arXiv preprint arXiv:2003.11277*. 2020
- Toda AA. Susceptible-infected-recovered (sir) dynamics of covid-19 and economic impact. *arXiv preprint arXiv:2003.11221*. 2020
- Caccavo D. Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model. *medRxiv*. 2020
- Siwiak MM, Szczesny P, Siwiak MP. From a single host to global spread. The global mobility based modelling of the COVID-19 pandemic implies higher infection and lower detection rates than current estimates. *medRxiv*. 2020
- Zareie B, Roshani A, Mansournia MA, Rasouli MA, Moradi G. A model for COVID-19 prediction in Iran based on China parameters. *medRxiv*. 2020
- Teles P. Predicting the evolution Of SARS-Covid-2 in Portugal using an adapted SIR Model previously used in South Korea for the MERS outbreak. *arXiv preprint https://arXiv:2003.10047*. 2020
- Russo L, Anastassopoulou C, Tsakris A, Bifulco GN, Campana EF, Toraldo G, Siettos C. Tracing DAY-ZERO and forecasting the fade out of the COVID-19 outbreak in Lombardy, Italy: a compartmental modelling and numerical optimization approach. *medRxiv*. 2020
- Liu P, Beeler P, Chakrabarty RK. COVID-19 progression timeline and effectiveness of response-to-spread interventions across the United States. *medRxiv*. 2020
- Nadim SS, Ghosh I, Chattopadhyay J. Short-term predictions and prevention strategies for COVID-2019: a model based study. *arXiv preprint https://arXiv:2003.08150*. 2020
- Hossain M, Junus A, Zhu X, Jia P, Wen TH, Pfeiffer D, Yuan HY. The effects of border control and quarantine measures on global spread of COVID-19. In: Alvin and Zhu, Xiaolin and Jia, Pengfei and Wen, Tzai-Hung and Pfeiffer, Dirk and Yuan, Hsiang-Yu, The effects of border control and quarantine measures on global spread of COVID-19 .2020 (March 2, 2020).

39. Rocha Filho TM, dos Santos FSG, Gomes VB, Rocha TA, Croda JH, Ramalho WM, Araujo WN Expected impact of COVID-19 outbreak in a major metropolitan area in Brazil. medRxiv. 2020.
40. Traini MC, Caponi C, De Socio GV. Modelling the epidemic 2019-nCoV event in Italy: a preliminary note. medRxiv. 2020.
41. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med. 2020;1–6.
42. Wangping J, Ke H, Yang S, Wenzhe C, Shengshu W, Shanshan Y, Miao L. Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. Front Med. 2020;7:169.
43. Li L, Yang Z, Dang Z, Meng C, Huang J, Meng H, Shao Y. Propagation analysis and prediction of the COVID-19. Infect Dis Model. 2020;5:282–92.
44. La S, Bogoch II, Ruktanonchai N, Watts AG, Li Y, Yu J, Lv X, Yang W, Hongjie Y, Khan K, Li Z. Assessing spread risk of Wuhan novel coronavirus within and beyond China, January–April 2020: a travel network-based modelling study. 2020.
45. Zhu X, Zhang A, Xu S, Jia P, Tan X, Tian J, Wei T, Quan Z, Yu J. Spatially explicit modeling of 2019-nCoV epidemic trend based on mobile phone data in mainland China. medRxiv. 2020.
46. Volpert V, Banerjee M, Petrovskii S. On a quarantine model of coronavirus infection and data analysis. Math Modell Nat Phenomena. 2020;15:24.
47. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the novel coronavirus (2019-nCoV) outbreak. medRxiv. 2020.
48. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Eurosurveillance. 2020;25(10):2000199.
49. Bayham J, Fenichel EP. The impact of school closure for COVID-19 on the US healthcare workforce and the net mortality effects. medRxiv. 2020.
50. Giuliani D, Dickson MM, Espa G, Santi F. Modelling and predicting the spread of coronavirus (COVID-19) infection in NUTS-3 Italian regions. arXiv preprint [arXiv:2003.06664](https://arxiv.org/abs/2003.06664). 2020.
51. Lu J. A new, simple projection model for COVID-19 pandemic. medRxiv. 2020.
52. Webb G. Predicting the number of reported and unreported cases for the COVID-19 epidemic in South Korea, Italy, France and Germany. medRxiv. 2020.
53. Victor AO. Mathematical predictions for Covid-19 as a global pandemic. medRxiv. 2020.
54. Wang H, Zhang Y, Lu S, Wang S. Tracking and forecasting milestone moments of the epidemic in the early-outbreak: framework and applications to the COVID-19. medRxiv. 2020.
55. Botha AE, Dednam W. A simple iterative map forecast of the COVID-19 pandemic. arXiv preprint [arXiv:2003.10532](https://arxiv.org/abs/2003.10532). 2020.
56. Coelho FC, Lana RM, Cruz OG, Villela D, Bastos LS, Pastore y Piontti A, Davis JT, Vespignani A, Codeco C, Gomes MF. Assessing the potential impacts of COVID-19 in Brasil: mobility, morbidity and impact to the health system. medRxiv. 2020.
57. Weber A, Ianelli F, Goncalves S. Trend analysis of the COVID-19 pandemic in China and the rest of the world. arXiv preprint <https://arxiv.org/abs/2003.09032>. 2020.
58. Sameni R. Mathematical modeling of epidemic diseases; a case study of the COVID-19 coronavirus. arXiv preprint <https://arxiv.org/abs/2003.11371>. 2020.
59. Long C, Ying Q, Fu X, Li Z, Gao Y. Forecasting the cumulative number of COVID-19 deaths in China: a Boltzmann function-based modeling study. medRxiv. 2020.
60. Dowd JB, Andriano L, Brazel DM, Rotondi V, Block P, Ding X, Mills MC. Demographic science aids in understanding the spread and fatality rates of COVID-19. Proceed nat acad sci 2020;117(18):9696–9698.
61. He X, Lau EH, Wu P, Deng X, Wang J, Hao X, Lau YC, Wong JY, Guan Y, Tan X, Mo X. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med. 2020;26(5):672–675.
62. Giannakeas V, Bhatia D, Warkentin MT, Bogoch I, Stall NM. Estimating the maximum daily number of incident COVID-19 cases manageable by a healthcare system. medRxiv. 2020.
63. Banerjee A, Pasea L, Harris S, Gonzalez-Izquierdo A, Torralbo A, Shallcross L, Noursadeghi M, Pillay D, Pagel C, Wong WK, Langenberg C. Estimating excess 1-year mortality from COVID-19 according to underlying conditions and age in England: a rapid analysis using NHS health records in 3.8 million adults. medRxiv. 2020.
64. Siegenfeld AF, Bar-Yam Y. Eliminating COVID-19: a community-based analysis. arXiv preprint <https://arxiv.org/abs/2003.10086>. 2020.
65. Chen B, Liang H, Yuan X, Hu Y, Xu M, Zhao Y, Zhang B, Tian F, Zhu X. Roles of meteorological conditions in COVID-19 transmission on a worldwide scale. MedRxiv. 2020.
66. Ma Y, Zhao Y, Liu J, He X, Wang B, Fu S, Yan J, Niu J, Luo B. Effects of temperature variation and humidity on the mortality of COVID-19 in Wuhan. medRxiv. 2020.
67. Shi P, Dong Y, Yan H, Li X, Zhao C, Liu W, He M, Tang S, Xi S. The impact of temperature and absolute humidity on the coronavirus disease 2019 (COVID-19) outbreak-evidence from China. MedRxiv. 2020.
68. Fong SJ, Li G, Dey N, Crespo RG, Herrera-Viedma E. Composite monte carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction. Appl Soft Comput. 2020;106282.
69. Batista M. Estimation of the final size of the second phase of the coronavirus COVID-19 epidemic by the logistic model.
70. Hu Z, Ge Q, Li S, Jin L, Xiong M. Evaluating the effect of public health intervention on the global-wide spread trajectory of Covid-19. medRxiv. 2020.
71. Jia L, Li K, Jiang Y, Guo X. Prediction and analysis of coronavirus disease 2019. arXiv preprint <https://arxiv.org/abs/2003.05447>. 2020.
72. Kumar J, Hembram KPSS. Epidemiological study of novel coronavirus (COVID-19). 2020 arXiv preprint <https://arxiv.org/abs/2003.11376>.
73. DeCaprio D, Gartner J, Burgess T, Kothari S, Sayed S. Building a COVID-19 vulnerability index. arXiv preprint <https://arxiv.org/abs/2003.07347>. 2020.
74. Santosh K, Das D, Pal U. Truncated inception net: COVID-19 outbreak screening using chest X-rays. PREPRINT (Version 1) available at Research Square 3, 2020.
75. Santosh KC. AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multi-tudinal/multimodal data. J Med Syst. 2020;44(5)1–5 <https://doi.org/10.1007/s10916-020-01562-1>.
76. Dey N, Rajinikant V, Fong SJ, Kaiser MS, Mahmud M. Social-group-optimization assisted kapur's entropy and morphological segmentation for automated detection of COVID-19 infection from computed tomography images. 2020.
77. Wagh CS, Mahalle PN, Wagh SJ. Epidemic peak for COVID-19 in India, 2020. Preprints 2020, 2020050176 (<https://doi.org/10.20944/preprints202005.0176.v1>).
78. Rajinikanth V, Dey N, Raj ANJ, Hassanien AE, Santosh KC, Raja N. Harmony-search and otsu based system for coronavirus disease (COVID-19) detection using lung CT scan images. 2020 arXiv preprint [arXiv:2004.03431](https://arxiv.org/abs/2004.03431).
79. Bhapkar HR, Mahalle P, Dhotre PS. Virus graph and COVID-19 pandemic: a graph theory approach. Preprints 2020, 2020040507 (<https://doi.org/10.20944/preprints202004.0507.v1>).

80. Bullock J, Pham KH, Lam CS, Luengo-Oroz M. Mapping the landscape of artificial intelligence applications against COVID-19. arXiv preprint [arXiv:2003.11336](https://arxiv.org/abs/2003.11336). 2020.
81. Mahalle PN, Sable NP, Mahalle NP, Shinde GR Data analytics: COVID-19 prediction using multimodal data. Preprints 2020, 2020040257 (<https://doi.org/10.20944/preprints202004.0257.v1>).
82. Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Rabczuk T, Atkinson PM. Covid-19 outbreak prediction with machine learning. Available at SSRN 3580188. 2020.
83. COVID-19 in India: guidance from the IndiaSIM Model- March 24, 2020. <https://cddep.org/covid-19/>.
84. Dey N, Fong S, Song W, Cho K Forecasting energy consumption from smart home sensor network by deep learning. In: International conference on smart trends for information technology and computer communications 2017 (pp. 255–265). Springer, Singapore.
85. Hu S, Liu M, Fong S, Song W, Dey N, Wong R. Forecasting China future MNP by deep learning. In: Behavior engineering and applications 2018 (pp. 169–210). Springer, Cham.
86. Singh N, Mohanty SR. Short term price forecasting using adaptive generalized neuron model. *Int J Ambient Comput Intell (IJACI)*. 2018;9(3):44–56.

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