

Prediction and Spread Visualization of the Novel Coronavirus Disease COVID-19 Using Machine Learning

Zahra Taheri

July 2020

The current destructive pandemic of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1], was first reported in Wuhan, China, in December 2019 [2, 3]. The outbreak has affected millions of people around the world and the number of infections and mortalities has been growing at an alarming rate. In such a situation, forecasting and proper study of the pattern of disease spread can inspire design better strategies to make more efficient decisions. Moreover, such studies play an important role in achieving accurate predictions.

Machine learning has numerous tools that can be used for visualization and prediction, and nowadays it is used worldwide for study of the pattern of COVID-19 spread, e.g., see [4, 5, 6]. One of the main focus of the study in this project is to use machine learning techniques to analyze and visualize the spreading of the virus country-wise as well as globally during a specific period of time by considering confirmed cases, recovered cases and fatalities.

The global impact of the outbreak on various aspects of life has been the focus of many studies, e.g., see [7, 8, 9, 10]. On the other hand, a pandemic can be forecast by considering a variety of parameters such as the impact of environmental factors, quarantine, age, gender and a lot more.

The forecasting accuracy depends on the availability of proper data to base its predictions and provide an estimate of uncertainty. A challenge to use machine learning techniques for the current outbreak is that the datasets are not yet standardized by any standardization organization and the statistical anomalies are not considered. Also, the appropriate selection of parameters and the selection of the best machine learning model for prediction are other challenges involved in training a model.

In this project, we are going to perform Linear regression, Support vector machine, Multilayer perceptron, Ensemble methods, etc., on the Johns Hopkins University's COVID-19 data to anticipate the future effects of COVID-19 pandemic in Iran and some other countries. Moreover, we are going to study the impact of some parameters such as geographic conditions, economic statistics, population statistics, life expectancy, etc., in prediction of COVID-19 spread.

REFERENCES

- [1] World Health Organization (WHO), Naming the coronavirus disease (COVID-19) and the virus that causes it.
- [2] World Health Organization (WHO), Novel Coronavirus-China, Retrieved 9 April 2020.
- [3] 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.
- [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] Jia, L., Li, K., Jiang, Y., and Guo, X., Prediction and analysis of coronavirus disease 2019, arXiv preprint, (2020).

Keywords and Phrases: Coronavirus, COVID-19, Machine learning, Pandemic, Predictive modelling, Visualization.

Contact: zahra.taheri518@gmail.com

- [6] Kumar, J., and Hembram, K. P. S. S., Epidemiological study of novel coronavirus (COVID-19), arXiv preprint, (2020).
- [7] Walker, P., et al., Report 12: The global impact of COVID-19 and strategies for mitigation and suppression, (2020).
- [8] Alon, T. M., et al., The impact of COVID-19 on gender equality, National Bureau of Economic Research, (2020), no. w26947.
- [9] 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.
- [10] Fernandes, N., Economic effects of coronavirus outbreak (COVID-19) on the world economy, Available at SSRN 3557504, (2020).