Chapter 3
COVID-19 Epidemic Analysis and Prediction Using Machine Learning Algorithms

Arun Solanki and Tarana Singh

Abstract  COVID-19 is a real problem, and it is spreading like a forest fire. The data of this pandemic is time-series data. The models that can handle time-series data are the ARIMA model, the Holt-Winter model, the SARIMAX model, polynomial regression, and LSTM. These models have been applied to COVID-19 data, and the results are discussed with significance. This chapter used three types of datasets. The primary dataset is the 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE (https://github.com/CSSEGISandData/COVID-19). The second dataset is used from Worldometers website (https://www.worldometers.info/), and third is from Kaggle. The SARIMAX model produced 0.236 as the MAPE value, while the Holt-Winter model produced 0.249. The polynomial regression model shows that the accuracy of the model approximated for the tenth day is 85% in the prediction of the number of affected cases and the number of deaths. The LSTM model used the ADAM optimizer and calculated the root mean square error. The prediction error for training is 6.45, and the calculated overall error is 5.34.

Keywords  COVID-19 · ARIMA · SARIMAX · Polynomial regression · Long short-term memory · World Health Organization
3.1 Introduction

The new coronavirus disease 2019 (COVID-19) pandemic started in Wuhan, China, in December 2019 and is a genuine general medical issue around the world. The infection that causes the COVID-19 pandemic is called severe acute respiratory syndrome [1, 2]. Coronaviruses (CoV) are an enormous class of infections that cause sicknesses brought about by cold, for example, the Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV). The COVID-19 is another species of covid family which found in 2019 [3, 4]. Coronavirus is a zoonotic sickness because of sullying by human creatures. Studies show that the SARS-CoV infection contaminates civet people and the MERS-CoV infection taints dromedary people. The COVID-19 infection is accepted to be contracted by people from bats. Unhealthy respiratory diseases starting with one individual and then onto the next quickly spread the scourge [5, 6]. COVID-19 shows mild to moderate manifestations in about 82% of cases, while different cases are serious or extreme. As per data received on 30th March 2020 from WHO the number of instances of coronavirus was roughly 335,403 of which 14,611 died & 97,636 recovered and rest 223,156 were active cases [3]. While 95% of infected patients live poorly, the remaining 5% suffer from serious or severe illness. Signs of infection include respiratory symptoms, fever, cough, and dyspnoea. In more severe cases, the infection can cause pneumonia, severe acute respiratory syndrome, septic shock, multiple organ failure, and death. The infection rate in men is higher than that in women, and it is determined that no children aged 0–9 die. The respiration rate of COVID-19 pneumonia was shown to be faster compared to healthy people [3, 7, 8]. In many developed countries, the demand for intensive care units is also increasing at the same time, and the medical system is on the verge of collapse [9]. There are many patients in the intensive care unit with exacerbated COVID-19 pneumonia. The World Health Organization reports that the pandemic situation is divided into four stages. The first stage begins with a case reported about people traveling to the affected area, while the second stage is reported locally by family, friends, and other people in contact with other affected areas [10]. Currently, what is affected can be tracked. The situation worsened in the third stage, after which the cause of the infection became untraceable and was transmitted to people who had never travelled or who had no contact with the infected person. In this situation, an immediate national blockade is needed to reduce social contact between individuals and control the rate of infection. Figure 3.1 shows that the WHO detailed that COVID-19 happened in countless affirmed cases in the USA, Italy, Spain, and China. To top it all off, stage 4 gets pervasive and wild [11–13]. Figure 3.2 shows all phases of the COVID-19 epidemic. So far, some countries have entered the fourth stage. China is the primary nation to encounter the fourth phase of COVID-19 infection. However, the source of the infection is accepted to be Wuhan in China which affects the other developed nations (the USA, Italy, Spain, the UK, etc.). These nations are presently in the fourth phase of contamination and face a larger number of diseases and deaths than China [14–16]. In China, the exponential increase in
confirmed cases reached a saturation point, at which point the number of cases stopped increasing. This is due to the fact that the number of susceptible people exposed to the virus is significantly reduced [17]. It is possible to isolate infected people and reduce social contact between them, and the Chinese government has begun a period of complete blockade, thereby reducing the likelihood of further transmission.

Machine learning algorithms [18–22] play an important role in epidemiological analysis and prediction. Machine learning techniques can help mitigate epidemic patterns when large amounts of epidemiological data are present [23–25]. Therefore, you can take early steps to prevent the spread of the virus. This study used real-time information from the Johns Hopkins dashboard to observe the daily behaviour of domestic COVID-19 and predict future accessibility using machine learning and deep learning models. Since the Spanish flu epidemic in the early twentieth century,
the COVID-19 pandemic is the largest pandemic, affecting approximately three million people worldwide. Pandemics have a serious impact on the population and economy of many countries. From a machine learning perspective, COVID-19 can easily be considered a random forest when mapping diseases or viruses that have recently affected human species to the decision tree. The spread of this deadly virus is unimaginable, but it is not impossible to calculate. It is therefore of great interest to be able to model this epidemiological data and provide information and better adaptation strategies to contain the effects of the virus. That is good for the people and the economy of the country [26–29]. This study uses existing data and various machine learning models to try to predict the epidemiologic spread of the virus in India [30].

The whole chapter is organized in five different sections. Section 3.1 contains the introduction. In Sect. 3.2, related work is discussed briefly. In Sect. 3.3, different models have been discussed. In Sect. 3.4, results analysis has been done. Section 3.5 discusses the final conclusion and future work.

### 3.2 Related Work

Over the past decade, digital technology has played a key role in key issues in the health sector, including disease prevention, and today’s global health emergencies also seek technical support to resolve COVID-19. In the article, the author
highlighted in trendy digital technologies, for example, the Internet of Things (IoT), big data analytics, artificial intelligence (AI), deep learning, and technology. Blockchain is being created to screen, distinguish, and forestall scourges and decide the materialness of the technique and the effect of the epidemic on the health sector [9].

In a study, Benvenuto et al. [10] proposed an autoregressive integrated moving average (ARIMA) model for foreseeing the spread of COVID-19. In light of an investigation of the predominance and frequency of COVID-19, the author anticipated different boundaries throughout the following 2 days of his article. The examination additionally indicated correlation maps of epidemic incidence and prevalence and ARIMA prediction maps. Authors [11] proposed a time-series method to dissect the occurrence model and evaluated duplicate number of the COVID-19 pestilence. Having directed factual examination, talked about scourge drifts, and featured the current epidemiological phase of the district, various arrangements can be chosen to react to COVID-19 pandemics in various areas. In the current circumstance, it is important to comprehend the early methods of contamination so as to plan and screen viable security measures. Authors proposed a significant logical model of the spread of SARS-CoV-2 utilizing different informational collections to examine COVID-19 episodes since Wuhan. In this way, they investigated the potential for outbreaks of the disease outside Wuhan [12]. Several studies have recently been conducted on the epidemiology of COVID-19 using exploratory data analysis (EDA) based on the different datasets available. These studies focus primarily on cases of diagnosis, death, and rehabilitation in Wuhan and other parts of the world to understand the plans for suspicious threats and subsequent containment activities [13]. The study by Lauer et al. raised the important question of COVID-19 latency. They investigated 181 confirmed cases, determined that the incubation period varied from 5 to 14 days, and planned better surveillance and surveillance activities accordingly [14]. In a recent study, Singer analysed data from 25 affected counties to track short-term forecasts for the COVID-19 epidemic. In this study, where there is a disease-specific infection rate, the exponential law differs according to the regular or explosive growth of the exponential law. Based on this understanding, the author analyses the effects of blockades in the world [15]. Traditionally, mathematical modelling has been used epidemiologically to model viruses and infectious diseases. The compartment model is one of these models. The most popular model is the SIR model (S stands for susceptibility, I represents infectious dose, and R represents the number of individuals who recovered or died (or immunized)). However, the authors believe that by leveraging the large amount of data currently available, machine learning can provide better predictions [24].

Based on the literature above, it is clear that exploratory data analysis can do a good job of understanding current trends in the epidemic, but there is still plenty of development and room for development. Test effective predictive models based on machine learning so that you can adopt active strategies. Make sure it meets your current needs.
3.3 Model Discussions

The study compares the predictions of multiple different models. These include the traditional statistical time-series forecasting models like ARIMA, SERIMAX, and Holt-Winter’s exponential smoothing model [31–33]. Another approach is to use machine learning-based regression models. To demonstrate this approach, the study uses a regularized regression model—the ridge regression model. Deep learning is another exciting new field of research in machine learning [34–36]. Since the data in question is time series or sequential, it is not possible to model it using traditional feed-forward neural networks that do not have any concept of time [37–39]. Recurrent neural networks (RNN) are neural networks that have a feedback loop that enables them to model sequential data. Long short-term memory (LSTM) is a more developed form of a recurrent neural network that utilizes the concept of attention and has been shown to provide better results than RNNs in many applications [40–42]. An LSTM-based approach has also been used to predict the COVID 19 spread in this study [43–46].

3.3.1 Ridge Regression

The study is using a dataset prepared by the Johns Hopkins University Centre for Systems Science and Engineering (JHU CSSE) and the crowdsourced data from the Covid19india.org tracker [47, 48]. The study first discussed the nature of epidemic growth and analyses the initial exponential growth of the virus and calculates the growth ratio and growth factor of the epidemic for the case of India. The study also makes predictions for the expected number of cumulative confirmed cases using a ridge regression model and uses the growth factor and growth as features in this model [49–52].

3.3.1.1 Initial Exponential Epidemic Growth

Epidemics initially follow exponential growth due to the nature of viruses where the people infected with a virus in turn infect others. This growth is measured using a positive constant called the growth factor of the epidemic [53]. However, this exponential growth is not sustained throughout the epidemic, due to factors like intervention, a large portion of the population getting infected, etc., and starts to stabilize at a growth factor of 1 before gradually decreasing [54, 55]. This point is known as the Inflection point for a curve and can be visualized using a logistic curve. The study observes this phenomenon in the graph of the growth of the virus in China, as shown in Fig. 3.3.

Comparing this with the growth of COVID 19 in India as of April 22, shown in Fig. 3.4, it can be seen that the virus is still in the initial stages.
The question now becomes whether the growth of the virus has reached the inflection point. This can be analysed using the growth factor [56].

### 3.3.1.2 Growth Factor

The growth factor is a constant factor by which the total cases would be multiplied in the case of exponential growth. The growth factor on day $N$ is the number of confirmed cases on day $N$ minus confirmed cases on day $N - 1$ divided by the
number of confirmed cases on day \( N - 1 \) minus confirmed cases on day \( N - 2 \). It is a measure of whether the spread of the virus is growing or not [57]. A value of greater than 1 is associated with growth, and values of less than 1 are associated with decline. A growth factor of 1 is the inflection point, and at this point, the spread is not growing. Figure 3.5 shows the updated growth factor of the virus in India. As of April 24, the growth factor for the last 10 days is around 1.05, and it can be seen that the growth of the virus has not stabilized at the inflection point but is much lower than the mean growth factor up to this point indicating that the initial exponential growth might be over [58].

A linear regression on the growth factor points to a downward trend, but the margin of error is high as defined by a 95% confidence interval, shown in Fig. 3.6.
3.3.1.3 Growth Ratio of Virus

The growth ratio on day $N$ is the number of confirmed cases on day $N$ divided by the number of confirmed cases on day $N - 1$ [59]. Looking at the graph in Fig. 3.7, there is an unmistakable downward trend to the growth ratio, with a 6% growth in the number of cases occurring on April 24 compared to highs of 25–30% since the start of reliable testing. After performing a linear regression on the growth ratio, a strong downward trend can be seen in Fig. 3.8a, and in Fig. 3.8b, a linear relationship can be seen in the residual plot confirming that such an assumption is reasonable.

3.3.2 Regularized Ridge Regression

The approach discussed here attempts to model the spread of COVID-19 as a regression problem. A linear regression model in machine learning attempts to model the target or dependent variable $Y$, the total number of confirmed cases of COVID-19 in...
this case, as a linear relationship between some predictor features $X_p$ and the regression coefficients $\beta_p$ [60]. However, it is not reasonable to expect that complex data such as epidemiological data would have a simple linear relationship with some predictor features. An approach to model complex non-linear data as a regression task is to add polynomial combinations of the features as predictor variables in the form of $X_{p_d}$. This method is known as polynomial regression [61]. A downside of polynomial regression is that it can easily overfit the training data and results in low generalization performance. A low generalization error is desirable for a machine learning model as it means that the model is able to provide accurate predictions on unseen data. The solution is to use a method that applies a regularization penalty to reduce this overfitting. Ridge regression applies half of the squared l2 norms of model parameters as a regularization term or penalty to reduce overfitting and improve generalization performance [62]. This penalty is controlled using the hyperparameter alpha and reduces in-sample training accuracy but will help the model generalize better in the absence of a large number of features and training data [63].

The study uses days since the first case, along with the growth factor and growth ratio as features for the ridge regression model. The growth factor and growth ratio for predictions are also estimated using a ridge regression with polynomial features of degree 2 and alpha 5 and 2, respectively [64–66], and is shown in Fig. 3.9.

Using the features discussed above and polynomial features of degree 3 and alpha 10, the model is fit to the currently available data and achieves an R2 score of 0.80 or 80% in-sample accuracy and an MSE of 13,631, as shown in Fig. 3.10.

### 3.3.3 LSTM Model

LSTM is an RNN architecture which contains memory, i.e. it remembers the state of the model at different stages. It is most suited for time-series prediction when the time gap length is unknown [54]. Thus, LSTM can be proved to be the well-suited
model for the COVID-19 spread in the world. This section investigates and estimates the number of new cases in India as compared to all over the world. The dataset is taken from Kaggle which contains four columns which are new_cases, new_deaths, total_cases, and total_deaths. The dataset contains the country-wise data describing the above columns for each country on an everyday basis. The data is available from December 31, 2019, to March 25, 2020. The dataset for India is depicted in Fig. 3.11.

The number of actual cases based on new cases can be represented by a graph shown in Fig. 3.12. The number of new cases depends upon the number of active cases, as this pandemic is communicable and spreads from an infected person. Thus, the greater the number of active cases, the greater the number of new cases. The growth rate of actual cases, therefore, becomes exponential.

- Training the model: The model is built to train the dataset in order to predict the number of new cases which can be infected by COVID-19 based on the data provided. The LSTM model is a recurrent neural network whose architecture is shown in Fig. 3.13.
The model is trained on the dataset for 2000 iterations in each epoch. A total of 16 epochs were run for complete training. The training loss came out to be 0.0081. The loss during the training is shown in Fig. 3.14.

3.3.4 **SARIMA Model**

The SARIMA model stands for seasonal autoregressive integrated moving average or seasonal ARIMA and belongs to the family of the ARIMA model. It is an extension of the ARIMA model which deals with the univariate time-series data
containing the seasonality component. It adds three new hyperparameters to specify the autoregression (AR), differencing (I), and moving average (MA) for the seasonal component of the series, as well as an additional parameter for the period of the seasonality. The SARIMAX model is nothing but a SARIMA model having an exogenous variable represented by $X$. This work is an attempt to calculate the rate of spread of this virus and thus to predict the deaths, recoveries, and confirmed cases, so that it may help us to prepare better and survive. The dataset used in this research work is the 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data
Repository by Johns Hopkins CSSE (https://github.com/CSSEGISandData/COVID-19). As the dataset used had trend and seasonality in it, the models that have been used here are the SARIMAX model and the Holt-Winter model. Figure 3.15 displays the values for deaths, confirmed cases, and recovered cases for India.

This is no news that COVID-19 has turned into a pandemic because it is communicable and can spread by the medium of contact.

It means that the number of new cases or future cases is in one way or another related to the present cases. This is a prime reason why the past data can be tuned into a machine learning model and then use it to predict new cases. There is an upward trend in the dataset which can be observed from graphical representation of the data. One such representation has been displayed in Fig. 3.16.

- Model Training: The model is trained on a time period ranging from January 22, 2020, to March 23, 2020. A summary of the training is displayed in Fig. 3.17 in order to understand the various parameters of training.

The stationarity of the time series was checked with the help of augmented Dickey-Fuller test, and the $p$-value for different datasets (deaths, recovered, and confirmed) lied in the range of 0.00–0.07.
3.4 Result and Discussion

3.4.1 Scenario 1: Ridge Regression—Regularized Regression Model

Using the model to predict the total confirmed cases, we see a maximum or plateau of 140,000 confirmed cases before the total number of active cases starts to slow down around the middle of September. This is shown in Fig. 3.18. The predictions also indicate that the inflection point of the virus will be reached around July. The results appear to follow similar curves when compared to China and Italy. Upon observation the middle portion of the graph corresponding to slowing growth and the crossover to the inflection point seems prolonged. This can be due to a lack of features in the training data and the high predictive power of the days since the first case feature. It is a shortcoming of the model and is basis for future work.

![Fig. 3.17 Model fit summary for recovered cases](image)
3.4.2 Scenario 2: LSTM Model

This data is trained by using the LSTM model. The error is calculated using root mean square error, and the optimizer used is ADAM. The prediction error for training came out to be 6.45, and the overall calculated error is calculated to be 5.34. The number of new cases of people infected with coronavirus is predicted by the LSTM model, which gave the results as shown in Fig. 3.19.

The comparative study of the predicted cases as compared to the actual cases in India is shown in Fig. 3.20. This is done to estimate the accuracy of the model.

![Fig. 3.18 Predictions for coronavirus cases in India using ridge regression](image1)

![Fig. 3.19 Future prediction of the number of new cases](image2)
3.4.3 \textit{Scenario 3: SARIMA Model}

The model has then been used in order to make predictions for deaths and confirmed and recovered cases for the time period of April 1, 2020, to April 23, 2020. These predictions have been depicted in Fig. 3.21. Mean absolute percentage error (MAPE) has been used to measure the accuracy of predictions made by the model. MAPE is inversely proportional to the accuracy of a model, i.e. the higher the value of MAPE, the lower is the model accuracy and vice versa (Fig. 3.22).

3.5 \textit{Conclusion}

The conclusion is divided according to the model discussed in the chapter, i.e. regression, LSTM, and SARIMA models.
The predicted values for confirmed cases, deaths and recovered cases

|         | Deaths | Confirmed_cases | Recovered_cases |
|---------|--------|----------------|-----------------|
| 2020-04-01 | 10     | 10             | 41              |
| 2020-04-02 | 9      | 9              | 43              |
| 2020-04-03 | 11     | 11             | 47              |
| 2020-04-04 | 11     | 11             | 47              |
| 2020-04-05 | 11     | 11             | 47              |
| 2020-04-06 | 12     | 12             | 47              |
| 2020-04-07 | 12     | 12             | 47              |
| 2020-04-08 | 12     | 12             | 55              |
| 2020-04-09 | 12     | 12             | 55              |
| 2020-04-10 | 13     | 13             | 56              |
| 2020-04-11 | 13     | 13             | 56              |
| 2020-04-12 | 14     | 14             | 57              |
| 2020-04-13 | 15     | 15             | 61              |
| 2020-04-14 | 14     | 14             | 64              |
| 2020-04-15 | 16     | 16             | 67              |

Predicted vs actual value for recovered cases
3.5.1 Regression

The study has analysed the growth of COVID-19 in India and performed ridge regression to predict the growth of cases in the country. The model is fit to the currently available data with an R2 score of 0.80 or 80% in-sample accuracy. The analysis of the results shows that the growth of the virus has still not stabilized, and the expected total cases are 140,000 by the middle of September. The total active cases are expected to decrease after this point. The lack of features to fit the model on is a limitation of the study and an opportunity for future work.

3.5.2 LSTM

The study has analysed the growth of COVID-19 in India and used the LSTM model to predict the number of new cases in the country. The analysis of the results shows that the growth of the virus has still not stabilized, and the expected total cases will be 2000 new cases every day by the end of July. The extension of relaxation in lockdown will greatly affect the spread of this pandemic.

3.5.3 SARIMA

The predictions made by model have the MAPE value ranging between 0.38 and 0.08 for different cases. The COVID-19 virus is spreading like a forest fire, and if sufficient measures are not taken, it could get out of hand. The various parameters of the SARIMA(X) model, like the seasonal order, etc., can be tweaked to decrease the MAPE value and make more accurate predictions.

References

1. Stoecklin, S.B., Rolland, P., Silue, Y., Mailles, A., Campese, C., Simondon, A., Mechain, M., et al.: First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020. Euro Surveill. 25(6), 2000094 (2020)
2. Github repository. 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE. https://github.com/CSSEGISandData/COVID-19. Accessed 31 Mar 2020
3. WHO coronaviruses (COVID-19). https://www.who.int/emergencies/diseases/novel-coronavirus-2019. Accessed 30 Mar 2020
4. Prasad, S., Potdar, V., Cherian, S., Abraham, P., Basu, A.: Transmission electron microscopy imaging of SARS-CoV-2. Indian J. Med. Res. 151, 241–243 (2020)
5. Leppard, G.G.: Evaluation of electron microscope techniques for the description of aquatic colloids. In: Environmental Particles, pp. 231–289. CRC Press, Boca Raton (2019)
6. Jia, L., Li, K., Jiang, Y., Guo, X.: Prediction and analysis of coronavirus disease 2019. arXiv preprint arXiv:2003.05447 (2020)
7. Kalipe, G., Gautham, V., Behera, R.K.: Predicting malarial outbreak using machine learning and deep learning approach: a review and analysis. In: 2018 International Conference on Information Technology (ICIT), pp. 33–38. IEEE (2018)
8. Singh, R., Singh, R., Bhatia, A.: Sentiment analysis using machine learning technique to predict outbreaks and epidemics. Int. J. Adv. Sci. Res. 3(2), 19–24 (2018)
9. Ting, D.S.W., Carin, L., Dzau, V., Wong, T.Y.: Digital technology and COVID-19. Nat. Med. 26(4), 459–461 (2020)
10. Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., Ciccozzi, M.: Application of the ARIMA model on the COVID-19 epidemic dataset. Data in brief (2020): 105340. It is made available under a CC-BY-NC-ND 4.0 International license. (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. medRxiv preprint. https://doi.org/10.1101/2020.04.08.20057679. This version posted April 11, 2020. The copyright holder for this preprint 10 Narinder Singh Punn et al
11. Deb, S., Majumdar, M.: A time series method to analyze incidence pattern and estimate reproduction number of COVID-19. arXiv preprint arXiv:2003.10655 (2020)
12. Kucharski, A.J., Russell, T.W., Diamond, C., Yang, L., Edmunds, J., Funk, S., Eggo, R.M., et al.: Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect. Dis. 20(5), 553–558 (2020)
13. Dey, S.K., Rahman, M.M., Siddiqui, U.R., Howlader, A.: Analyzing the epidemiological outbreak of COVID-19: a visual exploratory data analysis (EDA) approach. J. Med. Virol. 92(6), 632–638 (2020)
14. Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A.S., Reich, N.G., Lessler, J.: The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann. Intern. Med. 172, 577 (2020)
15. Singer, H.M.: Short-term predictions of country-specific Covid-19 infection rates based on power law scaling exponents. arXiv preprint arXiv:2003.11997 (2020)
16. Awad, M., Khanna, R.: Support vector regression. In: Efficient Learning Machines, pp. 67–80. Apress, Berkeley, CA (2015)
17. De Castro, Y., Gamboa, F., Henrion, D., Hess, R., Lasserre, J.-B.: Approximate optimal designs for multivariate polynomial regression. Ann. Stat. 47(1), 127–155 (2019)
18. Pandey, S., Solanki, A.: Music instrument recognition using deep convolutional neural networks. Int. J. Inf. Technol. 13(3), 129–149 (2019)
19. Rajput, R., Solanki, A.: Real-time analysis of tweets using machine learning and semantic analysis. In: International Conference on Communication and Computing Systems (ICCCS2016), Taylor and Francis, at Dronacharya College of Engineering, Gurgaon, 9–11 Sept, vol. 138(25), pp. 687–692 (2016)
20. Ahuja, R., Solanki, A.: Movie recommender system using K-means clustering and K-nearest neighbor. In: Accepted for Publication in Confluence-2019: 9th International Conference on Cloud Computing, Data Science & Engineering, Amity University, Noida, vol. 1231(21), pp. 25–38 (2019)
21. Tayal, A., Kose, U., Solanki, A., Nayyar, A., Saucedo, J.A.M.: Efficiency analysis for stochastic dynamic facility layout problem using meta-heuristic, data envelopment analysis and machine learning. Comput. Intell. 36(1), 172–202 (2019)
22. Tayal, A., Solanki, A., Singh, S.P.: Integrated frame work for identifying sustainable manufacturing layouts based on big data, machine learning, meta-heuristic and data envelopment analysis. Sustain. Cities Soc. 62, 102383. https://doi.org/10.1016/j.scs.2020.102383
23. Greff, K., Srivastava, R.K., Koutnik, J., Steunebrink, B.R., Schmidhuber, J.: LSTM: a search space odyssey. IEEE Trans. Neural Netw. Learn. Syst. 28(10), 2222–2232 (2016)
24. Chen, Y.-C., et al.: A time-dependent SIR model for COVID-19 with undetectable infected persons. arXiv:2003.00122 (2020)
25. Chowell, G., Sattenspiel, L., Bansal, S., Viboud, C.: Mathematical models to characterize early epidemic growth: a review. Phys. Life Rev. 18, 66–97 (2016)
26. Ma, J.: Estimating epidemic exponential growth rate and basic reproduction number. Infect. Dis. Modell. 5, 129–141 (2020)
27. Brauer, F.: Age of infection models and the final size relation. Math. Biosci. Eng. 5, 681–690 (2008)
28. Davoudi, B., Pourbohloul, B., Miller, J.C., Meza, R., Meyers, L.A., Earn, D.J.D.: Early real-time estimation of infectious disease reproduction number. arXiv 0905 0728 (2009)
29. Diekmann, O., Heesterbeek, J.A.P.: Mathematical Epidemiology of Infectious Diseases. Wiley, Chichester (2000)
30. Yan, P., Feng, Z.: Variability order of the latent and the infectious periods in a deterministic SEIR epidemic model and evaluation of control effectiveness. Math. Biosci. 224(1), 43–52 (2010)
31. Chowell, G., Brauer, F.: The basic reproduction number of infectious diseases: computation and estimation using compartmental epidemic models. In: Chowell, G., Hyman, J.M., Bettencourt, L.M.A., Castillo-Chavez, C. (eds.) Mathematical and Statistical Estimation Approaches in Epidemiology. Springer, Dordrecht (2009)
32. Chowell, G., Ammon, C.E., Hengartner, N.W., Hyman, J.M.: Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions. J. Theor. Biol. 241, 193–204 (2006)
33. Nishiura, H., Chowell, G.: The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. In: Chowell, G., Hyman, J.M., Bettencourt, L.M.A., Castillo-Chavez, C. (eds.) Mathematical and Statistical Estimation Approaches in Epidemiology. Springer, Dordrecht (2009)
34. De Jong, M.C., Diekmann, O., Heesterbeek, J.A.: The computation of R0 for discrete-time epidemic models with dynamic heterogeneity. Math. Biosci. 119, 97–114 (1994)
35. Chowell, G., Nishiura, H.: Quantifying the transmission potential of pandemic influenza. Phys Life Rev. 5, 50–77 (2008)
36. Anderson, R.M., May, R.M.: Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford (1991)
37. Davidian, M., Giltinan, D.M.: Nonlinear models for repeated measurement data. In: Monographs on Statistics and Applied Probability 62. Chapman and Hall, New York (1995)
38. Ying, S., et al.: Spread and control of COVID-19 in China and their associations with population movement, public health emergency measures, and medical resources, p. 24 (2020)
39. Bai, Y., et al.: Presumed asymptomatic carrier transmission of COVID-19. J.A.M.A. 323(14), 1406–1407 (2020)
40. Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J.: LSTM: a search space odyssey. IEEE Trans. Neural Netw. Learn. Syst. 28(10), 2222–2232 (2017)
41. Chen, B., et al.: Data visualization analysis and simulation prediction for COVID-19. arXiv:2002.07906 (2020)
42. Lin, I.: Real-Time Data API for COVID-19 Epidemic (2020). Accessed 3 Feb 2020
43. Hong, X.: Predicting Alzheimer’s Disease Using LSTM. 35(18), 2034–2048 (2019)
44. Gers, F.A.: Learning to forget: continual prediction with LSTM. In: Proc. 9th Int. Conf. Artif. Neural Netw. (ICANN), pp. 850–855 (1999)
45. Chen, X., Yu, J., Wu, Z.: Temporally identity-aware SSD with attentional LSTM. IEEE Trans. Cybern. 18(11), 2228–2245 (2019)
46. Wang, L.: MCPL-Based FT-LSTM: medical representation learning-based clinical prediction model for time series events. 22(10), 1825–1840 (2019). Accessed 12 March 2019
47. Li, F., Tran, L., Thung, K.H., Ji, S., Shen, D., Li, J.: A robust deep model for improved classification of AD/MCI patients. IEEE J. Biomed. Health Inform. 19(5), 1610–1616 (2015)
48. Hou, B.: LSTM-Based Auto-Encoder Model for ECG Arrhythmias Classification References for Potential Neutralizing Antibodies Discovered for Novel Corona Virus Using Machine Learning
49. Computational predictions of protein structures associated with COVID-19/research/opensource/computational-predictions-of-protein-structures-associated-with-COVID-19. Accessed 9 Mar 2020
50. Seber, G.A.F., Lee, A.J.: Linear Regression Analysis Ser, Wiley Series in Probability and Statistics. Wiley (2003)
51. Lall, U., Moon, Y.I., Kwon, H.H., Bosworth, K.: Locally weighted polynomial regression: parameter choice and application to forecasts of the Great Salt Lake. Water Resour. Res. 42, W05422 (2006). https://doi.org/10.1029/2004WR003782
52. Cleveland, W.S., Devlin, S.J.: Locally weighted regression: an approach to regression analysis by local fitting. J. Am. Stat. Assoc. 83(403), 596–610 (1988)
53. Jakobsson, E., Farimani, A.B., Tajkhorshid, E., Aluru, N.: Combining physics-based and evolution-based methods to design antibodies against an evolving virus. Biophys. J. 118(3), 482a (2020). https://doi.org/10.1016/j.bpj.2019.11.2669
54. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. 9(8), 1735–1780 (1997)
55. Zheng, M., Song, L.: Novel antibody epitopes dominate the antigenicity of spike glycoprotein in SARS-CoV-2 compared to SARS-CoV. Cell. Mol. Immunol. 17(5), 536–538 (2020)
56. JCI Insight. Predicting the broadly neutralizing antibody susceptibility of the HIV reservoir https://insight.jci.org/articles/view/130153. Accessed 13 Mar 2020
57. Yoon, H., Macke, J., West, A.P., Foley, B., Bjorkman, P.J., Korber, B., Yusim, K.: CATNAP: a tool to compile, analyze and tally neutralizing antibody panels. Nucleic Acids Res. 43(W1), W213–W219 (2015). https://doi.org/10.1093/nar/gkv404
58. Ekiert, D.C., Bhabha, G., Elsliger, M.-A., Friesen, R.H.E., Jongeneelen, M., Throsby, M., Goudsmit, J., Wilson, I.A.: Antibody recognition of a highly conserved influenza virus epitope. Science. 324(5924), 246–251 (2009). https://doi.org/10.1126/science.1171491
59. Schmid, N., Eichenberger, A.P., Choutko, A., Riniker, S., Winger, M., Mark, A.E., van Gunsteren, W.F.: Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur. Biophys. J. 40(7), 843 (2011). https://doi.org/10.1007/s00249-011-0700-9
60. Darden, T., York, D.M., Pedersen, L.G.: Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98(12), 10089–10092 (1993). https://doi.org/10.1063/1.464397
61. Bakibayev, T., Kulzhanova, A.: Common movement prediction using polynomial regression. In: 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT), Almaty, Kazakhstan, pp. 1–4 (2018). https://doi.org/10.1109/ICAICT.2018.8747047
62. Omelchenko, A.V., Fedorov, O.V.: Polynomial regression coefficients estimation in finite differences space. In: 2015 25th International Conference Radioelektronika (RADIOELEKTRONIKA), Pardubice, pp. 257–260 (2015). https://doi.org/10.1109/RADIOELEK.2015.7129024
63. Chen, X.: Recursive local polynomial regression estimation and its applications. In: Proceedings of the 31st Chinese Control Conference, Hefei, pp. 2043–2048 (2012)
64. Su, W., Yang, H., Liu, Z.: System identification based on orthogonal polynomial regression analysis method. In: Proceedings of the 29th Chinese Control Conference, Beijing, pp. 1441–1446 (2010)
65. Rallis, M., Vazirgiannis, M.: Rank prediction in graphs with locally weighted polynomial regression and EM of polynomial mixture Models. In: 2011 International Conference on Advances in Social Networks Analysis and Mining, Kaohsiung, pp. 515–519 (2011). https://doi.org/10.1109/ASONAM.2011.44
66. Krzyzak, A., Pawlak, M.: Almost everywhere convergence of a recursive regression function estimate and classification (Corresp.). IEEE Trans. Inform. Theory. 30(1), 91–93 (1984). https://doi.org/10.1109/TIT.1984.1056833