Abstract

Samir Y Marzouk

A diagnostic genomic signal processing (GSP)-based system for automatic feature analysis and detection of COVID-19

Coronavirus Disease 2019 (COVID-19) is a sudden viral contagion that appeared at the end of last year in Wuhan city, the Chinese province of Hubei, China. The fast spread of COVID-19 has led to a dangerous threat to worldwide health. Also in the last two decades, several viral epidemics have been listed like the severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002/2003, the influenza H1N1 in 2009 and recently the Middle East respiratory syndrome coronavirus (MERS-CoV) which appeared in Saudi Arabia in 2012. In this research, an automated system is created to differentiate between the COVID-19, SARS-CoV and MERS-CoV epidemics by using their genomic sequences recorded in the NCBI GenBank in order to facilitate the diagnosis process and increase the accuracy of disease detection in less time. The Selected database contains 76 genes for each epidemic. Then, some features are extracted like a discrete Fourier transform (DFT), discrete cosine transform (DCT) and the seven moment invariants to two different classifiers. These classifiers are the k-nearest neighbor (KNN) algorithm and the trainable cascade-forward back propagation neural network where they give satisfying results to compare. To evaluate the performance of classifiers, there are some effective parameters calculated. They are accuracy (ACC), F1 score, error rate and Matthews correlation coefficient (MCC) that are 100%, 100%, 0 and 1, respectively, for the KNN algorithm and 98.89%, 98.34%, 0.0111 and 0.9754, respectively, for the cascade-forward network.