**Prediction and Monitoring of Covid-19 Using Machine Learning Algorithms**

**Abstract**

Global health has been threatened by the rapid spread of the new coronavirus illness 2019 (COVID-19). In order to keep track of the progress of the pandemic, spread forecasting utilizing artificial intelligence and machine learning approaches is essential. Project goal is to visualize the virus's worldwide and country-wide transmission, as well as perform Linear regression, Support vector machine, Ensemble methods, Multilayer perceptron and Recurrent neural network-LSTM, ARIMA and Prophet on the COVID-19 data. There are a number of new parameters that we'll be looking at to see how they affect our prediction of COVID-19 spread.

**Keywords:** COVID-19, Machine learning, Pandemic, Predictive modeling, Visualization.

# Introduction

Coronavirus disease 2019 (COVID-19) is a destructive pandemic caused by SARS-CoV-2, the current causative agent of the present destructive coronavirus disease pandemic (COVID-19) in China [14, 6]. Since the outbreak began, millions of people have been infected and hundreds of thousands have died around the world. More than 15 million confirmed COVID-19 cases have been reported to date in 187 countries. Predicting and studying the spread of disease in such a setting can inspire design.

decision-making processes that are more effective. It's vital to note that studies like these help us make accurate predictions.

Machine learning provides a wide range of visualization and prediction tools, and it is currently being utilized around the world to study the spread of COVID-19, for example, [4, 7, 8, 15, 11]. Machine learning techniques will be used to analyze and depict the spread of the virus around the world and in specific countries during a specific time period, taking into account confirmed cases, recoverable cases, and fatalities, respectively.

Several studies have examined the global impact of the pandemic on numerous elements of life, such as [12, 1, 5, 3]. According to [2, 9, 10], a pandemic can be predicted by taking into account a wide range of elements, such as the impact of environmental conditions, quarantine, age, gender, and more.

The accuracy of forecasting is dependent on the availability of accurate data on which to make its predictions and provide an estimate of uncertainty. The current outbreak's datasets haven't been standardized by any standardization organization, which makes it difficult to employ machine learning algorithms. In addition, choosing the right parameters and the optimum machine learning model to use for prediction are also difficult aspects of training a model.

Linear regression, Support vector machine, Ensemble methods, Multilayer perceptron, Recurrent neural network-LSTM, ARIMA and Prophet, etc. will be used in this project on the COVID-19 data from Johns Hopkins University to predict the future effects of the global COVID-19 pandemic, including Iran and other countries. Furthermore, we plan to investigate the influence of environmental factors, life expectancy, demographic statistics, and other variables on the dissemination of COVID-19.

# Experimental data and results

There are three time series with the number of daily confirmed cases, recovered cases, and deaths by nation supplied by the Johns Hopkins University Center for Systems Science and Engineering (JHU-CSSE). This set of data is automatically refreshed each day.

We used data from January 22, 2020, to July 21, 2020, for this project. Many processes were necessary to clean the data because the dataset was not standard and there were a lot of procedures involved in preprocessing. This was done meticulously, and appropriate dataframes, such as the following, were generated.







After exploring the data, we performed some visualizations on the data in order to get a better understanding of the data and how the pandemic is affecting all of us. For example, in Figure 1, we can see the latest status of cases in the world.



Figure 1

Also in Figure 2, we can see that the latest global recovery rate per 100 cases is 56*.*16 whereas the mortality rate per 100 cases is 4*.*08, that is a good news because at the start point of this project, the recovery rate was around 54 percent whereas the mortality rate was around 5 percent.

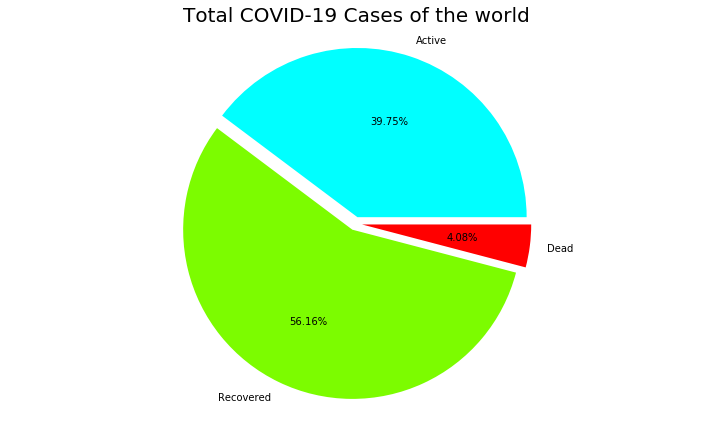


Figure 2

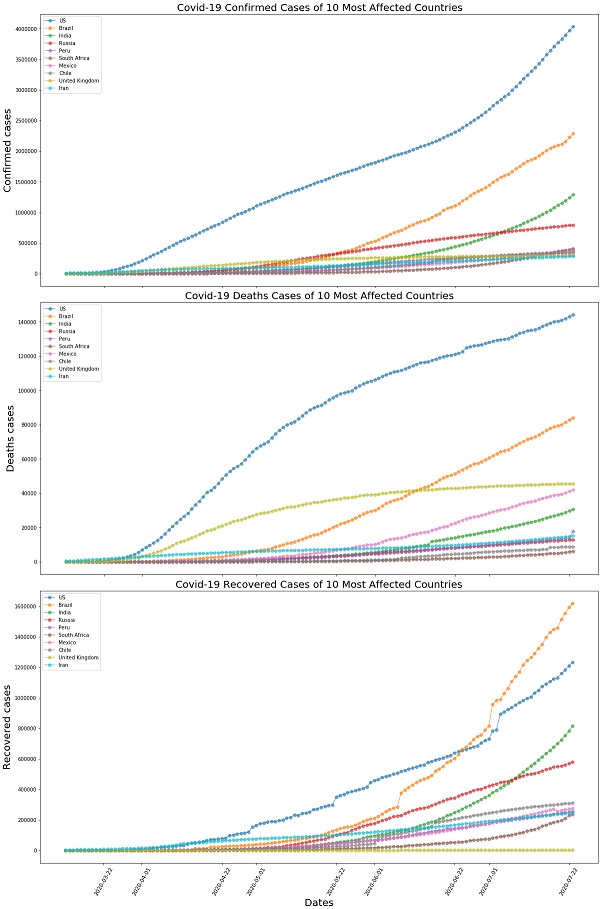


Figure 3 After visualization, we investigated data modeling and prediction based on univariate time series, using Linear regression, Support vector machine, Random forests, XGBoost, Multilayer perceptron (MLP), and a recurrent neural network, Long Short-Term Memory network (LSTM-RNN) to forcast the number of confirmed cases and deaths in the world and some other countries such as Iran. Some of our results are summarized in the following tables:

Table 1: Prediction errors of total confirmed cases of the world

|  |  |
| --- | --- |
| Regressor | RMSE |
| Support vector machine | 84855*.*25 |
| Linear regression | 880383*.*54 |
| Random Forests | 3254934*.*58 |
| XGBoost | 3172578*.*3 |

Table 2: Prediction errors of total deaths of the world

|  |  |
| --- | --- |
| Regressor | RMSE |
| Support vector machine | 140122*.*16 |
| Linear regression | 19337*.*55 |

Table 3: Accuracy of predicting the total cases of Iran using MLP and LSTM-RNN

|  |  |  |
| --- | --- | --- |
| Neural Network | MAPE | Accuracy(in percent) |
| MLP | 0*.*24124079849664185 | 99*.*99758759201504 |
| LSTM-RNN | 0*.*6490237741213257 | 99*.*99350976225878 |

Table 4: Accuracy of predicting the total deaths of Iran using MLP and LSTM-RNN

|  |  |  |
| --- | --- | --- |
| Neural Network | MAPE | Accuracy(in percent) |
| MLP | 0*.*10785990390804867 | 99*.*99892140096092 |
| LSTM-RNN | 1*.*1043418430881735 | 99*.*98895658156911 |

Table 5: Accuracy of predicting the total cases of the world using MLP and LSTM-RNN

|  |  |  |
| --- | --- | --- |
| Neural Network | MAPE | Accuracy(in percent) |
| MLP | 0*.*47257501056634127 | 99*.*99527424989434 |
| LSTM-RNN | 1*.*3621032695517894 | 99*.*98637896730449 |

Table 6: Accuracy of predicting the total deaths of the world using MLP and LSTM-RNN

|  |  |  |
| --- | --- | --- |
| Neural Network | MAPE | Accuracy(in percent) |
| MLP | 0*.*14795993064612276 | 99*.*99852040069354 |
| LSTM-RNN | 0*.*6681058141637979 | 99*.*99331894185836 |



Figure 4: Iran-Prediction of confirmed cases using Neural Networks



Figure 5: Iran-Prediction of deaths using Neural Networks

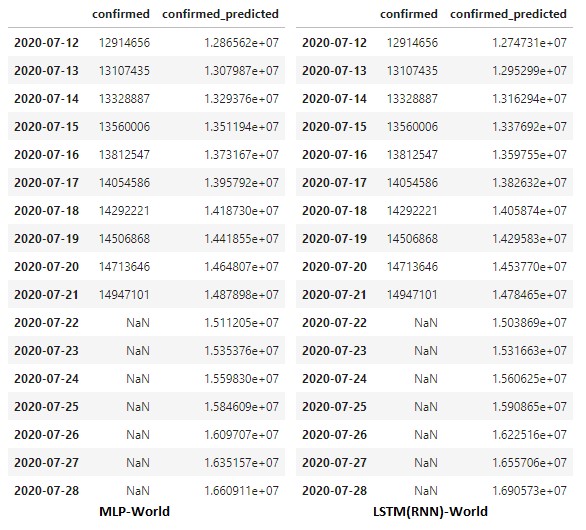


Figure 6: World-Prediction of confirmed cases using Neural Networks



Figure 7: World-Prediction of deaths using Neural Networks

Also, we did some predictions of confirmed cases and deaths related to 5 most affected countries and Iran using ARIMA and Prophet. As examples, Figures 8 and 9 show some of such predictions.

Moreover, we did some predictions on multivariate time series using Linear regression, Support vector machine, Ensemble methods, etc., and at the end, by examining the correlations between the features, we studied the impact of adding some new parameters such as life expectancy, GDP per capita, social support, freedom to make life choices, generosity, and population in prediction of COVID-19 spread.

# Conclusions and future works

As a conclusion based on the analysis of the observations, it seems that even though the total number of confirmed cases and deaths in the world are monotonically (almost exponentially) increasing, the recovery rate shows some increase whereas the mortality rate shows some decrease. On the other hand, by data modeling and prediction based on univariate time series, using Linear regression, Support vector machine, Random forests and XGBoost we concluded that Support vector machine and Random forests performed the best and the worst accuracy, respectively. Moreover, both of Multilayer perceptron and LSTM-RNN performed high accuracy, more than 99*.*98 in percent.

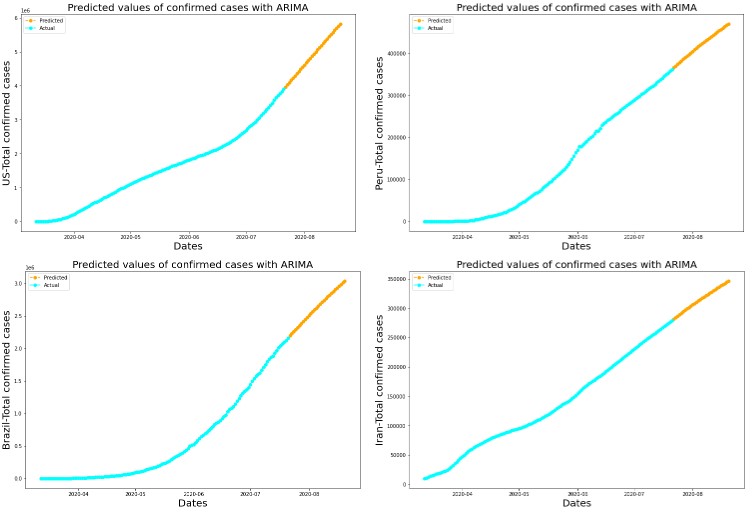


Figure 8: ARIMA-Confirmed

Furthermore, by examining the correlations between the features, it seems that there exist week correlations between the new parameters, life expectancy, GDP per capita, social support, freedom to make life choices, generosity, and the primary ones, confirmed, deaths, recovered, active cases, recovery rate and mortality rate. Also, it seems that the correlation between population and confirmed and, population and active cases is moderate (near 0*.*4).

As future works, by considering the population of each country, we may investigate the percentage of total populations that will be affected by COVID-19. Also, the impact of some other parameters in prediction of COVID-19 spread can be considered. Moreover, data modeling and prediction based on multivariate time series using Multilayer perceptron and LSTM-RNN can be considered.

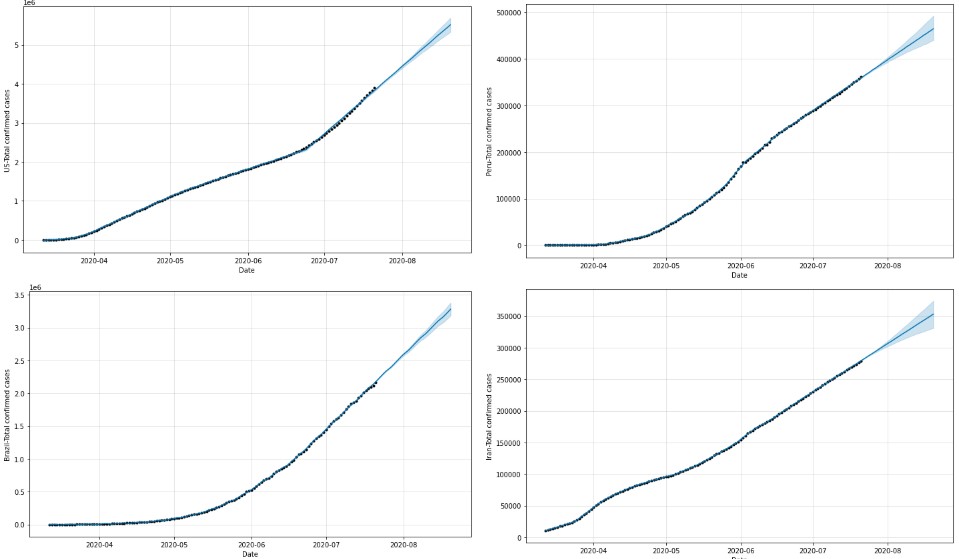


Figure 9: Prophet-Confirmed

# References

1. Alon, T. M., et al., The impact of COVID-19 on gender equality, National Bureau of Economic Research, (2020), no. w26947.
2. Chen, B., et al., Roles of meteorological conditions in COVID-19 transmission on a worldwide scale, MedRxiv, (2020).
3. Fernandes, N., Economic effects of coronavirus outbreak (COVID-19) on the world economy, Available at SSRN 3557504, (2020).
4. Fong, S. J., Li, G., and Dey, N., Finding an Accurate Early Forecasting Model from Small Dataset: A Case of 2019-nCoV Novel Coronavirus Outbreak, Int. j. interact. multimed. artif. intell., 6(2020), no. 1, 132–140.
5. Ho, C. S., Chee, C. Y., and Ho, R. C., Mental health strategies to combat the psychological impact of COVID-19 beyond paranoia and panic, Ann Acad Med Singapore, 49(2020), no. 1, 1–3.
6. Huang, C., Wang, Y., Li, X., et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, The Lancet , 395(2020), no. 10223, 497–506.
7. Jia, L., Li, K., Jiang, Y., and Guo, X., Prediction and analysis of coronavirus disease 2019, arXiv preprint, (2020).
8. Kumar, J., and Hembram, K. P. S. S., Epidemiological study of novel coronavirus (COVID-19), arXiv preprint, (2020).
9. Ma, Y., et al., Effects of temperature variation and humidity on the mortality of COVID19 in Wuhan, MedRxiv, (2020).
10. Shi, P., et al., The impact of temperature and absolute humidity on the coronavirus disease 2019 (COVID-19) outbreak-evidence from China, MedRxiv, (2020).
11. Sujath, R., et al., A machine learning forecasting model for COVID-19 pandemic in India, Stochastic Environmental Research and Risk Assessment, (2020), no. 34, 959– 972.
12. Walker, P., et al., Report 12: The global impact of COVID-19 and strategies for mitigation and suppression, (2020).
13. World Health Organization (WHO), [Naming the coronavirus disease (COVID–19).](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it)
14. World Health Organization (WHO), [Novel Coronavirus–China,](https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/) Retrieved 9 April 2020.
15. Yang, Z., et al., Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions, J Thorac Dis, 12(2020), no. 3, 165.