Enhancing COVID-19 Diagnosis based on Feature Extraction of Complete Blood Count Tests

Manar Ahmed Elfakahany, Mahmoud Adel Hassan, Mona A. Bayoumi, Omar M. Salim

Abstract

Early and effective identification of COVID-19 is fundamental for effective treatment and virus control. Classifying the COVID-19 disease is based on lab work of common complete blood count (CBC) data that became a focal point of interest. CBC testing is inexpensive, broadly available, and may provide useful indicators for identifying COVID-19. Although CBC provides useful diagnostic data, not every feature has the same influence on classification performance. In that belief, machine learning (ML) algorithms have been applied to study CBC datasets and help in screening COVID-19 in the persons who may catch the disease. In ML, feature selection (FS) is one of the greatest essential challenges. Several real-world datasets include several redundant or unnecessary features, which reduce a classifier's performance. In this paper, five ML algorithms are used to improve the classifier performance. FS based on filter, wrapper, and hybrid methods employed to extract the most significant features of COVID-19 from a CBC test for datasets from a well-known repository while pursuing the model performance. The most relevant CBC attributes that may indicate COVID-19 are identified. Hybrid FS techniques offer the best results with fewer features. The results show that CBC features are not universally transferable across populations.

Keywords: COVID-19, Machine Learning, Feature Selection, CBC test