



Role of different types of RNA molecules in the severity prediction of SARS-CoV-2 patients

Pratheeba Jeyanthan

Faculty of Engineering, University of Jaffna, Sri Lanka

ARTICLE INFO

Keywords:

Severity prediction
Classification algorithm
lncRNA
miRNA and mRNA
COVID-19 molecular data
Feature selection
Treatment stage

ABSTRACT

SARS-CoV-2 pandemic is the current threat of the world with enormous number of deceases. As most of the countries have constraints on resources, particularly for intensive care and oxygen, severity prediction with high accuracy is crucial. This prediction will help the medical society in the selection of patients with the need for these constrained resources. Literature shows that using clinical data in this study is the common trend and molecular data is rarely utilized in this prediction. As molecular data carry more disease related information, in this study, three different types of RNA molecules (lncRNA, miRNA and mRNA) of SARS-COV-2 patients are used to predict the severity stage and treatment stage of those patients. Using seven different machine learning algorithms along with several feature selection techniques shows that in both phenotypes, feature importance selected features provides the best accuracy along with random forest classifier. Further to this, it shows that in the severity stage prediction miRNA and lncRNA give the best performance, and lncRNA data gives the best in treatment stage prediction. As most of the studies related to molecular data uses mRNA data, this is an interesting finding.

1. Introduction

A pneumonia caused by a novel coronavirus (SARS-CoV-2) emerged in China in December 2019 and rapidly spread over the world [17]. Since then the world is suffering with this pandemic. It affects many aspects of lives including politics ([55].), economy [39], business [26], livelihood and psychology of people [75], healthcare worker's wellness [92] and education quality [80]. Also it has a great impact on researches, mainly in distinct areas of medical field [106].

One such study is analyzing the molecular data of COVID-19 patients. Literature shows that different kind of molecular data were measured on the patients for various biological studies and they were stored in public repositories. However, using those data in machine learning prediction seems to be very low. Even though previous studies show that molecular data carries much information about the diseases, and a huge number of such COVID-19 related data was made available for public, clinical features of the patients were used in bulk in machine learning studies of COVID-19 patients.

However, few interesting studies used molecular data in machine learning related studies. It includes cell marker identification of COVID-19 severity [19], severity related SARS-CoV-2 mutations [37] and methylation signatures for predicting severity of SARS-CoV-2 [56].

These studies identified several biomarkers related to SARS-CoV-2 severity in different level of gene expression. Further to this, [54] did a systematic review in the genome variation, origin tracing, and vaccine development for COVID-19. This review covers epidemiology, immunological pathogenesis, molecular mechanisms, structural biology, and approaches for drug or vaccine development for SARS-CoV-2.

Few other studies used transcriptome data in severity prediction [42], finding comorbidity interactions of COVID-19 with other diseases [63], comparing COVID-19 with other COVID infections [40,41], prolonged viral shedding prediction [43,44], and identifying therapeutic potentials [88] of SARS-CoV-2 patients. On top of this, transcriptome data was used along with machine learning in the prediction of ten different phenotypes of COVID-19 patients including time since onset, COVID-19 status, impact of age on COVID-19, hospitalization status and ICU status [43,44].

Single cell data was also used in the severity prediction of COVID-19 patients [72]. lncRNA [99,113], Protein [57,104], miRNA [27,31], epigenome data [15,91] and Multiomic data [89,113] also used in few researches related to COVID-19. However, targets of those studies differ in many directions including time series prediction of confirmed/death cases [12,34,82], prediction for need for ventilation [14], distinguish normal cough from COVID-19 cough [71] and discharge time prediction

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

<https://doi.org/10.1016/j.prp.2023.154311>

Received 18 October 2022; Received in revised form 11 January 2023; Accepted 14 January 2023

Available online 15 January 2023

0344-0338/© 2023 Elsevier GmbH. All rights reserved.