

CORRELATION BETWEEN SOCIO-DEMOGRAPHIC AND CLINICAL PARAMETERS WITH SEVERITY OF DISEASE IN COVID-19 PATIENTS

General Medicine

Dr. Vijaykumar B. Barge*	Assistant Professor, Dept. of General Medicine. RCSM, GMC Kolhapur *Corresponding Author
Dr Mahendra-kumar V. Bansode	Associate Professor and Head, Dept. of General Medicine. RCSM, GMC Kolhapur
Dr. Sunita Ramanand	Professor and Head, Dept of Pharmacology, RCSM, GMC Kolhapur
Divya Menon	Executive Director Kikito Ventures Pvt. Ltd
Ranjeet Ghatage	Head of Engineering & marketing Kikito Ventures Pvt. Ltd
Mohit Patil	Head of HR & IR Kikito Ventures Pvt. Ltd
Lakshyakumar Singh	Data scientist Kikito Ventures Pvt. Ltd
Vinod Patil	Technical Executive Kikito Ventures Pvt. Ltd
Radha Ghorpade	Health Officer Kikito Ventures Pvt. Ltd

ABSTRACT

Background: Coronavirus disease (COVID-19) is an infectious disease, which has spread rapidly throughout the world. Current investigation studies determined that artificial intelligence (AI) and machine learning (ML) techniques can play a key role in reducing the spread of the virus. ML application technologies on patients' data fall under a range of different research direction. One of the most important research directions is predicting the infection rate and mortality rate and building a model to classify patients based on their clinical findings. **Objectives:** To find out correlations between various socio-demographic and clinical parameters affecting severity levels of COVID-19 and developing a prediction model to predict severity of disease in an individual. **Methodology:** In the present study, 3146 laboratory confirmed cases of COVID-19 were included. This study covers the various analysis & prediction of the survival and the death of COVID-19-positive patients using demographic, vital signs, and chronic diseases, respectively. The study was performed using a retrospective dataset from various hospitals. The data was generated between April 2021 and September 2021. In this study, we utilized various methods of finding correlation. Data related to socio-demographic parameters and clinical parameters with time durations and Blood investigations included. The study established a machine-learning algorithmic approach that trained on hospital records of 3146 individuals. **Results:** There are 3146 COVID-19 patient records in the dataset in which the number of patients who "survived" and "deceased" were 2362 and 784 respectively. The dataset includes 61% males, 39% females, 57% urban, 43% rural and 22% with co-morbidities. The result of study shows that correlations among the socio-demographic variables with clinical parameters can predict severity with over 99%. The male as well as urban population are at a higher risk of disease. Diabetes Mellitus is the only comorbidity highly correlated with the severity. Out of total 147 below 45 years old death cases 100% non-vaccinated, 21% with comorbidities, 61% with late detection & Treatment. Late detection and treatment correlated with death in over 50% cases, regardless of severity. Out of total 37 mild-death cases 91% non-vaccinated, 40% with comorbidities, 51% with late detection & Treatment. Out of total 19 moderate death cases 89% non-vaccinated, 36% with comorbidities, 58% with late detection & Treatment. Out of total 727 Severe death cases 96% non-vaccinated, 21% with comorbidities, 56% with late detection & Treatment. **Conclusion:** The study conducted contributes in growing awareness in the general population regarding the factors affecting COVID-19 and helps in taking informed and necessary steps in overcoming it. This work lays the groundwork for doing trend analysis of various other diseases. The machine learning model developed in this work, predicts COVID-19 severity in subjects and will serve as a foundation for developing further models for other diseases as well by paving a path for following methodology to develop such systems.

KEYWORDS

Covid 19, severity of disease, severity prediction model, comorbidities

INTRODUCTION

Coronavirus disease (COVID-19) is an infectious disease, which has spread rapidly throughout the world. Psychological pressures on society & healthcare personnel have increased as this virus is highly contagious & there is no standardized treatment protocol worldwide. The pandemic has severely ravaged health systems, economic and social progress globally & emerged as a global emergency¹.

A reliable forecast of the pandemic outbreak is a key for targeted disease countermeasures and for the appropriate design of exit strategies. Current investigation studies determined that artificial intelligence (AI) and machine learning (ML) techniques can play a key role in reducing the spread of the virus^{2,3}. ML application technologies on patients' data fall under a range of different research direction. One of the most important research directions is predicting the infection rate and mortality rate and building a model to classify patients based on their clinical findings^{4,5}.

The aim of this research is to find correlations between various socio-demographic and clinical parameters affecting severity levels of

COVID-19 and to develop a prediction system to estimate the severity of the disease in COVID-19 patients, using risk factors that can also be monitored without visiting the hospital. Moreover, the study explores the impact of vital signs, chronic diseases, preliminary clinical investigations, and demographic features to determine the survival versus the mortality of COVID-19 patients.

METHODOLOGY

There are 3146 COVID-19 patient records in the dataset. This study covers the various analysis & prediction of the survival and the death of COVID-19-positive patients using demographic, vital signs, and chronic diseases, respectively. The sampling technique used was based on the 80-20 rule in which they selected 80% of the total data for training and remaining 20% for validating models after randomization of data.

The study was performed using a retrospective dataset from the CPR hospital, Morya hospital and Ambika laboratory, Kolhapur. The data was generated between April 2021 and September 2021. In this study, we utilized various methods of finding correlation. Data related to

socio-demographic parameters like age, sex, co-morbidities, clinical parameters with time durations like symptoms, number of Vaccination doses, HRCT score. Blood composition parameters like Hemoglobin, WBC, Platelet Count, Sr. Ferritin, Sr. LDH, Sr. Creatine, BUN, ALT (SGPT), AST (SGOT), Eosinophils, HCT, MCH, MCHC, MCV, Monocytes, RDW, Sr. Bilirubin (Direct), Sr. Bilirubin (Total), Sr. Potassium.

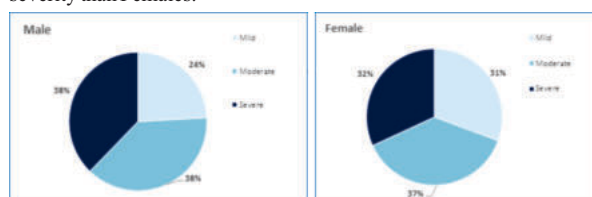
RESULTS

This section contains results obtained in this study and is divided into three sections covering correlation of different parameters with severity, mortality analysis of the subjects and prediction model to determine severity level of the subjects based on various health parameters. There are 3146 COVID-19 patient records in the dataset in which the number of patients who "survived" and "deceased" are 2362 and 784 respectively. The dataset includes 61% males, 39% females, 57% urban, 43% rural and 22% with co-morbidities.

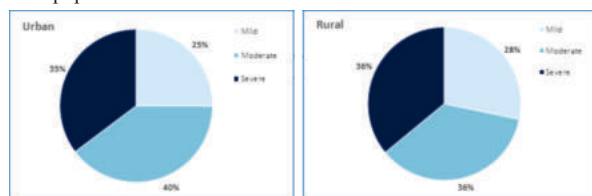
Correlations

This study brought to light several findings, not covered in literature. Key findings are covered next.

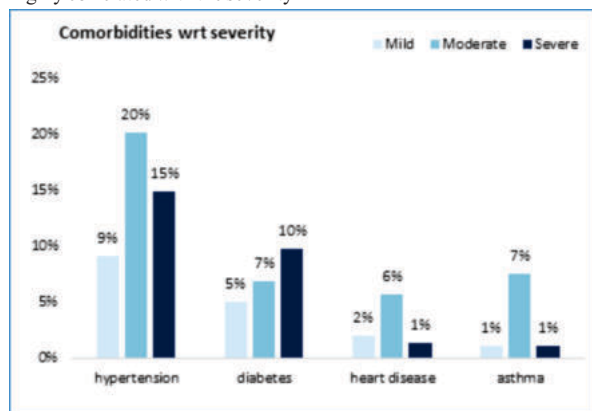
1. The Male population is at a higher risk of disease contraction and severity than Females.



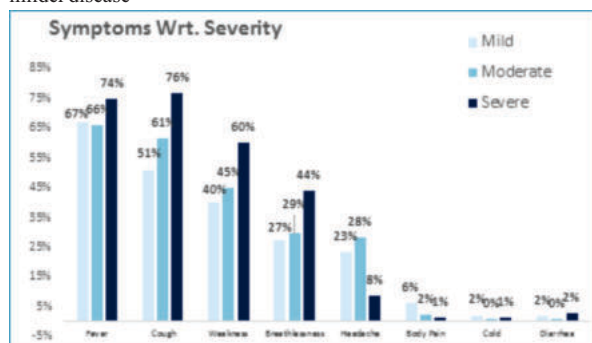
2. The Urban population is at higher risk of severity as compared to rural population.



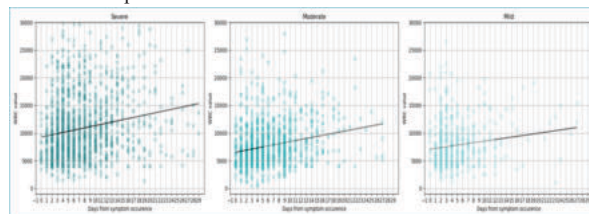
Among comorbidities, Diabetes Mellitus is the only comorbidity highly correlated with the severity



3. Among symptoms, headache and body ache are indicative of milder disease

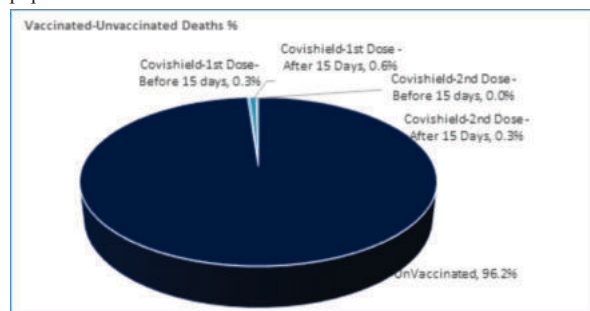


4. Time series studies of blood parameters indicate an inherently higher risk profile of severe patients which can further aid in early detection and prediction of severe diseases.



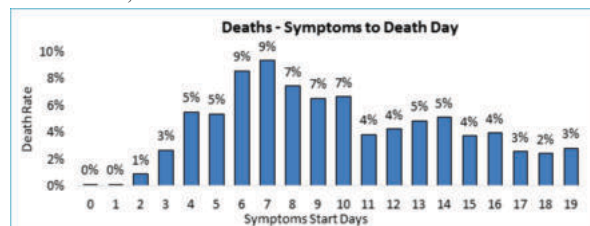
Vaccination Analysis

A huge proactive benefit of vaccination has been observed post both the vaccination doses, with 96% deaths recorded among non-vaccinated population and merely 4% among the vaccinated population.



Mortality Analysis

Death analysis indicates over 50% deaths attributable to late detection and treatment. There exists high correlation of death rate with Age. Out of total 147 below 45 years old death cases 100% non-vaccinated, 21% with comorbidities, 61% with late detection & Treatment. Late detection and treatment correlated with death in over 50% cases, regardless of severity. Out of total 37 mild-death cases 91% non-vaccinated, 40% with comorbidities, 51% with late detection & Treatment. Out of total 19 moderate death cases 89% non-vaccinated, 36% with comorbidities, 58% with late detection & Treatment. Out of total 27 Severe death cases 96% non-vaccinated, 21% with comorbidities, 56% with late detection & Treatment



Prediction Models

Over 100 iterations with different prediction model algorithms resulted in following algorithms coming on top and leading to high accuracies:

Sr. no.	Parameters	Best Classifier	Test Results			Accuracy %
			Severe	Moderate	Mild	
1	Age, Sex, Symptoms, Comorbidities, Vaccination Doses, HRCT Score, Time Durations	Gradient Boosting	Input 262 Pred 251	241 248	107 111	92.46
2	Age, Sex, Symptoms, Comorbidities, Vaccination Doses & Type, HRCT Score, Time Durations + Blood Params	Random Forest	Input 247 Pred 246	257 258	106 106	99.51

DISCUSSION

The study concluded correlations between various socio-demographic variables that gender, age, residency area, history of chronic diseases

and clinical parameters included symptoms, blood parameters, and vaccination status affected severity levels of COVID-19. So with help of this data developed a prediction model system to predict the severity of the disease. Real-time study of data aided with a prediction model can help reduce burden on healthcare system and save lives.

CONCLUSION

The study conducted contributes in growing awareness in the general population regarding the factors affecting COVID-19 and helps in taking informed and necessary steps in overcoming it. This work aligns with the current research of COVID-19 and will help in consolidating the known facts even further. The machine learning model developed in this work, predicts COVID-19 severity in subjects and will serve as a foundation for developing further models for other diseases as well by paving a path for following methodology to develop such systems.

Acknowledgement

This study at Menon Health Tech Pvt. Ltd., Kolhapur was made possible with the support of CPR Govt. Hospital Kolhapur, Morya hospital Kolhapur and Ambika Laboratory Kolhapur. We express our gratitude for their support, guidance and in ensuring smooth collection of data. We would also like to thank Dr. Vijay Kumar Barge, Dr. Sunita Ramanand and record room team from CPR hospital Kolhapur, Dr. Sanjay Kulkarni and Dr. R. S. Patil from Ambika Laboratory Kolhapur, Dr. Santosh Nimbalkar from Morya Hospital Kolhapur, Dr. Kunal Chavan from Primary Health Center Kotoli Kolhapur, who provided us with their invaluable insight in medical field and made this research possible. We would like to thank our whole data collection team which persevered to collect the data and invested over 2500 man-hours to make this endeavor possible

REFERENCES

1. Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine learning-based model to predict the disease severity and outcome in COVID-19 patients. Hindawi Sci Program. 2021
2. WHO Coronavirus (COVID-19) Dashboard. Global situation by WHO region. Available from URL: <https://covid19.who.int/> (cited on August 26, 2022)
3. WHO Coronavirus (COVID-19) Dashboard. India situation. Available from URL: (cited on August 26, 2022) <https://covid19.who.int/region/searo/country/in>
4. Status of Covid 19 in Maharashtra. Available from URL: (cited on August 26, 2022) <https://www.covid19maharashtrgov.in/mh-covid/dashboard>
5. Status of Covid 19 in Kolhapur, Maharashtra. Available from URL: (cited on August 26, 2022) https://www.covid19maharashtrgov.in/mh-covid/dashboard#?_trenddis=Kolhapur