2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS) | 978-1-7281-8524-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICIIS51140.2020.9342660

COVID-19 Infection Dynamics for India-Forecasting the Disease using SIR models

Narayana Darapaneni
Director - AIML
Great Learning/Northwestern University
Illinois, USA
darapaneni@gmail.com

Arjun Panwar
Student - AIML
Great Learning
Mumbai, India
arjunpanwar85@gmail.com

Anwesh Reddy Paduri
Research Assistant - AIML
Great Learning
Mumbai, India
anwesh@greatlearning.in

Ankit Patel
Student - AIML
Great Learning
Mumbai, India
ankit411631@gmail.com

Chaitanya Shah
Student - AIML
Great Learning
Mumbai, India
chaitanya28@gmail.com

Jigar Gada
Student - AIML
Great Learning
Mumbai, India
jigar.gada@gmail.com

Milind Majrekar
Student - AIML
Great Learning
Mumbai, India
Milind.Manjrekar@hotmail.com

***Abstract* -- A retrospective cohort study of novel Coronavirus disease (COVID-19) on India data and predicting the disease outcome (Infected & Recovered) using SIR compartment model.An existing literature survey of SIR modeling on coronavirus disease was probed and further improvised the finding and methodologies used in modeling were performed over the existing India data set. Numerous papers were surveyed and the model was trained to understand the optimal value of hyperparameters (β and γ) and the approach to forecast the coronavirus disease.As on 30th June 2020 a total of 215k Active cases and 17k deaths were reported. A sample(N) of 2% of the overall population (133 cr.) was considered and the initial date of infection to model the disease was 1st May 2020. Further comparison was made using 5% and 10% Susceptible population to check the model efficacy. The reported number of Initial Suspected(S0), Infected(I0) and Recovered(R0) cases were 25,965k, 24,755 & 9,065 respectively. The optimal values of Beta(β) and Gamma(γ) were estimated to be 0.093 and 0.055 respectively. The overall Infected MAPE(Mean percentage absolute error) was 12.23% & Recovered MAPE was 6.48% on 2% Susceptible population.The research presented the current trends of Covid-19 disease in India from 1st May'20 to 30th June'20. The trajectory of the disease was also forecasted for next days from 30th June 2020.Rapid growth of the disease has been seen in the month of June with maximum peak reaching on 164th Day (Mid Oct'20) with ~25.5 Lakh infected cases considering 2% Susceptible, ~64.75 lakh cases considering 5% Susceptible & ~1.3 cr. cases considering 10% Susceptible. The estimated R0 (β/γ) value is ~ 1.69. However further research and addition of compartments like Exposed, Critical can be made to improve the forecast.**

Keywords -- SIR model, Covid 19, Infected, Recovered, Maps, Reproduction number(R0).

I. INTRODUCTION

The epidemic of novel coronavirus disease named COVID-19 [4], is rapidly spreading worldwide and has wreaked havoc in various countries [13] [14]. This infectious disease outbreak is the occurrence of the disease that is not usually expected in a particular community, geographical region, or time period. The epidemiological parameters of this contagious epidemic have been intensively investigated and disclosed to global communities. Regularly, a rising irresistible infection includes quick spreading, jeopardizing the soundness of huge quantities of individuals, and subsequently requires prompt activity to forestall the malady at the network level. COVID-19 is brought about by another kind of coronavirus which was recently named 2019-nCoV by the World Health Organization (WHO). It is the seventh individual from the coronavirus family, along with MERSnCoV and SARS-nCoV, that can spread to people. The side effects of the contamination incorporate fever, hack, windedness, and looseness of the bowels. In more serious cases, COVID-19 [11] can cause pneumonia and even demise. The brooding time (Incubation period) [7] of COVID-19 can keep going for about fourteen days or more. During the time of idle contamination, the malady may in any case be irresistible. The infection can spread from individual to individual through respiratory beads and close contact. The objective is to explain the background and provide an introduction to the topic of modelling infectious diseases. Numerous examinations have utilized the transmission elements of infection in metapopulations patches where the all out population is partitioned into various discrete patches, every one of which is treated as being admirably mixed [15]. Bolker et al. [16] utilized such a model to exhibit the adequacy of immunization against measles flare-up. Arino and van nook Driessche [17] have proposed a model in which the number of inhabitants in a city is combined with different urban areas through a portability boundary that portrays the net pace of intercity movement [18]. Lloyd and Jansen [19] additionally utilized metapopulation of n-patches to pestilences' elements to see the examples of synchrony in flare-ups. A more included model with cross-fix disease was examined for different cases by Muroya et al [20], who took a gander at the worldwide security of an endemic. Wu et al. [21] utilized a SIR model to depict the transmission elements of COVID-19 and furthermore gauge the clinical seriousness for the coronavirus. To consider the elements of COVID-19, a stochastic transmission model additionally created by Kucharski et al. [22]. Investigation of viral elements utilizing numerical models have helped gain bits of knowledge into the comprehension of viral diseases, for example, tuberculosis, dengue, and zika infection [23].

The SIR [1] (Susceptible - Infectious - Recovered) model equations are derived and explained from scratch using real time data of India as a base country. Once the compartments of SIR [2] (Susceptible-Infectious-Recovered) are determined, modelling can be done using a variety of methods to estimate the error and peak day of the infection.

II. MATERIALS AND METHODS

A. Data Collection

We will use country-level data [1]- day-level data from validated sources such as Covid-19 India.org to model the disease, which records the number of infections, recovered, and death at a day-level [4]. The daily cases state the total number of Confirmed Cases, we calculated the daily Active case by subtracting from the Confirmed Cases the Cured(Recovered) and Deaths. Below is the sample data.

TABLE.1. SAMPLE SNIPPET OF THE DATA POINTS TO BE MODELLED.

Date	Cured	Deaths	Confirmed	Active
1-May-20	9065	1152	34972	24755
2-May-20	9951	1218	37157	25988
3-May-20	10887	1306	40124	27931
4-May-20	11762	1389	42836	29685
5-May-20	13161	1583	46711	31967

The data [6] was ordered in the ascending order of the Date variable and a window of 2 months' data beginning from 1st May 2020 to 30th June 2020 was subsetted.

B. Modelling

The SIR [1] model is a kind of compartmental model that explains the infectious disease dynamics. The model breaks down the population into compartments. It is predicted each compartment would have the same characteristics. The SIR [2] describes the model 's three segmented compartments. Susceptible is a group of people vulnerable to infectious exposure. When the infection occurs they should be careful. The contagious community represents the people infected. We will pass the disease on to susceptible individuals and can be recovered in a given time. People recovered get immunity so they are no longer susceptible to the same disease. SIR model is a framework [6] which describes how the number of individuals in each group can change over time.

Fig 1: The SIR model compartments

Variable definitions:

1) S(t) =Number of susceptible people on Day t.
2) I(t) =Number of people on day t exposed.
3) R(t) = Amount of retrieved people on day t.
4) N = Country level population.
5) ß = expected amount of people getting infected by an infectious person per day
6) D =The number of days the disease has, and will spread.
7) γ = Ratio of infected recovery per day(1 / D)
8) Ro = Total number of people getting infected by an infected person (R□ = β / γ).

For India, the R0 value of COVID-19 determined on the basis of earlier data is stated in range 1.5-4[11]. If the R0 goal is 0.5, the lower population limit required for the vaccine is 919.7 million and the upper limit will be 1.2071 billion. That remains a very vague figure so it remains an important part to calculate the value of R0.

Model SIR [10] helps one to classify the number of people with the ordinary differential equation in each compartment. β is a parameter which controls how much the disease can be transmitted by exposure. It is measured by the probability of touch and the possibility of transmission of disease. γ is a parameter which expresses how often the disease can be recovered in a given time. Below are the ordinary nonlinear differential equations that we'll use to suit the model.

$dS/dt = -\beta . I . S/N$ -----------------------------(1)

$dI/dt = \beta . I . S/N - \gamma . I$ ------------------------(2)

$dR/dt = \gamma . I$ -------------------------------------(3)

We should measure the existence of the disease in terms of the simple amount of replication of the infection (R0) (=β/γ). We will use odeint from the scipy.integrate python library to model the differential equation.

C. Statistical Analysis

We estimated the beta and gamma values using the SIR model over the historical data of India from 1st May 2020 to 30th June 2020.The current trend suggests the increase in rate of infected people in the month of June as compared to May. The recovery rate seems to be quite good and approximated at 50% in India.

Despite having the lockdown in place for 2 months the infectious disease has seen a drastic rise post the list starting from the early June. Below is the current trend of the disease in India.

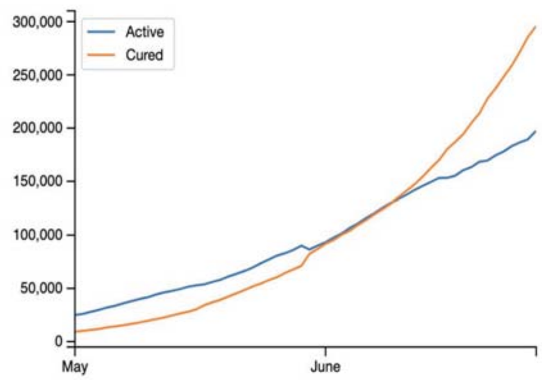

Fig.2. Trend of Active and Recovered cases from May 2020 to June 2020 in India

The Beta and Gamma [6] estimates from the list of predefined values declared were used to get the lowest MAPE. A grid search was performed over the data to come out with the optimal values of beta and gamma [7] giving the low error of MAPE on Infected and Recovered predictions. The model was built using Python as a platform to model the disease and plot the curves.

III. RESULTS AND DISCUSSION

Considering 2% of the Overall Population the SIR modelled the disease with **Beta=0.093** and **Gamma=0.055** with the overall fitted error of **12.23%** on Infected data. Below are the SIR model trends fitted with the data and the prediction of Infected, Suspected and Recovered.

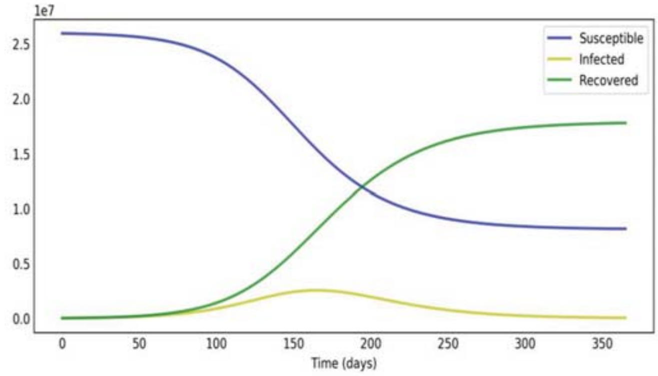

Fig.3. SIR model fitted on 2% susceptible population of India

From the SIR model it is clear that the highest peak of Infection is supposed to reach **164 days** from the date of modelling i.e. Mid of October with total Infected cases close to ~**25.5 Lakhs**. Using the estimates of Beta and Gamma we quantified the infected cases as below:

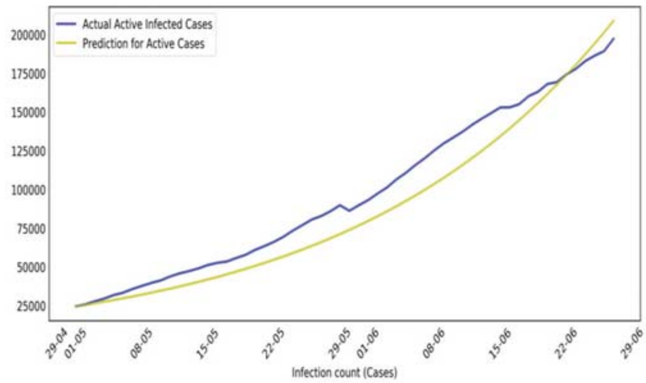

Fig.4. Prediction on the Active cases in India with 2% Susceptible

The recovery curve seems to fit well with the optimal values of Beta and gamma and the mean error reported on recovered cases was 6.48% which seems to be better than that observed in infected cases. Further improvement can be seen with using other compartments like Death and Critical and then modelling the disease separately for each compartment.

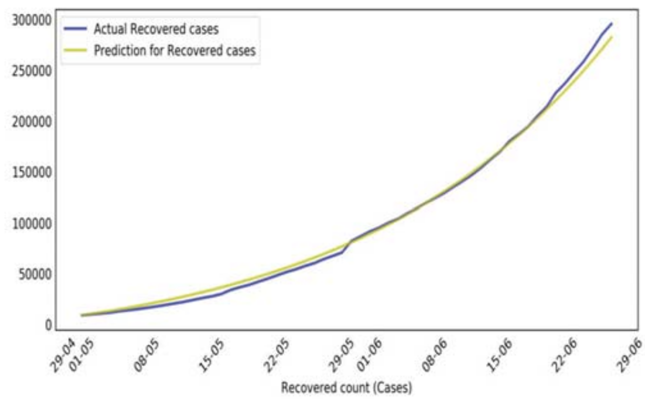

Fig.5. Prediction on the Recovered cases in India with 2% Susceptible

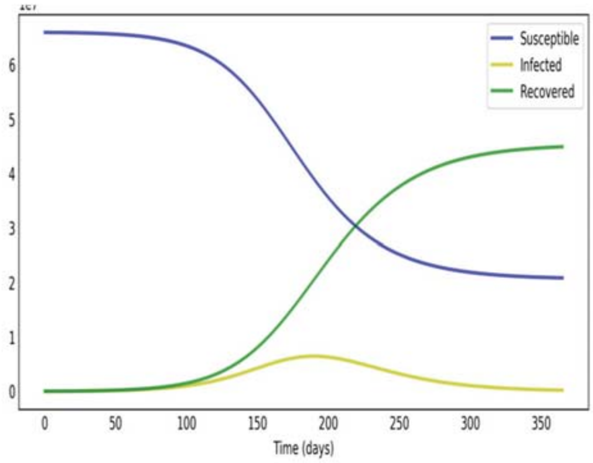

Fig.6. SIR model fitted on 5% susceptible population of India

With a 5% susceptible population the estimated days of reaching the highest peak is 189 days i.e. on 7th Nov 2020 with approx. **64.7 lakh** population being infected. The model fitted on a 5% susceptible population seems to capture variation with MAPE of 11.55% for infected cases.

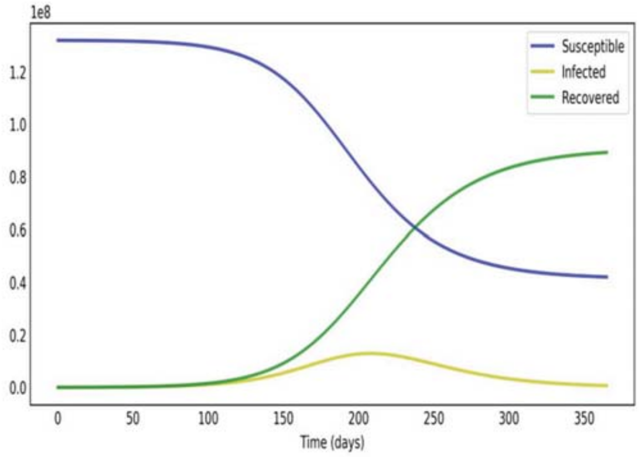

Fig .7. SIR model fitted on 10% susceptible population of India

Given the estimates from the SIR model we were able to capture the trend of infection by plotting the curves of the number of Infected and Recovered cases on a daily basis. Below table is the forecast of the number of infected cases and recovered cases considering the overall susceptible population to be 2%,5% and 10% respectively and comparison is made on the MAPE across them. There is a slight quadratic trend in terms of number of infected cases reaching approximately 4 lakh cases by mid-July with a slight variation across the 3 sets of susceptible populations. The recovered trend seems to be steeper than the infected curve.

TABLE.2. FORECASTED NUMBER OF INFECTED AND RECOVERED CASES IN INDIA.

Date	Infected cases prediction			Recovered cases prediction		
	2% susceptible	5% susceptible	10% susceptible	2% susceptible	5% susceptible	10% susceptible
01/07/2020	241,571	248,392	250,679	332,181	336,525	337,966
02/07/2020	250,427	257,821	260,304	345,747	350,482	352,055
03/07/2020	259,586	267,600	270,295	359,810	364,970	366,685
04/07/2020	269,057	277,740	280,664	374,386	380,006	381,876
05/07/2020	278,849	288,253	291,426	389,493	395,612	397,650
06/07/2020	288,972	299,153	302,594	405,150	411,808	414,028
07/07/2020	299,433	310,454	314,184	421,374	428,616	431,034
08/07/2020	310,243	322,168	326,211	438,184	446,059	448,692
09/07/2020	321,411	334,310	338,692	455,601	464,160	467,025
10/07/2020	332,947	346,895	351,642	473,643	482,943	486,059
11/07/2020	344,859	359,937	365,079	492,332	502,432	505,820
12/07/2020	357,157	373,452	379,021	511,689	522,653	526,337
13/07/2020	369,852	387,456	393,485	531,735	543,633	547,637
14/07/2020	382,951	401,964	408,490	552,492	565,400	569,749
15/07/2020	396,466	416,995	424,056	573,983	587,980	592,704

The error (MAPE) of the model fitted on 2%, 5% & 10% of the Susceptible population is tabulated below.

TABLE.3. COMPARISON OF THE ERROR RATE OF INFECTED AND RECOVERED CASES ACROSS 2%, 5%, AND 10% SUSCEPTIBLE.

Compartment	2% Susceptible	5% Susceptible	10% Susceptible
Infected	12.23%	11.55%	11.33%
Recovered	6.48%	6.80%	6.91%

The error rate for the infected case seems to be slightly high for 2% Susceptible population but on similar lines in case of 5% and 10%. Recovered error rate seems to be almost the same across all the three sets of the Susceptible population.

IV. CONCLUSION

This exploration introduced current patterns of COVID-19 episode In India from 1st May 2020 to 30th June 2020 as pictured in the Corona Tracker site. The direction of the flare-up is likewise determined by utilizing the SIR model, 365 days from 1st May 2020.COVID-19 is as yet an irresistible sickness with some indistinct or obscure properties, which implies precise SIR forecast must be acquired once the episode has been effectively contained. The episode spreads are to a great extent affected by every nation's arrangement and social obligation. This examination is as yet progressing the same number of more examinations with respect to this malady can be completed.

However, it fills in as the beginning stage to explore more top to bottom on addresses that spin around this worldwide pandemic.

REFERENCES

[1] R. Ranjan, “Predictions for COVID-19 outbreak in India using Epidemiological models,” bioRxiv, p. 2020.04.02.20051466, 2020.

[2] M. Batista, “Estimation of the final size of the coronavirus epidemic by the logistic model,” bioRxiv, p. 2020.02.16.20023606, 2020.

[3] S. Deb and M. Majumdar, “A time series method to analyze incidence pattern and estimate reproduction number of COVID-19,” arXiv [stat.AP], 2020.

[4] S. L. Chang, N. Harding, C. Zachreson, O. M. Cliff, and M. Prokopenko, “Modelling transmission and control of the COVID-19 pandemic in Australia,” arXiv [q-bio.PE], 2020.

[5] J. Ma, J. Dushoff, B. M. Bolker, and D. J. D. Earn, “Estimating initial epidemic growth rates,” Bull. Math. Biol., vol. 76, no. 1, pp. 245–260, 2014.

[6] Moorthy, A. M. Henao Restrepo, M.-P. Preziosi, and S. Swaminathan, “Data sharing for novel coronavirus (COVID-19),” Bull. World Health Organ., vol. 98, no. 3, p. 150, 2020.

[7] Z. Xiao et al., “Examining the incubation period distributions of COVID-19 on Chinese patients with different travel histories,” J. Infect. Dev. Ctries., vol. 14, no. 4, pp. 323–327, 2020.

[8] S. Dey, “Coronavirus: 15 lakh who flew back to India between Jan 18-Mar 23 under lens,” Times Of India, 28-Mar-2020. [Online]. Available: https://timesofindia.indiatimes.com/india/15-lakh-who-flew-back-to-india-between-jan-18-mar-23-under-lens/articleshow/74855553.cms. [Accessed: 06-Nov-2020].

[9] J. Ma, “Estimating epidemic exponential growth rate and basic reproduction number,” Infect. Dis. Model., vol. 5, pp. 129–141, 2020.

[10] Simha, R. V. Prasad, and S. Narayana, “A simple Stochastic SIR model for COVID 19 Infection Dynamics for Karnataka: Learning from Europe,” arXiv [q-bio.PE], 2020.

[11] F. Wu et al., “A new coronavirus associated with human respiratory disease in China,” Nature, vol. 579, no. 7798, pp. 265–269, 2020.

[12] S. Mandal et al., “Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India: A mathematical model-based approach,” Indian J. Med. Res., vol. 151, no. 2 & 3, pp. 190–199, 2020.

[13] “Coronavirus disease (COVID-19) – World Health Organization,” Who.int. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019. [Accessed: 06-Nov-2020].

[14] N. Zhu et al., “A novel Coronavirus from patients with pneumonia in China, 2019,” N. Engl. J. Med., vol. 382, no. 8, pp. 727–733, 2020.

[15] M. S. Bartlett, “Monte Carlo studies in ecology and epidemiology,” Berkeley.edu. [Online]. Available: https://digitalassets.lib.berkeley.edu/math/ucb/text/math_s4_v4_article-03.pdf. [Accessed: 06-Nov-2020].

[16] B. M. Bolker and B. T. Grenfell, “Impact of vaccination on the spatial correlation and persistence of measles dynamics,” Proc. Natl. Acad. Sci. U. S. A., vol. 93, no. 22, pp. 12648–12653, 1996.

[17] J. Arino and P. van den Driessche, “A multi-city epidemic model,” Math. Popul. Stud., vol. 10, no. 3, pp. 175–193, 2003.

[18] J. Arino, R. Jordan, and P. van den Driessche, “Quarantine in a multi-species epidemic model with spatial dynamics,” Math. Biosci., vol. 206, no. 1, pp. 46–60, 2007.

[19] Europepmc.org. [Online]. Available: https://europepmc.org/article/med/14766090. [Accessed: 06-Nov-2020].

[20] Y. Muroya, Y. Enatsu, and T. Kuniya, “Global stability of extended multi-group sir epidemic models with patches through migration and cross patch infection,” Acta Math. Sci. Ser. B Engl. Ed., vol. 33, no. 2, pp. 341–361, 2013.

[21] J. T. Wu et al., “Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China,” Nat. Med., vol. 26, no. 4, pp. 506–510, 2020.

[22] J. Kucharski et al., “Early dynamics of transmission and control of COVID-19: a mathematical modelling study,” Lancet Infect. Dis., vol. 20, no. 5, pp. 553–558, 2020.

[23] D. K. Das, S. Khajanchi, and T. K. Kar, “The impact of the media awareness and optimal strategy on the prevalence of tuberculosis,” Appl. Math. Comput., vol. 366, no. 124732, p. 124732, 2020.

[24] Windarto, 1 Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya 60115, Indonesia, M. Altaf Khan, Fatmawati, and 2 Faculty of Natural and Agricultural Sciences, University of the Free state, South Africa, “Parameter

estimation and fractional derivatives of dengue transmission model," AIMS math., vol. 5, no. 3, pp. 2758–2779, 2020.

[25] E. Okyere, S. Olaniyi, and E. Bonyah, "Analysis of Zika virus dynamics with sexual transmission route using multiple optimal controls," Scientific African, vol. 9, no. e00532, p. e00532, 2020.

[26] S. Khajanchi, D. K. Das, and T. K. Kar, "Dynamics of tuberculosis transmission with exogenous reinfections and endogenous reactivation," Physica A: Statistical Mechanics and its Applications, vol. 497, no. C, pp. 52–71, 2018.