Mathematical Modelling and Prediction Tools for the COVID-19 Pandemic: A Review

Chin Kuan Ho, Seng Huat Ong, Kamarul Imran Musa, Choo Yee Ting, Chiung Ching Ho, Kwan Hoong Ng

Submitted to: JMIR Public Health and Surveillance
on: May 20, 2021

Table of Contents

Mathematical Modelling and Prediction Tools for the COVID-19 Pandemic: A Review

Chin Kuan Ho[1] BSc, MScIT, PhD; Seng Huat Ong[2] PhD; Kamarul Imran Musa[3] PhD; Choo Yee Ting[1] PhD; Chiung Ching Ho[4] PhD; Kwan Hoong Ng[5] PhD

[1]Multimedia University Cyberjaya MY
[2]UCSI University Kuala Lumpur MY
[3]Universiti Sains Malaysia Health Campus Kubang Kerian MY
[4]Sunway University Subang Jaya MY
[5]Universiti Malaya Kuala Lumpur MY

Corresponding Author:
Chin Kuan Ho BSc, MScIT, PhD
Multimedia University
Persiaran Multimedia
Cyberjaya
MY

Abstract

The latest threat to global health is the ongoing outbreak of the Coronavirus Disease 2019 (COVID-19). There are three main areas of modeling research, namely epidemiology, drug repurposing and vaccine design. The most important purpose of the models is to inform institutional and nationwide efforts to ensure patient safety. This study aimed to review COVID-19 modelling and prediction tools. Understanding these methods streamlines the strengths and limitations of each method. We researched the traditional model and the more current models that flourish during the pandemic. This understanding is the key to the proper use of specific models to achieve certain goals. Modeling approaches for COVID-19 can be very broadly categorized into phenomenological models and mechanistic models. Phenomenological approaches treat the modeling problem purely from an empirical perspective. From our survey, there are three major types of approaches under the phenomenological models: time-series analysis and forecasting, fractal-based models, and machine learning approaches. Mechanistic models consider the underlying mechanics of the epidemic. In this survey, compartmental models and agent-based models are categorized as mechanistic models. We studied 46 scientific articles (published between 22 February 2020 and 29 January 2021) that we think are representative of the scientific community's approaches in modeling and prediction. We highlight the challenges and limitations of modelling approaches such as the need for high quality data, and interpretable models. Finally, we list the desired features for developing robust and reliable modelling and prediction tools.

(JMIR Preprints 20/05/2021:30546)
DOI: https://doi.org/10.2196/preprints.30546

Original Manuscript

Mathematical Modelling and Prediction Tools for the COVID-19 Pandemic: A Review

Chin Kuan Ho[1], Seng Huat Ong[2], Kamarul Imran Musa[3], Choo Yee Ting[1], Chiung Ching Ho[4], Kwan Hoong Ng[5]

[1]Faculty of Computing and Informatics, Multimedia University
Cyberjaya, Selangor, Malaysia.

[2]Department of Actuarial Science and Applied Statistics, Faculty of Business and Management, UCSI University, Malaysia

[3]Department of Community Medicine, School of Medical Sciences, Universiti Sains Malaysia Health Campus, Kubang Kerian, Kelantan, Malaysia.

[4]Department of Computing and Information Systems, School of Science and Technology, Sunway University, Malaysia.

[5]Department of Biomedical Imaging, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia

A Review of Mathematical Modelling and Prediction Tools for the COVID-19 Pandemic

Abstract

The latest threat to global health is the ongoing outbreak of the Coronavirus Disease 2019 (COVID-19). There are three main areas of modeling research, namely epidemiology, drug repurposing and vaccine design. The most important purpose of the models is to inform institutional and nationwide efforts to ensure patient safety. This study aimed to review COVID-19 modelling and prediction tools. Understanding these methods streamlines the strengths and limitations of each method. We researched the traditional model and the more current models that flourish during the pandemic. This understanding is the key to the proper use of specific models to achieve certain goals. Modeling approaches for COVID-19 can be very broadly categorized into phenomenological models and mechanistic models. Phenomenological approaches treat the modeling problem purely from an empirical perspective. From our survey, there are three major types of approaches under the phenomenological models: time-series analysis and forecasting, fractal-based models, and machine learning approaches. Mechanistic models consider the underlying mechanics of the epidemic. In this survey, compartmental models and agent-based models are categorized as mechanistic models. We studied 46 scientific articles (published between 22 February 2020 and 29 January 2021) that we think are representative of the scientific community's approaches in modeling and prediction. We highlight the challenges and limitations of modelling approaches such as the need for high quality data, and interpretable models. Finally, we list the desired features for developing robust and reliable modelling and prediction tools.

Keywords: infectious disease, mathematical modeling, prediction, SARS-CoV-2, COVID-19

Introduction

The latest threat to global health is the ongoing outbreak of the respiratory disease that was recently given the name Coronavirus Disease 2019 (COVID-19). COVID-19 was recognized in December 2019.1 It was rapidly shown to be caused by a novel coronavirus that is structurally related to the virus that causes severe acute respiratory syndrome (SARS). As in the two preceding instances of emergence of coronavirus disease in the past 18 years — SARS (2002 and 2003) and Middle East respiratory syndrome (MERS) (2012 to the present) — the COVID-19 outbreak has posed critical challenges for the public health, research, and medical communities [1–3]. The new coronavirus shares about 82% of its genome with the one producing the 2003 outbreak (SARS CoV-1)[4]. Since December 2019, the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has produced an outbreak of pulmonary disease which has soon become a global pandemic, known as COronaVIrus Disease-19 (COVID-19). Globally, 4.1 million new cases were reported in the final week of January 2021, and the number of new deaths reported was just under 96 000. This brings the cumulative numbers to over 98.2 million reported cases and over 2.1 million deaths globally since the start of the pandemic [5].

Epidemiological analysis is a well-established area of interdisciplinary research in which medical, statistical, computational and mathematical elements are integrated to analyse and forecast the progression of a disease. We should differentiate these studies in ‘peace time’, when there is no health emergency, from those of ‘war times’, when the emergency of a pandemic is pressing for results with limited data, in a constantly changing landscape and having to make a lot of assumptions [4]. There are three main aims of modeling research against SARS CoV-2 and COVID-19: (1) epidemiology; (2) drug repurposing; and (3) vaccine design [4]. The most important use of the models is to inform institutional and nationwide efforts to ensure patient safety.

Models of infection cases can inform the implementation of public policy measures designed to reduce the spread of the infection, such as social distancing and closure of nonessential businesses. They can also help guide institutional efforts to care for patients by ensuring adequate hospital beds, PPE, testing resources, and staffing. Predicting surges in infected caseloads can help determine whether outpatient visits and operations should be scaled back to ensure adequate resources to care for the infected or allowed to grow to pre-pandemic levels [6]. In addition to that, these models have been used to predict the course of the pandemic, inform disaster preparedness planning, forecast the economic outlook, and allocate scarce resources, such as personal protective equipment (PPE) and testing supplies [6].

The efficiency of transmission for any respiratory virus has important implications for containment and mitigation strategies. The current study indicates an estimated basic reproduction number (R_0) of 2.2, which means that, on average, each infected person spreads the infection to an additional two persons [2].

The early modern model is the basic compartmental model. The deterministic mathematical models used to analyze the coronavirus rest on the shoulders of a pair of giants: A.G. McKendrick and W.O. Kermack, the first a military doctor and the second a chemist [7]. They developed the first compartmental mathematical model for the study of the evolution of an epidemic. In compartmental models, the population is divided into different compartments. The basic model having the compartments Susceptible (S), Infectious (I), and Recovered (R). However, these simple, classic SIR models have obvious limitations. They unrealistically assume that the population is uniform and homogeneously mixing, whereas it is known that mixing depends on many factors including age (children usually have more adequate contacts per day than adults). Moreover, different geographic and social-economic groups have different contact rates [8].

Approaches that don’t account for how disease transmission occurs are generally not well-suited for long-term predictions about epidemiologic dynamics (such as when the peak will occur and whether resurgence will happen) or for inference about intervention efficacy [9,10]. The models should provide this information to be useful: a) purpose and time frame of the model b) the basic model assumptions, c) the uncertainty, d) data used, and e) Is the model general, or does it reflect a particular context?

Objectives of this review

Given the above, the present study aimed to develop a comprehensive review of COVID-19 modelling methods. Understanding these methods streamlines the strengths and limitations of each method. We researched the traditional model and the more current models that flourish during the COVID-19 pandemic to achieve the objective of this review. This understanding is the key to match the right model to the intended goal of using the model. Not all models work the same way and not all models provide the same answers. The diligent assessment of the available COVID-19 models and the proper use of the models is so paramount that the wrong models will not only derail the control and prevention of COVID-19 transmission but will result in more COVID-19 complications and deaths and distrust of people for the infectious disease models.

Methods

Search Strategy

References for this review were identified through searches of the following main databases: IEEEXplore, ScienceDirect, SpringerNature, and PubMed for articles published from 1 January 2020, to 31 January 2021. We use a combination of the search terms “COVID-19 ”, “transmission”, “spread”, “modelling”, and “prediction”. Other relevant articles were identified through searches in Google Scholar. Selected articles resulting from these searches and relevant references cited in those articles were reviewed. The 46 articles that were selected to be featured in this paper are those that we think will represent well the respective model category.

Categorization of Modelling Approaches

The COVID-19 epidemic data comes in the form of a time series $y(1)$, $y(2)$,… $y(\mathrm{t})$ where $y(t)$ gives the number of newly infected cases at time t. Plotting the number of new cases against time will produce what is known as the epidemic curve. In principle, the task of modeling the epidemic curve entails learning a mathematical model from the time series up to time k (usually called the training set). The model is then used to forecast the number of newly infected cases in a window of w time points, called the forecast horizon.

For a COVID-19 model to be useful in informing decisions, it must be practical, robust yet mathematically vigorous. It must attempt to capture as accurately as possible the disease transmission dynamics that are influenced by some of the following factors:

- The efficiency of transmission of the COVID-19 virus.
- Intervention measures, both pharmaceutical and non-pharmaceutical, are put in place by the authorities.
- Spatio-temporal behavior and movements of the population.
- Weather patterns.

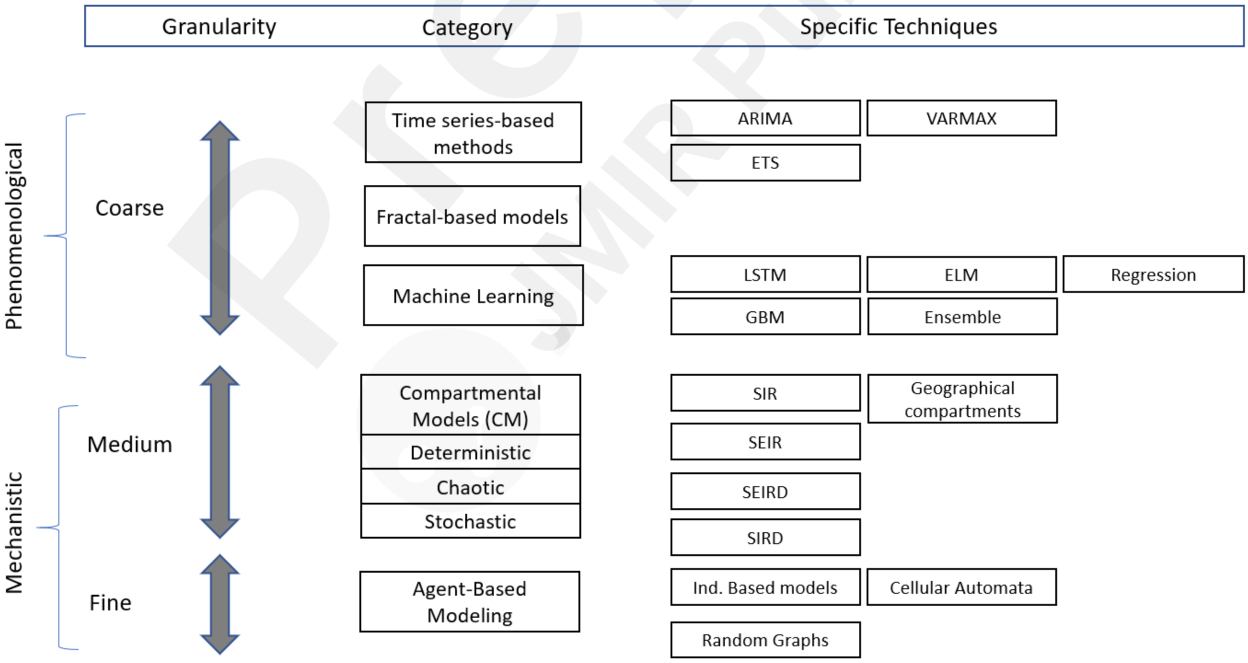
Public health officials can use a good model to:

- Estimate key transmission parameters of COVID-19.
- Assess and optimize the impact of control strategies.
- Gain insights into how different transmission pathways contribute to the epidemic curve.
- Generate short and long-term forecasts.

Figure 1 shows our categorization of the COVID-19 published work that we surveyed. Modeling approaches for COVID-19 can be very broadly categorized into phenomenological models and mechanistic models. Phenomenological approaches treat the modeling problem purely from an empirical perspective. They do not explicitly model the underlying mechanics that gave rise to the epidemic curve. From our survey, there are three major types of approaches under the phenomenological models: time-series analysis and forecasting, fractal-based models, machine learning approaches. Mechanistic models consider the underlying mechanics of the epidemic. They model the spread of the epidemic as dynamical systems described by a series of differential equations. In this survey, compartmental models and agent-based models are categorized as mechanistic models.

COVID-19 models can also be categorized by their granularity. Agent-based models provide a fine-grain abstraction because they use an individual as an *identifiable* basic unit in the model. Agent-based models can arguably give a more faithful representation of the problem since the behavior of each individual or group of individuals can be customized in the simulation depending on environmental factors. On the other hand, time series, fractal and machine learning approaches are coarse-grain models. They primarily model the relationship between input features and the response variable. Though these input features may be related to the process generating the epidemic curve, they are usually aggregated. In the middle of the two extremes (coarse-grain and fine-grain, we find compartmental models that divide the population into groups. Each group contains individuals with similar health characteristics. Compartmental models are widely used in epidemiology where the patient is the community and individuals are managed collectively, fitting the concept of compartments.

Figure. 1. Categorization of reviewed COVID-19 modeling approaches.



We reviewed 46 scientific articles that we think are representative of the scientific community's approaches towards the modeling and prediction of the spread of COVID-19. In Table 1, we summarize each study by describing the modeling approach along with the claimed novelty, the date of publication (or the date when a paper is made available online by the publisher), the region or country of study, the date of study and important observations from the results. For compartmental models, we provide a brief description of the approach because simply mentioning the compartments is not intuitive.

Table 1. List of COVID-19 modeling studies surveyed.

Category	**Study**	**Publication Date**	**Country/Region; (Date of Study)**	**Modeling Approach**
Time series-based methods	Benvenuto et al. (2020) [11]	1 Apr 2020	Multiple countries (20 Jan – 10 Feb 2020)	Model: ARIMA
	Singh et al. (2020) [12]	11 May 2020	Italy, Spain, France, UK, US (21 Jan – 11 Apr 2020)	Model: ARIMA with wavelet decomposition pre-processing.
	Chintapuldi et al. (2020) [13]	1 June 2020	Italy (15 Feb – 31 Mar 2020)	Model: SARIMA (Seasonal ARIMA)
	Ankarali et al. (2020) [14]	18 June 2020	Multiple countries (31 Dec 2019 – 15 Mar 2020)	Models: ARIMA, Simple Exponential Smoothing, Holt's Two Parameter, Brown's Double Exponential Smoothing
	Agosto & Giudici (2020) [15]	16 July 2020	China, South Korea, Iran, Italy (20 Jan – 31 Mar 2020)	Model: Poisson autoregressive
	Katris (2021) [16]	2 Oct 2020	Greece (5 Mar – 14 Aug 2020)	Models: ARIMA, Holt and Artificial Neural Network (1 input node)
	Jamdade & Jamdade (2021) [17]	11 Dec 2020	Philippines (31 Dec 2019 - 13 Oct 2020)	Model: VARMAX (Vector Autoregression Moving-Average with Exogenous Regressors)
Fractal-based models	Păcurar & Necula (2020) [18]	3 July 2020	Romania (27 Feb 2020 up to 98 days) Italy (31 Jan 2020 up to 127 days) Spain (1 Feb 2020 up to 125 days) Germany (1 Feb 2020 up to 126 days) Sweden (28 Jan up to 130 days)	Model: Fractal interpolation to reconstruct the epidemic curve.
	Abbasi et al. (2020) [19]	11 July 2020	Sao Paolo, Brazil (26 Feb – 10 May 2020) Europe (21 Jan – 14 May 2020)	Model: Fractal SIR Model

Machine Learning-based methods	Roosa et al. (2020) [20]	22 Feb 2020	Guangdong and Zhejiang (22 Jan – 13 Feb 2020)	Models: Generalized Logistic Model, Richards model.
	Ribeiro et al. (2020) [21]	1 May 2020	Brazil (1 Mar – 19 Apr 2020)	Models: Support Vector Regression, stacked ensemble, ARIMA, CUBIST Regression, Ridge Regression and Random Forests.
	Chimmula & Zhang (2020) [22]	8 May 2020	Canada (22 Jan – 31 Mar 2020)	Model: LSTM as a means to model and forecast time series.
	Chen et al. (2020) [23]	15 May 2020	US (22 Jan – 6 Apr 2020)	Model: Logistic growth model
	Sanchez-Caballero et al. (2020) [24]	20 May 2020	China, Italy, Spain (22 Jan – 30 Mar 2020)	Verhulst equation (Logictic function)
	Rustam et al. (2020) [25]	25 May 2020	Australia, Canada, Algeria, Afghanistan (22 Jan – 27 Mar 2020)	Models: Linear regression, SVM, LASSO, Exponential Smoothing.
	B.Yan et al. (2020) [26]	30 June 2020	Henan, Zhejiang, Hubei (16 Feb – 16 Apr 2020)	Model: LSTM (Long Short-Term Memory)
	Khan et al. (2020) [27]	1 July 2020	China (21 Jan – 3 Feb 2020)	Model: Deep Extreme Learning Machine
	Fokas et al. (2020) [28]	13 July 2020	Italy, France, Spain, Germany (Period of 0 to 120 days after 500 cases were reported.)	Model: BiLSTM (LSTM: Long Short-Term Memory)
	Aviv-Sharon & Aharoni (2020) [29]	24 July 2020	Chinese mainland (21 Jan – 24 Feb, 2020), Iran (26 Feb – 31 Mar 2020), Philippines (19 Feb – 24 Mar 2020).	Model: Generalized Logistic Model
	Ramchandani et al. (2020) [30]	14 Sept 2020	US (5 Apr – 28 June 2020)	Model: Deep Neural Network
	Shrivastav & Jha (2020) [31]	1 Oct 2020	India (28 Mar – 22 Apr 2020)	Model: Gradient Boosting Machine (GBM)
	Mehta et al. (2020) [32]	11 Sept 2020	USA (14 Mar – 31 Mar 2020)	XGBoost
	Pinter et al. (2020) [33]	2 June 2020	Hungury (4 Mar – 28 April 2020)	Network-based fuzzy inference system (ANFIS) and multi-layered perceptron-imperialist competitive algorithm (MLP-

				ICA)
	Suzuki et al. (2020) [34]	14 May 2020	South Korea	XGBoost
	Ardabili et al. (2020) [35]	1 Oct 2020	Italy, Germany, Iran, USA, and China for a 30 days duration	Models: mathematical models (i.e., logistic, linear, logarithmic, quadratic, cubic, compound, power, and exponential) were fitted using GA, PSO, and GWO. Time-series prediction using MLP and ANFIS
Compartmental Models	Fanelli and Piazza (2020) [36]	12 Mar 2020	China, Italy, France (22 Jan – 15 Mar 2020)	Model: SIRD.
	He, Tang, & Rong (2020) [37]	16 Mar 2020	Hubei Province, China (11 Jan – 13 Feb 2020)	Model: SEIHR (H: Hospitalized). Stochastic compartmental models using Binomial distributions. MCMC method was used to estimate the parameters of the model.
	Anastassopoulou et al. (2020) [38]	31 Mar 2020	Hubei Province, China (11 Jan – 10 Feb 20920)	Model: SIRD. Wrapper approach that uses an optimization algorithm tp wrap around around the SIRD model.
	Liu et al. (2020) [39]	12 May 2020	Wuhan, China (8 Dec 2019 – 18 Mar 2020)	Model: SEIR. The 'I' compartment is further divided into two groups: infected individuals who are hospitalized (therefore preventing contact with the 'S' group); and infected individuals who were not hospitalized.
	He, Peng, & Sun (2020) [40]	18 June 2020	Hubei Province, China (24 Jan – 12 Apr 2020)	Model: SEIR. Particle Swarm Optimization (PSO) is used to estimate two parameters: 1. rate of infectious to hospitalized, and 2. recovery rate of quarantined infected individuals. The paper also investigates the dynamics of Covid-19 spread from a chaotic system perspective.
	Croccolo & Roman (2020) [41]	3 July 2020	US (No specific time framework indicated)	Model: Random graphs with an underlying lattice model. Each

				site in the lattice corresponds to membership of one of four categories S, I, R, D.
	Vigueri et al. (2021) [42]	15 July 2020	Lombardy, Italy (Feb – Apr 2020)	Model: SEIRD. Four parameters representing diffusion terms were introduced into the compartmental PDEs, where each new parameter is attached to the S, E, I, and R compartments.
	Wang & Yamamoto (2020) [43]	15 July 2020	Arizona, US. (24 Apr – 19 June 2020)	Model: Compartmental-oriented. The work models three regions in Arizona as compartments in the model. The model is represented by a series of partial differential equations (PDEs). Novelty: inclusion of Google mobility data into the model to account for the effects of human activities.
	Barbarossa et al. (2020) [44]	4 Sept 2020	Germany (6 Apr – 14 Apr 2020)	Model: SEIRD.
	Al-Khani et al. (2020) [45]	20 Sept 2020	Saudi Arabia (2 Mar – 19 May 2020)	Model: SEIR
	Postavaru et al. (2020) [46]	3 Oct 2020	China (1 Dec 2019 - 23 Apr 2020) South Korea (9 Feb - 23 Apr 2020)	Model: Fractional order SEIR.
	Hussain et al. (2020) [47]	14 Oct 2020	US (1 Apr – 19 July 2020)	Model: SIR. The sizes of the S, I and R compartments are formulated by a system of stochastic equations. Each compartment is modeled using a Poisson process.
	Baek et al. (2020) [48]	26 Oct 2020	South Korea (24 Feb – 13 Mar 2020)	Model: SEIR
	Karaivanov (2020) [49]	29 Oct 2020	Not specified.	Model: SIR. Network-augmented SIR (NSIR) model where a social network structure is embedded and simulated for each compartment. A variety of network topologies were studied.
	Law et al. (2020)	10 Dec 2020	Malaysia	Model:

	[50]		(18 Mar – 7 May 2020)	SIR, with the addition of three parameters to account for the early depleting dynamics.
	Liu (2021) [51]	24 Dec 2020	New York City (9 Feb – 26 July 2020) Wuhan (8 Dec -29 Mar 2020)	Model: Variation of SIR called AIR. Recursive relations are used to model the relationship between the A, I and R comparments. (A: Actual active infective infectious group; I: Cummulative infected groip, R: Cummulative recovered group
	Fiacchini & Alamir (2021) [52]	29 Jan 2021	13 Regions in mainland France (15 March - 15 June 2020)	Model: SIRD. A two parameter model is derived from the SIRD model. The problem of modeling is formulated and solved as a least squares minimization problem with non-convex constraints. The model was fitted to the data concerning the number of deaths.
Agent-based Models	Cuevas (2020) [53]	20 May 2020	Not specified	Model: Uses ABM to study the potential spreading pattern in faciltiies (eg. School). The work basically reports the output of the model for the various scenarios under study.
	Hoertel et al. (2020) [54]	14 July 2020	France (15 Apr – 15 May 2020)	Model: Uses ABM to assess the impact of non-pharmaceutical interventions on cummulative number of infectives, mortality and on ICU bed occupancy.
	Chang et al. (2020) [55]	11 Nov 2020	Australia (3 Feb – 21 Aug 2020)	Model: Agent-based Model (ABM). High-resolution stochastic agent model. Each agent represents an individual in the population. Agents are calibrated to the Australian census data (2016). The authors used the model to assess their effectiveness of several intervention strategies.
	Truszkowska et al. (2021) [56]	18 Jan 2021	New Rochelle, New York (22 Feb – 14 July 2020)	Model: Agent-based Model (ABM) Each individual in the population is modeled as an agent. The model can incorporate a variety of elements such as different testing strategies, pharmaceutical

				and non-pharmaceutical interventions, modeling different groups of people, modeling of expert knowledge.

Results: Analyses of Selected Works

Time Series Analysis and Forecasting Approaches

Due to the time structure of the data, an appropriate model for the Covid-19 pandemic is the time series model. An accurate forecast of the spread of Covid-19 infections helps control the disease and determine if measures to curb the pandemic are effective. In the past year or so, various time series models have been proposed for the prediction and study of the evolution of the disease. In time series analysis, the two most widely used models are the exponential smoothing and ARIMA (Autoregressive Integrated Moving Average) models. Exponential smoothing models examine the trend and seasonality in the data and used weighted averages of past observations with more recent observations having higher weightage. ARIMA models are based on the concept of stationarity (roughly speaking this means certain properties of the time series are time-independent) and are concerned with the autocorrelations in the data.

To facilitate the discussion, we first give some basic definitions for the exponential smoothing and ARIMA time series models. Exponential smoothing methods use weighted averages of past observations where the weights decay exponentially so that the more recent observations receive larger weights. Given observations $x_0, x_1, x_2, \ldots, x_t$, the next best value $\hat{x}_{t+1\vee t}$ is given by

$$\hat{x}_{t+1\vee t} = \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \ldots + \alpha(1-\alpha)^{t-1} x_1 + (1-\alpha)^t x_0$$

where α is the smoothing parameter $0<\alpha<1$. Simple exponential smoothing is also represented as

$$\hat{x}_{t+l\vee t} = s_t, x_0 = s_0,, l = 1,2,3,\ldots$$
$$s_t = \alpha x_t + (1-\alpha)s_{t-1}$$

If there is no trend or seasonal pattern, simple exponential smoothing is a suitable method. The simple exponential smoothing was extended to double exponential smoothing to include trend [57,58]. The above simple exponential smoothing equations are expanded by adding a trend equation:

$$\hat{x}_{t+l\vee t} = s_t + l a_t, l = 1,2,3,\ldots$$
$$s_t = \alpha x_t + (1-\alpha)(s_{t-1} + a_{t-1})$$
$$a_t = \beta(s_t - s_{t-1}) + (1-\beta)a_{t-1}$$

where a_t is the estimate of trend and β is the trend smoothing parameter, $0<\beta<1$. A further extension that incorporates seasonal patterns is the Holt-Winters model [59].

A time series is a quantity measured sequentially in time over some interval. Mathematically, a time series is a stochastic process $\{X_t, t \in T\}$ where X_t is a random variable and T is an index set. The Autoregressive (AR), Moving Average (MA) and Autoregressive Moving Average (ARMA) are basic time series models, which form the building blocks of more complicated models such as the Autoregressive Integrated Moving Average (ARIMA) and other families of models like GARCH (Generalised Autoregressive Conditional Heteroskedastic). A time series is an Autoregressive (AR) model of order *p*, $AR(p)$ if

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \omega_t$$

where $\alpha_p \neq 0$ and $\{\omega_t\}$ is white noise (error term) with mean zero and variance σ^2. A moving average model of order *q*, $MA(q)$, is regressed on past errors and is defined as

$$X_t = \beta_1 \omega_{t-1} + \beta_2 \omega_{t-2} + \ldots + \beta_q \omega_{t-q} + \omega_{t.}$$

By combining the AR and MA models we have the autoregressive moving average model of order (*p*,*q*), $ARMA(p,q)$, given by

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + \omega_t + \beta_1 \omega_{t-1} + \beta_2 \omega_{t-2} + \ldots + \beta_q \omega_{t-q}.$$

Let $\nabla X_t = X_t - X_{t-1}$ denote differencing once and $\nabla^d X_t$ is differencing *d* times. A time series is an *ARIMA*(*p*,*d*,*q*) model, if $\nabla^d X_t$ is an autoregressive moving average of order (*p*,*q*), *ARMA*(*p*,*q*). In other words, we combine differencing with *ARMA*(*p*,*q*) model where *d* is the degree of differencing. That is, ARIMA models are the general class of models for a time series that can be made to be stationary through the operation of differencing. A time series is said to be stationary if its statistical properties do not change over time. The aforementioned models do not cater to seasonality. A seasonal ARIMA model (SARIMA) is constructed by including another ARIMA model for the seasonal effects. This is represented multiplicatively as

$$ARIMA(p,d,q) \times (P,D,Q)_S$$

where (p,d,q) is the non-seasonal part and $(P,D,Q)_S$ is the seasonal part with *S* denoting the seasonal time span or time intervals. There is some overlap between the class of exponential smoothing models and ARIMA models. For example, the simple exponential smoothing model is *ARIMA*(0, 1,1).

In the analysis of Covid-19 pandemic data, many researchers have employed and compared time series models ranging from the simple moving average, simple exponential smoothing and Holt's model [60] to ARIMA [11,13] and their variants (VARMAX, SARIMA). Some researchers also conducted their analyses with modified machine learning models to include the time-series nature of the data [22]. Due to the differing conditions and methods to mitigate the spread of the disease in each region and country, many of these research work examined country-specific analysis and data-driven modelling. The Simple Exponential Smoothing, Holt's Two Parameter, Brown's Double Exponential Smoothing Models and *ARIMA*(*p, d, q*) models were considered for COVID-19 data from twenty-five countries [14]. The COVID-19 data were collected for at least twenty days and six indicators of the outbreak (Cumulative Cases, Cumulative Deaths, Daily Cases, Daily Deaths, Cumulative Recovered, Active Cases). The authors also examined trends and seasonal effects. Cumulative death was found to be the highest in Italy and Spain. Clearly the statistical analyses are based on limited data and thus the forecast is short-term.

Another study utilized a much bigger data set collected after a sixty-day lockdown in Italy [13]. ARIMA models were considered and the R package *forecast* [61] was employed in the data analysis. For the registered and recovered cases the authors reported prediction accuracy of 93.75% and 84.4% respectively.

Mixed approaches have also been developed. A combination of short-term forecasting models based on autoregressive integrated moving average (ARIMA), and other machine learning models: cubist regression (CUBIST), random forest (RF), ridge regression (RIDGE), support vector regression (SVR), and stacking-ensemble learning [21]. The objective is to compare the predictive performance of the time series model and the machine learning models. Forecasting is done one, three, and six days ahead for COVID-19 cumulative confirmed cases in ten Brazilian states with a high daily incidence of infection. Based on their analysis, the models ranked in order of accuracy are as follows: SVR, stacking-ensemble learning, ARIMA, CUBIST, RIDGE, and RF models. Another work that focused on the short-term forecasting of Brazilian COVID-19 cases uses the Holt model [60]. The *holt* function in the R package *forecast* [61] was used in the statistical analysis. It is found that the Holt model performed well in forecasting COVID-19 cases in Brazil and São Paulo and Minas Gerais states, but with underestimation for the Rio de Janeiro state.

In a short paper [11], ARIMA model statistical analysis was conducted on the daily prevalence data of COVID-19 obtained from Johns Hopkins University web site. The ARIMA (1,0,3) and ARIMA (1,0,4) were used in the analysis. The authors showed that both prevalence and incidence of COVID-19 are not influenced by seasonality. A vector ARMA with exogenous variables (VARMAX) time series model was used to analyze the transmission dynamics of COVID-19 in the Philippines [17]. Model parameters were estimated by ordinary least squares. The authors also conducted model simulation and found that simulated infected values are in close agreement with the actual number of COVID-19 cases.

Concerning the outbreak spread forecasting in Greece [16], the approach taken is a data-driven approach to formulate a procedure based on time series models to track COVID-19 outbreak. Exponential smoothing and ARIMA models have been combined with machine learning methods like Multivariate Adaptive Regression Splines (MARS) and Feed-Forward Artificial Neural Networks (ANN) for improved accuracy. The study then examined a time series for SIR (Susceptible-Infected-Recovered) model. The SIR model is a popular model for infectious diseases. Apparently, there is difficulty in applying the SIR model to time series data.

The COVID-19 incidence data from the UK and US was used to theoretically study the connection between the polynomial representation and time series representation of data with an illustration of goodness-of-fit for both classes of representation [62]. The author showed that all finite degree polynomials can be represented by finite order AR time-series, but not conversely. There exist infinitely many AR time-series of finite orders that are not representable by finite order polynomials. The error-trend-seasonal (ETS) model was utilised to forecast the prevalence and mortality of COVID-19 in USA, UK, Russia, and India [63]. The ETS model is exponential smoothing under a nonlinear framework. The predictive performance of the ETS model was compared with the ARIMA model. They examined several specific $ARIMA(p, d, q)$ and ETS models. The authors concluded that the ETS models have smaller forecasting error rates than ARIMA models for all the datasets.

Another hybrid model involves discrete wavelet decomposition of the dataset of COVID-19 deaths, which splits the data into component series [12]. ARIMA models are then applied to each of the component series for making predictions of number of deaths. The rationale for using wavelets for forecasting purposes is because the method provides more refinement and flexibility when compared with the traditional methods. The authors used the dataset of daily deaths due to the COVID-19 from Italy, Spain, France, the United Kingdom (UK), and the United States of America (USA) to illustrate the performance of the hybrid and ARIMA models. The hybrid Wavelet-ARIMA model is shown to perform much better than the established ARIMA model. The hybrid Wavelet-ARIMA model gives forecasts with an error reduction of about fifty percent for France, UK, and USA relative to the

ARIMA model. For Italy and Spain, the error reduction is about eighty five percent.

There is also a different type of autoregressive model[15] by assuming a Poisson distribution $Poi(\lambda_t)$ or the number of infected at time *t*, Y_t, conditional on information up to *t* -1, F_{t-1}, and has a log-linear autoregressive intensity, that is,

$$\log(\lambda_t) = \alpha_1 \log(1+Y_{t-1}) + \alpha_2 \log(\lambda_{t-1}) + \varepsilon$$

where $\alpha_1, \alpha_2, \varepsilon \in R$. The α component represents the short-term dependence on the previous time point while the β component is the representation of long-term dependence, the trend component. This specification of a log-linear autoregressive intensity allows for negative dependence. The authors illustrated their Poisson autoregressive (PAR) model with Covid-19 data from China, Iran, South Korea and Italy available from WHO daily reports. The predictive performance of the proposed PAR model is compared with the exponential growth model and the PAR model with covariate, and it is seen that PAR model has the better performance.

Machine Learning Models

Machine learning approaches formulates the COVID-19 modeling task as a regression analysis problem. This is a type of supervised learning that models the relationship between several features and a continuous target variable. Here, the continuous target variable is taken to be the number of reported daily or cumulative CVOID-19 cases. Machine learning models used for COVID-19 modeling can be broadly grouped into the following three categories:

i. **The logistic growth model.** A short-term forecast of cumulative reported cases for Guangdong and Zhejiang, China was produced using the generalized logistic growth model and the Richards growth model [64]. The model over-predicted the epidemic curve. There could be difficulty to produce a better forecast as the data used was up until 13 February 2020, during the early stages of the epidemic. Another attempt used the Richards model with an R^2 value of above 0.99 [29]. In another study, a five-parameter logistic growth model was used to fit the US data.[23] A visual inspection of the results showed that the model has large errors when forecasting new daily cases after the change point of the time series. The Verhulst equation can be used to model the number of infected persons in several countries [24]. The authors reported and R^2 value of 0.9969. An examination of the forecasts for daily cases reveals that it is common for the model to over-estimate the number of cases after the change point, but before the peak; while nearing the peak, the model tends to underestimate the number of cases.

ii. **Decision Tree-based Models.** Several attempts at modeling have used state-of-the-art tree models, namely Gradient Boosting Machines (GBM) and Extreme Gradient Boosting (XGBoost). In principle, gradient boosting minimizes the loss function by adding weak learners into an ensemble of learners using gradient descent. XGBoost is a variant of gradient boosting that implements a couple of enhancements. Firstly, the use of second-order gradients of the loss function, and secondly the use of advanced regularization. Gradient Boosting as a machine learning technique has also been proposed to explore the effect of minimum temperature, maximum temperature, minimum humidity, and maximum humidity on the COVID-19 infection count [31]. A variation of Gradient Boosting technique, namely, Extreme Gradient Boosting (XGBoost), was employed to predict the occurrences of COVID-

19 cases of the unaffected county in the United States, resulting in significant results [32]. In a study involving Korea [34], XGBoost was used to perform binary classification on whether each of 17 provinces in Korea has more than 100 total COVID-19 cases in the coming 24 days of their period of study.

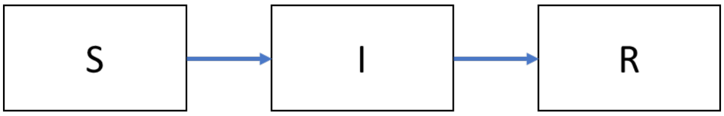
iii. **Deep Learning Models.** The long short-term memory networks, in short LSTM [65] is one of the preferred deep learning approaches to modelling the COVID-19 epidemic curve. The LSTM improves upon the standard recurrent neural networks (which suffers from the diminishing gradient problem) by adding a memory cell so that information can be maintained in memory for a longer period of time. Once can treat the problem as a time series analysis problem and use the standard LSTM for modeling and forecasting [26]. The LSTM model was found to be more accurate in producing forecasts compared to the logistic and Hill equations. Another study modeled the number of persons infected using the Ricatti equation. A variant of LSTM called the Bidirectional LSTM is then used to model and forecast this number [28]. A bidirectional LSTM trains two instead of one LSTM on the input sequence. The first is trained on the input sequence as-is while the second is trained on a reserved copy of the input sequence. The authors reported that the BiLSTM is capable of predicting accurately the evolution of the cumulative infectives. Another deep learning approach called Deep Extreme Learning Machine (DELM) was also reported [27]. DELM is essentially a stack of supervised autoencoders. This approach gave an accuracy of 97.59%. Another notable approach is the DeepCOVIDNet that outputs a probability distribution to predict the range of new cases on a given date in the future [30].

In addition to these three groups of machine learning models, there have been attempts at using mode conventional machine learning techniques. Techniques such as decision trees, random forests, logistic model trees, and Naive Bayes classifiers were used to model the development of COVID-19 based on 482 records of cases from Wuhan, China [66]. The models obtained accuracies between 87.44% and 90.83%, suggesting that the role of environmental factors does contribute to the spread of COVID-19. The Support Vector Machine (SVM) depicted good accuracy in modeling the spread of COVID-19 cases in India.[67] Apart from that, the Multi-Layer Perceptron-based approach is also being applied by several researchers in predicting the outbreak of COVID-19 [33,35].

Compartmental Models

Compartmental models started from the work of Kermack & Mckendrick [7]. The general model divides the population under study into compartments (or groups). The simplest compartmental model is the SIR model with three compartments; S:Susceptible (the group of people who are not yet infected and are at risk of infection), I: Infective (the group of people who have been infected), R:Removed (the group of people who either have recovered and received immunity, or those who are deceased). The sequence of the compartments, in this case, S-I-R denotes the flow of between the compartments (see Figure 2).

Figure 2. Flow chart for the SIR model.



The choice of which compartments to include in a model depends on the characteristics of the particular disease being modeled and the purpose of the model. The passively immune class LI and the latent period class E are often omitted, because they are not crucial for the susceptible-infective interaction. Acronyms for epidemiology models are often based on the flow patterns between the compartments such as LISEIR, MSEIRS, SEIR, SEIRS, SIR, SIRS, SEI, SEIS, SI, and SIS [7,8].

The transfer rates per unit time between compartments are described by a series of differential equations. Denote *S*, *I* and *R* the sizes of the respective compartments. For the basic SIR model, the differential equations are:

$$S' = -\beta SI$$
$$I' = \beta SI - \alpha I$$
$$R' = \alpha I$$

The two parameters of the model are described below:

i. β controls the rate of contact within the population. The number of susceptible individuals infected per unit time is βSI
ii. Infectives leave the *I* compartment at a rate of αI per unit time.

The quantity $\beta S_0/\alpha$ is called the *basic reproduction number*, usually denoted R_0. R_0 is the number of secondary infections caused by a single infective introduced into a wholly susceptible population. R_0 is an important parameter that determines if there is an epidemic. $R_0 > 1$ indicates that there will be an epidemic R_0 while < 1 indicates the epidemic will die out.

In the papers surveyed, a common assumption is that there is no entry into the population. In other words, birth and migration are negligible. Departure from the population is through death caused by COVID-19. The use of SIR model or any extension of SIR with the ‘R’ compartment will permit the assumption that COVID-19 confers immunity.

In our study, we found that the most commonly used compartmental models COVID-19 are the SIR, SEIR, SIRD, and SEIRD models. The ‘E’ compartment represents people who have been infected but are not themselves infectious yet. The ‘D’ compartment represents the deceased.

A key criticism of compartmental models is that each compartment is homogeneous. This assumption discounts the spatio-temporal and topological properties of the real world. The work that we surveyed for COVID-19 attempts to improve compartmental models in several ways, which we describe below.

Innovation and improvements to the base models are made in the following directions:

i. Using of sub-compartments within the main compartments. The 'I' compartment was divided into two groups - infected individuals who are hospitalized, and infected individuals who were not hospitalized [39]. The use of sub-compartments can be used to group members of the population who are subjected to different intervention measures.

ii. Treating the transition of individuals from one compartment to another as a stochastic process [37,47]. This is to introduce randomness into an otherwise homogenous compartmental model.

iii. Using optimization algorithms that wrap around the model to find good or optimal values for the model parameters [40]. In addition to the SEIR model, Particle Swarm Optimization (PSO) was used to estimate two parameters: the rate of infectious to hospitalized, and the recovery rate of quarantined infected individuals. Another example used an optimization algorithm that wraps around the SIRD model [38].

iv. Incorporating extra parameters to make the base model adapt to specific behaviors. A representative work under this category introduced four parameters representing diffusion terms into the compartmental PDEs, where each new parameter is attached to the S, E, I, and R compartments [42]. Another approach made use of the SIR model with the addition of three parameters to account for the early depleting dynamics of the epidemic [50].

v. Embedding structures into compartments to model real-world spatial properties and socio-economic networks. Croccolo & Roman (2020) applied Random graphs were applied with an underlying lattice model where each site in the lattice corresponds to membership of one of four categories S, I, R, D [41]. Another study proposed a network-augmented SIR (NSIR) model where a social network structure is embedded and simulated for each compartment [49]. A variety of network topologies were studied.

There are a couple of other studies worth highlighting. The modelling problem can be addressed through the lens of Ockham's Razor, motivating the derivation of a two-parameter model from the SIRD model [52]. The model was solved as a least-squares minimization problem with non-convex constraints. There is also a model called AIR, which is loosely based on the SIR model [51]. The author developed an analytical solution to the model in the form of recursive relations. These relations were used to model the relationship between the A, I and R compartments. (A: Actual active infective infectious group; I: Cummulative infected group, R: Cummulative recovered group).

The works cited above have demonstrated or claimed the efficacy of their proposed compartmental models in modeling and forecasting the COVID-19 epidemic within the interval of study. Compartmental models are appealing because of their simplicity. It is also easily adaptable to specific situations – compartments can be added; structures and constraints can be embedded within a compartment. A large number of compartmental models for COVID-19 are deterministic. Nevertheless, there is increasing interest in modeling the transition from one compartment to another as stochastic processes [37,47]. A study has found non-linear dynamics, including chaotic behavior in their SEIR model [40].

Agent-Based Models (ABM)

Unlike compartmental models that use a group or population as the basic modeling unit, agent-based models use the individual as the basic unit. It simulates individuals interacting in various settings. The spread of COVID-19 is estimated through agent interactions. Agent-based models are usually configured using data from activity surveys, census data, mobile phone location data, and traffic data. This provides high-resolution modeling as the behavior of each individual can be programmed. Agents are placed in a virtual environment which can be in two or three dimensions. Spatial characteristics of the area under study can be abstracted and incorporated into the model. In the simplest form, agents can have states corresponding to the groups in the compartmental model: susceptible, exposed, infected, removed. A more fine-grain model can be obtained by refining the states. For example, the state 'infected' can be divided into two states: 'symptomatic' and ‘asymptomatic’.

The level of abstraction will determine the computational complexity of the simulation. It may be feasible to study the risk of transmission of COVID-19 in a medical facility or a town by incorporating very detailed spatial and physical characteristics. However, the same study done on a regional or country scale may necessitate a higher level of abstraction. There is a trade-off between model resolution and computational complexity.

An agent-based model was used to study the risk of COVID-19 transmission in facilities [53]. In this context, facilities can be taken as referring to hospitals and schools. It was demonstrated that ABM can be conveniently used to study the effects of varying levels of prevention measures on the disease spread. Here, the ABM model enables the administrator of a facility to estimate the maximum capacity of individuals that can be supported without posing a high risk of COVID-19 transmission.

Another studyused ABM to assess the impact of non-pharmaceutical interventions on the cumulative number of infectives, mortality, and in-ICU bed occupancy in France [54]. The study found that physical distancing and mask-wearing combined with the protection of vulnerable groups will lead to better outcomes concerning disease control. However, it was pointed out that the model may not capture all components driving the viral transmission. In addition to this, the epidemic data used may be an underestimate due to limited availability of testing or asymptomatic cases.

An ABM has also been developed for Australia using stochastic agents [55]. Each agent represents an individual in the population. Agents are calibrated to the Australian census data (2016). The authors used the model to assess their effectiveness of several intervention strategies – varying levels of social distancing, school closures, home quarantine, international air travel restrictions and case isolation. When comparing with the actual Australian epidemic curve within the duration of the study, the model tends to overestimate the tail-end of the window.

In a similar study but at a smaller scale for the city of New Rochelle [56], an ABM was developed to incorporate a variety of elements such as different testing strategies, pharmaceutical, and non-pharmaceutical interventions, modeling different groups of people, modeling of expert knowledge. The model closely matches the initial progression of the disease. It was highlighted that the model lacks agent mobility, random contacts. There was also an oversimplistic assumption that hospitals have infinite beds and ICUs.

In general, studies applying ABM to COVID-19 show the following key advantages of ABM:

i. The ability to model individual human behaviors. This can potentially give a more realistic estimate of the impact of preventive measures such as social distancing, wearing of face masks, and the washing of hands.

ii. The capability to define and model the structural and spatio-temporal properties of the real world. An ABM can incorporate elements of the building or city infrastructure and traffic flow as a proxy for human mobility. An ABM model also typically defines a social network structure. This social network models physical contact and interactions among people.

iii. With ABMs, it is possible to compute the reproductive numbers (basic and effective) at the scale of individuals. Super-spreaders have very large individual reproductive numbers. This will skew the value of R_0 which is the mean of the population, making it challenging to forecast the trajectory of the COVID-19 spread. An ABM can address this issue as reproductive numbers can be estimated for individuals.

A commonly cited drawback of the ABM is that it requires large computational resources for carrying out the simulation. Secondly, building a good ABM model requires detailed data of movements, demography, and relevant geospatial data. The use of some of these data may raise privacy concerns.

Discussion

Since there are a great number of models and a variety of approaches, one issue facing a statistical analyst is the selection of an appropriate model. A paradigm that is adopted in mathematical modelling is known as the Occam's Razor, or the law of parsimony, which is the problem-solving principle that "entities should not be multiplied without necessity". This principle has been stated in various ways; for example, "when you have two competing theories that make the same predictions, the simpler one is the better." For mathematical models, it means that if there are two or more models that work about equally well, choose the simplest model [52]. Parsimonious models have been found to be useful in the understanding of the pandemic [68].

Time series models are widely employed in epidemiological studies due to their dynamic nature and suitability for time-structured data. Regarding the time or temporal nature of time series models, data could be recorded hourly, daily, weekly or monthly basis. This flexibility permits short to long-term analysis. Since the COVID-19 pandemic spreads rather quickly, a short-period (daily) analysis is of importance in tracking the progression of the disease.

Most of the time series analyses do not directly model the integer nature of the pandemic data (number of infected, deaths, recovered), with an exception [15], and account for the effects of intervention (lockdown, precautionary measures, vaccination etc). More comprehensive statistical models could be developed to include factors like healthcare system, economy, climatic conditions, population density that will influence the spread of COVID-19. Virtually all the time series models used do not go beyond ARIMA. Extensions like ARCH and GARCH to reflect volatility in the data could be considered.

Despite the huge success of machine learning techniques, the use of black-box models do not inspire confidence when used on high-stakes applications such as COVID-19 forecasting. Even though techniques have been created to explain black-box models, it may be better to use models that are interpretable by design [69]. Before a machine learning model can be widely acceptable for forecasting COVID-19 (or any epidemic) trajectory, the model has to be interpretable or explainable. Interpretability is the extent to which cause and effect can be understood within the model while explainability refers to the extent to which the inner workings of a model can be explained in human terms (so that it can be understood by the subject matter experts)

The COVID-19 pandemic is a black swan event [70]. At the initial onset of the pandemic, there was hardly any knowledge of the parameters to characterize the transmission dynamics of COVID-19. Attempts to seed the COVID-19 models using parameters from what was thought of as the next best thing, SARS did not produce good models because we now know that both SARS and COVID-19 are very different. From a machine learning perspective, this is known as *transfer learning* - using a model developed for a task as a starting point for a model on a new task. Challenges in obtaining quality data, especially in the early stages of the pandemic hamper the accuracy of models. Two common issues frequently highlighted are the under-reporting of cases due to the limited availability or access to testing facilities, and the number of asymptomatic cases. Another challenge related to data is what is known as concept drift [71]. A concept drift occurs when the statistical properties of the target variable change over time in unpredictable ways. The challenge then becomes detecting the occurrence of concept drifts and re-training the model.

Another significant observation from the work surveyed is that models are usually able to match earlier data but their forecasts diverge for future time points. This means that one can train any class of models to minimize errors on training data; however, the differences in the forecast of new cases highly depend on the assumptions taken by the modeling approach. This will further cast doubt on the usefulness of the models in such critical applications. One common approach to unify these different assumptions is to use the ensemble approach to model learning.

For a model to be useful in advising policy and intervention measures, it is expected to have the following properties:

i. The ability to perform good forecasts for different time horizons. Different types of models may be useful at different stages of the pandemic [68]. In particular, exponential growth models for the initial stage, a self-exciting branching process when one is still interested in the individual count data progressing into the development of the pandemic, and a macroscopic mean-field model approaching the peak of the pandemic. Comparing across approaches, the classical time series forecasting techniques have been demonstrated to perform better under certain circumstances compared to machine learning approaches, including deep learning [72]. Specifically, it was demonstrated that the ETS (Error, Trend, Seasonal) and ARIMA are better at one-step forecasting on univariate data

ii. In contrast to black-box models, the ability to provide an explanation to serve as a basis for public health professionals to understand and accept the forecast. The area is known as Explainable Artificial Intelligence [73]. Understanding how and why a model arrives at a certain output is important in high-stakes use cases such as advising policy and intervention measures.

iii. The ability to allow for fast scenario planning and ‘what-if’ analyses – public health authorities need a tool can quickly simulate the impact of various pharmaceutical and non-pharmaceutical interventions across a variety of spatio-temporal characteristics. This means the tool must be able to simulate or account for different behavioural responses to intervention measures. The tool must be able to assess the potential for global spread before significant case numbers had been reported internationally; and able to quantify the expected disease severity and burden of COVID-19 [74].

iv. A model that encourages a participatory approach. A participatory approach model uses international collaboration to develop a dynamic infectious disease model that addressed a global need [75]. A participatory approach is key for policymakers to fully appreciate the uncertainties subjacent to assumed parameter values, implemented mechanisms of action and general model structure. The immediate consequence of that understanding is clarity on the relevance of critical data in circumventing uncertainties, and the understanding that continual validation frameworks are key to guarantee the best possible policy is implemented at all times.

A promising approach lies at the intersection of the compartmental model and the agent-based model (ABM). An example under this category of models is a compartmental model where each compartment is modeled and simulated using ABM. In this approach, individuals (or small groups of homogeneous people) are the basic unit of study. The interactions between individuals are governed by their social network structure, temporal geospatial characteristics of the environment, and to a certain extent cultural nuances. There are two major model design considerations. Firstly, selecting a suitable level of abstraction for ABM to ensure the emergent behaviour is as close as possible to the real world. Having a more fine-grained ABM model will incur computational overhead without improvement in model performance. Secondly, translating human behaviours into a set of mathematical rules. This is challenging as behaviours can be unpredictable.

Declaration of interests

We declare no competing interests.

References

1. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: Recent insights into emerging coronaviruses [Internet]. Nature Reviews Microbiology. Nature Publishing Group; 2016 [cited 2021 Mar 5]. p. 523–534. PMID:27344959
2. Fauci AS, Lane HC, Redfield RR. Covid-19 — Navigating the Uncharted. New England Journal of Medicine [Internet] Massachusetts Medical Society; 2020 Mar 26 [cited 2021 Mar 5];382(13):1268–1269. PMID:32109011
3. WHO | Pneumonia of unknown cause – China [Internet]. [cited 2021 Mar 5]. Available from: https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/
4. Estrada E. COVID-19 and SARS-CoV-2. Modeling the present, looking at the future. Physics Reports. Elsevier B.V.; 2020. p. 1–51. [doi: 10.1016/j.physrep.2020.07.005]
5. Weekly epidemiological update - 27 January 2021 [Internet]. [cited 2021 Mar 5]. Available from: https://www.who.int/publications/m/item/weekly-epidemiological-update---27-january-2021
6. Challener DW, Dowdy SC, O'Horo JC. Analytics and Prediction Modeling During the COVID-19 Pandemic [Internet]. Mayo Clinic Proceedings. Elsevier Ltd; 2020 [cited 2021 Mar 5]. p. S8–S10. PMID:32807519
7. Kermack W 0, Mckendrick AG. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character [Internet] The Royal Society; 1927 Aug 1 [cited 2021 Mar 5];115(772):700–721. [doi: 10.1098/rspa.1927.0118]
8. Hethcote HW. The Mathematics of Infectious Diseases * [Internet]. Society for Industrial and Applied Mathematics. 2000. Available from: https://epubs.siam.org/page/terms

9. Holmdahl I, Buckee C. Wrong but Useful — What Covid-19 Epidemiologic Models Can and Cannot Tell Us. New England Journal of Medicine [Internet] Massachusetts Medical Society; 2020 Jul 23 [cited 2021 Mar 5];383(4):303–305. [doi: 10.1056/NEJMp2016822]
10. Jewell NP, Lewnard JA, Jewell BL. Caution Warranted: Using the Institute for Health Metrics and Evaluation Model for Predicting the Course of the COVID-19 Pandemic [Internet]. Annals of internal medicine. NLM (Medline); 2020 [cited 2021 Mar 5]. p. 226–227. PMID:32289150
11. Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. Data in Brief Elsevier Inc.; 2020 Apr 1;29. [doi: 10.1016/j.dib.2020.105340]
12. Singh S, Parmar KS, Kumar J, Makkhan SJS. Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19. Chaos, Solitons and Fractals Elsevier Ltd; 2020 Jun 1;135. PMID:32395038
13. Chintalapudi N, Battineni G, Amenta F. COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach. Journal of Microbiology, Immunology and Infection Elsevier Ltd; 2020 Jun 1;53(3):396–403. PMID:32305271
14. Ankaralı H, Erarslan N, Pasin Ö, Al-Mahmood AK. Modeling and short-term forecasts of indicators for COVID-19 outbreak in 25 countries at the end of march. Bangladesh Journal of Medical Science Ibn Sina Trust; 2020;19(Special issue):6–20. [doi: 10.3329/bjms.v19i0.47611]
15. Agosto A, Giudici P. COVID-19 Contagion Dynamics. 2020; [doi: 10.3390/risks8030077]
16. Katris C. A time series-based statistical approach for outbreak spread forecasting: Application of COVID-19 in Greece. Expert Systems with Applications Elsevier Ltd; 2021 Mar 15;166. [doi: 10.1016/j.eswa.2020.114077]
17. Gautam Jamdade P, Gautamrao Jamdade S. Modeling and prediction of COVID-19 spread in the Philippines by October 13, 2020, by using the VARMAX time series method with preventive measures. Results in Physics Elsevier B.V.; 2021 Jan 1;20. [doi: 10.1016/j.rinp.2020.103694]
18. Păcurar CM, Necula BR. An analysis of COVID-19 spread based on fractal interpolation and fractal dimension. Chaos, Solitons and Fractals Elsevier Ltd; 2020 Oct 1;139. [doi: 10.1016/j.chaos.2020.110073]
19. Abbasi M, Bollini AL, Castillo JLB, Deppman A, Guidio JP, Matuoka PT, Meirelles AD, Policarpo JMP, Ramos AAGF, Simionatto S, Varona ARP, Andrade-II E, Panjeh H, Trevisan LA. Fractal signatures of the COVID-19 spread. Chaos, Solitons and Fractals Elsevier Ltd; 2020 Nov 1;140. [doi: 10.1016/j.chaos.2020.110119]
20. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, Yan P, Chowell G. Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang, China: February 13–23, 2020. Journal of Clinical Medicine MDPI AG; 2020 Feb 22;9(2):596. [doi: 10.3390/jcm9020596]
21. Ribeiro MHDM, da Silva RG, Mariani VC, Coelho L dos S. Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for Brazil. Chaos, Solitons and Fractals Elsevier Ltd; 2020 Jun 1;135. [doi: 10.1016/j.chaos.2020.109853]
22. Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. Chaos, Solitons and Fractals Elsevier Ltd; 2020 Jun 1;135. PMID:32390691
23. Chen D-G, Chen X, Chen JK. Reconstructing and forecasting the COVID-19 epidemic in the United States using a 5-parameter logistic growth model. Global Health Research and Policy Springer Science and Business Media LLC; 2020 Dec;5(1). [doi: 10.1186/s41256-020-00152-5]
24. Sanchez-Caballero S, Selles MA, Peydro MA, Perez-Bernabeu E. An Efficient COVID-19 Prediction Model Validated with the Cases of China, Italy and Spain: Total or Partial Lockdowns? Journal of Clinical Medicine MDPI AG; 2020 May 20;9(5):1547. PMID:32443871
25. Rustam F, Reshi AA, Mehmood A, Ullah S, On BW, Aslam W, Choi GS. COVID-19 Future Forecasting Using Supervised Machine Learning Models. IEEE Access Institute of Electrical and Electronics Engineers Inc.; 2020;8:101489–101499. [doi: 10.1109/ACCESS.2020.2997311]

26. Yan B, Wang J, Zhang Z, Tang X, Zhou Y, Zheng G, Zou Q, Lu Y, Liu B, Tu W, Xiong N. An improved method for the fitting and prediction of the number of covid-19 confirmed cases based on LSTM. Computers, Materials and Continua Tech Science Press; 2020 Jun 30;64(3):1473–1490. [doi: 10.32604/cmc.2020.011317]
27. Khan MA, Abbas S, Khan KM, Ghamdi MAA, Rehman A. Intelligent forecasting model of covid-19 novel coronavirus outbreak empowered with deep extreme learning machine. Computers, Materials and Continua Tech Science Press; 2020 Jun 30;64(3):1329–1342. [doi: 10.32604/cmc.2020.011155]
28. Fokas AS, Dikaios N, Kastis GA. Mathematical models and deep learning for predicting the number of individuals reported to be infected with SARS-CoV-2. Journal of The Royal Society Interface The Royal Society; 2020 Aug;17(169):20200494. PMID:32752997
29. Aviv-Sharon E, Aharoni A. Generalized logistic growth modeling of the COVID-19 pandemic in Asia. Infectious Disease Modelling KeAi Communications Co.; 2020 Jan 1;5:502–509. [doi: 10.1016/j.idm.2020.07.003]
30. Ramchandani A, Fan C, Mostafavi A. DeepCOVIDNet: An Interpretable Deep Learning Model for Predictive Surveillance of COVID-19 Using Heterogeneous Features and Their Interactions. IEEE Access Institute of Electrical and Electronics Engineers Inc.; 2020;8:159915–159930. [doi: 10.1109/ACCESS.2020.3019989]
31. Shrivastav LK, Jha SK. A gradient boosting machine learning approach in modeling the impact of temperature and humidity on the transmission rate of COVID-19 in India. Applied Intelligence Springer; 2020; [doi: 10.1007/s10489-020-01997-6]
32. Mehta M, Julaiti J, Griffin P, Kumara S. Early stage machine learning–based prediction of US county vulnerability to the COVID-19 pandemic: Machine learning approach. JMIR Public Health and Surveillance JMIR Publications Inc.; 2020 Jul 1;6(3). [doi: 10.2196/19446]
33. Pinter G, Felde I, Mosavi A, Ghamisi P, Gloaguen R. COVID-19 pandemic prediction for Hungary; A hybrid machine learning approach. Mathematics MDPI AG; 2020;8(6). [doi: 10.3390/math8060890]
34. Suzuki Y, Suzuki A, Nakamura S, Ishikawa T, Kinoshita A. Machine learning model estimating number of COVID-19 infection cases over coming 24 days in every province of South Korea (XGBoost and MultiOutputRegressor). medRxiv. medRxiv; 2020. [doi: 10.1101/2020.05.10.20097527]
35. Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Rabczuk T, Atkinson PM. COVID-19 outbreak prediction with machine learning. Algorithms MDPI AG; 2020;13(10). [doi: 10.3390/a13100249]
36. Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos, Solitons and Fractals Elsevier Ltd; 2020 May 1;134. [doi: 10.1016/j.chaos.2020.109761]
37. He S, Tang S, Rong L. A discrete stochastic model of the COVID-19 outbreak: Forecast and control. Mathematical Biosciences and Engineering American Institute of Mathematical Sciences; 2020 Mar 16;17(4):2792–2804. PMID:32987496
38. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS ONE Public Library of Science; 2020;15(3). PMID:32231374
39. Liu M, Ning J, Du Y, Cao J, Zhang D, Wang J, Chen M. Modelling the evolution trajectory of COVID-19 in Wuhan, China: experience and suggestions. Public Health Elsevier B.V.; 2020 Jun 1;183:76–80. PMID:32442842
40. He S, Peng Y, Sun K. SEIR modeling of the COVID-19 and its dynamics. Nonlinear Dynamics Springer Science and Business Media B.V.; 2020 Aug 1;101(3):1667–1680. [doi: 10.1007/s11071-020-05743-y]
41. Croccolo F, Roman HE. Spreading of infections on random graphs: A percolation-type model for COVID-19. Chaos, Solitons and Fractals Elsevier Ltd; 2020 Oct 1;139. [doi: 10.1016/j.chaos.2020.110077]
42. Viguerie A, Lorenzo G, Auricchio F, Baroli D, Hughes TJR, Patton A, Reali A, Yankeelov TE,

Veneziani A. Simulating the spread of COVID-19 via a spatially-resolved susceptible–exposed–infected–recovered–deceased (SEIRD) model with heterogeneous diffusion. Applied Mathematics Letters Elsevier Ltd; 2021 Jan 1;111. [doi: 10.1016/j.aml.2020.106617]

43. Wang H, Yamamoto N. Using a partial differential equation with Google Mobility data to predict COVID-19 in Arizona. Mathematical Biosciences and Engineering American Institute of Mathematical Sciences; 2020 Jul 15;17(5):4891–4904. PMID:33120533
44. Barbarossa MV, Fuhrmann J, Meinke JH, Krieg S, Varma HV, Castelletti N, Lippert T. Modeling the spread of COVID-19 in Germany: Early assessment and possible scenarios. PLoS ONE Public Library of Science; 2020 Sep 1;15(9 september). PMID:32886696
45. Al-Khani AM, Khalifa MA, Almazrou A, Saquib N. The SARS-CoV-2 pandemic course in Saudi Arabia: A dynamic epidemiological model. Infectious Disease Modelling KeAi Communications Co.; 2020 Jan 1;5:766–771. [doi: 10.1016/j.idm.2020.09.006]
46. Postavaru O, Anton SR, Toma A. COVID-19 pandemic and chaos theory. Mathematics and Computers in Simulation Elsevier B.V.; 2021 Mar 1;181:138–149. [doi: 10.1016/j.matcom.2020.09.029]
47. Hussain S, Zeb A, Rasheed A, Saeed T. Stochastic mathematical model for the spread and control of Corona virus. Advances in Difference Equations Springer Science and Business Media Deutschland GmbH; 2020 Dec 1;2020(1). [doi: 10.1186/s13662-020-03029-6]
48. Baek YJ, Lee T, Cho Y, Hyun JH, Kim MH, Sohn Y, Kim JH, Ahn JY, Jeong SJ, Ku NS, Yeom JS, Lee J, Choi JY. A mathematical model of COVID-19 transmission in a tertiary hospital and assessment of the effects of different intervention strategies. PLoS ONE Public Library of Science; 2020 Oct 1;15(10). PMID:33104736
49. Karaivanov A. A social network model of COVID-19. PLoS ONE Public Library of Science; 2020 Oct 1;15(10 October). PMID:33119621
50. Law KB, Peariasamy KM, Gill BS, Singh S, Sundram BM, Rajendran K, Dass SC, Lee YL, Goh PP, Ibrahim H, Abdullah NH. Tracking the early depleting transmission dynamics of COVID-19 with a time-varying SIR model. Scientific Reports [Internet] 2020 Dec 10;10(1):21721. [doi: 10.1038/s41598-020-78739-8]
51. Liu X. A simple, SIR-like but individual-based epidemic model: Application in comparison of COVID-19 in New York City and Wuhan. Results in Physics Elsevier BV; 2021 Jan;20:103712. [doi: 10.1016/j.rinp.2020.103712]
52. Fiacchini M, Alamir M. The Ockham's razor applied to COVID-19 model fitting French data. Annual Reviews in Control [Internet] 2021 Jan; [doi: 10.1016/j.arcontrol.2021.01.002]
53. Cuevas E. An agent-based model to evaluate the COVID-19 transmission risks in facilities. Computers in Biology and Medicine Elsevier Ltd; 2020 Jun 1;121. PMID:32568667
54. Hoertel N, Blachier M, Blanco C, Olfson M, Massetti M, Rico MS, Limosin F, Leleu H. A stochastic agent-based model of the SARS-CoV-2 epidemic in France. Nature Medicine 2020;26(9):1417–1421. PMID:32665655
55. Chang SL, Harding N, Zachreson C, Cliff OM, Prokopenko M. Modelling transmission and control of the COVID-19 pandemic in Australia. Nature Communications Nature Research; 2020 Dec 1;11(1). PMID:33177507
56. Truszkowska A, Behring B, Hasanyan J, Zino L, Butail S, Caroppo E, Jiang ZP, Rizzo A, Porfiri M. High-Resolution Agent-Based Modeling of COVID-19 Spreading in a Small Town. Advanced Theory and Simulations Wiley-VCH Verlag; 2021; [doi: 10.1002/adts.202000277]
57. Holt CC. Forecasting seasonals and trends by exponentially weighted averages (O.N.R. Memorandum No. 52). Carnegie Institute of Technology, Pittsburgh; 1957.
58. Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. International Journal of Forecasting 2004;20:5–10. [doi: 10.1016/j.ijforecast.2003.09.015]
59. Winters PR. Forecasting Sales by Exponentially Weighted Moving Averages. Management Science 1960;6(3):324–342.

60. Martinez EZ, Aragon DC, Nunes AA. Short-term forecasting of daily COVID-19 cases in Brazil by using the holt's model. Revista da Sociedade Brasileira de Medicina Tropical Sociedade Brasileira de Medicina Tropical; 2020;53:1–5. PMID:32520235
61. Hyndman RJ. Automatic Time Series Forecasting : The forecast Package for R. 2008;27(3).
62. Nandi AK. Data Modeling with Polynomial Representations and Autoregressive Time-Series Representations, and Their Connections. IEEE Access Institute of Electrical and Electronics Engineers Inc.; 2020;8:110412–110424. [doi: 10.1109/ACCESS.2020.3000860]
63. Wang Y, Xu C, Yao S, Zhao Y, Li Y, Wang L, Zhao X. Estimating the prevalence and mortality of coronavirus disease 2019 (COVID-19) in the USA, the UK, Russia, and India. Infection and Drug Resistance Dove Medical Press Ltd; 2020;13:3335–3350. PMID:33061481
64. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman JM, Yan P, Chowell G. Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infectious Disease Modelling KeAi Communications Co.; 2020 Jan 1;5:256–263. [doi: 10.1016/j.idm.2020.02.002]
65. Hochreiter S. Long Short-Term Memory. 1997;1780:1735–1780.
66. Keshavarzi A. Coronavirus Infectious Disease (COVID-19) Modeling: Evidence of Geographical Signals [Internet]. Available from: https://ssrn.com/abstract=3568425
67. Arun SS, Neelakanta Iyer G. On the Analysis of COVID19 - Novel Corona Viral Disease Pandemic Spread Data Using Machine Learning Techniques. 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) [Internet] IEEE; 2020. p. 1222–1227. [doi: 10.1109/ICICCS48265.2020.9121027]
68. Bertozzi AL, Franco E, Mohler G, Short MB, Sledge D. The challenges of modeling and forecasting the spread of COVID-19. Proceedings of the National Academy of Sciences of the United States of America National Academy of Sciences; 2020 Jul 21;117(29):16732–16738. PMID:32616574
69. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence [Internet] Springer US; 2019;1(5):206–215. [doi: 10.1038/s42256-019-0048-x]
70. Cirillo P, Taleb NN. Tail risk of contagious diseases. Nature Physics. Nature Research; 2020. p. 606–613. [doi: 10.1038/s41567-020-0921-x]
71. Lu J, Liu A, Dong F, Gu F, Gama J, Zhang G. Learning under concept drift: A review. IEEE Trans Knowledge and Data Engineering 2019;31(12):2346–2363.
72. Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLoS ONE 2018;13(3):1–26. PMID:29584784
73. Barredo Arrieta A, Díaz-Rodríguez N, del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F. Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion [Internet] Elsevier B.V.; 2020;58(October 2019):82–115. [doi: 10.1016/j.inffus.2019.12.012]
74. Meehan MT, Rojas DP, Adekunle AI, Adegboye OA, Caldwell JM, Turek E, Williams BM, Marais BJ, Trauer JM, Mcbryde ES. Modelling insights into the COVID-19 pandemic. Paediatric Respiratory Reviews 2020;35(January):64–69.
75. Aguas R, White L, Hupert N, Shretta R, Pan-Ngum W, Celhay O, Moldokmatova A, Arifi F, Mirzazadeh A, Sharifi H, Adib K, Sahak MN, Franco C, Coutinho R. Modelling the COVID-19 pandemic in context: an international participatory approach. BMJ Global Health 2020;5(12):e003126. [doi: 10.1136/bmjgh-2020-003126]