

PREDICTION OF COMMUNITY TRANSMISSION LEVEL OF COVID-19 USING MACHINE LEARNING ALGORITHMS BASED ON THE CDC SOCIAL VULNERABILITY INDEX

Saviz Saei, Yibin Wang, Mohammad Marufuzzaman, Nazanin Morshedlou, Haifeng Wang

Department of Industrial and Systems Engineering
Mississippi State University, Mississippi State, MS, 39762 USA

Corresponding Author: Haifeng Wang
Email: wang@ise.msstate.edu
DOI: https://doi.org/10.34107/ LWWJ5713168

ABSTRACT

Response to hazardous events is crucial in every community, whether natural or anthropogenic disasters. Social Vulnerability Index (SVI) helps people who need support. Social vulnerability refers to the number of adverse effects of external stress, including natural causes or disease outbreaks like the Coronavirus Disease 19 (COVID-19) pandemic on human health. The SVI dataset possesses California state of the US, subdivisions of counties of 15 features into four groups as related themes (i.e., socioeconomic status; household composition and disability; minority status and language; and housing type and transportation). In addition to the SVI dataset, the recent COVID-19 data tracker for each county posted by the Centers for Disease Control and Prevention (CDC) shows the new cases per 100,000 persons in the last seven days. The transmission values are low, moderate, substantial, and high. The impact of SVI on COVID-19 attracts the attention of researchers to find the relationships between SVI and COVID-19 incidence. This paper aims to incorporate SVI data and the incidence in the urban and rural areas of the United States using eight machine learning algorithms for COVID-19 transmission level classification. The experimental results show the proper prediction based on the community transmission level of COVID-19 by considering the features of SVI. Among all used machine learning methods, Random Forest achieved the best performance based on the percentage of various performance metrics accuracy and F1-score.

Keywords: Social Vulnerability Index(SVI); Coronavirus Disease 19(COVID-19); Machine learning algorithms; Prediction

INTRODUCTION

Since December 2019, Wuhan, China, has experienced Coronavirus Disease 19 (COVID-19). The world has been battling the COVID-19 pandemic with over 537,591,764 susceptibles and 6,319,395 deaths. This virus, claimed to be originated from bats [1], spreads quickly through contact and droplet, airborne, and fomite transmissions [2,3]. On February 6, 2020, the first COVID-19 death occurred at a home in Santa Clara County, California [4]. With the 19,436,826 nationwide cases and 208,339 deaths in only six months, it became evident that the world was in the most deadly virus outbreak since the Spanish flu pandemic in 1918 [5].

Recent studies analyzed the risk factors affecting the distribution of the COVID-19 pandemic [6-8]. Such risk factors reflect the correlations between pandemic patterns and topographic (e.g., spatial), environmental (e.g., road density and temperature), demographic (e.g., race and gender), socioeconomic (e.g., household income and health status), and behavioral (e.g., smoking), characteristics [9-15]. Particularly focused on sociodemographic factors, Johns Hopkins Medicine described how catching the coronavirus can be a significant risk for people over age 65 years, with weak immune systems, and have certain chronic diseases, such as asthma, cardiovascular conditions, and diabetes [13]. In another example, American Community Survey (ACS) [14] relates the chance of exposure to COVID-19 and