

# Modelling Covid-19: Transmission Dynamics Using Machine Learning Techniques



Vijay Kumar, Jyoti Chawla, Rajeev Kumar, and Arti Saxena

## 1 Background

Pandemic disease Covid-19 is affecting the day-to-day life of people across the globe from the past few months. Scientists across various disciplines ranging from molecular biology to applied mathematics have teamed up for the assessment and control of this rapidly spreading virus. Mathematical models play a significant role in assessing, predicting and proposing potential outbreaks [1]. Mathematical models in pandemic of diseases have been recognized as an effective tool in analysing the propagation of infectious diseases and to test the complex dynamics of diseases to propose test strategies [2–4]. Epidemic coronavirus (Covid-19) transmission rate is found to be different in different regions and that may be due to different factors such as climate change, movement of individuals from one region to another, population density, different types of immune system, population pyramid, and antibiotics resistance. Mathematical modelling may can predict the disease transmission and mortality with recognition of the possible reasons of transmission of disease. It may also analyse the effect of intervention strategies for optimal control of transmission rate [5]. Many researchers have implemented the model on the situations for control of infectious diseases [6–8]. As per the data available till date, the transmission rate of Covid-19 virus is seen to be different in different countries. It has been observed that Covid-19 outcome trends in terms of number of infected individuals depend on various factors [9, 10]. Even with all healthcare facilities on place, the challenge is to decrease the spread and doubling time of diseases. It is very important that optimal interventions should be on place as per the severity of problem.

---

V. Kumar (✉) · J. Chawla · R. Kumar · A. Saxena

Manav Rachna International Institute of Research and Studies, Faridabad, Haryana, India

## 2 Objective

In this chapter, we have proposed a framework which employs machine learning to study the transmission of Covid-19. The objective of study is to highlight the transmission dynamics of the virus and monitor the transmission among top five countries with highest number of infected persons as on May 31, 2020 and predict the situation further. Linear regression techniques have been used for the purpose of analysis and prediction.

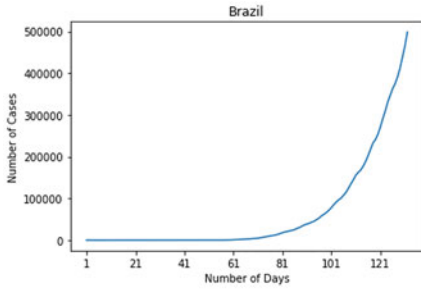
## 3 Methodology

Python has been used as the main programming language for analysis, and forecasting. Data have been taken from a reliable source (<https://www.kaggle.com/imdevskp/corona-virus-report>) [11]. Data have been pre-processed and recorded from 1 January 2020 to 31 May 2020. Experimental results have been illustrated by means of graphs and tables since the inception of the disease in the country. The fitting of the model is assessed by means of  $R^2$  statistics and residual.

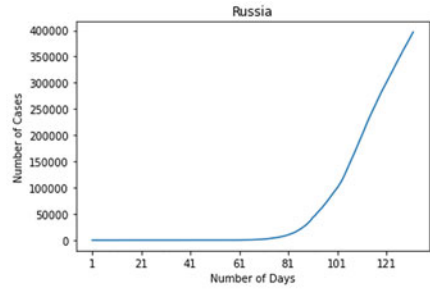
## 4 Results

The most affected 5 countries Brazil, Russia, Spain, the UK and the USA have been considered. After removing zero values, data were considered from 21 January 2020 for the purpose of analysis. The shape of data set is (5, 133), that is the profiling of number of active cases in 5 countries and 133 days. Profile of all countries and their comparison is shown in Figs. 1a–e and 2.

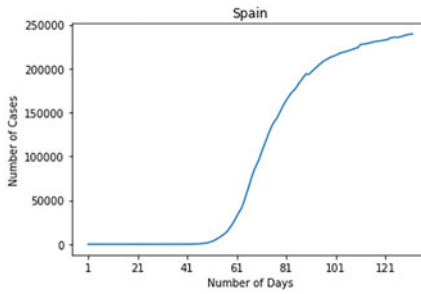
To calculate a good measure with 132 days, we have analysed the spread in an interval of 33 days as depicted in Figs. 3, 4, 5, 6 and 7 for all countries. Figure 3 represents the spread of Covid-19 from day 1 to 133 days in Brazil. From Fig. 3, it can be seen that from day 1 to day 45, there was no case of Covid-19 and it was increased from day 45 till the end of study (Fig. 3a–d). Figure 4 represents the spread of Covid-19 from day 1 to day 133 in Russia. From Fig. 4, it can be seen that from day 1 to day 11, there was no case of Covid-19, but after day 11, there was a drastic change in the number of cases and increased regularly till the end of the study (Fig. 4a–d). Figure 5 represents the spread of Covid-19 from day 1 to day 133 in Spain. From Fig. 5, it can be seen that from day 1 to day 11, there was no case of Covid-19, but after day 11, there was also a drastic change in the number of cases and increased regularly till the end of the study (Fig. 5a–d). Figure 6 represents the spread of Covid-19 from day 1 to day 133 in the UK. From Fig. 6, it can be seen that from day 1 to day 10, there was no case of Covid-19, but after day 10, there was a



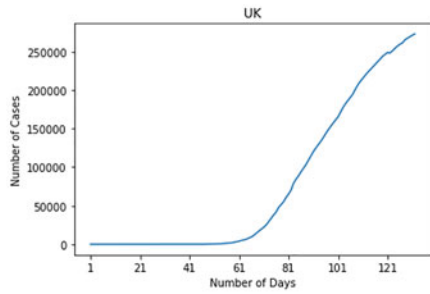
(a)



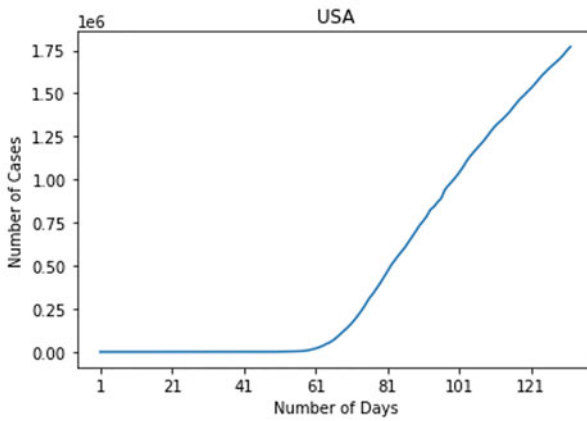
(b)



(c)



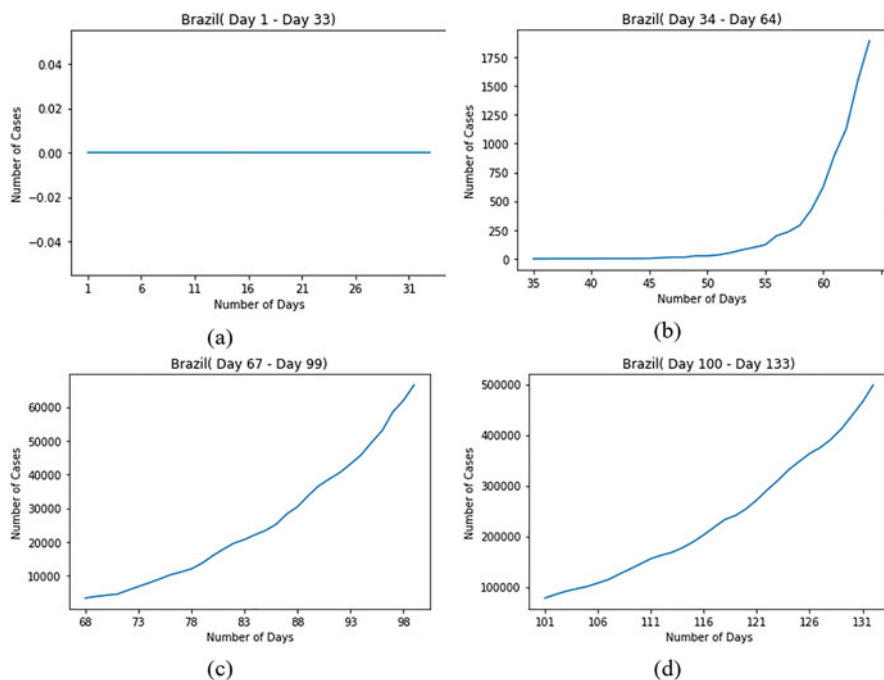
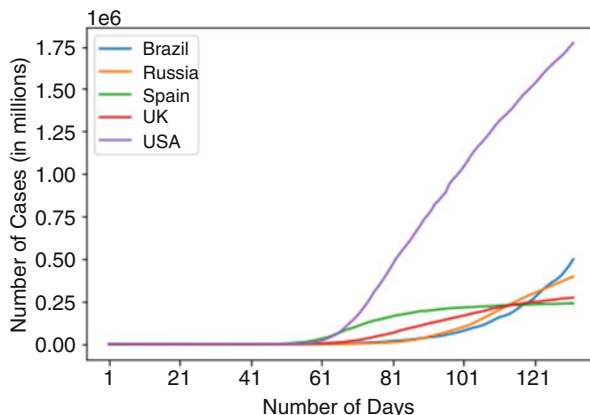
(d)



(e)

**Fig. 1** Profile of different countries: (a) Brazil, (b) Russia, (c) Spain, (d) UK and (e) USA

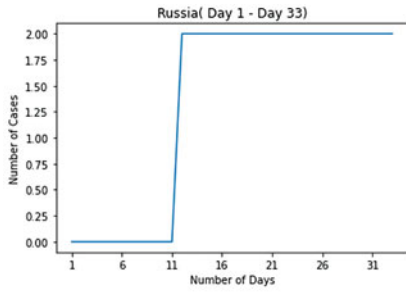
**Fig. 2** Profile trends of all selected five countries from day one to last day



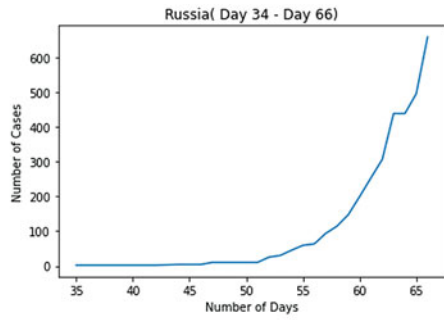
**Fig. 3** Spread graph of Covid-19 from day one to last day in Brazil

change in the number of cases and increased regularly till the end of the study (Fig. 6a-d).

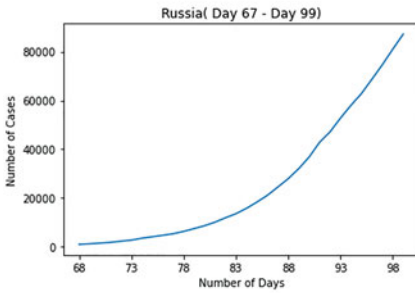
Out of all these countries, the number of cases in the USA was totally different. Figure 7 represents the spread of Covid-19 from day 1 to day 133 in the USA. From Fig. 7, it can be seen that from day 1, cases were spread and increased day by day (Fig. 7a-d).



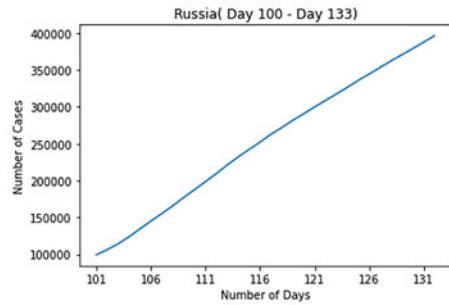
(a)



(b)

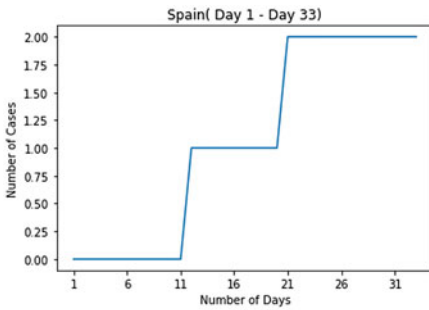


(c)

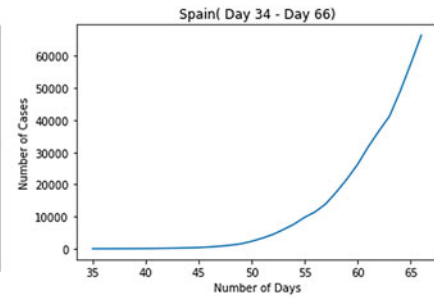


(d)

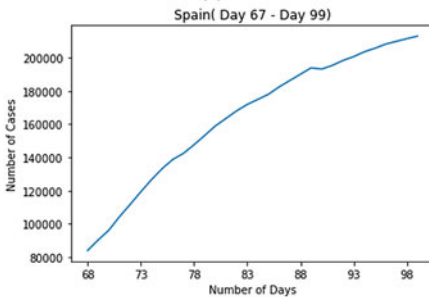
**Fig. 4** Spread graph of Covid-19 from day 1 to day 133 in Russia



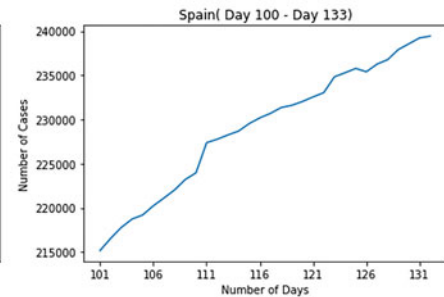
(a)



(b)

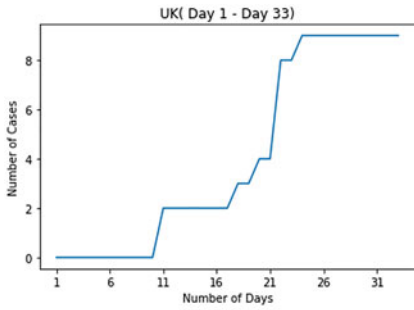


(c)

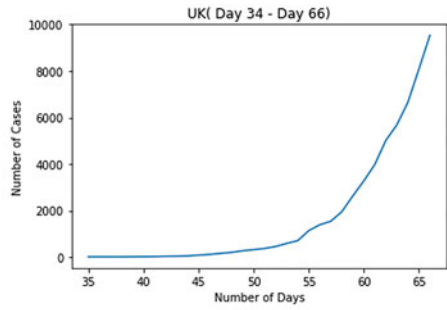


(d)

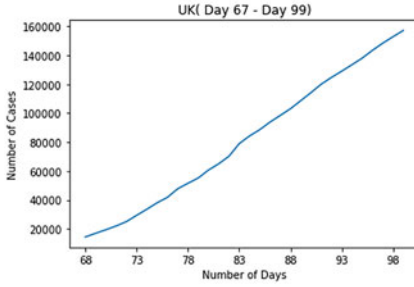
**Fig. 5** Spread graph of Covid-19 from day 1 to day 133 in Spain



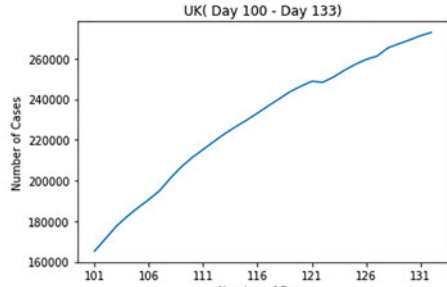
(a)



(b)

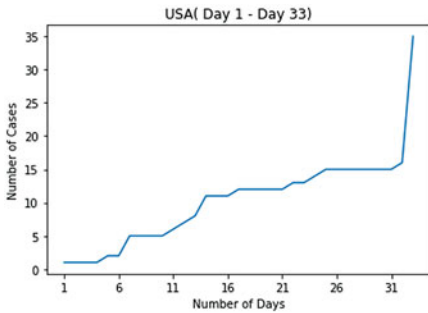


(c)

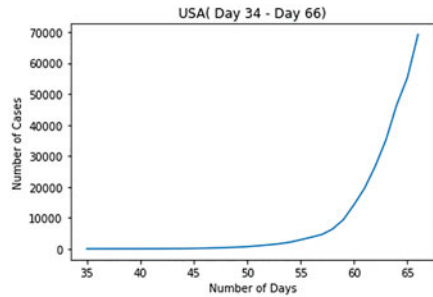


(d)

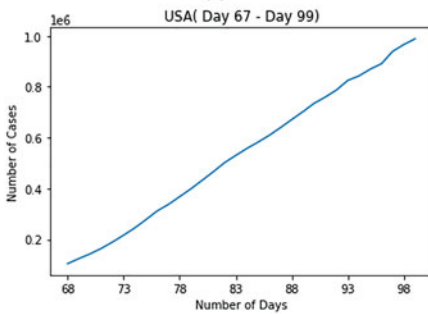
**Fig. 6** Spread graph of Covid-19 from day 1 to day 133 in UK



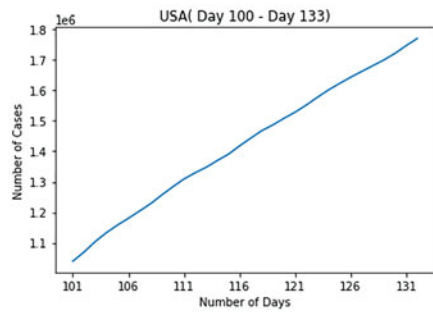
(a)



(b)



(c)

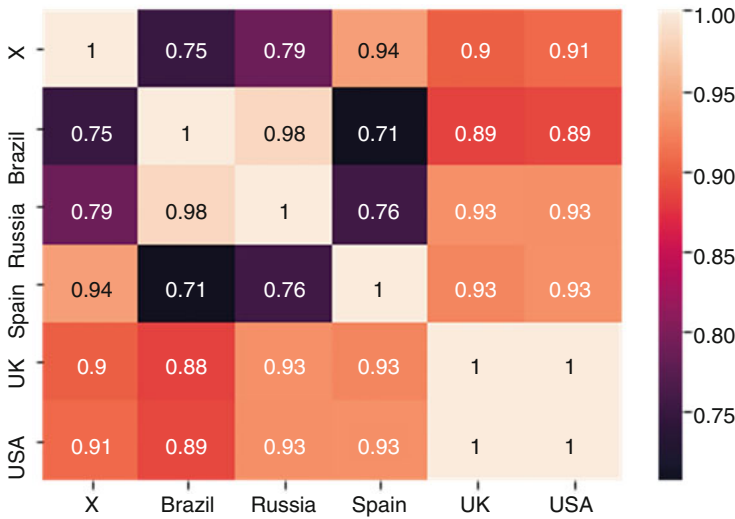


(d)

**Fig. 7** Spread graph of Covid-19 from day 1 to day 133 in USA

First derivative curve was also analysed for all the countries (Fig. 8). First derivative curves of different countries show the rate changes of the number of cases with respect to the days.

Maximum infection rates calculated are 33274, 11656, 9181, 8719 and 48529 for Brazil, Russia, Spain, the UK and the USA, respectively. Scatter Plot and Correlation matrix are shown below:



R square values for linear regression are 0.680, 0.694, 0.869, 0.824 and 0.831 for Brazil, Russia, Spain, the UK and the USA, respectively. Then, a hybrid model has been proposed for the prediction of accurate daily cases. First of all, original data values in terms of model no. 1 were fitted and then re-fitting was done on the resultant values by model no. 2 to improve the R<sup>2</sup> value.

Mathematical models 1 and 2 were applied to get the final results.

**Mathematical Model 1:**

The nature of data is exponential, the proposed model  $y = a \cdot b^x$

Its normal equations are:

$$A \cdot n + B \sum_{i=1}^n x_i = \sum_{i=1}^n Y_i$$

$$A \sum_{i=1}^n x_i + B \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i Y_i$$

Where,  $Y = \log_{10}(y)$ ,  $a = 10^A$  and  $b = 10^B$

From this model, the fitted equation of the countries is shown in Table 1:

**Table 1** Fitted equations from model 1

| Countries | $y = a b^x$                         |
|-----------|-------------------------------------|
| Brazil    | $y = (33.3892) (1.0908506)^x$       |
| Russia    | $y = (51.28389622) (1.088954485)^x$ |
| Spain     | $y = (591.0982511) (1.012786858)^x$ |
| UK        | $y = (98.10369544) (1.055804358)^x$ |
| USA       | $y = (1.662316149) (1.136061032)^x$ |

**Table 2** Fitted equations from model 2

| Countries | $y = A + B x + C x^2 + D x^3$                   |
|-----------|---|
| Brazil    | $y = - 3911 + 747 x - 29.71 x^2 + 0.333 x^3$    |
| Russia    | $y = - 2755 + 575.9x - 24.73 x^2 + 0.307 x^3$   |
| Spain     | $y = 590.2 + 7.692 x + 0.038 x^2 + 0.00 x^3$    |
| UK        | $y = - 164.7 + 58.95 x - 2.242 x^2 + 0.033 x^3$ |
| USA       | $y = - 6122 + 1157 x - 45.90 x^2 + 0.488 x^3$   |

**Mathematical Model 2:**

The proposed model

$$y = A + B x + C x^2 + D x^3$$

Its normal equations are:

$$A n + B \sum_{i=1}^n x_i + C \sum_{i=1}^n x_i^2 + D \sum_{i=1}^n x_i^3 = \sum_{i=1}^n y_i$$

$$A \sum_{i=1}^n x_i + B \sum_{i=1}^n x_i^2 + C \sum_{i=1}^n x_i^3 + D \sum_{i=1}^n x_i^4 = \sum_{i=1}^n x_i y_i$$

$$A \sum_{i=1}^n x_i^2 + B \sum_{i=1}^n x_i^3 + C \sum_{i=1}^n x_i^4 + D \sum_{i=1}^n x_i^5 = \sum_{i=1}^n x_i^2 y_i$$

$$A \sum_{i=1}^n x_i^3 + B \sum_{i=1}^n x_i^4 + C \sum_{i=1}^n x_i^5 + D \sum_{i=1}^n x_i^6 = \sum_{i=1}^n x_i^3 y_i$$

From this model, the fitted equation of the countries is given in (Table 2):

After using the models, the improved R square values are 0.983, 0.988, 1.00, 0.99 and 0.947 for Brazil, Russia, Spain, the UK and the USA, respectively.



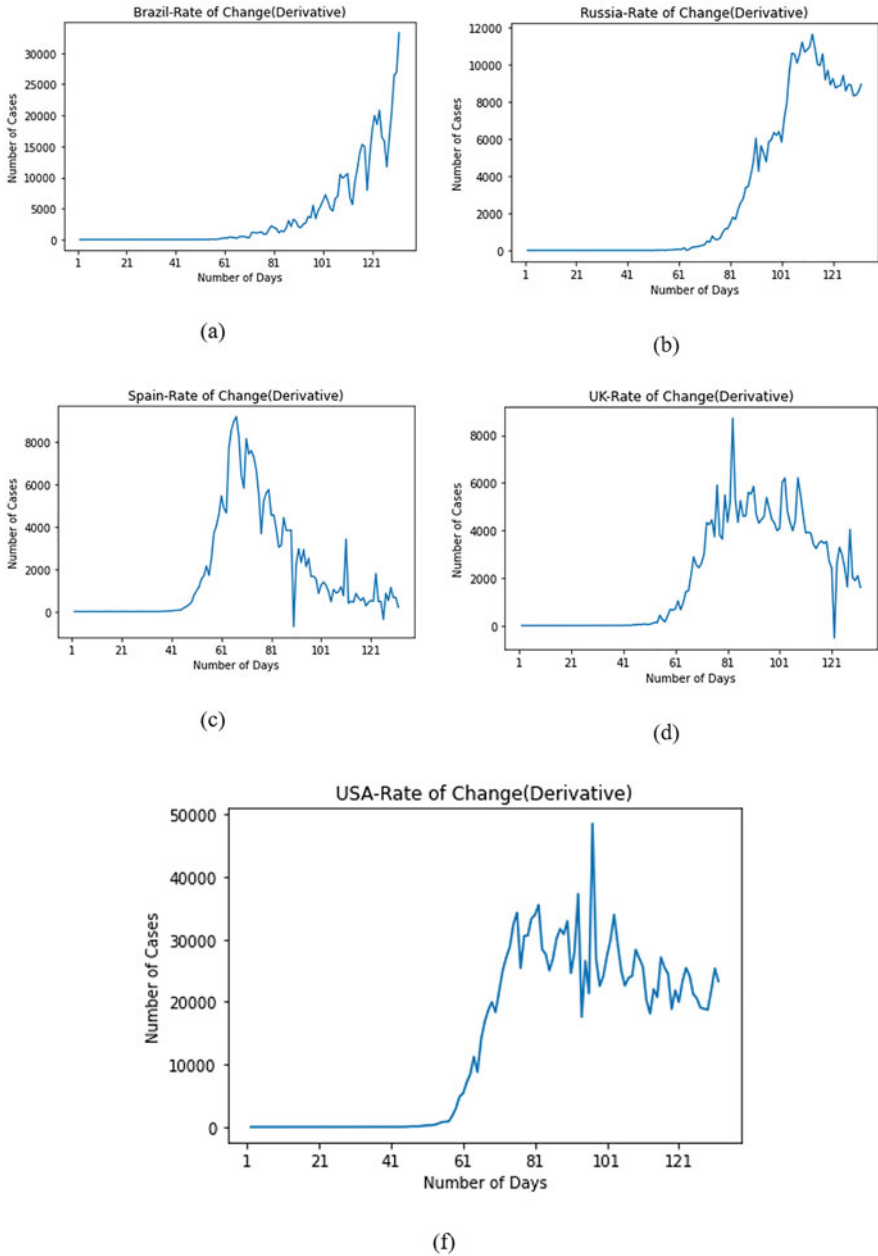


Fig. 8 First derivative curve: (a) Brazil, (b) Russia, (c) Spain, (d) UK and (e) USA

## 5 Conclusion

The prediction model obtained is based on the trend of the data with highest  $R^2$  value and minimum residual. The model helps the authorities to make necessary arrangements during the emergency.

## References

1. B. Ivorra, B. Martínez-López, J.M. Sánchez-Vizcaíno, Á.M. Ramos, Mathematical formulation and validation of the Be-FAST model for classical swine fever virus spread between and within farms. *Ann. Oper. Res.* **219**(1), 25–47 (2014)
2. F. Brauer, C. Castillo-Chavez, C. Castillo-Chavez, *Mathematical models in population biology and epidemiology*, vol 2 (Springer, New York, 2012), p. 508
3. C.I. Siettos, L. Russo, Mathematical modeling of infectious disease dynamics. *Virulence* **4**(4), 295–306 (2013)
4. C.E. Walters, M.M. Meslé, I.M. Hall, Modelling the global spread of diseases: A review of current practice and capability. *Epidemics* **25**, 1–8 (2018)
5. M.J. Keeling, L. Danon, Mathematical modelling of infectious diseases. *Br. Med. Bull.* **92**(1), 33 (2009)
6. Z. Feng, Final and peak epidemic sizes for SEIR models with quarantine and isolation. *Math. Biosci. Eng.* **4**(4), 675 (2007)
7. J. Li, N. Cui, Dynamic analysis of an SEIR model with distinct incidence for exposed and infectives. *Sci. World J.* **2013**, 1 (2013)
8. M.A. Safi, M. Imran, A.B. Gumel, Threshold dynamics of a non-autonomous SEIRS model with quarantine and isolation. *Theory Biosci.* **131**(1), 19–30 (2012)
9. M. Coccia, Factors determining the diffusion of COVID-19 and suggested strategy to prevent future accelerated viral infectivity similar to COVID. *Sci. Total Environ.* **729**, 138474 (2020)
10. L. Li, Z. Yang, Z. Dang, C. Meng, J. Huang, H. Meng, et al., Propagation analysis and prediction of the COVID-19. *Infect. Dis. Model.* **5**, 282–292 (2020)
11. <https://www.kaggle.com/imdevskp/corona-virus-report>