Prediction of COVID-19 cases in India using SIR Model

Suman Kumar
Great Learning
PGP-AIML
Pune, India
sumankumarsinha@yahoo.com

Bhavik Parikh
Great Learning
PGP-AIML
Pune, India
bhavikpar@gmail.com

Tushar Khoche
Great Learning
PGP-AIML
Pune, India
tusharkhoche@gmail.com

Atul Hatgaonkar
Great Learning
PGP-AIML
Pune, India
ahatgoankar@gmail.com

***Abstract*—Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered corona virus. COVID-19 pandemic spread has impacted the whole world with several other consequences such as loss of lives, economic slowdowns, etc. Various measures are taken by the governments of different countries across the globe to control the spread. Some of the measures like social distancing, lock-down, mandatory use of personal protective equipment are to name a few. COVID-19 has also taken the medical fraternity by surprise as standard protocol, medicines for treatment did not exist.**

This study aims to find the rise of COVID-19 cases in India, its peak, and when it would flatten. The study also focuses on the demand of the medical system (such as beds, oxygen, etc.) to handle the crisis.

Keywords—Covid-19, SIR, Covid Prediction, Covid Deaths, Reproduction Rate

I. Introduction

The COVID-19 an acronym for ``Coronavirus Disease-2019", is a respiratory illness caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), a contagious virus belonging to a family of single-stranded, positive-sense RNA viruses known as coronaviridae [1].

The COVID-19 outbreak came to light on 31 December 2019 when 27 cases of pneumonia of unknown etiology were reported at the WHO's country office in China.

For the entire timeline of events, kindly refer to Fig. 1 The epicenter of the outbreak was linked to Wuhan's wholesale market for seafood and other exotic animals, including snakes, bats, and marmots [30]. A new strain of a highly contagious β-coronavirus, SARS-CoV-2, has been deemed responsible for the rapid outbreak of COVID-19. Distinguishing characteristics of the virus include its extremely contagious nature and relatively long (1-14 days) incubation period. During this period, a person can be infected by the virus and not show any symptoms at all.

Therefore, people infected with the disease may unknowingly serve as silent carriers of the virus, contributing to a high basic reproductive number for the COVID-19 virus. While some studies indicate that SARS-CoV-2 could be susceptible to heat and ultraviolet (UV) light [31], there is no specific treatment for the infection to date, and the management protocols for the disease are still evolving.

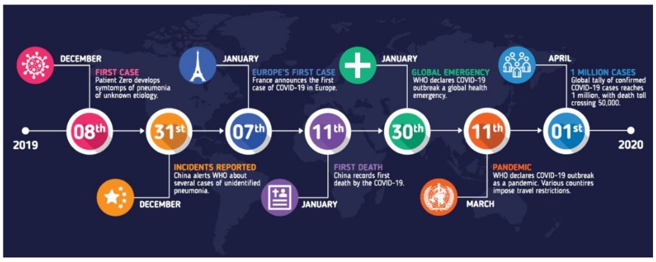

Fig. 1. A timeline of the COVID-19 pandemic origin

The world is still in the midst of COVID-19 pandemic. India isn't unaffected [Fig. 2] and the current trend is showing a formation of a pattern of the second wave (with a downward trend) in the country. The growth of COVID-19 has forced the government, medical fraternity, scientists to develop measures to stop the outbreak. Several measures were proposed, implemented to contain the consequences of the pandemic [1].

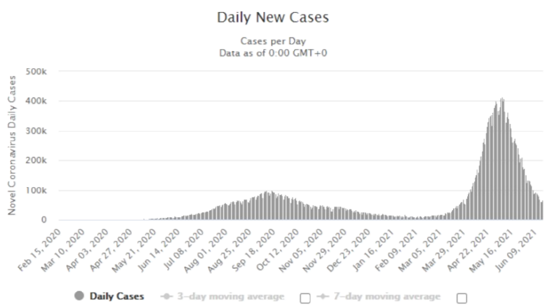

Fig. 2. Daily New Cases trend of India

Cities put under lock-down, partial lock-down, or night curfews was aimed to put a pause on the spread of the virus. However, aggressive tracking, tracing, testing, and vaccination was recognized as a key focus area to contain the spread.

Scientists have proposed and implemented various technologies/measure to reduce the negative impact of the pandemic and to accelerate the recovery phase [1].

Effect of Testing - Testing can help people determine if they are infected with SARS-CoV-2, regardless of whether they have symptoms and whether they are at risk of spreading the infection to others. A positive test early in the course of the illness enables individuals to isolate themselves - reducing the chances that they will infect others and allowing them to seek treatment earlier, which likely reduces disease severity and the risk of long-term disability, or death.

A high test rate for COVID-19 has emerged as the sure-shot way of reducing the impact of the pandemic across the world. Lock-downs has helped in physical distancing, but a sure way of limiting the growth of the infection spread have been to test significantly, isolate infected patients and treat them. Within India, States like Kerala had tested at a high rate relative to its population initially. Kerala managed to do so by maximizing contact tracing and local surveillance. This early testing strategy helped the State to reduce the total number of infections over time.

One of the reasons for fewer people to seek COVID-19 tests is cost. This can be ramp-ed up by offering testing free of charge or paid via health insurance or by making the test affordable to all.

Effect of Vaccination: Vaccines save millions of lives each year. Vaccines work by training and preparing the body's natural defenses – the immune system – to recognize and fight off the viruses and bacteria they target.

Aggressive vaccination can help to control the prevailing situation of rising cases of COVID-19 across the country. [24].

Impact of Testing & Vaccination on COVID-19 spread: USA, UK, Israel, South Korea, and the United Arab Emirates are leading the world in vaccine coverage. These nations have vaccinated a large part of their populations. Studies indicate a sharp drop in hospitalizations among vaccinated older people.

In this study, we extensively used the SIR model to arrive at the result and conclusions. The SIR model (also known as Susceptible, Infected, Recovered Model), was coined by Kermack and McKendrick [2], which laid the foundation for the Dynamic Model of Infectious Diseases.

The SIR epidemiological model is one of the oldest and has the most significant consequences of biological science. The model assumes that S, I, and R represent the proportion of the population in the susceptible, infected, or recovered state.

The Susceptible-Infectious-Recovered (SIR) model is known to have an exact semi-analytical solution [12]. The SIR model incorporating time-based parameters and AI algorithms was also used to study the spread of COVID-19 in South Korea [23] and other countries.

There have been a few recent works studying different transmission models for COVID-19 such [26] which developed an agent-based model to reproduce the characteristics of COVID-19 transmission or [25] which proposed a mobility-based model to measure COVID-19 growth rate ratio for a particular given day.

Another recent work by Systrom [28] has presented a Bayesian prediction approach to obtain confidence intervals for R_t . However, Systrom's work builds on [27], where the definition of infection rate Rt is not based on a time-varying contact rate of the SIR model. Instead, their approach estimates infection rate probabilistically based on the number of new cases alone.

Mathematical models [12] for disease epidemic are either deterministic or stochastic, where the first is considered as a thermodynamic limit of the other.

Mathematical modeling of epidemics broadly consists of three types, Statistical methods for epidemic surveillance, Mechanistic State-Space model and Empirical Learning models. Mechanistic State-Space models have outperformed the other two in describing respiratory diseases such as MERS and SARS. Mechanistic State-Space Models are classified [Fig. 3] as "Continuum" models, "Markov Chain" models, "Complex Network" models and "Agent Based Simulations".

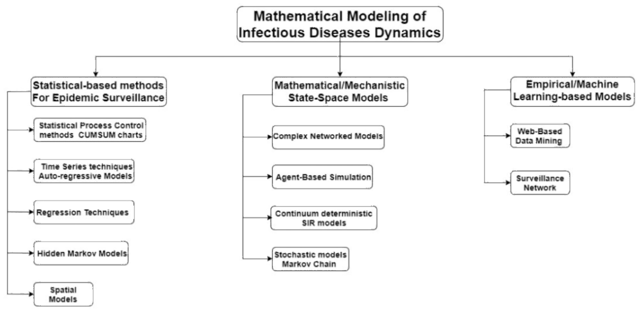

Fig. 3. Classification of Mathematical Models in Epidemiology

The last century has seen a plethora of outbreaks and epidemics[1]. While coronaviruses such as SARS-CoV & MERS-CoV have been responsible for a majority of these outbreaks [Table 1], different types of influenza viruses, such as H1N1, H2N2, and H3N2, have been at the helm of all the four pandemics in the past 105 years.

The H1N1 virus alone has been responsible for two pandemics - 1) the Spanish Flu of 1918-1919 and the 2) Swine flu in 2009-2010, while the H2N2 and H3N2 influenza viruses have been responsible for the Asian Flu of 1957-1958, and the Hong Kong Flu of 1968-1969, respectively.

Disease	Causative Agent	Year(s)	Death Toll	Classificiation
Spanish Flu	H1N1	1918-1919	~50 million	Pandemic
Asian Flu	H2N2	1957-1958	~1.1 million	Pandemic
Hong Kong Flu	H3N2	1968-1969	~1 million	Pandemic
SARS	SARS-CoV	2002-2004	774	Outbreak
Swine Flu	H1N1 (new strain)	2009-2010	~151,700 to 575,400	Pandemic
MERS	MERS-CoV	2012-present	871*	Outbreak
Asian Lineage Avian Influenza	H7N9	2013-2017	~605	Epidemic
Ebola Virus Disease (EVD)	Zaire ebolavirus	2014-2016	11,325	Epidemic
COVID-19	SARS-CoV-2	2019-present	217,769*	Pandemic

*as of 30 April 2020 Data Source: WHO, CDC

Table 1: Major viral diseases (1915 – present)

In other case, a simulation of the SIR model [29] has been created to identify the effective reproduction number (R_e), and then through time series analysis using Prophet model [Fig. 4], the conclusion has been drawn for the number of days it will take to vaccinate enough population to achieve herd immunity.

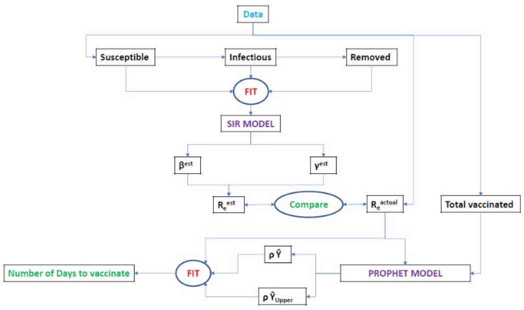

Fig. 4. SIR and Prophet Model

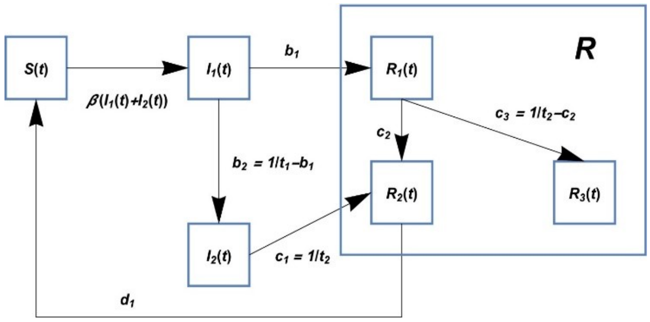

Fig. 5. SIIR Model

Another [23] study aims at analyzing the novel coronavirus infection (COVID-19) spread in South Korea using the susceptible-infected-recovered (SIR) model. It is assumed that each parameter in the SIR model is a function of time so that important parameters, such as the basic reproduction number (R_0), can be computed more delicately. Using neural networks, a method is proposed to find the best time-varying parameters and the solution for the model simultaneously. Moreover, using time-dependent parameters, traditional numerical algorithms, such as the Runge-Kutta methods, can successfully approximate the SIR model while fitting the COVID-19 data, thus modeling the propagation patterns of COVID-19 more precisely.

Based on the SIR model, study [9] evaluates the efficiency of certain measures regarding new coronavirus (COVID-19). The data from the Provincial Health Committee and WHO are used to establish SIR model and draw graphs for evaluation. Simulations and pictures of SIR-F model are cited to predict the efficiency of certain measures. This article analyzes how the epidemic was regulated in different regions. The result shows that it is necessary to introduce strict regulations and take immediate actions so as to reduce the further spread of the virus.

A new mathematical model [7] called SIIR model is constructed to describe the spread of infection by taking account of the characteristics of COVID-19 and is verified by the data from Japan. The following features of COVID-19: (a) there exist pre-symptomatic individuals who have infectivity even during the incubation period, (b) there exist asymptomatic individuals who can freely move around and play crucial roles in the spread of infection, and (c) the duration of immunity may be finite, are incorporated into the SIIR model.

The SIIR model [Fig. 5.] has the advantage of being able to explicitly handle asymptomatic individuals who are delayed in discovery or are extremely difficult to be discovered in the real world.

It is shown that the conditions for herd immunity in the SIIR model become more severe than those in the SIR model; that is, the presence of asymptomatic individuals increases herd immunity threshold (HIT).

Structure of the SIIR model: S(t) as susceptible population, I1(t) as pre-symptomatic population (infectious), I2(t) as asymptomatic population (infectious), R1(t) as symptomatic population (quarantined and not infectious), R2(t) as recovered population (recovered with antibody and not infectious), and R3(t) as fatalities because of COVID-19 (not infectious).

Another paper [5] proposes a model for predicting COVID-19 using the SIR and machine learning for smart health care and the well-being of the citizens of KSA. Knowing the number of susceptible, infected, and recovered cases each day is critical for mathematical modeling to be able to identify the behavioral effects of the pandemic. It forecasts the situation for the upcoming 700 days. The proposed system predicts whether COVID-19 will spread in the population or die out in the long run.

Mathematical analysis and simulation results are presented as a means to forecast the progress of the outbreak and its possible end for three types of scenarios: “no actions,” “lockdown,” and “new medicines.” The effect of interventions like lockdown and new medicines is compared with the “no actions” scenario. The lockdown case delays the peak point by decreasing the infection and affects the area equality rule of the infected curves. On the other side, new medicines have a significant impact on infected curve by decreasing the number of infected people about time. Available forecast data on COVID-19 using simulations predict that the highest level of cases might occur between 15 and 30 November 2020.

To reduce the impact of COVID-19 pandemic most countries have implemented several counter-measures to control the virus spread including school and border closing, shutting down public transport and workplace and restrictions on gathering. In this research work [19], authors propose a deep-learning prediction model for evaluating and predicting the impact of various lockdown policies on daily COVID-19 cases. This is achieved by first clustering countries having similar lockdown policies, then training a prediction model based on the daily cases of the countries in each cluster along with the data describing their lockdown policies.

Once the model is trained, it can be used to evaluate several scenarios associated to lockdown policies and investigate their impact on the predicted COVID cases. The evaluation experiments, conducted on Qatar as a use case, show that the proposed approach achieved competitive prediction accuracy. Additionally, findings highlighted that lifting restrictions particularly on schools and border opening would result in significant increase in the number of cases during the study period.

II. Materials and Methods

A. Methodology

The methodology primarily used is based on the SIR model to understand

a) the numbers of the suspected cases in the coming days, and

b) map the need of medical facilities requirement

A simple mathematical description of the spread of a disease in a population is achieved through the so-called SIR model, which divides the (fixed) population of N individuals into three "compartments" which may vary as a function of time, t:

S(t) are those susceptible but not yet infected with the disease

I(t) is the number of infectious individuals;

R(t) are those individuals who have recovered from the disease and now have immunity to it.

The SIR model describes the change in the population of each of these compartments in terms of two parameters, β and γ.

β describes the effective contact rate of the disease: an infected individual comes into contact with βN other individuals per unit time (of which the fraction that is susceptible to contracting the disease is S/N).

γ is the mean recovery rate: that is, $1/\gamma$ is the mean period of time during which an infected individual can pass it on.

The differential equations describing this model were first derived by Kermack and McKendrick [2] as:

$S/dt = - \beta * S * I / N$

$dI/dt = \beta * S * I/N - \gamma * I$

$dR/dt = \gamma * I$

where N is the total number of population

B. Number of Suspected cases

The SIR model was used to predict the number of suspected cases in the future. In this model, the

Effective contact rate β is transmission rate * contact rate

For an example, if 10% is transmission rate and 5 contacts a day then it is 0.1 * 5 = 0.50

Rate of recovery, γ is 1 / 16 number of days a person is contagious, which is 0.0625

$R_0 = \beta / \gamma$

R_0 can be defined as a product of the contact rate β and infection time. Replacing the values:

$R_0 = 0.50 / 0.0625$

$R_0 = 8$ which is very large

when R_0 is less than 1 then the curve flattens.

C. Need of Medical Facilities

The enormous impact of the COVID−19 pandemic has surprised many hospitals beginning March 2020. However, pandemics and natural disasters typically come with a sudden influx of unforeseen patients, which almost instantly pushes the boundaries of a hospital's capacity.

It soon became apparent that the capacity of hospital beds was on the verge of coming under great pressure. This behaviour was also observed during several waves in several countries.

Besides a shift of regular beds to specific COVID-19 beds with special hygiene measures, pressure on the number of beds arose primarily from the need to foresee sufficient capacity in the Intensive Care Unit (ICU). Indeed, while approximately 9 to 11% of admitted COVID-19 patients needed advanced life-supporting measures [3].

The different stages of COVID-19 pandemic can be described by two key variables i.e. ICU patients and deaths in hospitals [17].

Daily updates of the models with real data can allow forecasting some key indicators for decision-making. These are bed allocation, hospital stay time, oxygen requirement, doubling time, rate of renewal, the maximum daily rate of change (positive/negative), halfway points, maximum plateaus, asymptotic conditions, and dates and time intervals when some key thresholds are overtaken.

This data can be used by medical practitioners and decision-makers to predict the trends on both short-term and long-term horizons.

III. Results

A. Model Behavior and Predictions

The dataset from source (https://api.covid19india.org/) was used for the study The dataset provides India level time-series data for Confirmed, Recovered and Deceased cases. The dataset is updated on daily basis.

The Infected, Recovered progresssion [Fig. 6.] over a period of time based on the dataset using the model.

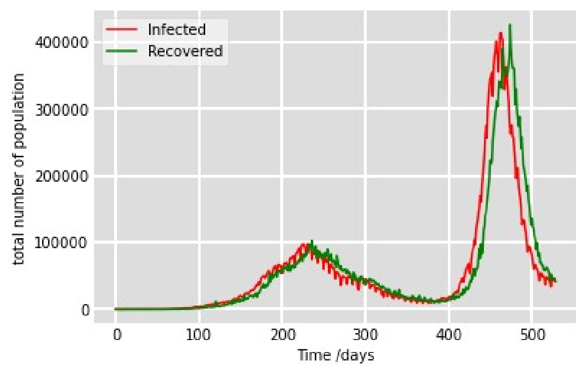

Fig. 6. Infected, Recovered progression

The model created using the SIR algorithm can predict the values which are very close to the actual values [Fig 7].

The data-set used for the analysis has data from 30/Jan/2020 onward. The model prediction suggests the peak on 435th days (from start date). The model also suggests that the curve would flatten on the 550th day.

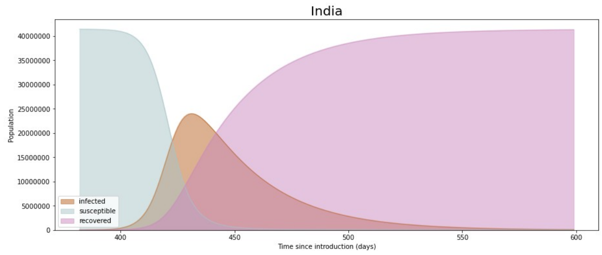

Fig. 7. Model Prediction

B. *Assumptions*

Assumptions made while developing the model was as below:

- The population of India; the model assumes the population of the country at the start of the year 2020. However, the data set used didn't have the total population of the country factored in.
- Data provided in the data set is consistent and reliable.
- The effect of testing and vaccination is not incorporated in the model training.
- The total population of India is assumed as 138 crores (approx). For simplicity of modelling, we have accepted homogenous appropriation of the Indian population that doesn't includes variations in population or the urban provincial variations.
- The count of the Susceptible population is assumed as 3% of the total population.

C. *Medical Intervention/Support Prediction*

The model created can be used to calculate[2] and plan the capacity required for normal, critical (ICU) hospital beds for treatment based on infection rate.

IV. Discussions and Conclusions

Through this study, we tried to observe the peak and decline of COVID-19 cases using the SIR model. The SIR model also helped to predict the rise of cases and curve flattening. This study was focused on India, however the same can be extended to any country.

Our results show that the SIR model is more suitable to predict the epidemic trend due to the spread of the disease as it can accommodate surges and be adjusted to the recorded data.

Although SIR model was used for our analysis, however other mathematical models such as SEIR, SEIR-D can also be suitable for COVID 19 like disease.

We also understand that none of the models can serve as a magic band to forecast what is going to happen in the coming days, but these models or sets of models can provide insight on aspects of the pandemic that might be currently unknown.

We have already started seeing a downward trend in India and the same was also predicted by the model. The analysis and results published here can be used to fight against the virus and waves.

References

[1] V. Chamola, V. Hassija, V. Gupta, and M. Guizani, "A comprehensive review of the COVID-19 pandemic and the role of IoT, drones, AI, blockchain, and 5G in managing its impact," IEEE Access, vol. 8, pp. 90225–90265, 2020.

[2] I. Z. Kiss, J. C. Miller, and P. L. Simon, Mathematics of epidemics on networks: From exact to approximate models, 1st ed. Cham, Switzerland: Springer International Publishing, 2017.

[3] A. Remuzzi and G. Remuzzi, "COVID-19 and Italy: what next?," Lancet, vol. 395, no. 10231, pp. 1225–1228, 2020.

[4] I. Cooper, A. Mondal, and C. G. Antonopoulos, "A SIR model assumption for the spread of COVID-19 in different communities," Chaos Solitons Fractals, vol. 139, no. 110057, p. 110057, 2020.

[5] S. A. Alanazi, M. M. Kamruzzaman, M. Alruwaili, N. Alshammari, S. A. Alqahtani, and A. Karime, "Measuring and preventing COVID-19 using the SIR model and machine learning in smart health care," J. Healthc. Eng., vol. 2020, p. 8857346, 2020.

[6] M. Venkatasen et al., "Forecasting of the SARS-CoV-2 epidemic in India using SIR model, flatten curve and herd immunity," J. Ambient Intell. Humaniz. Comput., pp. 1–9, 2020.

[7] M. Tomochi and M. Kono, "A mathematical model for COVID-19 pandemic-SIIR model: Effects of asymptomatic individuals," J. Gen. Fam. Med., vol. 22, no. 1, pp. 5–14, 2021.

[8] lisphilar, "COVID-19 data with SIR model," Kaggle.com, 19-May-2021. [Online]. Available: https://www.kaggle.com/lisphilar/covid-19-data-with-sir-model. [Accessed: 22-May-2021].

[9] L. Wang, "Evaluation of COVID-19 epidemic based on SIR model," in Proceedings of the 2020 International Symposium on Artificial Intelligence in Medical Sciences, 2020.

[10] V. Vytla, S. K. Ramakuri, A. Peddi, K. Kalyan Srinivas, and N. Nithish Ragav, "Mathematical models for predicting covid-19 pandemic: A review," J. Phys. Conf. Ser., vol. 1797, no. 1, p. 012009, 2021.

[11] M. K. Kakkar, M. Sood, B. Sharma, and J. Bhatti, "Mathematical modeling and forecasting the spread of covid-using python," Ijstr.org. [Online]. Available: https://www.ijstr.org/final-print/nov2020/Mathematical-Modeling-And-Forecasting-The-Spread-Of-Covid-19-Using-Python.pdf. [Accessed: 22-May-2021].

[12] J. C. Miller, "Mathematical models of SIR disease spread with combined non-sexual and sexual transmission routes," Infect. Dis. Model., vol. 2, no. 1, pp. 35–55, 2017.

[13] A. Atkeson, K. Kopecky, and T. Zha, "Estimating and forecasting disease scenarios for COVID-19 with an SIR model," National Bureau of Economic Research, Cambridge, MA, 2020.

[14] T. Xin, "The model of COVID-19 pandemic," in 2020 International Conference on Computing and Data Science (CDS), 2020, pp. 429–432.

[15] M. Deschepper, K. Eeckloo, S. Malfait, D. Benoit, S. Callens, and S. Vansteelandt, "Prediction of hospital bed capacity during the COVID− 19 pandemic," BMC Health Serv. Res., vol. 21, no. 1, 2021.

[16] D. Manca, D. Caldiroli, and E. Storti, "A simplified math approach to predict ICU beds and mortality rate for hospital emergency planning under Covid-19 pandemic," Comput. Chem. Eng., vol. 140, no. 106945, p. 106945, 2020.

[17] H. Nishiura et al., "Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19)," Int. J. Infect. Dis., vol. 94, pp. 154–155, 2020.

[18] D. Majra, J. Benson, J. Pitts, and J. Stebbing, "SARS-CoV-2(COVID-19) superspreader events," J. Infect., vol. 82, no. 1, pp.36–40, 2021.

[19] A. B. Said, A. Erradi, H. Aly, and A. Mohamed, "A deep-learning model for evaluating and predicting the impact of lockdown policies on COVID-19 cases," arXiv [cs.SI], 2020.

[20] S. Bugalia, V. P. Bajiya, J. P. Tripathi, M.-T. Li, and G.-Q. Sun, "Mathematical modeling of COVID-19 transmission: the roles of intervention strategies and lockdown," Math. Biosci. Eng., vol. 17, no. 5, pp. 5961–5986, 2020.

[21] American Scientific Research Journal for Engineering, Technology, & Sciences (ASRJETS). (n.d.). View of numerical study of Kermack-mckendrik SIR model to predict the outbreak of Ebola virus diseases

using Euler and fourth order Runge-Kutta methods. Retrieved July 10, 2021, from Asrjetsjournal.org website: https://asrjetsjournal.org/index.php/American_Scientific_Journal/article/view/3424/1279

[22] N. Zaki and E. A. Mohamed, “The estimations of the COVID-19 incubation period: A scoping reviews of the literature,” J. Infect. Public Health, 2021

[23] H. Jo, H. Son, S. Y. Jung, and H. J. Hwang, “Analysis of COVID-19 spreading in South Korea using the SIR model with time-dependent parameters and deep learning,”.

[24] Accelerating a safe and effective COVID-19 vaccine. (n.d.). Retrieved July 10, 2021, from Who.int website: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/accelerating-a-safe-and-effective-covid-19-vaccine

[25] Marshall Maximilian Dong Ensheng Squire Marietta Badr Hamada, Du Hongru and Gardner Lauren. 2020. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. The Lancet Infectious Diseases (2020).

[26] Sheryl L. Chang, Nathan Harding, Cameron Zachreson, Oliver M. Cliff, and Mikhail Prokopenko. 2020. Modelling transmission and control of the COVID-19 pandemic in Australia. arXiv:q-bio.PE/2003.10218

[27] Luís Bettencourt and Ruy Ribeiro. 2008. Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases. PloS one 3 (02 2008), e2185. https://doi.org/10.1371/journal.pone.0002185

[28] Kevin Systrom. 2020. The MetricWe Need to Manage COVID-19. http://systrom.com/blog/the-metric-we-need-to-manage-covid-19/

[29] Darapaneni, N., Dhua, S., Khare, N., Ayush, K., Karthikeyan, Ghodke, S, Paduri, A. R. (2021). Forecasting vaccination drive in India for herd immunity using SIR and prophet model. 2021 IEEE World AI IoT Congress (AIIoT), 0021–0027. IEEE.

[30] C. Sohrabi, Z. Alsa_, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosi_dis, and R. Agha, ``World health organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *Int.J. Surgery*, vol. 76, pp. 71_76, Apr. 2020.

[31] M. Cascella, M. Rajnik, A. Cuomo, S. C. Dulebohn, and R. Di Napoli, ``Features, evaluation and treatment coronavirus (COVID-19) [updated 2020 Apr 6]," in *StatPearls [Internet]*. Treasure Island, FL, USA: StatPearls Publishing, Jan. 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK554776/