COVID-19 Forecasting using Multivariate Linear Regression

R. Suganya (✉ rsuganya@tce.edu)

Thiagarajar College of Engineering, Madurai <https://orcid.org/0000-0003-1874-6479>

R.Arunadevi

Thiagarajar College of Engineering, Madurai

Sayed M.Buhari

Abdulaziz University, Saudi Arabia <https://orcid.org/0000-0002-7536-4066>

Research Article

Keywords: COVID-19, Forecasting spread, Multivariate Linear Regression, Top affected countries, Machine learning

Posted Date: September 9th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-71963/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe respiratory syndrome coronavirus 2 (SARS-CoV-2). It was first identified in December 2019 in Wuhan, the capital of China's Hubei province. The objective of this research is to propose a forecasting model using the COVID-19 available dataset from top affected regions across the world using machine learning algorithms. Machine Learning algorithms help us achieve this objective. Regression models are one of the supervised machine learning techniques to classify large-scale data. This research aims to apply Multivariate Linear Regression to predict the number of confirmed and death COVID-19 cases for a span of one and two weeks. The experimental results explain 99% variability in prediction with the R-squared statistics scores of 0.992. The algorithms are evaluated using the error matrix such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and accuracy for top affected regions across the world.

1. Introduction

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and was first identified in December 2019 in Wuhan, the capital of China's Hubei province. High fever, cough, sore throat, headache, fatigue, muscle pain, and shortness of breath are the initial symptoms of COVID-19. Mathematical analysis of infectious diseases using SIR epidemic and endemic models have been studied in the past literature [18]. Luo [21] indicates that certain measures taken by governments are based on various predictions. These predictions suggest the hospital needs, future deaths, infection peaks and others. Several regions across the world have gone into a state of lockdown, including travel restrictions, to prevent the spread of this deadly virus [10], [17]. Despite lockdown, massive amounts of death and confirmed cases occurred across the world. Data from countries like China, Japan and South Korea indicates that even though lockdowns proposed by governments have caused reduction in the spread of COVID-19 but there are other important factors which could not be neglected. One such factor is wearing masks in public, which have currently been accepted as a norm [27]. Social distancing, considered as an important factor that has contributed towards the control of the spread of this deadly virus, is expected to be prolonged in the near future [19]. The social distancing and lockdown are not just enough to prevent the spread of COVID-19 across the world. Forecasting cumulative death and confirmed cases are very useful to know and take preparatory measures to prevent the massive amount of deaths and confirmed cases across the world.

The objective of this research is to predict the cumulative count of future death and confirmed cases in top affected regions across the world using Multivariate Linear Regression. Regression models are known to be more robust towards noise provided there is aggregation of response and exposure variables [29]. The experimental results obtained from this paper are compared based on the error metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), R-Squared value, and Root Mean Squared Error (RMSE) to validate the efficiency of these proposed techniques. It explains 99% of the accuracy using Multivariate Linear Regression. These results are taken from top affected regions across the world such as the United

States of America, United Kingdom, Spain, Italy and India (countries in World) along with Maharashtra, Tamil Nadu, Gujarat, Uttar Pradesh and Delhi (states in India) and Chennai (district in Tamil Nadu).

2. Related Works

Discussion about COVID-19 has taken the world by surprise that almost all walks of life has been impacted due to it. Different medical specializations have been diverted towards caring for COVID-19 patients. One such situation arises for radiologists, who currently undergo down time in their medical professional. Fessell et al. [12] discusses about how to make the down time useful and meaningful to our lives. They discuss about the importance of self-care, like applying the emergency masks in the airline flight before applying to your dependents, and how stress is created due to the current news outlets. Re-thinking about those books that are bought before but not read and those parts of your home that are not well maintained are certain thought proving entities as of now [12]. Reduction in acute coronary syndrome in current times

[27] is attributed towards reduction in air pollution. The propaganda towards advising general public to be wary of air pollution did not gain popularity but the current widespread usage of N95 masks to prevent the spread of COVID-19 has addressed multiple issues [27]. Prior research has shown the impact of usage of masks to protect humans from air pollutants like PM_{2.5} [26], indicating the size and spread of such pollutants. Comparison of the presence of air pollutants with SARS-CoV-2 in the nasal epithelium and upper respiratory tract [14] suggests masks could be one factor to control the spread of this deadly virus.

Recent study on the curve of COVID-19 using Levitt's Metric suggest that the peak could be obtained somewhere within two to four weeks from 13th July 2020 [3]. Levitt's Metric suggests the closeness of $H(t)$ to 1 as an indicator of no more new deaths. This study considers only one parameter, either Covid deaths or Covid cases at a time. Luo [21] indicates that predictions made by many researchers around the globe have tend to be outside even the next day's 95% confidence interval. Predictive monitoring, which monitors predictions of crucial future events and considers the differences not as errors but as changes, is adopted [21]. Regression is used in predictive monitoring and only those countries with R^2 value greater than 0.8 is accepted and reported. Date to reach the last expected case, along with 99% and 97% of such a scenario is also reported. Author cautions that the predictions might go wrong due to efforts taken by the respective governments.

Many researchers have predicted the final size of COVID-19 [15, 23, 6, 28, 31]. Logistic growth model and the classic susceptible-infected-recoverable (SIR) model are used to estimate the final size COVID-19 [7]. The efficiency of the proposed logistic regression is justified by higher coefficient of determination (0.996) and lower p-value (< 0.000). Hashem et al. [10] has shown the spread of COVID-19 using machine learning techniques. This spread can be used not only to indicate the spread but also forecast, trace and diagnose COVID-19. Such a framework can be used in smart cities around the world. Prior to COVID-19, Cardiovascular mortality on a time series for Canada is studied using generalized additive

models (GAM), distributed lag nonlinear models (DLNM) and autoregressive-moving average (ARMA) [1]. Petropoulos has applied exponential smoothing based simple time series modeling to forecast the pattern of COVID-19 [8]. Related works are summarized in Table 1.

Likewise, in this article, we tend to predict the future for those top affected regions hoping that there is downward trend in the near future that can show some green light to the much affected stressful life these days.

One important implication on the data obtained is due to false negatives in the RT-PCR tests. This might be due to presence of the virus in the lower respiratory tracts than the upper respiratory track from which the sample was collected [32, 20]. Thus, there is a possibility of having discrepancy between the active cases and the confirmed cases. Thus, considering one parameter for prediction might affect the outcome of the prediction. This work considers multiple parameters as described in section 3.

3. Dataset

The COVID-19 dataset is collected from kaggle repository as a csv file. There are eight parameters such as region, date, confirmed cases, death, after one week confirmed and death cases and after two week confirmed and death cases (table 1). Data preprocessing is an important stage to identify missing and redundant data in the dataset [10]. Any missing data (NAN) are to be removed from the dataset to bring uniformity in the dataset. However in this dataset, there were no missing values and unnecessary values. Table 2 represents the description of the dataset.

In this paper, eleven top affected regions across the world are taken to apply this proposed methodology such as India, Spain, Italy, United Kingdom, United States of America, Tamil Nadu, Maharashtra, Gujarat, Delhi, Uttar

Pradesh and Chennai. In the world, the United States of America, United Kingdom, Italy, Spain and India are top affected regions for COVID-19. In India, at the time of this writing, Maharashtra, Tamil Nadu, Gujarat, Uttar Pradesh and Delhi are some of the top affected regions for COVID-19. Similarly, in Tamil Nadu, Chennai is the top affected district for COVID-19. Graphical representation leads to good visibility of COVID-19 impact. Figure 1 shows the cumulative number of confirmed cases across the world for COVID-19. Figure 2 represents the cumulative number of death cases across the world for COVID-19. For India, Figure 3 explains the confirmed cases is high for Maharashtra and India compared to Gujarat and Uttar Pradesh. Whereas cumulative number of death cases for Tamilnadu is gradually increases next to Maharashtra compared to other states in India (Figure 4)

4. Proposed Methodology

The proposed methodology is shown in Figure 5. The COVID-19 dataset is identified by kaggle repository. The main objective of this paper is to forecast the future cumulative death and confirmed cases for top affected regions across the world. Regression and classification algorithms are some of the

popular Supervised Learning techniques. In general, the Regression models are used to predict the continuous target values but classification models are used to predict the discrete target variables. In this paper, dependent variables (four target or output variables) are the continuous target variables. So, regression models are one of the suitable ways for this proposed methodology. After checking the unnecessary values and missing data in the dataset, we analyse the data to select the independent variables (features). Cumulative number of confirmed and death count (two independent variables) are used in this methodology. In the Multivariate Linear Regression Algorithm, multiple independent variables are used to predict the multiple target variables. Multivariate Linear Regression, one of the supervised learning techniques, is the extension of single Linear Regression. It is used to estimate the relationship between the two or more independent variables and dependent variables. These algorithms were implemented with the help of Python's scikit-learn and tensorflow library. Error metrics are used to evaluate the performance of the regression algorithm. Evaluation parameters such as Accuracy, R^2 score, Mean Absolute Error, Mean Squared Error and Root Mean Squared Error.

Independent Variables (Input): Cumulative number of confirmed cases and death

Dependent variables (target or Output): After one week Confirmed cases, After one week death cases, After two week Confirmed cases, and After two week Death cases

The above dependent and independent variables represent there are two independent variables (input) and four dependent variables (target or output). So Multivariate Linear Regression (Multi target variables) is a suitable way to handle this proposed methodology.

Multivariate regression is one of the regression techniques that estimates the single linear regression model with more than one dependent (target) variable. When there are more than one dependent and independent variables, that model is called Multivariate Multiple Linear Regression.

$$Y_{ik} = b_{0k} + \sum_{j=1}^n b_{jk} X_{ij} + e_{ik} \quad (1)$$

for $i \in \{1, \dots, n\}$ and $k \in \{1, \dots, m\}$

Where, $Y_{ik} \in R$ is the k-th real valued response for the i-th observation

$b_{0k} \in R$ is the regression intercept for k-th response.

$b_{jk} \in R$ is the j-th predictors regression slope for k-th response.

$x_{ij} \in R$ is the i-th predictor for the i-th observation.

The model is multivariate because $m > 1$ (dependent variables) and this proposed model is Multivariate Multi Linear Regression because $m > 1$ and $p > 1$ (more than one dependent and independent variables).

5. Evaluation Parameters

In Regression models, Error Metrics is the best way to evaluate the performance analysis. There are various error metrics used to check the result such as Mean Absolute Error (MAE), Mean Square Error (MSE), R^2 score and Root Mean Square Error (RMSE). Table 3 represents the performance analysis for COVID-19 world dataset. Error Metrics is used to evaluate the performance analysis for this methodology. In table 3, Accuracy is 99.7%, Mean Absolute Error is 79818 and R^2 score is 99.28% using Multivariate Linear Regression (MvLR). In those cases, overfitting might occur. So the dataset is divided into three parts such as train, test and validate data. Using Validation data, we evaluate error metrics and check for overfitting. In Validation data, accuracy is 99%. There is no overfitting in this Proposed Methodology. Finally Multivariate Linear Regression is the suitable algorithm to forecast the cumulative number of confirmed and death cases for top affected regions across the world for COVID-19.

6. Results And Discussion

Results from the Multivariate Linear regression analysis was conducted with July-21 Cumulative confirmed and death cases as input and predicted for July-28 and August-5 Cumulative death and confirmed cases count as output for the top eleven infected regions across the world using Multivariate Linear Regression (MLR). Results are tabulated in Table 4.

The figure 6 represent the forecasting the spread of COVID-19 (Cumulative Confirmed Cases) across the World. In India 1359221 and 1903457 Cumulative number of confirmed cases on July 28 and August 5. In United States of America (USA) 4428455 and 4905644 Cumulative number of confirmed cases on July 28 and August 5. In United Kingdom (UK) 322611 and 365408 Cumulative number of confirmed cases on July 28 and August 5. In Spain 279847 and 300939 Cumulative number of confirmed cases on July 28 and August 5. In Italy 245815 and 246179 Cumulative number of confirmed cases on July 28 and August 5. The figure 7 represent the forecasting the spread of COVID-19 (Cumulative Confirmed Cases) across the World. In India 34962 and 50048 Cumulative number of death cases on July 28 and August 5. In United States of America (USA) 151229 and 179077 Cumulative number of death cases on July 28 and August 5. In United Kingdom (UK) 48867 and 53181 Cumulative number of death cases on July 28 and August 5. In Spain 30411 and 33596 Cumulative number of death cases on July 28 and August 5. In Italy 35389 and 35611 Cumulative number of death cases on July 28 and August 5. The figure 8 represent the forecasting the spread of COVID-19 (Cumulative Confirmed Cases) across the India. In Tamil Nadu 3232 and 4442 Cumulative number of death cases on July 28 and August 5. In Uttar Pradesh (UP) 1490 and 2475 Cumulative number of death cases on July 28 and August 5 is shown in figure 8. In Maharashtra 14673 and 20859 Cumulative number of death cases on July 28 and August 5. In Delhi 4225 and 5290 Cumulative number of death cases on July 28 and August 5. In Gujarat 2121 and 2197 Cumulative number of death cases on July 28 and August 5 is shown in figure 9.

The figures (fig 6 fig 7) represent the forecasting the spread of COVID–19 (Cumulative Confirmed Cases) across the World. In Tamil Nadu 217822 and 291515 Cumulative number of confirmed cases on July 28 and August 5. In Uttar Pradesh (UP) 62287 and 95773 Cumulative number of confirmed cases on July 28 and August 5. In Maharashtra 382336 and 512465 Cumulative number of confirmed cases on July 28 and August 5. In Delhi 141010 and 174828 Cumulative number of confirmed cases on July 28 and August 5. In Gujarat 55559 and 72911 Cumulative number of confirmed cases on July 28 and August 5.

7. Conclusion

This article proposed to utilize the Regression models for COVID–19 epidemic analysis using COVID–19 dataset from Kaggle COVID–19 repository. The results show that Multivariate Linear Regression (MvLR) yielded a good accuracy in forecasting the COVID–19 pandemic virus. These experiments are evaluated by using various error metrics such as Mean Absolute Error (MAE), R-Squared Score, Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). This proposed methodology will predict cumulative number of death rate and cumulative number of confirmed cases (after one and two weeks) using the cumulative number of confirmed cases /active cases and cumulative number of death cases each day. It will be helpful for the government and health care sector (hospitals and pharmacies) to prevent the spread of COVID–19. The study can further be extended using deep learning models to forecast the spread of COVID–19 with efficiency.

Declarations

Competing interests

The authors declare no competing interests.

References

1. Aggregating the response in time series regression models, applied to weather-related cardiovascular mortality - PubMed. URL <https://pubmed.ncbi.nlm.nih.gov/29438931/>
2. COVID-19 URL <https://covid19.gleamproject.org/>. <https://covid19.gleamproject.org/>, last accessed on 28/07/2020
3. COVID-19: When will it end? Two months or so, says Levitt's Model - The URL <https://thefederal.com/pandemic-1-mn-in-india/when-will-covid-19-pandemic-end-two-months-or-so-says-levitts-model/>
4. URL <https://www.covidanalytics.io/projections>. <https://www.covidanalytics.io/projections>, last accessed on 28/07/2020
5. CRISPR-based surveillance for COVID-19 using genomically- comprehensive machine learning design URL <https://www.biorxiv.org/content/10.1101/2020.02.26.967026v2>

6. Data-based analysis, modelling and forecasting of the COVID-19 URL
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0230405>.
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0230405>, last accessed on 26/07/2020
7. Estimation of the final size of the COVID-19 epidemic medRxiv. URL
<https://www.medrxiv.org/content/10.1101/2020.02.16.20023606v5>
8. Forecasting the novel coronavirus COVID-19. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0231236>. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0231236>, last accessed on 26/07/2020
9. IHME COVID-19 Projections. URL <https://covid19.healthdata.org/>. Library Catalog: covid19.healthdata.org
10. A Machine Learning Solution Framework for Combatting COVID-19 in Smart Cities from Multiple Dimensions medRxiv. URL <https://www.medrxiv.org/content/10.1101/2020.05.18.20105577v3>
11. Mortality from COVID-19 in 12 countries and 6 states of the United States medRxiv. URL <https://www.medrxiv.org/content/10.1101/2020.04.17.20069161v1>
12. Navigating “Down Time” During COVID-19. URL
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7158834/>.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7158834/>, last accessed on 28/07/2020
13. Projections. URL <https://covid-19.tacc.utexas.edu/projections/>. <https://covid-19.tacc.utexas.edu/projections/>, last accessed on 28/07/2020
14. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes Nature Medicine. URL <https://www.nature.com/articles/s41591-020-0868-6>
15. Simulating the infected population and spread trend of 2019- nCov under different policy by EIR model medRxiv. URL <https://www.medrxiv.org/content/10.1101/2020.02.10.20021519v1>
16. COVID-19 scenarios for the United States (2020). URL <http://www.healthdata.org/research-article/covid-19-scenarios-united-states>. Library Catalog: [healthdata.org](http://www.healthdata.org) Simulating the infected population and spread trend of 2019- nCov under different policy by EIR model URL <https://www.medrxiv.org/content/10.1101/2020.02.10.20021519v1>
17. Chinazzi, M., Davis, J.T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., Piontti, Py., Mu, K., Rossi, L., Sun, K., Viboud, C., Xiong, X., Yu, H., Halloran, M.E., Longini, I.M., Vespignani, A.: The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 368(6489), 395–400 (2020). DOI 10.1126/science.aba9757. URL <https://science.sciencemag.org/content/368/6489/395>. Publisher: American Association for the Advancement of Science Section: Research Article
18. Hethcote, W.: The Mathematics of Infectious Diseases. SIAM Review 42(4), 599–653 (2000). URL <https://www.jstor.org/stable/2653135>. Publisher: Society for Industrial and Applied Mathematics
19. Kissler, M., Tedijanto, C., Goldstein, E., Grad, Y.H., Lipsitch, M.: Project- ing the transmission dynamics of SARS-CoV-2 through the postpandemic pe- riod. Science 368(6493), 860–868 (2020). DOI

- 10.1126/science.abb5793. URL <https://science.sciencemag.org/content/368/6493/860>. Publisher: American Association for the Advancement of Science Section: Report
20. Li, Q., Guan, X., Wu, P., Wang, , Zhou, L., Tong, Y., Ren, R., Leung, K.S., Lau, E.H., Wong, J.Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T.T., Wu, J.T., Gao, G.F., Cowling, B.J., Yang, B., Leung, G.M., Feng, Z.: Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine* 382(13), 1199–1207 (2020). DOI 10.1056/NEJMoa2001316. URL <https://doi.org/10.1056/NEJMoa2001316>. Publisher: Massachusetts Medical Society eprint: <https://doi.org/10.1056/NEJMoa2001316>
 21. Luo, J.: Predictive Monitoring of COVID-19 p. 12
 22. Ndiaye, M., Tendeng, L., Seck, D.: Analysis of the COVID-19 pandemic by SIR model and machine learning technics for forecasting. arXiv:2004.01574 [math, q-bio, stat] (2020). URL <http://arxiv.org/abs/2004.01574>. ArXiv: 2004.01574
 23. Nesteruk, I.: Statistics based predictions of coronavirus 2019-nCoV spreading in main- land China. medRxiv p. 2020.02.12.20021931 (2020). DOI 10.1101/2020.02.12.20021931. URL <https://www.medrxiv.org/content/10.1101/2020.02.12.20021931v1>. Publisher: Cold Spring Harbor Laboratory Press
 24. Pandey, , Gautam, V., Jain, C., Syal, P., Sharma, H., Bhagat, K., Pal, R., Dhin- gra, L.S., Arushi, Patel, L., Agarwal, M., Agrawal, S., Arora, M., Rana, B., Ku- maraguru, P., Sethi, T.: A Machine Learning Application for Raising WASH Aware- ness in the Times of COVID-19 Pandemic. arXiv:2003.07074 [cs] (2020). URL <http://arxiv.org/abs/2003.07074>. ArXiv: 2003.07074
 25. Pourhomayoun, , Shakibi, M.: Predicting Mortality Risk in Patients with COVID-19 Using Artificial Intelligence to Help Medical Decision-Making. medRxiv p. 2020.03.30.20047308 (2020). DOI 10.1101/2020.03.30.20047308. URL <https://www.medrxiv.org/content/10.1101/2020.03.30.20047308v1>. Publisher: Cold Spring Harbor Laboratory Press
 26. Rajagopalan, , Al-Kindi, S.G., Brook, R.D.: Air Pollution and Cardiovascu- lar Disease: JACC State-of-the-Art Review. *Journal of the American College of Cardiology* 72(17), 2054–2070 (2018). DOI 10.1016/j.jacc.2018.07.099. URL <https://www.onlinejacc.org/content/72/17/2054>. Publisher: Journal of the American College of Cardiology Section: The Present and Future
 27. Rajagopalan, S., Huang, S., Brook, R.D.: Flattening the curve in COVID-19 using personalised protective equipment: lessons from air pollution. *Heart* (2020). URL <https://dx.doi.org/10.1136/heartjnl-2020-317104>
 28. Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J.M., Yan, , Chow- ell, G.: Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang, China: February 13–23, 2020. *Journal of Clinical Medicine* 9(2), 596 (2020). DOI 10.3390/jcm9020596. URL

<https://www.mdpi.com/2077-0383/9/2/596>. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute

29. Sarmento, M., Verburg, T.G., Almeida, S.M., Freitas, M.C., Wolterbeek, H.T.: Robustness of different regression modelling strategies in epidemiology: a time-series analysis of hospital admissions and air pollutants in Lisbon (1999–2004). *Environmetrics* 22(1), 86–97 (2011). DOI 10.1002/env.1031. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/env.1031>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.1031>
30. Wang, , Wong, A.: COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images. arXiv:2003.09871 [cs, eess] (2020). URL <http://arxiv.org/abs/2003.09871>. ArXiv: 2003.09871
31. Wu, T., Leung, K., Leung, G.M.: Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet* 395(10225), 689–697 (2020). DOI 10.1016/S0140-6736(20)30260-9. URL <http://www.sciencedirect.com/science/article/pii/S0140673620302609>
32. Xiao, S.Y., Wu, , Liu, H.: Evolving status of the 2019 novel coronavirus infection: Proposal of conventional serologic assays for disease diagnosis and infection monitoring. *Journal of Medical Virology* 92(5), 464–467 (2020). DOI 10.1002/jmv.25702. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.25702>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmv.25702>
33. Yan, , Zhang, H.T., Xiao, Y., Wang, M., Sun, C., Liang, J., Li, S., Zhang, M., Guo, Y., Xiao, Y., Tang, X., Cao, H., Tan, X., Huang, N., Jiao, B., Luo, A., Cao, Z., Xu, H., Yuan, Y.: Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medRxiv p. 2020.02.27.20028027 (2020). DOI 10.1101/2020.02.27.20028027. URL <https://www.medrxiv.org/content/10.1101/2020.02.27.20028027v2>. Publisher: Cold Spring Harbor Laboratory Press
34. Yang, , Zeng, Z., Wang, K., Wong, S.S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., Xie, Y., Liu, B., Wang, Z., Zhang, S., Wang, Y., Zhong, N., He, J.: Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease* 12(3), 165–174 (2020). DOI 10.21037/jtd.2020.02.64. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7139011/>
35. Zou, D., Wang, , Xu, P., Chen, J., Zhang, W., Gu, Q.: Epidemic Model Guided Machine Learning for COVID-19 Forecasts in the United States. medRxiv p. 2020.05.24.20111989 (2020). DOI 10.1101/2020.05.24.20111989. URL <https://www.medrxiv.org/content/10.1101/2020.05.24.20111989v1>. Publisher: Cold Spring Harbor Laboratory Press

Tables

Due to technical limitations, the tables are provided in the Supplementary Files section.

Figures

Figure 1

Cumulative number of Confirmed cases across the World for COVID-19

Figure 2

Cumulative number of Death cases across the World for COVID-19

Figure 3

Cumulative number of Confirmed cases across the India for COVID-19

Figure 4

Cumulative number of Death cases across the India for COVID-19

Figure 5

Proposed Methodology

Figure 6

Forecasting the spread of COVID-19 (Cumulative confirmed cases) across the World on July 28 to August 5

Figure 7

Forecasting the spread of COVID-19 (Cumulative death cases) across the World on July 28 to August 5

Figure 8

Forecasting the spread of COVID-19 (Cumulative death cases) across the India on July 28 to August 5

Figure 9

Forecasting the spread of COVID-19 (Cumulative confirmed cases) across the India on July 28 to August 5

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Tables.pdf](#)