Chapter 10
Machine Learning-Based Categorization of COVID-19 Patients

Tanvi Arora
https://orcid.org/0000-0002-8454-6005
CGC College of Engineering, India

ABSTRACT

The world has been put to a standstill by the COVID-19 pandemic, which has been caused by the SARS-CoV-2 (initially called 2019-nCoV) infecting agent. Moreover, this pandemic is spreading like a wildfire. Even the developed nations are running short of hospital beds and ventilators to treat the critically ill. Considering the total population of the world and the pace at which this pandemic is spreading, it not possible to hospitalize all the positive patients with intensive care facilities. In the chapter, the authors present a machine learning-based approach that will categorize the COVID-19 positive patients into five different categories, namely asymptomatic, mild, moderate, severe, and critical. The proposed system is capable of classifying the COVID-19-affected patients into five distinct categories using selected features of age, gender, ALT, hemoglobin, WBC, heart disease, hypertension, fever, muscle ache, shortness of breath with 97.5% accuracy.

INTRODUCTION

The *SARS-CoV-2* (initially called 2019-nCoV) infecting agent is zoonotic and is spreading from one infectious person to another. Moreover, the symptoms of the infection becomes evident over a while, till that time the carrier of the virus

DOI: 10.4018/978-1-7998-9012-6.ch010

unknowingly comes in contact with several other persons. To date no firm vaccine has been developed to curtail the effect of this disease, prevention is the only viable escape. Since this disease emerged at such a large scale over a short duration of time, and has spread the world over. The world is not prepared to handle this pandemic.

Coronaviruses are a family of viruses that are known since the 1960s and they can infect humans as well as animals. They have been seen to cause respiratory infections. Towards the end of the year 2019, all of a sudden there was a surge of pneumonia patients in the Wuhan city of Hubei province of China and most of the patients that reported that diseases had a connection with the seafood market of Wuhan. The disease has been named as "2019 novel Coronavirus"(Suganthan, 2019).

This novel virus has been recognized as similar to SARS-CoV(Severe Acute Respiratory Syndrome Coronavirus) based upon the phylogeny and taxonomy of the past studies therefore it has been renamed as SARS-CoV-2(Gorbalenya, 2020).

The COVID 19 infection can spread like a wildfire from one human to human. Therefore whoever is found positive with COVID 19, he is isolated from others to stop the spread of the disease. The persons who have come in contact with the COVID 19 patient are also put in quarantine, as a precautionary measure, as the symptoms of the infection does not appear instantaneously, the virus is believed to have an average incubation period of 5 days and the maximum period of 14 days. The patients thus put in quarantine are said to be the suspected cases. On the contrary the patients who exhibit pneumonia-like symptoms like fever, short breathlessness, dry cough are also termed as suspected cases and are also monitored by putting in isolation(Zhang et al., 2020).

The COVID 19 infection is confirmed by carrying out a laboratory test of the suspected case, by taking a swap of the patient's nose or the back of the throat as the nose and back of the throat are the hot sites where the virus is replicating. The swabs are then put into a solution that helps to release the cells from the swab, then the genetic code of the COVID 19 is matched with the cells captured from the swab. (Tian et al., 2020)(Xiong et al., 2020).

Considering the total population of the world and the pace at which this pandemic is spreading, it not just possible to hospitalize all the positive patients with intensive care facilities In the proposed work we are presenting a machine learning-based approach, that will categorize the COVID 19 tested positive patients into five different categories namely Asymptomatic, Mild, Moderate, Severe, and Critical. Then based upon the severity of the patient and availability of the space in the Intensive Care Units, the Asymptomatic, Mild and Moderate patients can be kept in isolation without intensive care equipment and the severe and critical patients can be kept in the ICU's. In this study a dataset of 3567 COVID 19 patients has been taken from the hospitals located in Wuhan, China. The Random Forest-based machine learning model has been trained using the medical records of 2567 patients and the rest of the