

Comprehensive Machine Learning Analysis on the Phenotypes of COVID-19 Patients Using Transcriptome Data

Pratheeba Jeyanthan

Department of Computer Engineering, University of Jaffna, Sri Lanka.

E-mail: pratheeba@eng.jfn.ac.lk

Abstract

Purpose: Evolving technologies allow us to measure human molecular data in a wide reach. Those data are extensively used by researchers in many studies and help in advancements of medical field. Transcriptome, proteome, metabolome, and epigenome are few such molecular data. This study utilizes the transcriptome data of COVID-19 patients to uncover the dysregulated genes in the SARS-COV-2.

Method: Selected genes are used in machine learning models to predict various phenotypes of those patients. Ten different phenotypes are studied here such as time since onset, COVID-19 status, connection between age and COVID-19, hospitalization status and ICU status, using classification models. Further, this study compares molecular characterization of COVID-19 patients with other respiratory diseases.

Results: Gene ontology analysis on the selected features shows that they are highly related to viral infection. Features are selected using two methods and selected features are individually used in the classification of patients using six different machine learning algorithms. For each of the selected phenotype, results are compared to find the best prediction model.

Conclusion: Even though, there are not any significant differences between the feature selection methods, random forest and SVM performs very well throughout all the phenotype studies.

Keywords: COVID-19, Transcriptome data, Phenotype analysis, Machine learning models, Respiratory diseases, Dysregulated genes.

Received: 08/11/2021

Revised: 08/02/2022

Accepted: 28/02/2022

Introduction

Due to its high mortality rate and high spreading rate, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) poses a critical challenge to public health (Li, Liu, Yu, Tang, & Tang, 2020). Even though studies show that the vaccines reduce the severity of the disease, the world continues to suffer in controlling the spread of this disease (Christie, et al., 2021). As of 27th August 2021, there are 214, 468, 601 total confirmed cases and 4, 470, 969 total deaths reported to World Health Organization. Further, they reported that 4, 953, 887, 422 vaccine doses have been administered till 24 August 2021. This pandemic has a great impact not only on health, but also on economics, politics, social and other important aspects of many countries (Tisdell, 2020; Padhan & Prabheesh, 2021).

