frontiers | Frontiers in Digital Health

TYPE Methods
PUBLISHED 28 July 2023
DOI 10.3389/fdgth.2023.1193467

OPEN ACCESS

EDITED BY
Heysem Kaya,
Utrecht University, Netherlands

REVIEWED BY
Parisis Gallos,
National and Kapodistrian University of Athens, Greece
Sanne Abeln,
Utrecht University, Netherlands

*CORRESPONDENCE
Debarshi Datta
✉ dr.debarshidatta@gmail.com

RECEIVED 24 March 2023
ACCEPTED 12 July 2023
PUBLISHED 28 July 2023

CITATION
Datta D, George Dalmida S, Martinez L, Newman D, Hashemi J, Khoshgoftaar TM, Shorten C, Sareli C and Eckardt P (2023) Using machine learning to identify patient characteristics to predict mortality of in-patients with COVID-19 in South Florida. Front. Digit. Health 5:1193467. doi: 10.3389/fdgth.2023.1193467

Using machine learning to identify patient characteristics to predict mortality of in-patients with COVID-19 in South Florida

Debarshi Datta[1]*, Safiya George Dalmida[1], Laurie Martinez[1], David Newman[1], Javad Hashemi[2], Taghi M. Khoshgoftaar[2], Connor Shorten[2], Candice Sareli[3] and Paula Eckardt[3]

[1]Christine E. Lynn College of Nursing, Florida Atlantic University, Boca Raton, FL, United States, [2]College of Engineering & Computer Science, Florida Atlantic University, Boca Raton, FL, United States, [3]Memorial Healthcare System, Hollywood, FL, United States

Introduction: The SARS-CoV-2 (COVID-19) pandemic has created substantial health and economic burdens in the US and worldwide. As new variants continuously emerge, predicting critical clinical events in the context of relevant individual risks is a promising option for reducing the overall burden of COVID-19. This study aims to train an AI-driven decision support system that helps build a model to understand the most important features that predict the "mortality" of patients hospitalized with COVID-19.

Methods: We conducted a retrospective analysis of "5,371" patients hospitalized for COVID-19-related symptoms from the South Florida Memorial Health Care System between March 14th, 2020, and January 16th, 2021. A data set comprising patients' sociodemographic characteristics, pre-existing health information, and medication was analyzed. We trained Random Forest classifier to predict "mortality" for patients hospitalized with COVID-19.

Results: Based on the interpretability of the model, age emerged as the primary predictor of "mortality", followed by diarrhea, diabetes, hypertension, BMI, early stages of kidney disease, smoking status, sex, pneumonia, and race in descending order of importance. Notably, individuals aged over 65 years (referred to as "older adults"), males, Whites, Hispanics, and current smokers were identified as being at higher risk of death. Additionally, BMI, specifically in the overweight and obese categories, significantly predicted "mortality". These findings indicated that the model effectively learned from various categories, such as patients' sociodemographic characteristics, pre-hospital comorbidities, and medications, with a predominant focus on characterizing pre-hospital comorbidities. Consequently, the model demonstrated the ability to predict "mortality" with transparency and reliability.

Conclusion: AI can potentially provide healthcare workers with the ability to stratify patients and streamline optimal care solutions when time is of the essence and resources are limited. This work sets the platform for future work that forecasts patient responses to treatments at various levels of disease severity and assesses health disparities and patient conditions that promote improved health care in a broader context. This study contributed to one of the first predictive analyses applying AI/ML techniques to COVID-19 data using a vast sample from South Florida.

KEYWORDS

COVID-19 pandemic, random forest classifier, gini index, feature analysis and prediction, SHAP (Shapley additive explanation), SMOTE (Synthetic minority over-sampling techniques), AI/ML, caring data science