Analysis of COVID-19 Pandemic in India – Using Machine Learning Models

Shivakumar Nethani[1] and Dr. P. U. Anitha[2], P. Ramana[3]

Assistant Professor, Department of Computer Science and Engineering[1,2,3]

Hyderabad Institute of Technology and Management, Hyderabad, Telangana, India[1,3]

Christu Jyothy Institute of Technology and Sciencet, Janagaon, Telangana, India[2]

Correspondence should be addressed to Shivakumar Nethani; shivakumar.kumar093@gmail.com

Correspondence should be addressed to Anitha Podishetty; anithapodishetty123@gmail.com

Correspondence should be addressed to Rama Pochaboina; pochaboinarama@gmail.com

Abstract: *The early and reliable detection of COVID-19 infected patients is essential to prevent and limit its outbreak. The PCR tests for COVID-19 detection are not available in many countries, and also, there are genuine concerns about their reliability and performance. Motivated by these shortcomings, this article proposes a deep uncertainty-aware transfer learning framework for COVID-19 detection using medical Images. One of the practical and topical applications in the current scenario is to use the power of Machine Learning to study various aspects of the on going pandemic (COVID-19), since the entire world is in its grip. COVID-19 was declared a global pandemic on 11th March 2020 by WHO. Worldwide more than 43 Million people have contracted this viral disease and more than 1.1 million people have succumbed to it (as on 27th October 2020). Our approach uses Classification from Supervised Learning techniques to solve this problem. The efficacy of this approach could be used to scale and develop automated systems that could predict the likeliness of Covid-19 based on laboratory tests that are readily accessible. From the features presented to us in the dataset, we are able to predict with 87.0 - 97.4 percent accuracy at a 95 percent confidence level that a patient is suffering from Covid-19 when biomarkers are taken into consideration. The number of people affected by COVID-19 in India is increasing at a fast pace and currently India has the second highest number of cases and third highest casualties in the world. Four different Machine Learning algorithms namely Random Forest Regression, Multiple Linear Regression, Support Vector Regression and Lasso Regression have been considered. A Kaggle dataset consisting of figures of confirmed cases, patients recovered, and people that have died due to COVID-19 across India over a particular period of time has been used. The results of this study indicate that Random Forest Regression provides the most accurate results whereas Support Vector Regression is least accurate.*

Keywords: COVID-19, Lasso Regression, Multiple Linear Regression, Support Vector Machine, Machine Learning, Classification, Supervised Learning, Random forest Classifier

REFERENCES

[1]. (2020). Coronavirus Disease 2019 (COVID-19). Accessed: Feb. 27, 2020. [Online]. Available: https://www.who.int/emergencies/ diseases/novel-coronavirus-2019/situati%on-reports.

[2]. C. P. E. R. E. Novel, "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China", Zhonghua LiuXing Bing Xue Za Zhi= Zhonghua Liuxingbingxue Zazhi, vol. 41, no. 2, p. 145, 2020.

[3]. A guide to WHO's guidance on Covid-19. (n.d.). Retrieved from https://www.who.int/news-room/feature-stories/detail/a-guide-to- who-s-guidance.

[4]. M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang, "Disease prediction by machine learning over big data from healthcare communities", IEEE Access, vol. 5, pp. 8869-8879,2017.

[5]. Emergency Approval to Use Diagnostic Test for Corona-Virus. Accessed: Jan. 23, 2020. [Online]. Available: https://www. reuters.com/article/us-china-health-cdc/u-s-cdc-seeks-emergency-%approval-to-use-diagnostic-test-for-coronavirus-idUSKBN1ZM2XS

[6]. Andre Filipe de Moraes Batista, J. L. (April-14-2020). COVID-19 diagnosis prediction in emergency care patients: a machine learning approach.

[7]. A. Tomar and N. Gupta, "Prediction for the spread of COVID- 19 in India and effectiveness of preventive measures", Science of the Total Environment, vol. 728, 2020,138762.

[8]. B. Huang et al., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," Lancet, vol. 395, pp. 497–506, May 2020.

[9]. Complete Blood Count.(n.d.). Retrieved from https://www.mayoclinic.org/tests-procedures/complete-blood-count/about/pac-20384919

[10]. M. Barstugan., U. Ozkaya and S. Ozturk, "Coronavirus (COVID-19) classification using CT images by machine learning method", 2020 arXiv Preprint arXiv: 2003.09424.

[11]. S. Hassanpour et al., "Deep learning for classification of colorectal polyps on whole-slide images," J. Pathol. Informat., vol. 8, no. 1, p. 30, 2017.

[12]. Diagnosis of COVID-19 and its clinical spectrum. (n.d.). Retrieved from Kaggle: https://www.kaggle.com/einsteindata4u/covid19

[13]. Symptoms of Coronavirus.(n.d.). Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html

[14]. A. K. Jaiswal, P. Tiwari, S. Kumar, D. Gupta, A. Khanna, and J. J. P. C. Rodrigues, "Identifying pneumonia in chest X-rays: A deep learning approach," Measurement, vol. 145, pp. 511–518, Oct. 2019.

[15]. COVID-19inIndia, Available online: https://www.kaggle.com/sudalairajkumar/covid19-in- India? Select=covid_19_india.csv.