Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Predictive Analysis of COVID-19 Pandemic in India Based on SIR-F Model

Gopinath M. P.
Vellore Institute of Technology: VIT University

Satyam S. C. (satyamc9090@gmail.com)
Vellore Institute of Technology: VIT University https://orcid.org/0000-0003-3547-8955

Jenil S. M.
Vellore Institute of Technology: VIT University

Shashank P.
Vellore Institute of Technology: VIT University

Research Article

Keywords: Susceptible-Infected-Recovered (SIR), Susceptible-Infected-Recovered-Fatal (SIR-F), COVID-19

Posted Date: June 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-636672/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Predictive Analysis of COVID-19 Pandemic in India Based on SIR-F Model

M. P. Gopinath[1] • S. C. Satyam[1] • S. M. Jenil[1] • P. Shasank[1]

[1]School of Computer Science and Engineering, VIT, Vellore, India

Abstract

Impact of COVID-19 has been devastating worldwide, it has disrupted lives of people, economy has fallen and millions of people have lost their jobs. Second peak of COVID-19 is making it even worse for many developing countries. In India, the second peak is reaching nearly 4 times the cases that were reported during the first peak, thus making it a challenge for the government to plan for the future without affecting the economy further. According to WHO, millions of enterprises are at existential threat, nearly half of the global workforce is at risk of losing their jobs and the entire food chain has been disrupted. In order to provide some assistance to the situation, this study aims to use the SIR-F Model, which is a variation of the SIR model. W. O. Kermack and A. G. McKendrick in 1927, proposed the model in which they classified a fixed population into three compartments: S(t), susceptible; I(t), Infected; R(t), Recovered. We have used the SIR-F model which differentiates between Recovered (meaning people infected and later recovered and thus now immune) and Fatality. We have simulated two scenarios including one in which we study the impact of medicine on future cases and also inspected various parameters which shed light on reasons behind increasing and decreasing the number of covid cases in India. In the future, this work can be extended further to develop a completely new model to consider those cases in which people have recovered but are still at a risk of reinfection.

Keywords Susceptible-Infected-Recovered (SIR), Susceptible-Infected-Recovered-Fatal (SIR-F), COVID-19

1. Introduction

Novel Coronavirus, prominently known as Covid-19 has been affecting our life since the most recent one year. People are taking all precautions and fighting the infection with everything accessible or available to us. We have a benefit over our precursors as we have progressed Information Technology and the force of machines to help us battle it.

Associations and governments everywhere in the world have been utilizing various measures which best suits them. It has unquestionably been successful to spread mindfulness and track the spread of Covid19. In any case, it has not yet assisted us with getting back to regularity as we keep on battling the spread of the infection. With new infection strains arising all around the planet, this work has gotten more troublesome. With the assistance of safeguards, for example, social separating, wearing covers, cleaning, and preventive estimates executed by the enlistment of immunizations, governments around the globe have doubtlessly had the option to contain the spread and furthermore the death rate. Be that as it may, the battle isn't finished at this point as the infection is still on the loose.

As of 30 May 2021, around 171 million individuals have been tainted by Coronavirus and around 3.5 million people have passed away due to the infection around the world. The infection's significant

casualties are nations like India, US, Brazil, Russia and European nations. At the point when we talk about India, we have recorded 27.8 million cases and 325 thousand deaths till date. That represents about 16.3% of the world's caseload and 9 percent of deaths worldwide (deaths caused by Covid-19) with India's population representing 17.7 percent of the total population in the world. This shows that the Indian medical care framework, however thought to be inefficient, has performed better compared to advanced western medical services frameworks.

There has been significant research on how to deal with the novel coronavirus. Numerous mathematical and predictive models have been built to forecast the future of the coronavirus disease, but many of them failed to produce predictions for long term scenarios. To provide better outcomes, there is a need of predictive models which are more efficient and accurate for long term prediction scenarios.

Models can be valuable instruments which, however, ought not to be overinterpreted, especially for long haul projections or inconspicuous attributes, for example, the specific date of a peak number of infected cases. In the first place, models should be dynamic and not fixed to take into account significant and unexpected impacts, which makes them just helpful for the time being if exact expectations are required. The second most important presumption ought to be plainly expressed and the affectability to these suppositions should be examined. Different variables that are now known or thought to be related with the pandemic, however excluded from the model, ought to be outlined along with their subjective ramifications for model execution. Third, instead of giving fixed, exact numbers, all figures from these models ought to be straightforward by revealing reaches (like CIs or vulnerability stretches) with the goal that the inconsistency and vulnerability of the forecasts is clear. It is pivotal that such spans represent all expected wellsprings of vulnerability, including information detailing blunders and variety and impacts of model misspecification, to the degree conceivable. Fourth, models should fuse proportions of their accuracy as better information opens up. On the off chance that the projection from a model varies from other distributed forecasts, it is imperative to determine such contrasts. Fifth, the public revealing of assessments from these models, in logical diaries and particularly in the media, should be suitably cautious and incorporate key provisions to evade the confusion that these forecasts address logical truth [1].

Models ought to likewise try to utilize the most ideal information for local predictions. It is impossible that plagues will follow indistinguishable ways altogether in areas of the world, in any event, when significant factors, for example, age appropriation are thought of. Local information ought to be utilized when that information becomes accessible with sensible precision. For projections of emergency clinic needs, information on clinical results among patients in neighbourhood settings are probably going to empower more precise ends than inadequately announced mortality information from across the world [1].

We have taken data from Kaggle and John Hopkins University Center for Systems Science and Engineering (JHU CSSE). After the surge of Covid-19 around the world there has been a lot of uncertainty about the future of the virus and its trend. We make an attempt to resolve this by making a predictive model which forecasts future covid cases. This kind of study is lacking in India and it would be helpful for concerned authorities to counter and prepare for future impact of the same. We tried to create a reasonable and realistic scenario through SIR-F model which is a variation of SIR Model to help associations and governments battle the Covid-19 pandemic, take informed steps and plan better for future pandemics and medical care crises.

2. Related Works

This section provides a brief literature review of SIR and various other models for future forecasting of covid-19 cases.

The SIR-D model is built on four differential equations to show the evolution of epidemiology. The model is based on the notion that on any given day, the total population can be classified into four categories: someone who is susceptible to infection (S), individuals who are already infected or active patients (I), anyone who has recovered (R), and those who have died (D). It is standard procedure to disregard the daily birth and death rate, generally known as demographics. The SIR(D) model also implies that the values S, I, R, and D are solely time-dependent and not location-dependent [3].

SEIR and Regression models are two widely utilised machine learning methods for virus infection forecasting around the world. Susceptible (S), Exposed (E), Infected (I), and Recovered (R) are the four primary components of the SEIR model. S stands for susceptible individuals (those who are able to contract the virus), E stands for exposed persons (those who came in contact to infected but are not yet infectious), I is for infected persons (those who can transfer the virus), and R stands for recovered individuals (those who are now immune to virus) [4].

Calculating the R_0 value is the most key component of this model. The value of R_0 indicates how infectious a virus is. It is the primary purpose of epidemiologists who are investigating a new case. In simple terms, R_0 determines the average number of individuals who can be affected over time by a single infected individual. If the value of R_0 is less than one, the spread is predicted to come to an end. If the value of R_0 equals 1, the spread is stable or endemic. If $R_0 > 1$, this indicates that the spread is widening in the absence of involvement [4].

To minimise COVID-19 mortality and healthcare demand, the Medical Research Council (MRC) Centre for Global Infectious Disease Analysis utilised a Non-Pharmaceutical Intervention (NPI) model [5]. SEIR was used in their NPI model. In an unmitigated scenario, the NPI model estimated 2.2 million deaths in the United States. Similarly, Columbia University's Severe COVID-19 model & Mapping Tool projected the number of serious cases, hospitalizations, intensive care, ICU use, and deaths for 3-week and 6-week periods beginning April 2 under various social distancing scenarios [6, 7]. CHIME: COVID-19 Hospital Impact Model for Epidemics [8] is the name of a prediction model developed by the University of Pennsylvania. Users can change the inputs and parameters in their model. They anticipated best- and the worst-case scenarios for total number of hospitalizations, ICU bed demand, ventilator demand, and number of days these needs would exceed hospital capabilities over the next three months in their estimates [9].

3. Methodology

3.1. The SIR Model

W.O Kermack and A. G. McKendrick in 1927 proposed a model in which a fixed population is divided into three zones namely, Susceptible S(t), Infected I(t) and Recovered R(t).

Let's understand with an example, Susceptible people might come in contact with Infected persons and thus be confirmed as Infected. Therefore, patient will shift from Susceptible zone to Infected zone. Later on, Infected patient will move to Recovered zone after recovering.

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

Variables:

S: Susceptible = Population – Confirmed Cases
I: Infected =Confirmed Cases - Recovered Cases – Fatal Cases
R: Recovered or Fatal = Recovered Cases + Fatal Cases

Parameters:
β: Effective contact rate [1/min]
ϒ: Recovery + Mortality rate [1/min]

After simplifying the model by removing the units of the variables from the ODE model.
Set $(S, I, R) = N \times (x, y, z)$ & $(T, \beta, \gamma) = (\tau t, \tau^{-1}\rho, \tau^{-1}\sigma)$.

N: refers to total population
τ: refers to a coefficient ([min], an integer to simplify).

3.2. The SIR-D Model

In SIR-D Model number of fatal cases and recovered cases are considered separately. Here, we use two variables Recovered (R) and Deaths (D), instead of "Recovered + Deaths" in the mathematical model.

$$\mathrm{S} \xrightarrow{\beta I} \mathrm{I} \xrightarrow{\gamma} \mathrm{R}$$
$$\mathrm{I} \xrightarrow{\alpha} \mathrm{D}$$

Variables:
S: Susceptible = Population – Confirmed Cases
I: Infected =Confirmed Cases - Recovered Cases – Fatal Cases
R: Recovered
D: Deaths

Parameters:
α: Mortality rate [1/min]
β: Effective contact rate [1/min]
ϒ: Recovery rate [1/min]

3.3. The SIR-F Model

Initially in COVID-19 outbreak many cases were discovered after they died, to consider this issue following change should be done "S + I → Fatal + I". This model can be termed as SIR-F model. When α_1 is 0 then there will be no difference between SIR-F and SIR-D model.

$$\mathrm{S} \xrightarrow{\beta I} \mathrm{S}^* \xrightarrow{\alpha_1} F$$
$$\mathrm{S}^* \xrightarrow{1-\alpha_1} \mathrm{I} \xrightarrow{\gamma} \mathrm{R}$$
$$\mathrm{I} \xrightarrow{\alpha_2} \mathrm{F}$$

Variables:
S: Susceptible = Population – Confirmed Cases
S*: Confirmed and uncategorised
I: Confirmed and categorised as Infected

R: Confirmed and categorised as recovered
F: Confirmed and categorized as Fatal

Parameters:
α_1: Direct fatality probability of S* (non-dimensional)
α_2: Mortality rate in Infected cases[1/min]
β: Effective contact rate [1/min]
ϒ: Recovery rate [1/min]

S*: Cases who are carriers of the disease but don't know about it including themselves, who either die later and are confirmed after death, while some are later moved to infected after being confirmed.

Setting $(S, I, R, F) = N \times (x, y, z, w)$ & $(T, \alpha_1, \alpha_2, \beta, \gamma) = (\tau t, \theta, \tau^{-1}\kappa, \tau^{-1}\rho, \tau^{-1}\sigma)$
After this ODE is,

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

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \rho(1-\theta)xy - (\sigma+\kappa)y$$

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

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \rho\theta xy + \kappa y$$

3.4. Reproduction Number for SIR-F Model

R_0, "R naught" average number of secondary infections that can be caused by an infected individual. When $x = \frac{1}{R_0}, \frac{\mathrm{d}y}{\mathrm{d}t} = 0$ the number of infected cases does not change.

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

4. Results and Discussion

4.1. Covid-19 Situation Analysis in India

As many researchers are talking about the coming third wave it is important to examine daily cases of India. Currently, as per Fig. 1, we can see the covid cases are decreasing, thus it is now an important aspect to calculate the R_0, Reproduction number value to get a broader perspective.

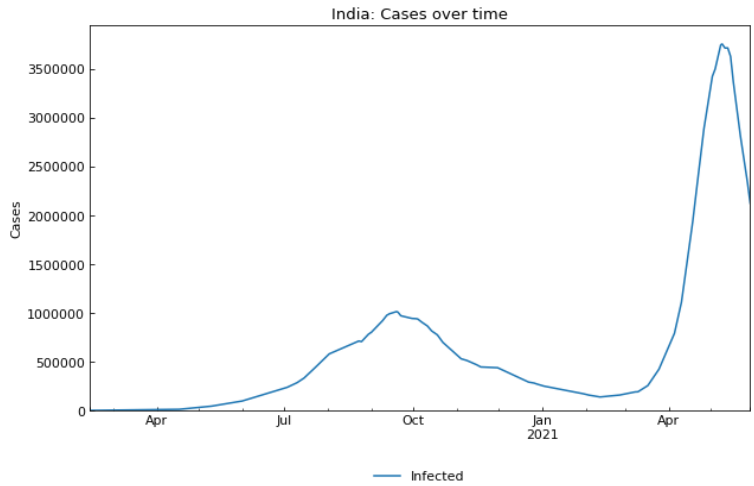

Fig. 1 Daily Cases in India overtime since first peak

India currently carrying out the largest vaccination program of the world and we can clearly see as per the Fig 2. Vaccination Drive is in full swing and thus it will have a positive impact and thus reducing further cases in India.

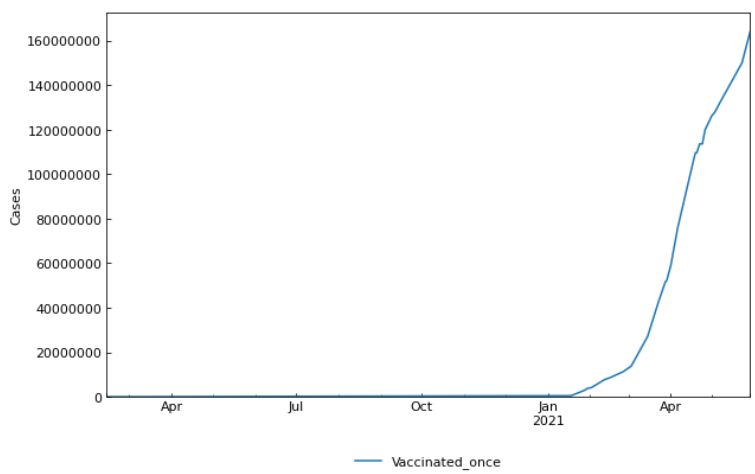

Fig. 2 Situation of total vaccination given in India

4.2. Evaluation of Estimation Accuracy

Accuracy of parameter estimation can be evaluated with RMSLE (Root Mean Squared Log Error) score.

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

Where *A* is the observed (actual) values, *P* is estimated (predicted) values. Variables are $S(i = 1)$, $I(i = 2)$, $R(i = 3)$ and $F(i = n = 4)$ for SIR-F model. When RMSLE score is low, hyperparameter estimation is highly accurate.

As we can see in Fig. 3 by visualising the accuracy on the last phase (19 May 2021 – 30 May 2021) of cases, the simulated cases and actual cases are quite accurate, especially accuracy of the Infected and Fatal cases increased over time.

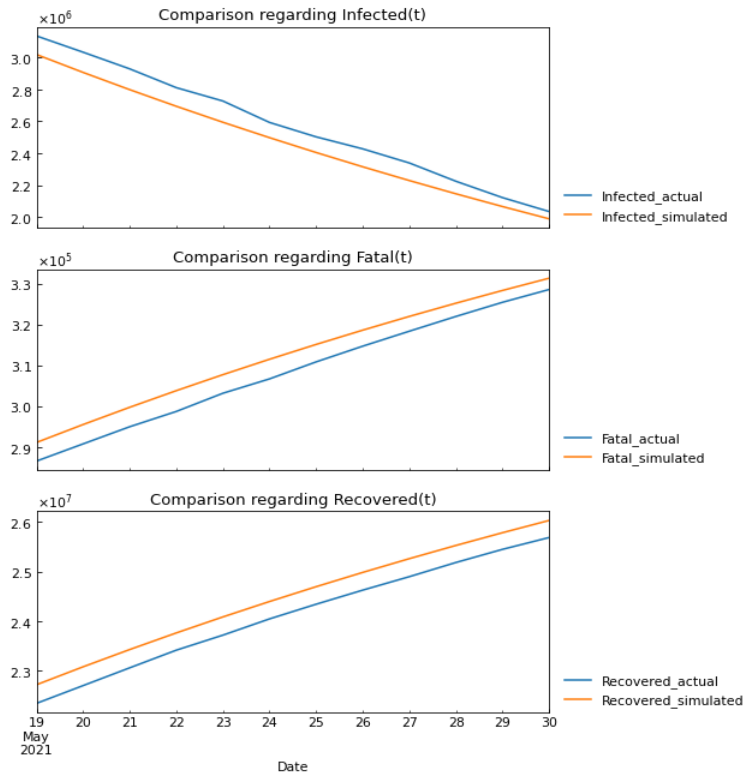

Fig. 3 Visualisation of Estimation Accuracy between 19 May 2021 – 30 May 2021

Total accuracy score received by ODE Models is MSLE = 0.049 and RMSLE = 0.221

4.3. Parameters Values

India is currently in Second wave, recovery period was large during the first peak because of the collapse of medical care system and No medicine was found till then.

We can observe in Fig. 4, Sigma values during first peak (Oct 2020) of India was less than the Sigma values during second peak (May 2021).

As cases are reducing very rapidly during the second peak it is expected to take less time to recover from the peak because of improved medical structure than before. The faster and more efficient vaccination drive will further have a positive impact.

Reproduction number indicates how many individuals can be transmitted from an infected host. If its value is less than one, then the virus is getting contained. We can clearly see from Fig. 5 that near the month of September and October last year (2020), reproduction number was nearing one and later went lower than one and downtrend was observed during the same peak.

Similarly, cases started increasing during month of march and thus simultaneously we can see its effects on Reproduction Number indicating value greater than one. Nearing the end of the month

of April, we can again see the Reproduction number falling to less than one, thus indicating further downtrend in daily cases.

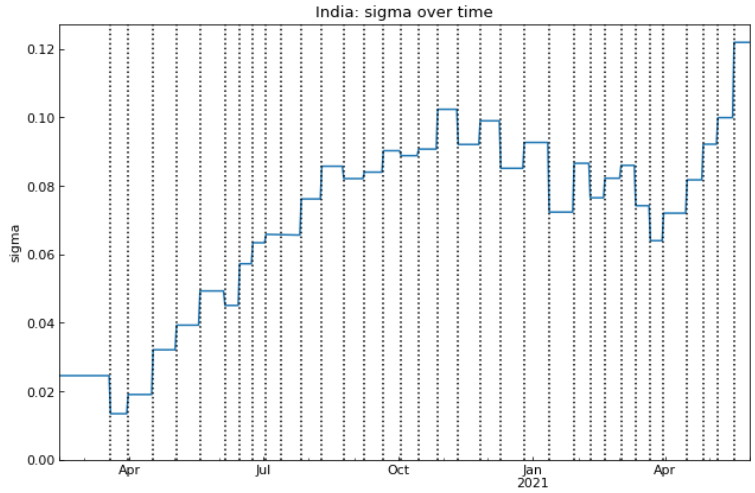

Fig. 4 Sigma values over time

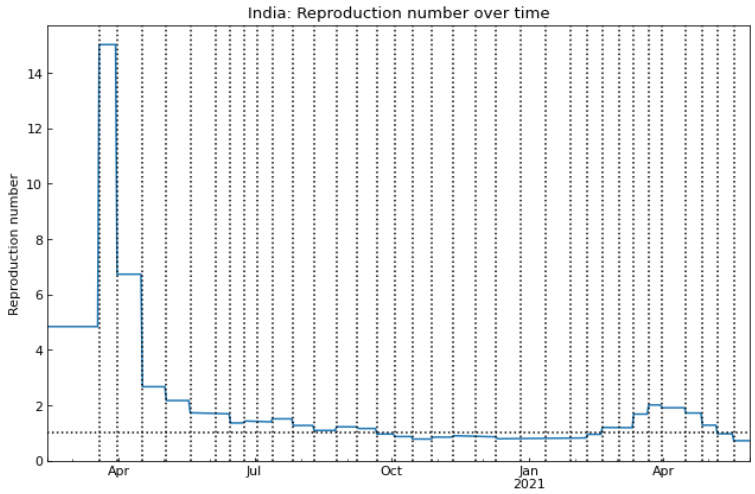

Fig. 5 Reproduction number over time

4.4. Analysing Simulated Cases

We simulated two scenarios; the first scenario named "main", where we assume parameter values will not change after the last phase (19 May 2021 – 30 May 2021). That is, if parameter values are not changed till 30 Jun 2021, then how many cases will be observed in the future. And now, as in after the first peak, medicines are also available. Thus, to consider the effect of medicine on total number of cases overtime we simulated the second scenario named "medicine". To simulate this scenario, we assume σ will change in future phases. We considered how many cases will be there in the near future after the last phase (30 May 2021) if σ will increase 1.2 times every 30 days.

In Fig. 6, the actual cases and the simulated (Main Scenario) cases are plotted. It can be clearly seen that the simulated cases are precisely overlapping the actual cases with little to no difference.

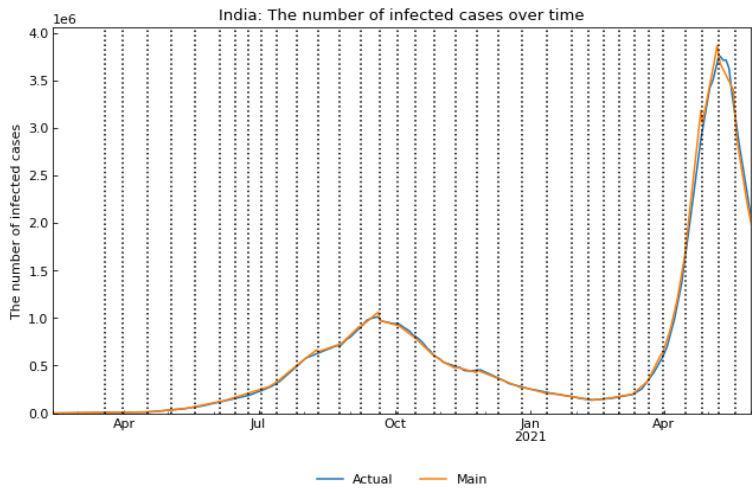

Fig. 6 Actual vs Simulated number of infected cases over time

In the "Main" Scenerio, we simulated the number of cases if parameter values are not changed for next 30 days. According to the model, by the end of June 2021 we will have a of total Infected cases around 6 lakhs, total fatal cases around 3.8 lakhs and total recovered cases around 3.04 crores. The actual numbers genertated by the model are shown in Table 1.

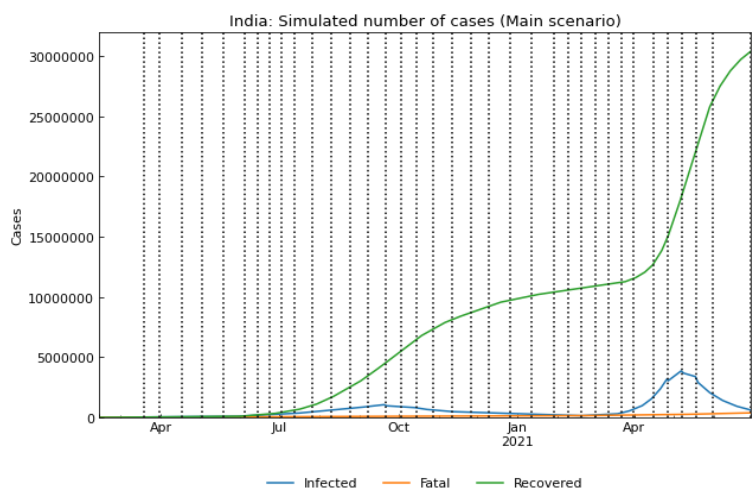

Fig. 7 Simulation of number of cases for Main scenario till 30 Jun 2021

Date	Infected	Fatal	Recovered
30 – Jun – 2021	608970	384784	30426937

Table 1 Simulated Number of Cases generated by model till 30 Jun 2021

After the optimistic results shown by the SIR-F model in the “Main” scenario we clearly hope medicine will make a huge impact in coming months as vaccine drive is on full swing, medical care centres are at a better place when compared to the first peak and thus recovery time is precisely expected to reduce and it is equally complemented by the model in the “Medicine” Scenario. In Fig. 8, we can clearly see what an impact Medicine will make on the future number of infected cases.

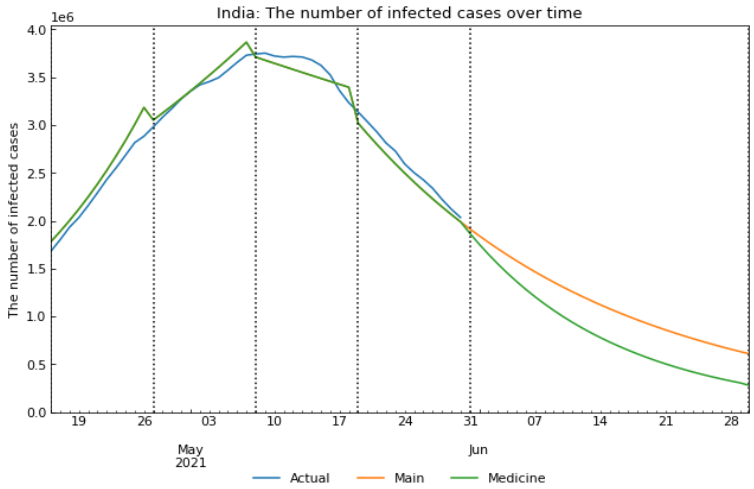

Fig. 8 Effect of medicine on number of infected cases when compared to "Main" Scenario

5. Conclusion

The COVID-19 Pandemic is a global challenge for all the governments and associations. We believe it can be contained using sufficient measures. We have an advantage over all the past epidemics and pandemics as we have better means to collect data and thus improve the capability to gain insights from it. We observed in the month of march that the reproduction number (R_0) hiked above the value one; hence there was a sudden surge in the number of infected cases. To contain the virus, reproduction number should be less than one, which we saw near the mid of month may (2021) thus showing the downtrend in daily number of cases. We are expecting better recovery period when compared to the first peak due to several reasons including availability of medicines, vaccination drive and better medical facilities when compared to the first peak. This was precisely complemented by the simulation carried out in the “Medicine” Scenario, in which a huge impact was observed. The SIR-F model generated very realistic results with RMSLE score of 0.221. When the RMSLE score is low, hyperparameter estimation is highly accurate.

SIR-F model generated optimistic results, but, in the future, this work can be extended to consider economic conditions, medical facilities, workplace restrictions, testing policy, international and national movement restrictions, contact tracing, vaccine effectiveness and other factors to complement the study and generate more efficient predictions. All countries are at different stages ; for example India is currently at its second peak while Japan is facing its fourth peak with each peak higher than the previous peak while the peak in US forms a different trend. Hence, we can use data from different countries to predict results for other countries. For example, data from Japan can be used to predict the third peak in India and thus it will help the government to take informed decisions and might even help to contain it before a huge surge.

The developed model could be very helpful for government and healthcare departments for combatting the COVID-19 pandemic. Experienced healthcare officials and researchers could even tweak several parameters such as σ (sigma) to enhance the results of the SIR-F model.

Funding

The authors have no relevant financial or non-financial interests to disclose.

Compliance with ethical standards

Conflict of interest: Authors have no conflict of interest to declare.

Ethical approval: This article does not contain any studies with human participants performed by any of the authors.

Availability of data and material

All the data generated in this research is made available at GitHub repository link1 - https://github.com/satyam9090/SIR-F-COVID-19/tree/main/data

Authors' contributions

All authors have contributed equally.

References

1. Kotwal A, Yadav AK, Yadav J, Kotwal J, Khune S. "Predictive models of COVID-19 in India: a rapid review." Medical Journal Armed Forces India. 2020 Oct 1;76(4):377-86.
2. Ndiaye BM, Tendeng L, Seck D. "Comparative prediction of confirmed cases with COVID-19 pandemic by machine learning, deterministic and stochastic SIR models." arXiv preprint arXiv:2004.13489. 2020 Apr 24.
3. Rajesh A, Pai H, Roy V, Samanta S, Ghosh S. "CoVID-19 prediction for India from the existing data and SIR (D) model study." medRxiv. 2020 Jan 1.
4. Gupta R, Pandey G, Chaudhary P, Pal SK. "Machine learning models for government to predict COVID-19 outbreak." Digital Government: Research and Practice. 2020 Aug 17;1(4):1-6.
5. Ferguson, Neil, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia et al. "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand." (2020).
6. Branas, Charles C., Andrew Rundle, Sen Pei, Wan Yang, Brendan G. Carr, Sarah Sims, Alexis Zebrowski et al. "Flattening the curve before it flattens us: hospital critical care capacity limits and mortality from novel coronavirus (SARS-CoV2) cases in US counties." *medRxiv* (2020).
7. Mapping tool: https://cuepi.shinyapps.io/COVID-19/. Columbia University (June 23, 2020, last accessed)
8. CHIME v1.1.5 (2020-04-08): https://penn-chime.phl.io. COVID-19 Hospital Impact Model for Epidemics (CHIME). University of Pennsylvania (April 2020)
9. Santosh KC. "COVID-19 prediction models and unexploited data." Journal of medical systems. 2020 Sep;44(9):1-4.
10. CovsirPhy Development Team (2020-2021), CovsirPhy version 2.20.2: Python library for COVID-19 analysis with phase-dependent SIR-derived ODE models, https://github.com/lisphilar/covid19-sir
11. Hirokazu Takaya (2020-2021), Kaggle Notebook, COVID-19 data with SIR model, https://www.kaggle.com/lisphilar/covid-19-data-with-sir-model