Chapter 10
COVID-19 Analysis, Prediction, and Misconceptions:
A Computational Machine Learning Model as a New Paradigm in Scientific Research

Balachandran Krishnan
https://orcid.org/0000-0002-9051-8801
CHRIST University (Deemed), India

Sujatha Arun Kokatnoor
CHRIST University (Deemed), India

Vandana Reddy
CHRIST University (Deemed), India

Boppuru Rudra Prathap
https://orcid.org/0000-0002-5161-4972
CHRIST University (Deemed), India

ABSTRACT

COVID-19 is an infectious disease of the newly discovered coronavirus (CoV). The importance and value of open access (OA) resources are critical in the context of the COVID-19 epidemic. OA aided in the development of a vaccine and informed public health actions necessary to stop the virus from spreading. Many publishers implicitly acknowledged that OA was vital to promote science in the fight against the disease. Accordingly, publishers have committed to OA publication and scholarly communication of disease-related scientific research. This chapter covers three issues based on the modeling of the CoV dataset. First, an exploratory data analysis is done to detect the hidden facts and the relevant information patterns about the affected, recovered, death cases caused by the CoV and the vaccination details. Second, a predictive model is developed using machine learning techniques to effectively predict the number of COVID-19 positive cases in India. In the last step, a hybrid computational model is developed to identify the misconceptions that are spread through social media networks.

DOI: 10.4018/978-1-7998-9805-4.ch010

INTRODUCTION

Scholarly journals have been turned into online publications/journals with the advent of the internet, and have developed numerous beneficial capabilities such as online submission, searching, indexing and referring to many items beyond merely citation referencing for improved scholarly communication. The importance and value of Open Access are critical in the face of the COVID-19 epidemic. Open access to scientific information and open data aids in the development of a vaccine and informs public health actions necessary to stop the virus from spreading. Open access resources keep citizens informed and educated about the virus, ensuring that they follow public health recommendations and allowing for distance study.

The novel coronavirus (COVID-19) was widely replicated in China at the end of 2019, infecting a substantial proportion of people. The coronavirus is a family of viruses capable of causing a variety of diseases that are life threatening to humans, including common and more severe forms of cold. The signs and symptoms of the disease may occur within two to 14 days after exposure. This time referred to as the incubation period is the time after exposure and before symptoms. The general signs and symptoms include fever, cough, tiredness, breathing difficulty, sore throat, running nose, headache and chest pain (Sear, R. F. et al., 2020). Other less common signs also include rash, nausea, vomiting and diarrhea. Some people may only have a few symptoms and some may not have any symptoms at all. These cases are referred to as cases, symptomatic and asymptomatic respectively.

As per the World Health Organization (WHO), data have shown that the virus spreads from person to person (about 6 feet or 2 meters) among the people in close contact. The virus spreads through respiratory droplets when someone is coughing, sneezing or talking. Such droplets may be inhaled or landed in a nearby person's mouth or nose. It can also spread when a person touches a surface and touches his or her mouth, nose or eyes, but this is not a major way of spreading the virus as per WHO reports (Saba, T. et al., 2021). In the case of symptoms (symptomatic), a person with the virus is the most infectious – and this is the time that they are most likely to transmit the virus – according to the Center for Disease Control and Prevention (CDC) trusted Source. But even before they start showing symptoms (asymptomatic) of the disease itself, someone can spread it.

India had the world's second highest (after the US), with 29.3 million cases of COVID-19 infections documented, and the third largest number of COVID-19 deaths (after the US and Brazil) with 367.081 deaths as of 12th June 2021. A second wave started in March 2021 with shortage of vaccines, hospital beds, oxygen cylinders and others in the various sections of the country being significantly larger than that of the first one. India led the globe in new and active cases by the end of April. In a 24-hour period on 30th April 2021, the country was the first to record more than 400,000 new cases. Health experts feel that India has underreported its data owing to a number of circumstances.

This chapter aims to study over time cumulative data on confirmed cases, deaths and recovered cases, and to analyze the transmission of this virus across India in the first step. It is feasible to acquire insight into how each state performed in COVID-19 using this data. During what time period was the particular condition successful, so that other Indian states might learn from their processes during that time period. In the second step, a predictive model is developed using the machine learning techniques to effectively predict the number of COVID_19 positive cases in India. AutoRegressive Integrated Moving Average (ARIMA), Seasonal Auto Regressive Integrated Moving Average with eXogenous factors (SARIMAX), FBProphet, Logistic Regression, Linear Regression, Ridge Regression, Decision Trees, Random Forest and Neural Networks are used for the predictive analysis in this chapter. In the last step,