

https://dspace.org/

The use of machine learning techniques in identifying gender differentials in COVID-19 hospitalizations, probabilities of hospitalization outcomes and hidden correlations with demographic and clinical factors

Malaatjie, Meghan Abigail

2024
https://hdl.handle.net/10539/37622

Downloaded from a DSpace repository. https://dspace.org/

The use of machine learning techniques in identifying gender differentials in COVID-19 hospitalizations, probabilities of hospitalization outcomes and hidden correlations with demographic and clinical factors

Meghan Malaatjie | 1155443

School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa| Email: 1155443@students.wits.ac.za

Abstract

Background:

Sex-differentiated data on hospitalisation frequency, case severity, pre-existing medical conditions, and mortality outcomes amongst Covid-19 hospitalised patients is needed but limited in Gauteng province, the epicentre of the Covid-19 pandemic in South Africa. This study aims to investigate whether Machine Learning techniques can provide insight into gender differentials in COVID-19 hospitalizations throughout the four waves of the pandemic, in the Gauteng province of South Africa.

Method:

A weak supervision learning algorithm was used to perform binary classification. The training of a DNN was performed on 14 features of patient characteristics (Demographic variables, presence of co-morbidity, care received upon admission and setting of care), to separate the two classes of data sets: a) severe disease class (a proxy measure of higher severity, which included those who died during admission or were admitted into an intensive care (ICU) or high care unit (HCU)), and b) less severe disease class.

Results:

The number of Covid-19 hospitalisations was highest in wave 3 for both males and females, and higher in females than males across all 4 waves. The observed difference in COVID-19 hospitalization frequency between men and women was the highest in the 20 - 40-year age group with a ratio of 1:3. There was a higher frequency of COVID-19 hospitalization for hypertension, diabetes, and HIV frequencies across all age groups.

Conclusion:

This study demonstrated the utility of machine learning for analysing multidimensional sex-disaggregated data to provide accurate, real-time information for public health monitoring of sex-differences in the Gauteng province.