Coronavirus Spreading Forecasts Based on Susceptible-Infected-Recovered and Linear Regression Model

Neha Tyagi, Naresh Dhull, Meenakshi Sharma, Vishal Jain, and Shashank Awasthi

Abstract An epidemic environment is an occurrence of disease that spreads rapidly and disturbs many persons at the same time. Well, transmission of viruses is a big problem of today’s era. Environmental conditions and habits of people are somehow responsible for the vulnerable conditions of the environment. In consequence, to find out the exact condition the study needs to know about the symptom of the person (like age, gender, current body temperature, dry cough duration, headache, and travel history). As per the study given by India Today magazine, a very first case of coronavirus was found in Kerala’s Thrissur district in India on January 30 2020. Thus, after 4–5 days, another two cases were found in Kerala too. Therefore, the state government took action and put these 3 patients in Isolation. Accordingly, with these 3 patients, the government quarantined their contacted persons (around 3400) too, who were suspected of symptoms of the coronavirus. Thereby, after the isolation and quarantine period, 3 patients among all were discharged upon recovery. Therefore, in this way, this study has to focus and trying to visualize the impact of isolation and quarantine on the health of the patients.

Keywords COVID-19 · Epidemic theory · SIR · Linear regression and coronavirus

N. Tyagi (✉) · S. Awasthi
G.L. Bajaj Institute of Technology and Management, Greater Noida, India
E-mail: nehacs1988@gmail.com

S. Awasthi
E-mail: shashankglbitm@gmail.com

N. Dhull · M. Sharma
Galgotias University, Greater Noida, India

V. Jain
BharatiVidhyapeeth’s Institute of Computer Applications and Management, New Delhi, India

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2020
A.-E. Hassanien et al. (eds.), Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach, Studies in Big Data 78, https://doi.org/10.1007/978-3-030-55258-9_1
1 Introduction

The 2020 year, which unstable the lives on the earth, will be considered in the history of pandemics. It was first reported by the Wuhan Health commission, in China on 31st December 2019. From the 31st December 2019, it blowouts all over the world and damage the lives of male, female, kids, etc. The pandemic COVID-19 may be a major world health threat. The novel coronavirus have been rumored because of the most damaging metabolism virus since the 1918 contagion pandemic. Consistent with WHO novel coronavirus state of affairs report as on March 2020, a complete of around 600, more than 150 confirmed cases and more than 300 deaths are rumored across the planet [1]. Therefore, World unfolds has been fast, with More than 150 countries currently having rumored a minimum of one case. Coronavirus sickness (COVID-19) is an associate in nursing communicable disease caused by severe acute metabolism syndrome coronavirus-2. Coronavirus belongs to a family of viruses that is to blame for unhealthiness starting from respiratory disease to deadly diseases as MERS and SARS that were 1st discovered in China (2002) and Saudi Arabia (2012).

In view of this, the novel Coronavirus or higher referred to as COVID-19 was rumored in a metropolis, China for the 1st time on thirty-first Gregorian calendar month 2019. Consistent with Jiang et al. the death rate for this virus has been calculable to be 9/5 except for the people 80, this has gone up to 0.086 whereas for those >80 it’s been noted to be 15% [2]. This has junction rectifier to aged persons higher than the age of 40 plus with underlying diseases like polygenic disorder, Parkinson’s sickness, and disorder to be thought-about at the best risk. Symptoms of this sickness will take 5–20 days to look and might vary from fever, cough, and shortness of breath to respiratory illness, nephropathy, and even death [1]. The spread is one individual to another via metabolism droplets among shut contact with the typical variety of individuals infected by a patient being 4–5% however the virus isn’t thought-about mobile [3]. Choice makers are benefited from a higher geared towards MC outputs complemented through min-max policies that foretell about the acute degrees of destiny possibilities with admiring to the epidemic [4].

Although, there exist an outsized variety of pieces of evidence wherever machine learning algorithms has proved to offer economical predictions in care [5–7]. Nsoesie et al. have provided a scientific review of approaches accustomed forecast the dynamics of contagion pandemic [8]. Similarly, they need reviewing analysis papers supported settled principle models, regression models, prediction rules, Bayesian network, SEIR model, ARIMA statement model, etc. Recent studies on COVID-19 embrace solely searching analysis of the offered restricted knowledge [9–11].

Here, the Prediction Module is projected for predicting the power of the patient to reply to treatment supported various factors e.g. age, infection stage, metabolic process failure, multi-organ failure, and therefore the treatment regimens. The Module implements the Whale improvement rule for choosing the foremost relevant patient options [12].
Thus, operative and effective vaccinum against novel coronavirus has not been unreal and thus a key half in managing this pandemic is to decrease the epidemic peak, conjointly referred to as flattening the epidemic curve. Here, the role of knowledge of scientists and data processing researchers is to integrate the connected data and technology to higher perceive the virus and its characteristics, which may facilitate in taking right selections and concrete arranges of actions. As a result in an even bigger image of taking aggressive measures in developing infrastructure, facilities, vaccines, and restraining similar to epidemic in the future. The objectives of this study area unit are following as designing a SIR (Susceptible, Infectious, Recovered) models to judge the unfold of sickness; Concluding the speed of unfolding of the sickness in the Republic of India &Forecasting of COVID-19 eruption exploitation SIR and Regression models.

Accordingly, the upcoming future is much unknown about this virus. The scientists, researchers, many peoples only can do predictions based on data given by the World health organization to the world.

The Introduction and Theory part given in Sect. 1, In order to this, the model used in this study illustrate in Sect. 2. In such a way, Analysis, performance evaluation and experimental results explicate in Sect. 3. Section 4 enlightens the Methodology discussion followed by the conclusion in Sect. 5.

2 Pandemic Theory

Koykul et al. examine about focal points and inconveniences of utilizing savvy lattice in circulated restoration vitality age [13]. Pipattanasomporn et al. talk about giving knowledge to a savvy network through a multi-operator framework [14]. The ideas remain identified with one another and bolster a ton of work just as parts. Generally the collaboration among supplier and performer is rapidly however here and there the change in encompassing may influence the communication of the supplier and the performer. Similarly, the plague hypothesis is likewise supportive to depict the idea of the spread of worm from source to a goal by characterizing the infection just as contortion or clamor in data. Likewise the survey study can say that the plague hypothesis depicts the proliferation of infection or mutilation or commotion in the data. Generally, the data consistently transmit or get regularly, so if the worm is engendered in a similar normal way, at that point the infection additionally proliferates similarly because the immediate contact characterizes the spread of infection using a typical frequency.

On condition that, there should be an occurrence of investigation of science, during the compelling contamination at the masses scale, two philosophies or ideas are portraying the engendering of disease which is characterized as the stochastic procedures and the deterministic techniques. Generally, a mass of n individuals is divided into a couple of compartments, and the spread of the illness is thought about (Fig. 1).
Fig. 1 Different types of epidemic models

3 SIR Model

The Kermack-McKendrick Model is very well known as the SIR model. Here for simulation, the study has used 3 variables, S–I–R to symbolize the total number of individuals in an epidemic environment at a given frame. Where susceptible individuals will depict by S, Infected individuals will depict by I and Recovered individuals or can say that the removed individuals from an epidemic environment that recovered by isolation or quarantine process by R.

Hence, these 3 kinds of persons may vary after some time frame. So, ‘t’ will represent that time frame. The final number of functions of a specific time frame will be S (t); I (t) & R (t). In the same way, the SIR model is dynamic because of the number of people changing their state after the specified interval of time ‘t’.

On that account, to find out the value of conversion rate of I and S, then have to use the formula \( \beta \frac{I}{N} \). \( \beta \) depicts the ordinary number of individuals contact per person per time that has to be multiplied by the disease spreading probability. Thus, the fraction of contact with individuals is depicted by \( I/N \).

Where \( \gamma \) is the conversion rate of S and I i.e. rate of recovery. D is the duration of infection, at that juncture \( \gamma \) will depict by \( 1/D \).

At this moment, a very important measurement of any disease prototypical is the R-Naught. It is the reproduction number of secondary contagions generated from one contagion individual in a total populace. As per the SIR model, \( R_0 = \frac{\beta N}{\gamma} \). The total population size will represent by \( N = S + I + R \). In order to find out the value of \( \beta \), then \( \beta = R_0 \gamma / N \). \( R_0 \) is property of breakout and does not change once calculated, as it’s calculated assuming inclined inhabitants. The efficient copy quantity, “R” represents the typical collection of secondary infections in an inhabitant that can come with immune individuals. That is the metric this is wanted to be minimized via vaccination campaigns [15].
Coronavirus Spreading Forecasts Based on Susceptible …

Thus, this is a very basic model but has effective assumptions. Very firstly, as a populace is shut and fixed, as such—nobody it included into the helpless gathering (means no births), all people that progress of from being contaminated to recuperated is for all time impervious to disease and there are no passings.

Similarly, as the populace considered as homogenous and just vary by their ailment state. At last third one, disease and that person’s “infectiveness” or capacity to defect powerless people, happens at the same time [16, 17]. This section describes the simulating part of SIR model.

3.1 Simulation of SIR (the Susceptible-Infected-Removed) Model Using R Packages

As anticipated, the simulation study has used the deSolve R package for this simulation prediction. The study has used different variables here $\gamma$ & $\beta$ & $t$ for several days. Here this research work had plotted the graph at 14 days for higher goals [15, 18].

Here, consider the R0 that is reproduction number. Because simulation found the data which is going to change day by day from COVID-19. As it a pandemic data onto the country, so, R0 represents the no. of cases produced by an infected person during the transmissible period. Thus, R0 is equal to cpd. So, with the help of R0, the value of $\beta$ and exact information about the isolation period of the COVID-19 data can be calculated. For example, initially consider as the transmissible duration is equivalent to the period of contamination. It is notable that if one person is infected by a coronavirus, so that person will not infect the other persons immediately. It will take between 3 and 5 days as an incubation period and able to flacking the symptoms of coronavirus. For this reason, have to find out the value of $\beta$, i.e. $\beta = R0/\gamma$, where, $R0 = 2.26$ [15, 18].

Here, initially set the data onto 3 infected individuals. Then we will run R function of 14 days. The Study says that if R0 > one, the infection rate is greater than the recovery rate, and thus the infection will grow throughout the population. If R0 < one, the infection fast will die out since people are health-giving faster than they are transmitting it [15, 18].

Accordingly, Fig. 2 represents the stepwise step model for Susceptible-Infected-Removed, where, S (Susceptible), I (Infected), and R(No. of recovery people). Here this study is checking the top ten results come after applying the model on the discussed case. Findings have plotted the visualization by using the ‘t’ parameter for 14 days [15, 18].

As a result, in Fig. 3, $\beta$ gives the typical choice of contacts in line with the particular person in line with time multiplied by way of the likelihood of the disease spreading in a touch between I and S and $\gamma$ is the conversion rate between I and S. Consequently, on the assumption of $S = 0.9899906$ I = $9.000750e−03$ & R = $1.008659e−03$, when $\beta = 1.4247$ and $\gamma = 0.14286$ [15, 18].

In order to this, Figs. 4 and 5, demonstrate the SIR model predictions in correspond
Fig. 2 Model for susceptible-infected-removed using R packages

\[
\text{SIR} \text{(14,0.32,1/7)} \\
\text{SIR} <- \text{function} \\
(\text{time, state, parameters}) (\text{with(as.list(c(state, parameters))}) \\
\text{dS} <- (-β*S*T) \\
\text{dI} <- ((β*S*I) - (γ*I)) \\
\text{dR} <- γ*I \\
\text{return list(dS,dR, dI)} \\
β = 1.427, γ = 0.14286
\]

Fig. 3 Depicting the susceptible-infected-removed (SIR) model by using $\beta$ and Susceptible (Blue), Infected (Red) and Removed (Green)

to the isolation period of the patients. This study also focuses on the isolation period which creates a great impact for the recovery of COVID-19 patients.
**Fig. 4** SIR model in respect of time of isolation of populace

**Fig. 5** SIR model representation in correspond to whole populace. The time span of simulation is 14 days
4 Linear Regression Model

Well, linear Regression approaches are numerical components of techniques that can approximation or are expecting a goal or variable that is dependent on the criteria of established variables. There are many types of Regression like Bayesian regression, polynomial regression, and linear regression, and so on. In this chapter regression type had used i.e. linear model for prediction of coronavirus pandemic. This model working is based on dependent and independent variables. The cost of slope and intercept is used to show the products of this model. Hence, in a linear regression model, $\beta_0$ and $\beta_1$ are used to signify an intercept & slope. Error rate will depict by $\epsilon$. $Y = \beta_0 + \beta_1x + \epsilon$ and $Y = \theta_0 + \theta_1x + \theta_2x^2 + \theta_3x^3 + \theta nx^n$, [19]. For this purpose, in the given equations depicts the value of coefficients that assigned to predictors & polynomial diploma has to be depicted by n. Figure 8 depicts the Linear regression model fitting line. By using this plotted figure author says that as the time goes the number of confirmed cases of COVID-19 also goes high.

Accordingly, for proposing the model, here author considers the pandemic COVID-19 as a case study/baseline from India. Hence, more than half of the India has already stopped commuting. India has replaced handshakes with Namaste and now, the new standard: social distance. Around the early of March month, WHO has been stated a COVID-19 is a pandemic.

By using Fig. 6 author gives outline of the processes used to predict the epidemic environment data (COVID-19) in India. Here, the Impact Analysis to study the epidemiological environment, the author has used nCov2019 data outbreak. It includes real-time data and historical data as well. Initially these simulations are deploying the packages to start the prediction work. By virtue of basic functions are used for extracting the data example as: get_nCov2019 () which is used to get the latest online information; load_nCov2019 () which is used to get the historical data; summary which is used to access the whole data and plot is used to visualize the whole data in the pictorial form. Subsequently, the deployment process, now install all the packages by using commands: Remotes::install_github (“GuangchuangYu/nCov2019”); require (nCov2019); require (dplyr). Initial Impression: Here author have fixed two Variables A and B. A $\leftarrow$ getnCov2019 () and B $\leftarrow$ load_nCov2019 (). Following this it will give the total confirmed cases of the country. Initially this study explores the required data then after prediction functions has applied on particular data [3]. Similarly, Fig. 7, elaborate that how COVID-19 data gets trained for predictions. Author gives learning from initial to N number of steps to find the optimal prediction results (Fig. 8).

![Deploy the packages](Exploring the Data)](Visuals Display)

**Fig. 6** Process followed for epidemic environment study
5 Comparison of SIR and Linear Regression Model

In Mentioned Table 1 is the Date-wise SIR, Linear Regression, and actual data comparison for 14 days; this chapter has used 14 days of training data onto prediction analysis. By using R packages for the SIR model depiction where with the help of this study found that $\beta = 1.4247$ and $\gamma = 0.14286$. Thereafter, Standard error in the linear regression model was 2.253 [17, 19]. The graphical assessment of Date-wise comparison of Susceptible-Infectious-Recovered, Linear Regression and actual data is given in Fig. 9. Subsequently, author finds out the linear regression model gives the
Table 1  Date-wise SIR, linear regression and actual data comparison for 14 days

| Date     | SIR  | Linear | Actual |
|----------|------|--------|--------|
| 25/03/20 | 648  | 604    | 562    |
| 26/03/20 | 720  | 740    | 649    |
| 27/03/20 | 874  | 820    | 724    |
| 28/03/20 | 974  | 923    | 873    |
| 29/03/20 | 1081 | 1040   | 979    |
| 30/03/20 | 1199 | 1172   | 1071   |
| 31/03/20 | 1333 | 1321   | 1251   |
| 1/4/2020 | 1485 | 1488   | 1397   |
| 2/4/2020 | 1654 | 1676   | 1965   |
| 3/4/2020 | 1837 | 1887   | 2301   |
| 4/4/2020 | 2038 | 2127   | 2902   |
| 5/4/2020 | 2264 | 2396   | 3374   |
| 6/4/2020 | 2520 | 2669   | 4067   |
| 7/4/2020 | 2807 | 3007   | 4421   |
| 8/4/2020 | 3122 | 3386   | 5194   |
| 9/4/2020 | 3465 | 3814   | 5734   |

Fig. 9  Comparative graph of susceptible-infectious- recovered and linear regression model of COVID-19 Data

very close results to the actual one scenario. Similarly, Fig. 10, shows the COVID-19, actual data depiction in India before the period of lockdown and after the period of lockdown. So, it shows the change in the number of confirmed cases too.
Fig. 10 Change in number of confirmed cases before and after lockdown period due to COVID-19 in India

6 Conclusion

This chapter focused on the COVID-19 initial case that occurs in India. Initially this chapter considered 3 cases of Kerala’s district on January 30 2020. Thereupon, the government took steps to isolate or quarantine those 3 as well as the people those who had contact with those 3 individuals. Later, that 3 people have been recovered by the process of isolation or quarantine. As a consequence, by using the SIR model in the particular case mentioned by author, exploration work found that the recovery rate has grown by using the isolation process for infected people. Initially set the value of $R$ is $0.000000e+00$ after the SIR model application will find $R = 1.008659e-03$. Therefore, it shows the Isolation and quarantine help to improve the health condition of COVID-19 infected people. Similarly, the problem has solved by using a linear regression model. Simulation results shows that the standard deviation between SIR and actual is ±15% and the standard deviation between linear regression and actual is ±7%. Thus, it has been concluded that linear regression model predicts the very near results of the actual data.

References

1. World Health Organization.: Coronavirus disease 2019 (COVID-19): situation report, 67 (2020)
2. Jiang, F., Deng, L., Zhang, L., Cai, Y., Cheung, C.W., Xia, Z.: Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). J. General Int. Med. 1–5 (2020)
3. Wu, Z., McGoogan, J.M.: Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA (2020)
4. Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction. Appl. Soft Comput. 106282 (2020)

5. Ye, Q.H., Qin, L.X., Forgues, M., He, P., Kim, J.W., Peng, A.C., Simon, R., Li, Y., Robles, A.I., Chen, Y., Ma, Z.C.: Predicting hepatitis B virus–positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. Nat. Med. 9(4), 416–423 (2003)

6. Mai, M.V., Krauthammer, M.: Controlling testing volume for respiratory viruses using machine learning and text mining. In: AMIA Annual Symposium Proceedings, vol. 2016, p. 1910. American Medical Informatics Association

7. Purcaro, G., Rees, C.A., Wieland-Alter, W.F., Schneider, M.J., Wang, X., Stefanuto, P.H., Wright, P.F., Enelow, R.I., Hill, J.E.: Volatile fingerprinting of human respiratory viruses from cell culture. J. Breath Res. 12(2), 026015 (2018)

8. Nsoesie, E.O., Brownstein, J.S., Ramakrishnan, N., Marathe, M.V.: A systematic review of studies on forecasting the dynamics of influenza outbreaks. Influenza Other Respir. Viruses 8(3), 309–316 (2014)

9. Pirouz, B., ShaffieeHaghshenas, S., ShaffieeHaghshenas, S., Piro, P.: Investigating a serious challenge in the sustainable development process: analysis of confirmed cases of COVID-19 (New Type of Coronavirus) through a binary classification using artificial intelligence and regression analysis. Sustainability 12(6), 2427 (2020)

10. More, G.D., Dunowska, M., Acke, E., Cave, N.J.: A serological survey of canine respiratory coronavirus in New Zealand. New Zealand Vet. J. 68(1), 54–59 (2020)

11. Wu, C., Chen, X., Cai, Y., Zhou, X., Xu, S., Huang, H., Zhang, L., Zhou, X., Du, C., Zhang, Y., Song, J.: Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. JAMA Intern. Med. (2020)

12. Diagnosis and Prediction model for COVID-19 patients response to treatment based on convolutional neural networks and whale optimization algorithm using CT image. medRxiv 2020.04.16.20063990; https://doi.org/10.1101/2020.04.16.20063990

13. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Comput. Netw. 38(4), 393–422 (2002)

14. Vahdat, A., Becker, D.: Epidemic routing for partially connected ad hoc networks. Duke University Technical Report CS-200006 (2000)

15. https://cran.rproject.org/web/packages/shinySIR/vignettes/Vignette.html

16. CSSEGISandData. CSSEGISandData/COVID-19 (2020) https://github.com/CSSEGISandData/COVID-19

17. Pandey, G., Chaudharya, P., Guptab, R., Pal, S.: SEIR and regression model based COVID-19 outbreak predictions in India (2020)

18. https://rpubs.com/choisy/sir

19. Gupta, R., Pal, S.K.: Trend analysis and forecasting of COVID-19 outbreak in India. Published online 30/03/2020 in preprint archive MedRxiv https://www.medrxiv.org/content/10.1101/2020.03.26.20044511v1 (2020)