Complex Systems Project: Math Modeling, analysis, and predictions for COVID-19

Mitodru Niyogi
Faculty of Mathematics and Computer Science
Heidelberg University
Heidelberg, Germany
mitodru.niyogi@stud.uni-heidelberg.de

Abstract—In December 2019, a novel coronavirus was found in a seafood wholesale market in Wuhan, China. World Health Organization (WHO) officially named this coronavirus as COVID-19. Since the first patient was hospitalized on December 12, 2019, China has reported a total of 78,824 confirmed COVID-19 cases and 2,788 deaths as of February 28, 2020. The COVID-19 has been successfully contained in China but is spreading all over the world. COVID-19 epidemic is prone to disrupt and crumble the existing health-care infrastructures in both the developed and developing world. COVID19 also impacts people's daily life and country's economic development. In this paper, we adopt mathematical epidemic models such as Susceptible-Infected-Recovery (SIR), Susceptible-Infected-Recovery-Fatality/Deaths (SIR-F) to simulate the epidemic on the data available for the entire world and future projections on the number of infections, deaths in six specific countries (Italy, France, Spain, Germany, USA, and India) across a time-frame of 7 days, 1 month, 3 months, and 3 years in future. We analyzed the epidemic by extending the SIR-F model with controlled parameters and simulating the behavior on our default case study data. We also fit other mathematical models such as exponential and logistic models to $C(t)$, the cumulative number of positive infections trajectory function. In the latter section, we also used statistical machine learning techniques such as Polynomial regression, support vector machine regression, and simple neural network such as multilayer perceptron to better understand and learns the underlying pattern of the real epidemic growth and the virus proliferation pattern. We found out that the predictions by the logistic model was underreported, i.e, the actual trajectory is more complex than the logistic model. However, we found out that different models found to be better in modeling the pandemic outbreak in respective countries. We also performed data analysis to project the infection, recovery, and death statistics from the real data and also calculated the growth factor of pandemic outbreak in the countries and grouping them respectively. To our future projections and analysis, we found out USA, followed by India are gonna be the most affected countries with each resulting into millions of positive infections cases and deaths.

Index Terms—COVID-19, epidemic modeling, data analysis, SIR/SIR-F modeling, machine learning, polynomial regression, support vector machine, logistic modeling, predictions

I. Introduction

The first case of the novel coronavirus (COVID-19) was first reported in the Hubei province in China, and thereafter the novel coronavirus (COVID-19) has been spreading all over the world, after two months of outbreak in China. Facing uncertainty, criticism, and irresolution in December 2019 and the first half of January 2020, China then responded efficiently and massively to this new disease outbreak by implementing unprecedented containment measures to the whole country, including lockdown the whole province of Hubei and putting most of other provinces in de-facto quarantine mode. As of March 10, one and a half month after the national battle against the COVID-19 epidemic, China has successfully contained the virus transmission within the country, with new daily confirmed cases in mainland China excluding Hubei in the single digit range, and with just double digit numbers in Hubei. In contrast, many other countries especially in Europe, Iran have fast increasing numbers of confirmed cases mainly for not reacting promptly at the beginning. As of March 10, 103 countries in addition to China have reported confirmed cases infected by COVID-19.pan After the mid of March, the condition worsens in USA that it finally surpasses the Europe as the pandemic epicenter by the end of mid April. In this paper, we estimate the basic reproduction number R0 and predict the future trajectory of the coronavirus (COVID-19) outbreak in the world and in mainly six countries: US, Italy, Germany, Spain, UK, India. We will use SIR-F model that is a customized ODE model derived from SIR model. In this paper, we focus on developing epidemic modeling and the use of machine learning algorithms for COVID-19 outbreak. We implemented SIR, SIR-D, SIR-F models to model the epidemic outbreak considering default initial parameters as well by dimensioning it with the real data. We will use SIR-F model that is a customized ODE model derived from SIR model. To evaluate the effect of measures, parameter estimation of SIR-F will be applied to subsets of time series data in each country. In the SIR-F modeling, we take consideration of control parameters modeling to study the impact of lockdown implementation and the availability of vaccination. We also studied case studies about what will happen if the spread of the virus is very quick, slow, improved health-care system parameters. In machine learning techniques, we also used logistic regression, polynomial regression, support vector machine (SVM) predictions as another part of prediction

Identify applicable funding agency here. If none, delete this.

algorithms in our work to compare the predictions by the ODEs epidemic models versus the machine learning models to simulate the novel virus outbreak and its trajectory status of the epidemics and future scenarios of the outbreak. Usually, an epidemic follows an exponential growth at an early stage (following the law of proportional growth), peaks and then the growth rate decays as countermeasures to hinder the transmission of the virus are introduced. We also performed data analysis to compare the number of confirmed cases in six countries, the growth factor of respective countries. We also implemented K-Means clustering on the mitigation measures taken by major countries to cluster the important measures in groups. Our analysis dissects the development of the epidemics in the countries of interest and the impact of the drastic control measures both at the aggregate level and within each province. We made projections on the development of the outbreak in the six key countries and the whole World, based on different scenarios provided by the results from different models. Our study employs simple models to quantitatively document the effects of the Chinese containment measures against the COVID-19, and provide informative implications for the coming pandemic.

II. Background

A. Epidemic Modeling

Mathematical models have the ability to project how infectious diseases progress to show the likely outcome of an epidemic and help inform public health interventions to take wise decisions for the safety of the people. Models use basic assumptions or collected statistics along with mathematics to find parameters for various infectious diseases and use those parameters to calculate the effects of different interventions, like mass vaccination programmes. The modeling helps to decide which intervention/s are required to avoid and what to trial with, or can predict future growth patterns, etc. In a deterministic model, individuals in the population are assigned to different subgroups or compartments, each representing a specific stage of the epidemic. The deterministic model is expressed as a set of differential equations and the transition rates from one stage to another are mathematically expressed as derivatives of the differential equations. These models assume that the population size in a compartment is differentiable with respect to time and that the epidemic process is deterministic. In epidemic modeling, the basic reproduction number (denoted by R_0) is used to measure how transferable a disease is. It determines the average number of people that a single infectious person will infect over the course of their infection. Thus, the reproduction number reflects whether the infection will spread exponentially, die out, or remain constant:

- if $R_0 > 1$, virus/disease will spread as each person on average infects more than one other person;
- if $R_0 < 1$, virus/disease will die out as each person infects fewer than one person on average;
- and if $R_0 = 1$, virus/disease becomes endemic, i.e, it moves throughout the population but not increase or decrease as each person will infect on average exactly one other person

In this section, we will introduce the methodology of different existing differential mathematical models for modeling an epidemic in a region.

1) SIR Model: SIR model [4] is a simple mathematical model to understand outbreak of infectious diseases.

- S: Susceptible (=All - Confirmed)
- I: Infected (=Confirmed - Recovered - Deaths)
- R: Recovered or fatal (=Recovered + Deaths)

Though R in SIR model is "Recovered and have immunity", we defined "R as Recovered or fatal". This is because mortality rate cannot be ignored in the real COVID-19 data.

Model:

$$\mathrm{S} \xrightarrow{\beta I} \mathrm{I} \xrightarrow{\gamma} \mathrm{R} \tag{1}$$

$$\tag{2}$$

β: Effective contact rate [1/min], γ: Recovery(+Mortality) rate [1/min]

Ordinary Differential Equation (ODEs):

$$\frac{\mathrm{d}S}{\mathrm{d}T} = -N^{-1}\beta SI \tag{3}$$

$$\frac{\mathrm{d}I}{\mathrm{d}T} = N^{-1}\beta SI - \gamma I \tag{4}$$

$$\frac{\mathrm{d}R}{\mathrm{d}T} = \gamma I \tag{5}$$

$$\tag{6}$$

Where $N = S + I + R$ is the total population, T is the elapsed time from the start date.

Non-dimensional SIR model To simplify the model, we will remove the units of the variables from ODE.

Set $(S, I, R) = N \times (x, y, z)$ and $(T, \beta, \gamma) = (\tau t, \tau^{-1}\rho, \tau^{-1}\sigma)$.

This results in the ODE

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\rho xy \tag{7}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \rho xy - \sigma y \tag{8}$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \sigma y \tag{9}$$

$$\tag{10}$$

Where N is the total population and τ is a coefficient ([min], is an integer to simplify).

The range of variables and parameters:

$$0 \leq (x, y, z, \rho, \sigma) \leq 1 \tag{11}$$

$$\tag{12}$$

$$1 \leq \tau \leq 1440 \tag{13}$$
$$\tag{14}$$

Basic reproduction number, Non-dimensional parameter, is defined as

$$R_0 = \rho\sigma^{-1} = \beta\gamma^{-1} \tag{15}$$

Estimated Mean Values of R_0: R_0 ("R naught") means "the average number of secondary infections caused by an infected host" 2.06: Zika in South America, 2015-2016 1.51: Ebola in Guinea, 2014 1.33: H1N1 influenza in South Africa, 2009 3.5 : SARS in 2002-2003 1.68: H2N2 influenza in US, 1957 3.8 : Fall wave of 1918 Spanish influenza in Genova 1.5 : Spring wave of 1918 Spanish influenza in Genova

When $x = \frac{1}{R_0}$, $\frac{\mathrm{d}y}{\mathrm{d}t} = 0$. <!–This means that the max value of confirmed $(= y + z)$ is $1 - \frac{1}{R_0}$.–>

Example of non-dimensional SIR model For example, set $R_0 = 2.5, \rho = 0.2$ and initial values $(x_{(0)}, y_{(0)}, z_{(0)}) = (0.999, 0.001, 0)$. SIR class was defined in "Preparation" section.

2) SIR-D model: Model:

$$\mathrm{S} \xrightarrow{\beta I} \mathrm{I} \xrightarrow{\gamma} \mathrm{R} \tag{16}$$
$$\mathrm{I} \xrightarrow{\alpha} \mathrm{D} \tag{17}$$
$$\tag{18}$$

α: Mortality rate [1/min], β: Effective contact rate [1/min], γ: Recovery rate [1/min]

Ordinary Differential Equation (ODEs):

$$\frac{\mathrm{d}S}{\mathrm{d}T} = -N^{-1}\beta SI \tag{19}$$
$$\frac{\mathrm{d}I}{\mathrm{d}T} = N^{-1}\beta SI - (\gamma + \alpha)I \tag{20}$$
$$\frac{\mathrm{d}R}{\mathrm{d}T} = \gamma I \tag{21}$$
$$\frac{\mathrm{d}D}{\mathrm{d}T} = \alpha I \tag{22}$$
$$\tag{23}$$

Where $N = S + I + R + D$ is the total population, T is the elapsed time from the start date.

Non-dimnsional SIR-D model Set $(S, I, R, D) = N \times (x, y, z, w)$ and $(T, \alpha, \beta, \gamma) = (\tau t, \tau^{-1}\kappa, \tau^{-1}\rho, \tau^{-1}\sigma)$. This results in the ODE

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\rho xy \tag{24}$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = \rho xy - (\sigma + \kappa)y \tag{25}$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = \sigma y \tag{26}$$
$$\frac{\mathrm{d}w}{\mathrm{d}t} = \kappa y \tag{27}$$
$$\tag{28}$$

Where N is the total population and τ is a coefficient ([min], is an integer to simplify).

The range of variables and parameters:

$$0 \leq (x, y, z, w, \kappa, \rho, \sigma) \leq 1 \tag{29}$$
$$\tag{30}$$

$$1 \leq \tau \leq 1440 \tag{31}$$
$$\tag{32}$$

Reproduction number can be defined as

$$R_0 = \rho(\sigma + \kappa)^{-1} = \beta(\gamma + \alpha)^{-1} \tag{33}$$

3) SIR-F model: Measurable variables: Confirmed $= I + R + F$, Recovered $= R$, Deaths $= F$

Model:

$$\mathrm{S} \xrightarrow{\beta I} \mathrm{S}^* \xrightarrow{\alpha_1} \mathrm{F} \tag{34}$$
$$\mathrm{S}^* \xrightarrow{1-\alpha_1} \mathrm{I} \xrightarrow{\gamma} \mathrm{R} \tag{35}$$
$$\mathrm{I} \xrightarrow{\alpha_2} \mathrm{F} \tag{36}$$
$$\tag{37}$$

α_1: Mortality rate of S^* cases, α_2: Mortality rate of I cases [1/min], β: Effective contact rate [1/min], γ: Recovery rate [1/min],

Note: When $\alpha_1 = 0$, SIR-F model is the same as SIR-D model.

Ordinary Differential Equation (ODEs):

$$\frac{\mathrm{d}S}{\mathrm{d}T} = -N^{-1}\beta SI \tag{38}$$
$$\frac{\mathrm{d}I}{\mathrm{d}T} = N^{-1}(1 - \alpha_1)\beta SI - (\gamma + \alpha_2)I \tag{39}$$
$$\frac{\mathrm{d}R}{\mathrm{d}T} = \gamma I \tag{40}$$
$$\frac{\mathrm{d}F}{\mathrm{d}T} = N^{-1}\alpha_1\beta SI + \alpha_2 I \tag{41}$$
$$\tag{42}$$

Where $N = S + I + R + F$ is the total population, T is the elapsed time from the start date.

Non-dimensional SIR-F model: Set $(S, I, R, F) = N \times (x, y, z, w)$ and $(T, \alpha_1, \alpha_2, \beta, \gamma) = (\tau t, \theta, \tau^{-1}\kappa, \tau^{-1}\rho, \tau^{-1}\sigma)$. This results in the ODE

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\rho xy \tag{43}$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = \rho(1 - \theta)xy - (\sigma + \kappa)y \tag{44}$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = \sigma y \tag{45}$$
$$\frac{\mathrm{d}w}{\mathrm{d}t} = \rho\theta xy + \kappa y \tag{46}$$
$$\tag{47}$$

Where N is the total population and τ is a coefficient ([min], is an integer to simplify).

The range of variables and parameters:

$$0 \leq (x, y, z, w, \theta, \kappa, \rho, \sigma) \leq 1 \tag{48}$$
$$\tag{49}$$

$$1 \leq \tau \leq 1440 \tag{50}$$

$$\tag{51}$$

Reproduction number can be defined as

$$R_0 = \rho(1-\theta)(\sigma+\kappa)^{-1} = \beta(1-\alpha_1)(\gamma+\alpha_2)^{-1} \tag{52}$$

III. Dataset

The following dataset have been used for our project:

A. COVID-19 containment and mitigation measures

The dataset has been taken from Kaggle [1]. The dataset has been compiled from Dataset [2] of COVID-19 containment and mitigation measures . The dataset attempts to cover all measures of national significance intended to reduce the transmission of COVID-19, in all nations. Each measure in the database has entries on:

- Country (and state for the US)
- Textual description of the measure
- Start date of measure
- End date (if available)
- URL to source of more information
- Systematic keyword labels (e.g. "travel ban" or "hygiene enforcement")

We used this dataset to cluster the descriptions of the measures taken using KMeans clustering algorithm in the Data Analysis section.

B. Novel Coronavirus Dataset

This datset has been scrapped from Johns Hopkins University's publicly available data for COVID-19 tracking throughout th world and made available on Kaggle. [3] This time series dataset has daily level information on the number of positive confirmed cases, deaths and recoveries regarding the novel coronavirus pandemic. As this is a time series data which means that the number of reported cases on any given day is the cumulative number of the previous count. The dataset has information available from 22 Jan, 2020 onward till present. The dataset has the following attributes as follows:

- Sno - Serial number
- ObservationDate - Date of the observation in MM/DD/YYYY
- Province/State - Province or state of the observation
- Country/Region - Country of observation
- Last Update - Time in UTC at which the row is updated for the given province or country
- Confirmed - Cumulative number of confirmed cases till that date
- Deaths - Cumulative number of of deaths till that date
- Recovered - Cumulative number of recovered cases till that date

[1] https://www.kaggle.com/paultimothymooney/covid19-containment-and-mitigation-measures

[2] http://epidemicforecasting.org/containment

[3] https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset

C. COVID-19 Global Forecasting (Week 4)

This dataset is taken from Kaggle challenge [4] where participants were asked to predict the cumulative number of confirmed COVID19 cases in various locations across the world, as well as the number of resulting fatalities, for future dates. The dataset has the following attributes:

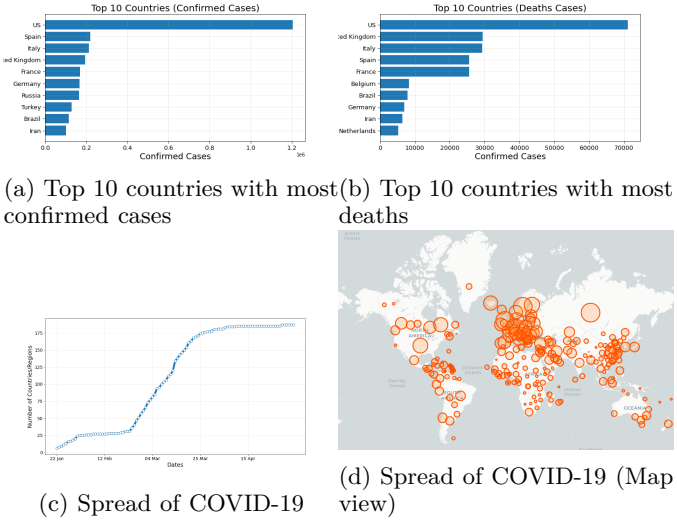
- Id
- Province_State Country_Region
- Date
- ConfirmedCases
- Fatalities

D. Population Pyramind Dataset

This dataset [5] has information about the population by gender for CoVid-19 affected countries.

IV. Data Analysis

In this section, we performed an exploratory analysis over six countries in specific and over the total number of global cases. We find out the global fatal rate, recovered rate per confirmed cases, fatal per recovered/fatal cases. We also estimated the Kernel density distribution of rates calculated. Furthermore, we reported the growth factor for six countries: USA, Italy, Spain, UK, Germany, India. In the Figure 1 we see the top 10 countries with highest number of confirmed COVID-19 cases, death cases, and the timeline of the spread of the novel virus.



(a) Top 10 countries with most confirmed cases

(b) Top 10 countries with most deaths

(c) Spread of COVID-19

(d) Spread of COVID-19 (Map view)

Fig. 1: Confirmed COVID-19 cases till May 05, 2020

Grouping by growth factor: In this section, we group the COVID-19 affected countries into three groups. In first group, the growth factor is larger than 1, in the second group the growth factor is less than 1 and in the third

[4] https://www.kaggle.com/c/covid19-global-forecasting-week-4/data

[5] https://www.kaggle.com/tanuprabhu/population-by-country-2020

group, the growth factor is 1. Let us calculate the growth factor of a country in the following equation

$$\text{Growth Factor} = \frac{\Delta C_n}{\Delta C_{n-1}} \tag{53}$$

where C is the number of confirmed cases. In our project, we have grouped countries based on growth factor as follows:

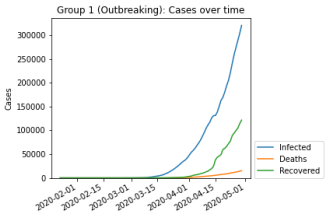
- Outbreaking: growth factor > 1 for the last 7 days
- Stopping: growth factor < 1 for the last 7 days
- At a crossroad: the others

Here, we have listed some of the countries in respective groups based on their growth factor as of May 05, 2020.

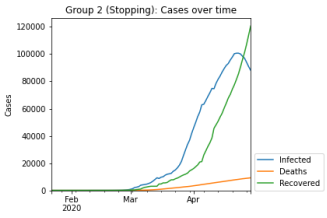
Group A: outbreaking countries: Mexico, Bahrain, Bulgaria, Estonia, Guatemala, Honduras, India, Latvia, Peru, Qatar, Russia, Senegal, Afghanistan, Brazil, Kenya, Saudi Arabia, Bangladesh, Armenia, Mali, Albania, Burma, Congo, Ecuador, Dominican Republic, Egypt, El Salvador, South Africa,

Group B: stopping countries: 'Chad, Equatorial Guinea, Nicaragua, Syria, Zimbabwe, Botswana, Trinidad and Tobago, Cameroon, Andorra, Hong Kong, Nepal, Central African Republic, Iran, Papua New Guinea, Angola, Ghana, West Bank and Gaza, Liberia, Liechtenstein, Benin, Libya, Bhutan, Turkey.'

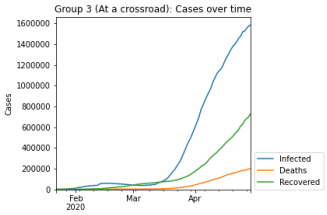
Group C: 'Argentina, Bolivia, Philippines, Uruguay, Venezuela, Cyprus, Eswatini, Georgia, Thailand, Zambia, Algeria, Austria, Bahamas, Bosnia and Herzegovina, Canada, Denmark, Ethiopia, Iraq, Slovenia, Ireland, Niger, Romania, South Korea, China, Colombia, France



(a) Group A: Outbreak



(b) Group B: Stopping



(c) Group C: At crossroads

Fig. 2: Grouping countries based on growth rate

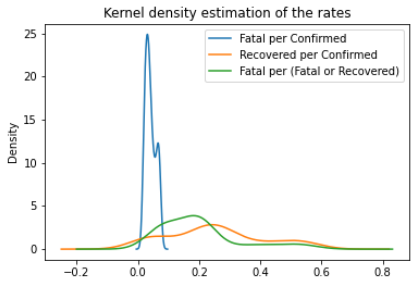


Fig. 5: Kernel density estimation of rates

number of cases over time.png number of cases over time.bb

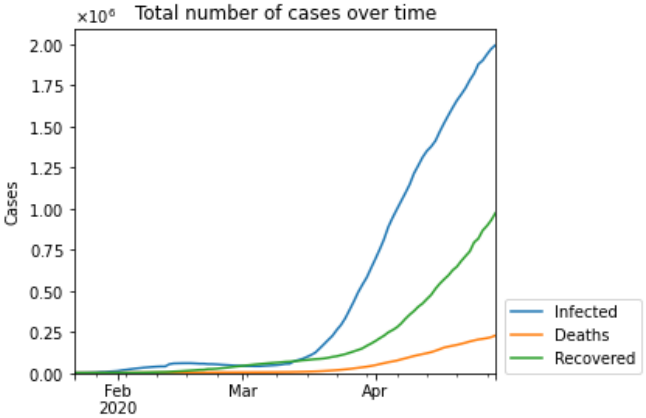


Fig. 3: Total number of cases over time

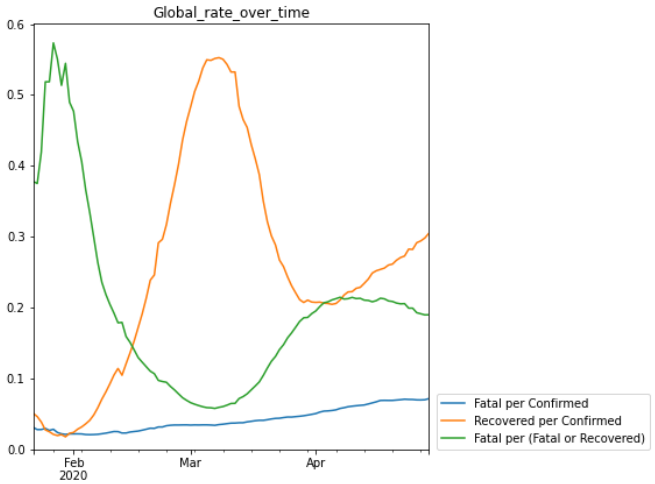
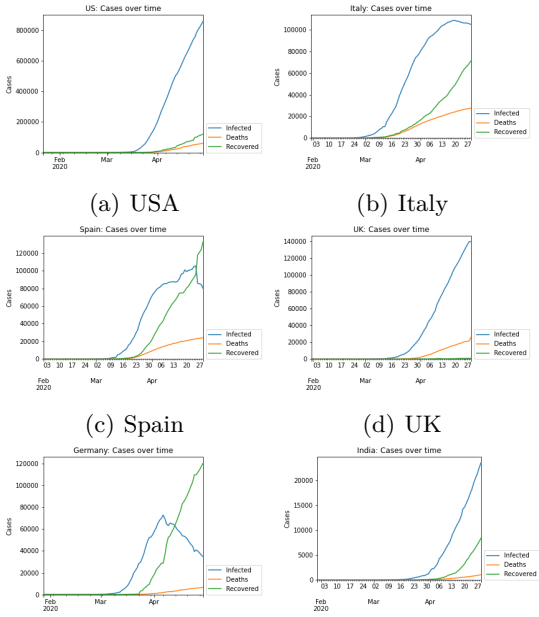


Fig. 4: Global rate over time

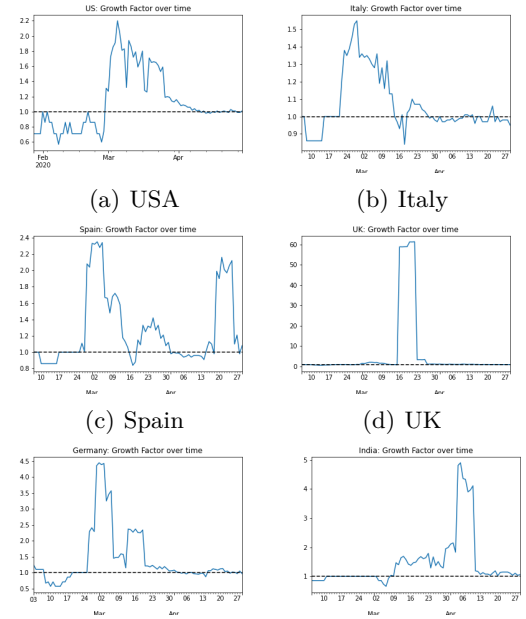


(a) USA (b) Italy

(c) Spain (d) UK

(e) Germany (f) India

Fig. 6: Infections, deaths, and recovery cases over different countries



(a) USA (b) Italy

(c) Spain (d) UK

(e) Germany (f) India

Fig. 7: Growth factor of different countries

A. Data Clustering

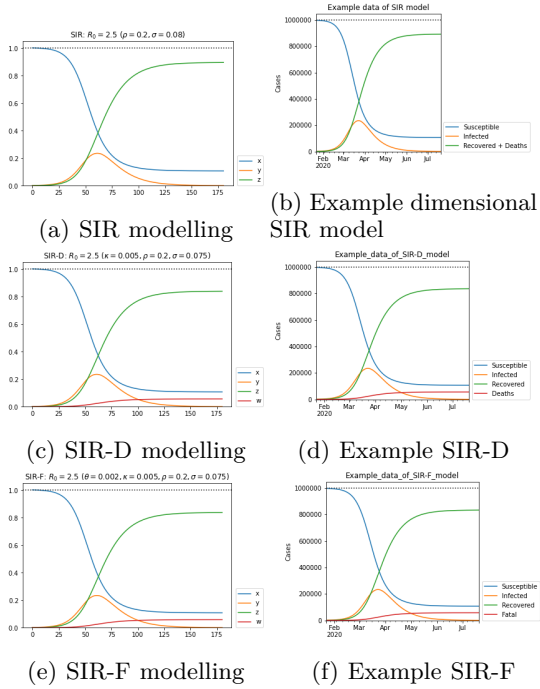
Here, we cluster the mitigation measures taken by countries regarding the covid outfall in their respective countries as provided by the dataset on mitigation measures. We clustered the sentences using KMeans algorithm into 10 clusters of words grouped together in the similar contextual measures as similar implementation measures to handle the outbreak in respective countries. The following table I lists the words clusters. To our investigation, we found out measures like testing, isolation, school closure, travel ban, mask use, public outdoor activity limitations, monetary support to businesses and people were mostly implemented by the affected countries.

V. Epidemic modelling

In this section, we will model the various epidemic models used in our project using initial defined values and also by dimensionalizing the data. In order to dimensionalize the various models with data, we assumed that the starting date of JHU assuming that start date is the first date of JHU dataset, $\tau = 1440$ [min] and total population $N = 1,000,000$. For example, we set the parameters $R_0 = 2.5, \rho = 0.2$ and initial values $(x_{(0)}, y_{(0)}, z_{(0)}) = (0.999, 0.001, 0)$ in SIR model.

SIR-D: For example, we set $R_0 = 2.5, \kappa = 0.005, \rho = 0.2$ and initial values $(x_{(0)}, y_{(0)}, z_{(0)}, w_{(0)}) = (0.999, 0.001, 0, 0)$.

SIR-F: For example, we set $R_0 = 2.5, \theta = 0.002, \kappa = 0.005, \rho = 0.2$ and initial values $(x_{(0)}, y_{(0)}, z_{(0)}, w_{(0)}) = (0.999, 0.001, 0, 0)$.



(a) SIR modelling (b) Example dimensional SIR model

(c) SIR-D modelling (d) Example SIR-D

(e) SIR-F modelling (f) Example SIR-F

Fig. 8: Epidemic Modelling

A. Hyperparameter optimization

In the previous section, we modeled the number of cases based on hypothesized parameter values. However, we do not know parameter values of the actual data. Here, we will estimate the $(\theta, \kappa, \rho, \sigma)$ values of the example data using hyperparameter optimization method by Optuna

Cluster	Measures (Words in clusters)
1	first case, symptom hotline, visa suspension, army deployed, sewing plain clothes masks, prisoner release, prohibition of medicine export
2	school closure, university closure, public green closure, closure of gathering places, airport closure, closure nonessential stores
3	outbound traveller ban, international traveller screening - all countries, international traveller quarantine- risk countries
4	court cancellation, activity cancellation, religious activity cancellation, very large event cancellation or postponement, sports cancellation
5	isolation allowance, compulsory isolation, social isolation, contact tracing, contact isolation - symptoms, contact isolation - no symptoms
6	test inpatients, test contacts, contacts traced total, test cohorts, test travellers, test symptomatic, test vulnerable
7	remote schooling, remote cultural content, mandated remote work, remote work, remote medical treatment
8	blanket isolation - symptoms, cluster isolation - no symptoms, cohort isolation - no symptoms, blanket announcement, blanket curfew
9	testing planned, testing numbers total, expansion of testing facilities, testing, end of testing, testing commenced, testing criteria
10	public statement in support of resuming business, essential public service, public facility cleaning, public mask and hygiene supply

TABLE I: Mitigation measures word clusters

package. We found the estimates of the following parameters using SIR-F model given τ be fixed as 1440 [min], and N=1000000. The estimates are as follows:

θ = 0.019857 κ =0.004552 ρ =0.203097, $sigma$ =0.075422 τ=1440 $R0$ =2.49, RMSLE=0.175833, α_1= 0.02, $1/alpha_2$ =219, $1/\beta$ =4, $1/\gamma$= 13

Root Mean Squared Log Error (RMSLE) score is given in the equation below:

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(log_{10}(A_i+1)-log_{10}(P_i+1))^2} \qquad (54)$$

Where A is observed (actual) values, P is estimated (predicted) values. Variables are S $(i=1)$, I $(i=2)$, R $(i=3)$ and F $(i=n)$ for SIR-F model. When RMSLE score is low, hyperparameter estimation is highly accurate.

Fig 9 shows the comparison of observed values and estimated values of the parameters of SIR-F modeling. The subplots shows the accuracy of each parameter by comparing the observed with the estimated values. The more the "v_observed" and "v_estimated" (v=y, z, w) are overlapping, the higher the accuracy of the estimates observed.

The number of exposed cases in latent period (E) and wating cases for confirmation (W) are un-measurable variables, but key variables as well as S, I, R, F. If E and W are large, outbreak will occur in the near future. Let's replace S $\xrightarrow{\beta I}$ S* with S $\xrightarrow{\beta_1(W+I)}$ E $\xrightarrow{\beta_2}$ W $\xrightarrow{\beta_3}$ S* because W also has infectivity.

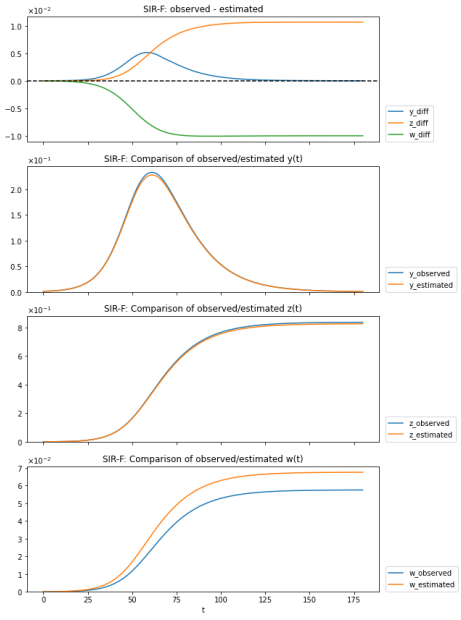


Fig. 9: SIR-F observed estimated parameters in comparison

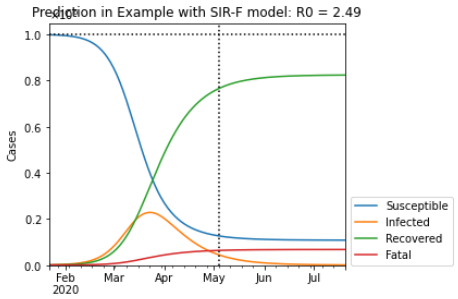


Fig. 10: Prediction of cases in next 6 months. The dotted line marks today.

VI. Factors of model parameters

Let us discuss various control factors of SIR-F modeling [3] that can affect the effective contact rate β_1 Please reconsider S $\xrightarrow{\beta_1(W+I)}$ E formula. Susceptible persons may contact with waiting/confirmed patients, and susceptible persons will be infected with COVID-19. The formula can be replaced with

$$S_q \overset{g_s}{\Longleftrightarrow} S_g \xrightarrow{f_1} \quad E^* \xrightarrow{e^{-h_2}} E \qquad (55)$$

$$E^* \xrightarrow{1-e^{-h_2}} R^* \qquad (56)$$

$$W_q \overset{g_w}{\Longleftrightarrow} W_g \qquad (57)$$

$$I_q \overset{g_i}{\Longleftrightarrow} I_g \qquad (58)$$

$$I_q \xrightarrow{q} I_{\hat{q}} \qquad (59)$$

$$(60)$$

whereas the following describes the equation variables:

S_q: Susceptible persons with self-quarantine

S_g: Susceptible persons with human contacts such as family members or friends, etc.

W_q: Patients who are waiting with self-quarantine

W_g: Waiting patients with family members or friends, etc.

I_q: Confirmed and un-recovered patients with self-quarantine

I_g: Confirmed and un-recovered patients with human contacts such as family members or friends, etc.

$I_{\hat{q}}$: Confirmed and un-recovered patients who were hospitalized

E^*: Just after being exposed to the virus

R^*: Being exposed to the virus, fought against the virus, recovered and immunes without confirmation

$$f_1 = v(W_g + I_g)(1-m)^2(1-w_e)^{w_n}e^{-h_1}sc \tag{61}$$

The control factors are as follows:

g_s: is the number of days in a week when a susceptible person goes out [day]

g_w: is the number of days in a week when waiting patient but un-quarantined goes out [day]

g_i: is the number of days in a week currently infected (confirmed) but un-quarantined persons go out [day]

q: is the quarantine rate of currently infected (confirmed) patients

v: is the probability of virus existence in a droplet

m: represents the rate of persons wearing masks effectively (depends logistically on supply of masks)

w_e: measures the virus reduction effect resulted due to regularly washing hands

w_n: is the number of times people washes their hands before touching their own faces after going out

h_1: Health condition (active rate of cellular immunity factors) of susceptible and contacted persons

h_2: Health condition (active rate of humoral immunity factors) of susceptible and contacted persons

c: is the number of contacts between susceptible persons and patients while on the go in a minute (depends on population density) [1/min]

δ: is the product of unknown real factors

The parameter β_1 in the math model can be calculated as follows:

$$\beta_1 = \frac{1}{49}[g_s\{g_w + g_i(1-q)\}v(1-m)^2(1-w_e)^{w_n}e^{-(h_1+h_2)}c\delta] \tag{62}$$

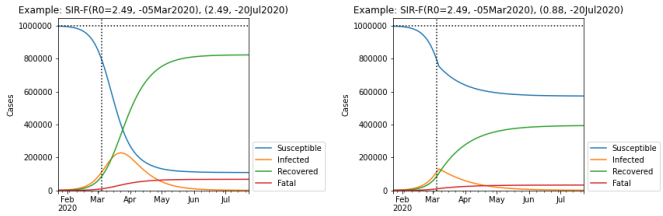
As an example, we will estimate the impact of lockdown. This means the control parameter g_s which measures the number of days a susceptible person goes out in a week will be minimized as an impact of lockdown. We grouped people “going out for” in 3 sections: school, work,and others. We can calculate weighted average of days with age composition of population from Population pyramid global dataset.

The action under consideration: If all schools and offices will be closed, g_s can be reduced. People will go out one day for other reasons instead of going to school/office.

Before the action of lockdown was taken, $g_s = 5.921853248424339$ and after the lockdown implementation, the value of g_s drops tp 2.082265659970043. The ratio β_after / $\beta_before = 0.35162398874441597$.

Impact of actions on β: In SIR-F model g_s is a control factor of β. Actions taken at 30th day: All schools and offices will be closed.

Predict the number of cases: with actions from 30th day: There is a delay between the time point of starting actions and that of appearing the effect. Because I is the main variable, the length of delay can be estimated as sum of latent period and waiting time for confirmation. This value [day] was calculated in ”SIR-F with exposed/waiting cases” section. Fig 11 shows the impact of lockdown



(a) If lockdown is not implemented on March 05, 2020

(b) If lockdown is implemented on March 05, 2020

Fig. 11: Impact of actions: lockdown implementation

on SIR-F modelling for future. The actions of lockdown results in:

- Total number of confirmed cases was decreased.
- Peak point of infected cases was delayed.

This basically shows that we need to fight with the virus for longer period of time. If there is no lockdown then we can see how the infections

VII. Control factors of recovery rate γ and mortality rate α_2

In this section, we reconsider the equation which determines the relationship between the infections and recovery, and infections to fatalities. $I \xrightarrow{\gamma} R$ and $I \xrightarrow{\alpha_2} F$. As the recovery of patient is dependent on the balance of immunity, the virulence, the effects of treatments. The above relationships can be replaced with: ,

$$I \xrightarrow{\bar{h}} I^\star \xrightarrow{\bar{s}} F^\star \xrightarrow{L^{-1}} F \tag{63}$$

$$I \xrightarrow{f_2} R^\star \xrightarrow{l^{-1}} R \tag{64}$$

$I^\star$: is the number of confirmed cases whose immune systems did not overcome virus multiplication, and without severe events

$F^\star$: is the number of confirmed cases whose immune systems did not overcome virus multiplication, and with severe events

R*: is the number of confirmed cases whose immune systems overcame virus multiplication or comfirmed cases whose severe events can be stopped

Where $f_2 = 1 - \bar{h}\ \bar{s}$

$\bar{h}$: is the rate of I whose immune systems do not overcame virus multiplication

$\bar{s}$: is the rate of I* who have severe events, including respiratory failure

L_i: is the inverse of F*'s mortality rate for people of i years old [min]

l_i: is the inverse of R*'s mortality rate for people of i years old [min]

P_i: is the number of people who are i years old

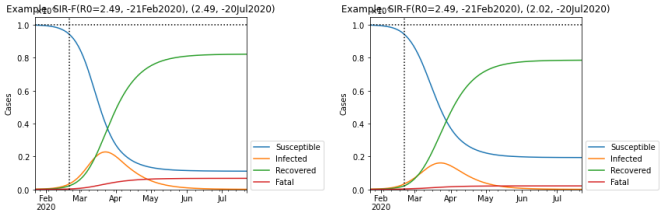
N: represents the total population

The mortality rate α_2 and the recovery rate γ can be formulated as:

$$\alpha_2 = \frac{\bar{h}\ \bar{s}}{N} \sum_{n=0}^{\infty} \frac{P_i}{L_i} \tag{65}$$

$$\gamma = \frac{1 - \bar{h}\ \bar{s}}{N} \sum_{n=0}^{\infty} \frac{P_i}{l_i} \tag{66}$$

Initially, we assume that $\bar{h} = 0.5$ and $\bar{s} = 0.5$. (Using population distribution data and case reports, $\bar{h}\ \bar{s}$ and $1 - \bar{h}\ \bar{s}$ can be calculated.)



(a) If medicine is not available (b) If medicine is available

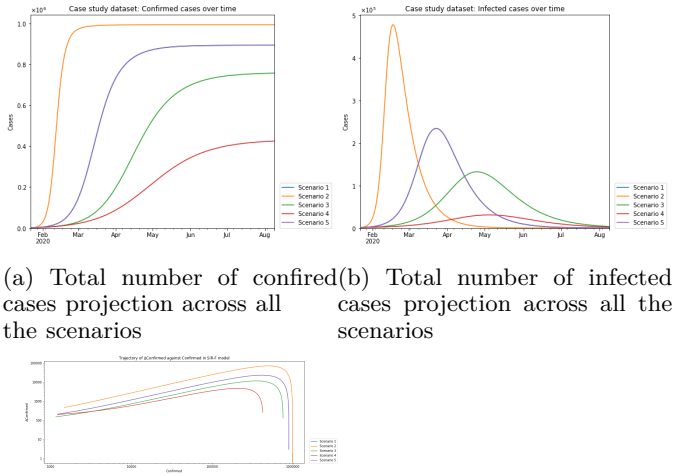
Fig. 12: Impact of actions: Availablity of medicines

VIII. Case studies analysis

In the previous section, we found that parameter values can be changed by actions. To predict the future, we need to recognize the parameter change from the actual records. Here, trend analysis method will be introduced. With the same initial values $(x_{(0)}, y_{(0)}, z_{(0)}, w_{(0)}) = (0.999, 0.001, 0, 0)$, we will create five SIR-F example datasets.

- Scenario 1: $(\theta, \kappa, \rho, \sigma) = (0.0002, 0.005, 0.20, 0.075)$
- Scenario 2: $(\theta, \kappa, \rho, \sigma) = (0.0002, 0.005, \underline{0.40}, 0.075)$ spread quickly
- Scenario 3: $(\theta, \kappa, \rho, \sigma) = (0.0002, 0.005, \underline{0.15}, 0.075)$, spread slowly
- Scenario 4: $(\theta, \kappa, \rho, \sigma) = (0.0002, \underline{0.003}, 0.20, \underline{0.150})$, improved health-care system
- Scenario 5: $(\theta, \kappa, \rho, \sigma) = (\underline{0.0000}, 0.005, 0.20, 0.075)$, equal to SIR-D model

Values are dimensionalized with total population $N = 1,000,000$ in the case studies example datasets. Fig. 13 draws the confirmed, infected cases projections for a period of 6 months based on different scenarios as discussed above. The fig 13 also plots the trajectory of the increase of confirmed cases against the confirmed cases for all the scenarios.



(a) Total number of confired cases projection across all the scenarios

(b) Total number of infected cases projection across all the scenarios

(c) Trajectory of ΔConfirmed against Confirmed in SIR-F model

Fig. 13: Case study analysis

A. ΔConfirmed vs. Confirmed in log-log plot

The number of new confirmed cases or the increase in the number of new cases ΔC can be desribed as

$$\Delta C = N^{-1}\beta(N - C)I \tag{67}$$

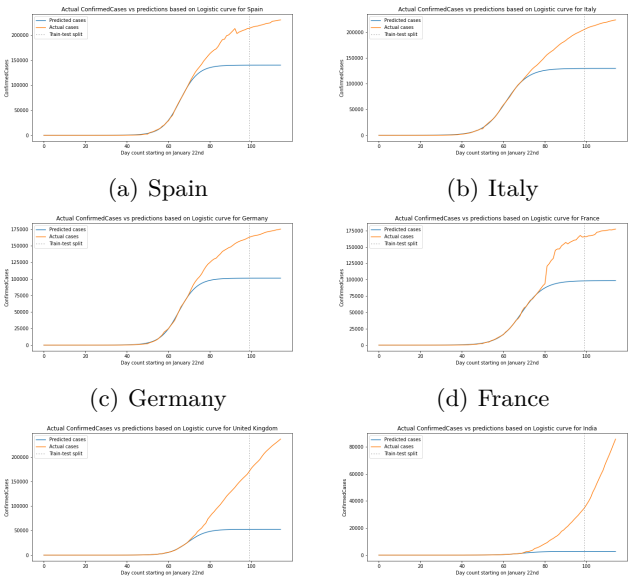
This is because $C = I + R + F$ and $S = N - C$ in SIR-F model. ΔC is determined by cummulative number of cases. In addition,C determines I when the parameters $(\alpha_1, \alpha_2, \beta, \gamma)$ are fixed. Then, $\Delta C = f(C)$ Plots of $(x, y) = (C, \Delta C)$ in log-log scale are shown in the Fig. 13 (c)

B. Curve fitting of $C(t)$

In this section, we will try to fit a curve to the function $C(t)$, which is sometimes described by logistic function and Gompertz function. In general, the fitting effect of Logistic model was found to be better the Gompertz function in this paper [2]. They also used these functions for prediction of the number of new confirmed cases in 2020. Thus, we chose to fit the logistc model to fit $C(t)$ and we also used the logistic model to predict the number of new cases.

$$\text{Logistic function}: g(t) = \frac{N}{1 + Ae^{-Bt}} \tag{68}$$

$$\text{Gompertz function}: h(t) = Ne^{-Ae^{-Bt}} \tag{69}$$



(a) Spain (b) Italy

(c) Germany (d) France

(e) United Kingdom (f) India

Fig. 14: Fitting logistic function and their predictions

We also fit the exponential growth and negative exponential growth function [8] on the case study data. $f(t)$ can be divided into stages;

- exponential growth function ($t \leq \text{argmax } \Delta C(t)$)
- negative exponential function ($otherwise$).

With constant (a, b, A, B, C),

$$f(t) = \begin{cases} ae^{bt} & (t \leq \text{argmax } \Delta C(t)) \\ C - Ae^{-Bt} & (otherwise) \end{cases} \tag{70}$$

Fig 15 shows the curve fitting of exponential and negative exponential growth function to the case scenarios. In Fig. 15 we also see some errors. However, these errors were found for curve fitting. This is because

$$\frac{\mathrm{d}C}{\mathrm{d}T} = \frac{\beta}{N} SI$$

$S \simeq S(0) := const.$ for $t \leq \text{argmax } \Delta C(t))$, but I is not proportinal to S in SIR-like model.

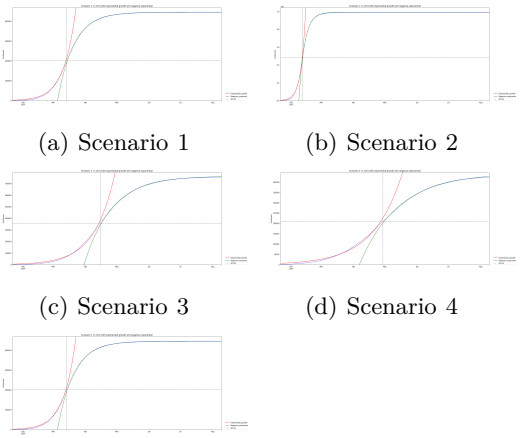
This implies that we cannot convert the differential equation to the following set of equations.

$$\frac{\mathrm{d}x}{\mathrm{d}t} = Bx \tag{71}$$

$$\text{i.e. } x(t) = Ae^{Bt} \tag{72}$$

C. Susceptible-Recovery plane

In this section, we discuss about the relationship between Susceptible and Recovery and draw a plot to show the trajectory of susceptible against the recovery cases in



(a) Scenario 1 (b) Scenario 2

(c) Scenario 3 (d) Scenario 4

(e) Scenario 5

Fig. 15: $(t, C(t))$ with exponential growth and negative exponential

SIR-F modeling of the case scenarios. According to SIR-F model, we have

$$\frac{\mathrm{d}S}{\mathrm{d}T} = -N^{-1}\beta SI \tag{73}$$

$$\frac{\mathrm{d}R}{\mathrm{d}T} = \gamma I \tag{74}$$

$$I > 0 \tag{75}$$

Then, we can derive

$$\frac{\mathrm{d}S}{\mathrm{d}R} = -\frac{\beta}{N\gamma} S \tag{76}$$

This leads to

$$S(R) = N \exp^{\frac{-R\beta}{N\gamma}} \tag{77}$$

$$\log(S(R)) = \log(N) - \frac{R\beta}{N\gamma} \tag{78}$$

With constant $a = \frac{\beta}{N\gamma}$ and constant $b = \log N$, The model can be reframed as follows:

$$\log S_{(R)} = -aR + b$$

Because $R(t)$ is a cummurative number, $R(t + \Delta t) \geq R(t)$ for all t and $\Delta t > 0$. Thus, slope of $\log S_{(R)}$ will be changed when SIR-F parameters are changed. We need to split the actual data, considering the change points of S-R line in log-scale. This logic will be used for actual data in scenario analysis section.

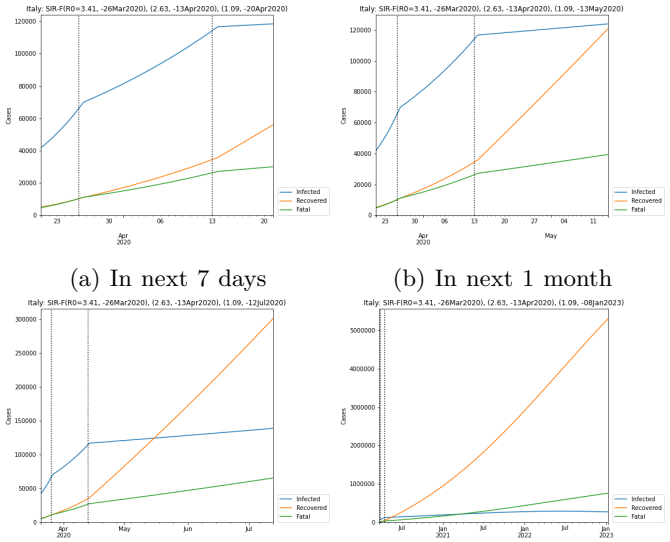
IX. Predictions

In this section, we predict the number of confirmed cases through logistic curve fitting for countries like India, UK, Spain, Italy, France, Germany. We also use predictions from SIR-F modeling using parameter estimation and also predict the number of infections, recoveries, and fatals for next 7 days, 1 month, 3 months, and next 3 years for USA and India. We also show the SIR-F graph to project the peak of the infections.

A. SIR-F modelling

Here, we show the predictions for the following countries from SIR-F model.

1) Italy: In the figure 16 we show the predictions for Italy. Italy was the first country outside Wuhan province in China to be the most affected by the virus. Thus, we have considered 3 phases of transmission of the virus epidemic in Italy. We observed that the infection curve grows flatten after mid April which reflects that the number of infections don't increase at dramatic rate as before. This may be owing to the strict national lockdown in Italy and we see that the recoveries count intersects the infection count after third week in May and the recoveries count increases steadily afterwards as shown in the subgraph (c). After July 2020, we notice that the margin between recoveries and infections going large which gives us the positive sign that the pandemic outbreak is almost over in Italy. In the subgraph (d) we see that the number of fatalities surpasses the new infection cases in May 2021. This may be due to the deaths caused by previous positive infections. The number of casualties is expected to rise to around 50000 by the end of July 2020, and may reach to 200000 deaths by the end of January 2023.

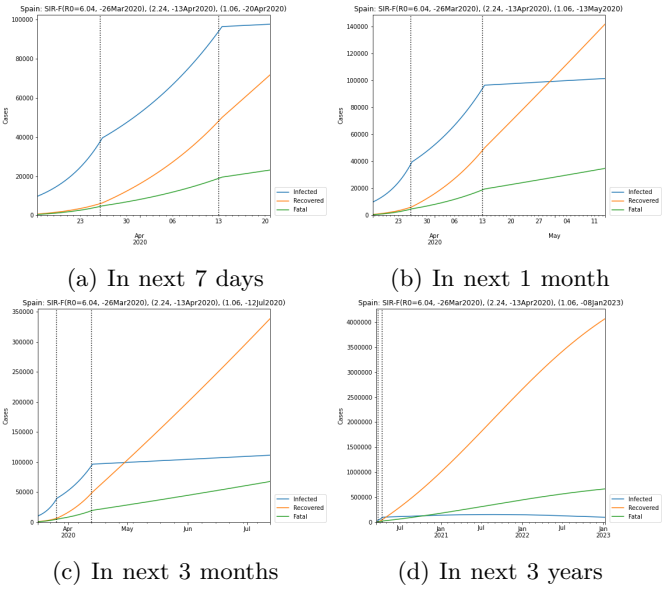


(a) In next 7 days (b) In next 1 month

(c) In next 3 months (d) In next 3 years

Fig. 16: SIR-F predictions for Italy

2) Spain: In Fig. 17, we forecast about the situation of Spain. We have considered 3 phases of outbreak in Spain. We see that in May 2020 onward, the recoveries count increases steadily and surpasses the new infection count as shown in the subgraph (c). From November 2020 onward the infection count goes down then the number of deaths and we see that the infection count decreases with a negative slope till the end of January 2023. The number of fatalities is expected to be around 50000 and the count of infections to be around 100000 by the end of July 2020 and the fatalities count to be around 500000 by the end of January 2023. From subgraph (d), we can observe that the infection graph is decreasing whereas the recovery curve is increasing.



(a) In next 7 days (b) In next 1 month

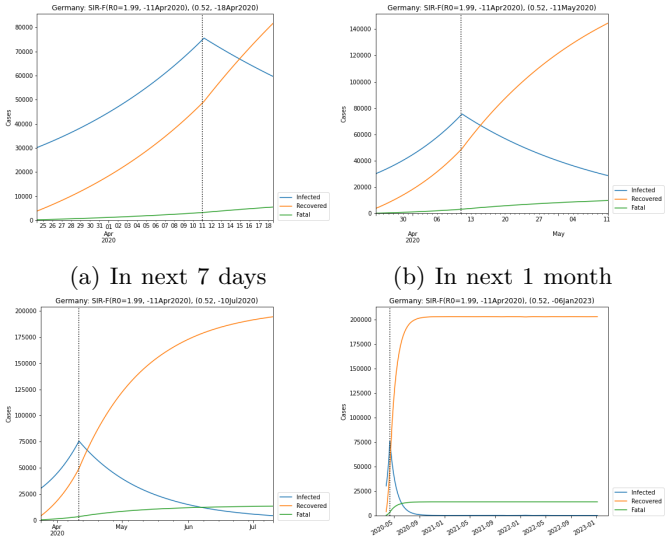
(c) In next 3 months (d) In next 3 years

Fig. 17: SIR-F predictions for Spain

3) Germany: In Fig. 18, we forecast about the condition of the outbreak of the virus in Germany. We see that the peak number of confirmed cases in Germany will be around 75000 and then decreases onward and the recoveries count increases exponentially as shown in the subgraph (c). The number of casualties is expected to be around 10000 till the end of July 2020. In the subgraph(d), it is evident that the fatalities number will almost become flat at the end of January 2023 with at most 15000 deaths. It is also nice to see how quickly the infection curve drops from the peak in Germany by September 2020. This shows us that Germany is the only country with the least casualties among other countries in our investigation. This may be due to strict measures adopted by the German authorities and rapid testing strategies and isolating the infected from the public in order to slow the growth of the pandemic unlike other countries.

4) France: In Fig. 19, we forecast about the virus outbreak and its projections in France. In subgraph (c), we see that the recoveries count surpasses the infections count after mid of June 2020 onward but we see that the fatalities count increases steadily from May to July 2020 at more than a double rate. The infection cases reaches around 200,000 after mid of July onward. The fatalities count can rise to more than 100000 by the end of July 2020 and may touch 5 million by January 2023.

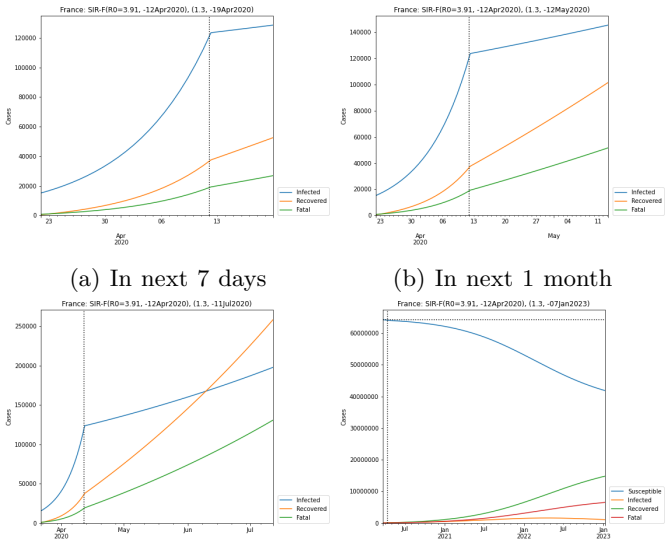
5) USA: In Fig. 20, we study about the SIRF modelling and its projections regarding the coronavirus outbreak in America. We see that the infection cases can rise upto 1 million cases till the mid of May as shown in the the subgraph (b). As shown in the subgraph (c), the number



(a) In next 7 days (b) In next 1 month

(c) In next 3 months (d) In next 3 years

Fig. 18: SIR-F predictions for Germany



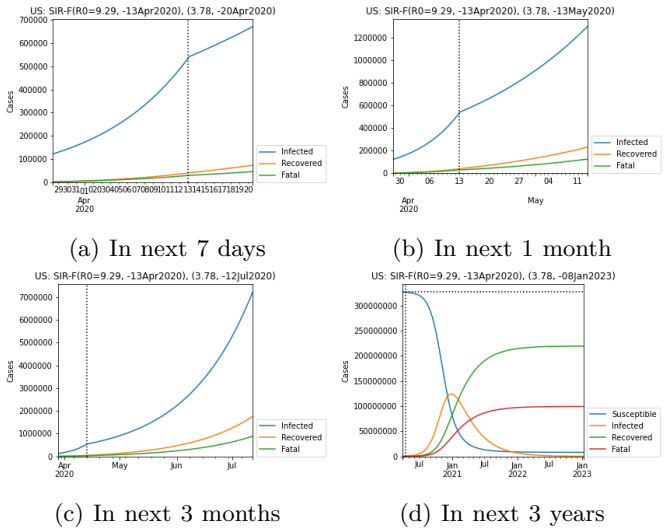
(a) In next 7 days (b) In next 1 month

(c) In next 3 months (d) In next 3 years

Fig. 19: SIR-F predictions for France

of cases is expected to reach 7 million by the end of July. As shown in the subgraph (d), the peak of the confirmed cases is expected to reach 125 million in USA by January 2021 and then the growth of new confirmed cases decreases almost at the same rate as it was increasing. After March 2021, the recovery curve surpasses the infections curve. We also noticed that the number of deaths may rise to around 500000 by the end of July and more than 50 million deaths by the end of January 2023.

6) India: In fig 21 we project the number of confirmed cases, casualties, recoveries till the end of January 2023 for the most populous and densely populated democracy in the world. The infection cases increases till the end

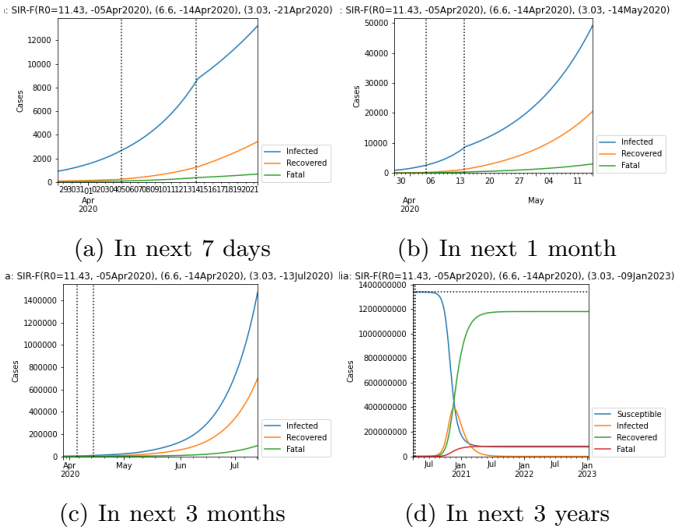


(a) In next 7 days (b) In next 1 month

(c) In next 3 months (d) In next 3 years

Fig. 20: SIR-F predictions for USA

of July as shown in the subgraph(c). Owing to a huge population of 1.3 billion people, the number of confirmed cases can reach upto 1.4 million by the end of July 2020. The peak of infections will go down after December 2020. It is expected that the number of confirmed cases can rise till 400 million in India by January 2021 and then the cases decreases considerably till January 2023. The causalities can be around 30000 till the end of July 2020 and can be as high as 40 million deaths in mid 2021. It is highly likely that India will not encounter such higher number of deaths because of the relatively younger population as the median age of India is 27 years. This can be inferred as it has been seen in various research and statistics that COVID-19 is comparatively less lethal to the younger people than the older ones and most of the casualties resulted in the deaths of the older people who are above 60 years old.



(a) In next 7 days (b) In next 1 month

(c) In next 3 months (d) In next 3 years

Fig. 21: SIR-F predictions for India

B. Machine learning

In this section, we used machine learning algorithms and neural network approach to predict the number of covid-19 infections in future. As the transmission progresses, the exponential regime is more suitable than a normal regression model. We found out that ordinary least square regression has resulted into worst predictions and highest mean squared error. We were aware of this limitation, and now alternative methods such as polynomial regression, support vector machine regression and linear perceptron neural network were tried out in order to capture the trajectory behavior.

Models considered in this section:

1) Support Vector Regression:: The Support Vector Regression (SVR) [5] uses the same principles as the SVM for classification, with only a few minor differences. In the case of regression, a margin of tolerance (epsilon) is set in approximation to the SVM which would have already requested from the problem. However, the main idea is always the same: to minimize error, individualizing the hyperplane which maximizes the margin, keeping in mind that part of the error is tolerated. We fit the non-linear polynomial kernel of degree 8 to transform the data into higher dimensional feature space in order to perform a linear separation. We set the following parameters: shrinking=True, kernel='poly', the co-efficient of the polynomial kernel, gamma=0.01, epsilon=1, degree=8, C=0.1 in SVR() function of Sklearn library. Here, epsilon specifies the epsilon-tube within which no penalty is associated in the training loss function with points predicted within a distance epsilon from the actual value.

2) Polynomial Regression: In statistics, polynomial regression [6] is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modeled as an nth degree polynomial in x. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted $E(y|x)$. Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear, in the sense that the regression function $E(y|x)$ is linear in the unknown parameters that are estimated from the data. For this reason, polynomial regression is considered to be a special case of multiple linear regression.

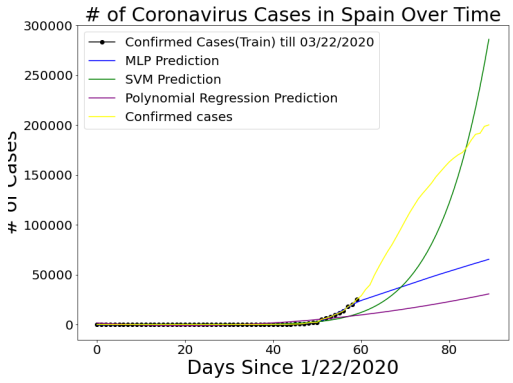
3) Multi-layer perceptron: A multilayer perceptron (MLP) [7] is a class of feedforward artificial neural network (ANN). An MLP has at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function to fire the neuron. MLP gets trained by backpropagation technique. The multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can approximate non linear distribution of data. In our MLP model, the input layer has the equal number of training data points (i.e, the number of days used for training purpose). In our case, the input layer has 60 neurons and there are two hidden layers comprising the half of input neurons, i.e, 30 neurons each in two hidden layers in our model. The output layer has neurons equal to the number of forecast days, 30 in our cases as we have prediced the number of confirmed cases for 30 days in future.

C. Analysis

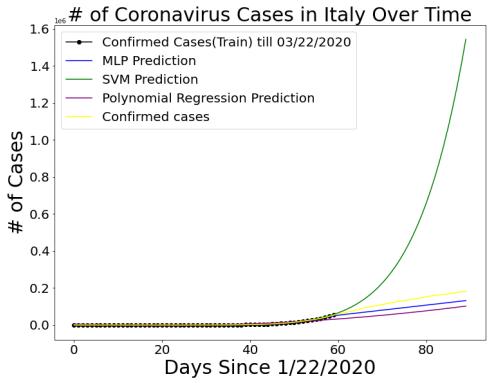
In the Fig. 22 we visualize the confirmed cases predictions of the various machine learning techniques against the actual number of cases in respective countries in the measured time frame. We have trained our model on two months data from 22/01/20 to 22/03/20 and then we have predicted for next 30days. In Spain, we can see from the plot that the real trajectory is closer to the SVR prediction rather than the polynomial regression of degree 8 and MLP prediction. In Italy, we see that the growth trajectory is well interpolated with more accurate predictions by the MLP model and polynomial regression model rather than the SVR predictions. In Germany,predictions are more accurate for MLP than the SVR and the polynomial regression respectively. However, the real growth trajectory is much faster than exponential growth. For France, the SVR prediction has been more accurate in approximating the real growth of cases trajectory whereas both the polynomial regression, and the MLP models have large MSE error as shown in the plot. For UK, we see that the SVR predictions have been the most accurate and thus closer to the real trajectory but the MLP and polynomial regression have resulted in very high MSE error. Following UK, we notice that our models have completely failed to learn the approximation function for USA. Thus, we see the highest MSE for the all the models used to model the growth distribution of North America. We can also see that the number of cases in USA are like 10 times more than any other country in that time frame. It is interesting to see that the SVR model has a negative exponential growth trajectory. India shows the least growth of confirmed cases in the given time-frame. It may be due to the strict lockdown implementation throughout India since 23rd March, 2020. At last, we predicted the total number of confirmed cases for the entire world. We see that SVM model overfits after 15 days from the 60th day onward. This is the only example where we see that our model overestimates the growth trajectory of the covid19 pandemic.

X. Conclusion

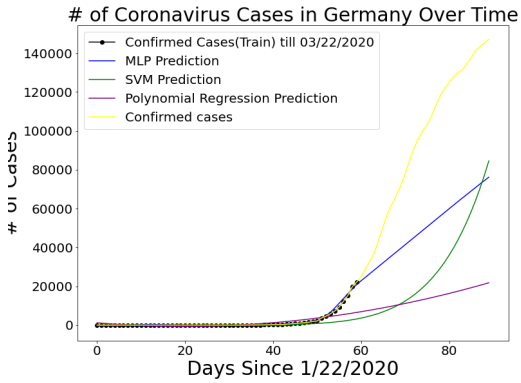
In this paper, we introduce the mathematical epidemic models to simulate the behavior of novel coronavirus (COVID-19) that emerged in Wuhan province in China mainland. We used models like SIR, SIR-F, SIRF-D models to simulate the impact of the epidemic and the prediction in future. We dealt with some case scenarios depending on various parameters influence. We also extended the SIR-F model with controlled parameters as to vividly



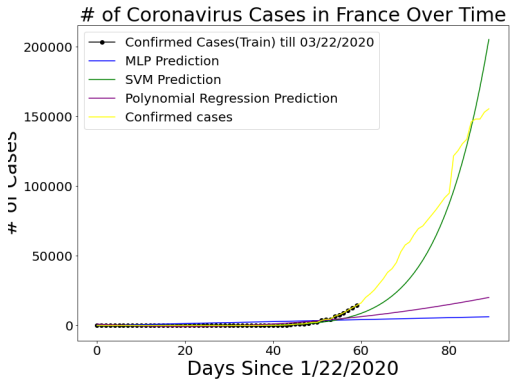
(a) Spain



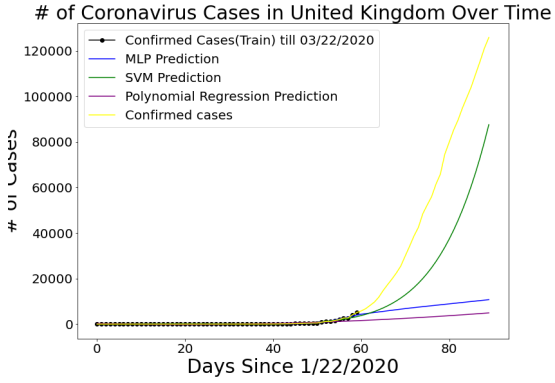
(b) Italy



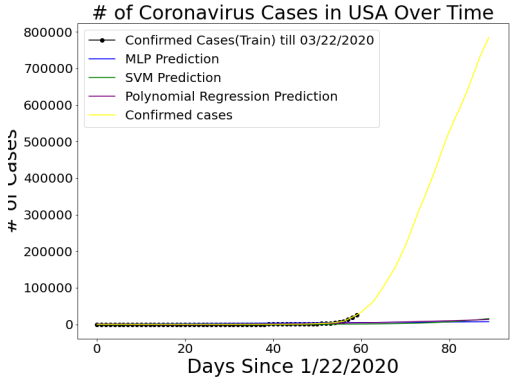
(c) Germany



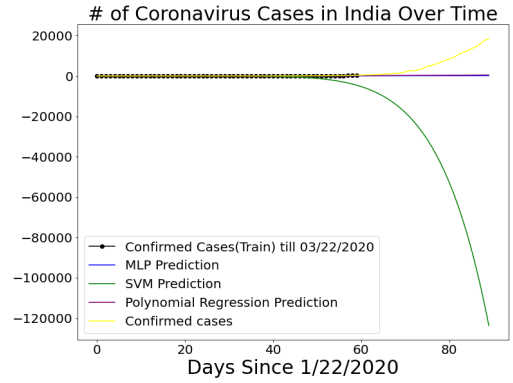
(d) France



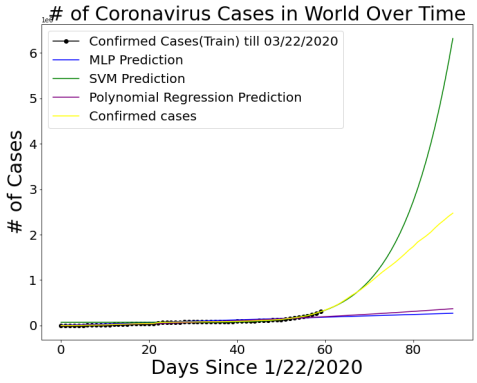
(e) United Kingdom



(f) USA



(g) India



(h) World

Fig. 22: Machine learning and neural network predictions in different countries. ML models: SVM Regression, Polynomial regression, and MLP neural model

make the math model more realistic by defining more control parameters. We also discuss the effect of lockdown implementation and what happens with the availability of vaccination in future. We found out that the lockdown implementation decreases the infection rate in a country to a large extent. We also vividly discussed about the SIR-F modeling and future projections as to predict the number of infections, recoveries, and deaths in a time-frame of 7 days, 1 month, 3 months, and after 3 years respectively in future for specific countries such as USA, UK, Spain, Italy, Germany, and India. We also used curve fitting technique to understand the trajectory pattern of the spread of the pandemic. We fit both exponential and logistic models to $C(t)$, the number of cases function. We selected machine learning algorithms such as logistic model to predict the number of cases in future. We found out that the predictions by the logistic model was underreported, i.e, the actual trajectory is more complex than the logistic model. Furthermore, we also implemented machine learning algorithms such as complex polynomial regression, support vector machine regression algorithms to represent the complex pattern of the data and use the model to simulate the effect of the pandemic in future in terms of virus proliferation and growth trajectory. We have observed that not a single machine learning model can be generalized for every country rather different machine learning model were found to be effective in equating the complex pattern of the pandemic for different countries. We see that the number of infections can rise upto 400 million in India and 125 million in USA by January 2021. Among European countries, we see that France with the most number of casualties (around 5 million) by January 2023 and Germany with the least number of fatalities. We also observed that the peak of the pandemic is almost over in Italy by July 2020. In future, we would like to extensively use the deep learning models such as recurrent neural networks to represent complex pattern of the pandemic time series data. We would also like to use time series prediction model such as Autoregressive Integrated Moving Average model (ARIMA), Exponential Smoothing method (ES), Grey Model (GM), Markov chain method (MC), etc. The ARIMA prediction model, which uses several differences to make it a stationary series, and then represent this sequence as a combination auto-regression about the sequence up to a certain point in the past.

References

[1] Ke Wu1, Didier Darcet, Qian Wang2,and Didier Sornette, "Generalized logistic growth modeling of the COVID-19 outbreak in 29 provinces in China and in the rest of the world,", CoRR abs/arXiv:2003.05681.

[2] Lin Jia, Kewen Li, Yu Jiang, and Xin Guo1 Ting zhao, "Prediction and analysis of Coronavirus Disease 2019,", CoRR abs/arXiv:2003.05447.

[3] Nakul Chitnis, "Introduction to SEIR Models,", http://indico.ictp.it/event/7960/session/3/contribution/19/material/slides/0.pdf [Accessed on 15th March, 2020]

[4] Balkew, Teshome Mogessie, "The SIR Model When S(t) is a Multi-Exponential Function," (2010). Electronic Theses and Dissertations. Paper 1747. https://dc.etsu.edu/etd/1747

[5] Mathworks, "Understanding Support Vector Machine Regression," 2020, accessed 25th March, 2020, https://de.mathworks.com/help/stats/understanding-support-vector-machine-regression.html.

[6] "Polynomial Regression," in Applied Regression Analysis 2020, The Pennsylvania State University, accessed on 17th March, 2020. https://online.stat.psu.edu/stat462/node/158/.

[7] Martin Riedmiller, Machine Learning: Multi Layer Perceptrons, accessed on 30th March, 2020, http://ml.informatik.uni-freiburg.de/former/_media/documents/teaching/ss09/ml/mlps.pdf.

[8] Chowell G, Sattenspiel L, Bansal S, Viboud C. Mathematical models to characterize early epidemic growth: A review. Phys Life Rev. 2016;18:66-97. doi:10.1016/j.plrev.2016.07.005