

Abstract No: SI-14

Prediction of the incubation period of COVID-19 patients using machine learning techniques

P. P. P. M. T. D. Rathnayake*, D. N. Wickramaarachchi and J. M. D. Senanayake

Department of Industrial Management, Faculty of Science, University of Kelaniya, Sri Lanka
thidasala.demintha@gmail.com*

Coronavirus Disease 2019 (COVID-19) is a respiratory infection caused by a newly discovered coronavirus. As of September 2020, within about eight months of this infectious disease being identified, more than thirty million cases and 950,000 deaths have been reported within two hundred countries and territories. The incubation period of COVID -19, is the time range between exposure to symptom onset. During this period, affected persons may not show symptoms of being infected but are still capable of transmitting the virus to others. It is very important to identify the incubation period accurately to decide quarantine periods and to develop policies. Based on the current findings, the incubation period ranges between 2 to 14 days. Since there is a range to the incubation period, almost all the suspected cases should undergo a quarantine period of 14 days, which sometimes leads to inefficient allocation of resources in some cases. Although there are many studies on assessing the incubation period, studies regarding the factors affecting the incubation period are limited. This study is primarily aimed at identifying the factors affecting the incubation period and to develop a model to classify the incubation period of suspected cases, using machine learning techniques. Publicly available patient records within different countries were used for the study. The gathered dataset consist of 500 patients records with the age ranging from 5 to 80 years. Out of those records, 285 were male and 215 were female. The dataset includes 205 patients from China, 51 patients from Japan, 36 patients from Malaysia, 24 patients from the United States, 41 patients from South Korea, 31 patients from France, 24 patients from Taiwan, 46 patients from Singapore, and 42 patients from other countries. The results indicate that factors such as patients' age, gender, geographic location, immunocompetent/immunocompromised state, direct/indirect contact with the affected patients, cause deviations to the incubation period. Chi-square test of independence and correlation analysis were used to identify the relationship among variables and to identify the factors which have the strongest relationship with the incubation period. Supervised learning classification algorithms such as Support Vector Machine, Naïve Bayes, Decision Tree, Logistic Regression, and Random Forest were compared in this study. Overall model performance was evaluated using the weighted average of the incubation classes. Random forest was selected as the best algorithm to classify the incubation period since it performed better than other algorithms achieving a 0.78 precision score, 0.84 recall score, and 0.80 F1 score. As the final step, AdaBoost algorithm was used to improve the performance of the Random Forest algorithm.

Keywords: AdaBoost, Boosted Random Forest, COVID-19, Incubation period