International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

To Predict Deaths Infect of Covid-19 with Machine Learning Using Machine Learning Algorithms

V Anitha

MSC Zoology,Kakatiya University Warangal,Telangana State,India.

anoo.vadlakonda@gmail.com

Abstract:

Prediction of the death among COVID-19 patients can help healthcare providers manage the patients better. We aimed to develop machine learning models to predict in-hospital death among these patients. We developed different models using different feature sets and datasets developed using the data balancing method. We used demographic and clinical data from a multicenter COVID-19 registry. Covid-19 is one of the deadliest viruses you've ever heard. Mutations in covid-19 make it either more deadly or more infectious. We have seen a lot of deaths from covid-19 while there is a higher wave of cases. The epidemic has overloaded advanced health care teams all around the world. WHO is actively looking into and reacting to this epidemic. The current statistical increase in the number of patients has prompted the use of AI approaches to foresee the probable result of a COVID affected patient that will benefit the heath care teams to make a decision on the manner of treatment to be administered. We can use historical data on covid-19 cases and deaths to predict the number of deaths in the future.

Keywords: Covid-19, Classification, CT, Machine Learning, WHO

I Interduction

In spite of more than 2 years since the COVID-19 pandemic and performing vaccination in many countries, the disease's prevalence and mortality have not slowed down, and many countries are still experiencing high peaks [1]. In addition, multiple mutations in the virus have become a new challenge to control the disease, leading to the spread of the disease and increased mortality [2–4]. Until April 16, 2022, more than 500 million cases of the disease and more than 6 million deaths due to COVID-19 have been reported globally, with more than 7 million cases and 140,000 deaths in Iran [1].

Since the beginning of the COVID-19 pandemic, one of the most critical challenges for the healthcare systems has been to increase the number of patients with severe symptoms and the growing demand for hospitalization. In developing countries, which do not have sufficient healthcare infrastructure, the increase in inpatients has put a lot of burden on the healthcare system. Moreover, numerous studies have reported various risk factors such as old age, male gender, and underlying medical conditions (such as hypertension, cardiovascular disease, diabetes, COPD, cancer, and obesity) for the deterioration of COVID-19 patients [5–9].

The use of modern and noninvasive methods to triage patients into specific and known categories at the early stages of the disease is beneficial [10]. One of these approaches is the use of predictive models based on machine learning [11, 12]. For example, developing predictive models based on

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

mortality risk factors can positively prevent mortality through controlling acute conditions and planning in intensive care units [13]. Furthermore, machine learning can classify patients based on the deteriorating risk and predict the likelihood of death to manage resources optimally [14, 15].

To date, several studies have been published on the application of machine learning to develop diagnostic models or predict the death of patients due to COVID-19 [14–23]. For example, several deep learning models have been reported to diagnose COVID-19 based on images [24]. In a study, researchers developed an enhanced fuzzy-based deep learning model to differentiate between COVID-19 and infectious pneumonia (no-COVID-19) based on portable CXRs and achieved up to 81% accuracy.

II Related Work

As per different papers available in literature, there are a few studies that focus on the trend analysis and forecasting for Indian region. The studies [5][6] on Indian region presents long term and short term trend, respectively. These studies use time series data from John Hopkins University database and present forecasting using ARIMA model, Exponential Smoothing methods, SEIR model and Regression Model. However network modelling and pattern mining are not attempted in these versions of the studies and that too at the regional level, hence the current study attempts to do that. Also, the studies in Indian region from the past are more focused on presenting time series analysis based on the overall data for Indian region rather than covering other sources of information apart from just considering the number of infected patients, so the need to analyze the patients background and information is required for the authorities to get better insight about the situation. Similarly, there are other mathematical models that were developed for analyzing the trends of COVID-19 outbreak in India. A model [7] for studying the impact of social distancing on age and gender of the patients in India was presented. It compared the country demographics amongst India, Italy and China and suggested the most vulnerable age categories and gender groups amongst all the nations. The study also predicted the rise of infected cases in India with different lockdown periods. Similarly, a network structure approach was used by one of the study [8] to see whether any specific node clusters were getting formed. But only travel data nodes were considered by the authors to check which the prominent regions are affecting Indian travelers coming back to the India. Also, the study presented the SIR model to see the rate of spread of the Corona Virus amongst patients in India. Analysis on the testing labs and infrastructure was also presented by earlier authors. Work of medical doctors and frontline health workers was also presented by some studies [9]. It was found that in India, the role of health workers was less stressed as the spread stage of corona virus was still in phase two or the phase of local transmission rather than the community transmission as compared to other nations like Italy, Spain and USA. However, it was also claimed that Indian healthcare infrastructure is not very strong as per the WHO guidelines and in case of community spread, the Indian government may find it difficult to manage the spread. Some detailed discussion on the nature of the Corona Virus was also presented by some studies [10][11].

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

III Methodology

1. Imputing Missing Variables

Because of the data quality controls in the registry, the database had a low rate of missing data. The 28 variables had a missing rate below 4%. In machine learning, data imputation is a standard approach to improve the models' performance. Different methods such as imputation with mean, median, or mode are common. We imputed the missing values with the mean for age and the highest frequency of values for nonnumerical variables as well [11, 3].

2. Features and Feature Selection

The outcome measure of the study is in-hospital mortality until discharge which is collected as binary (yes/no). The dataset contains 60 input variables. Age and the number of comorbidities are numerical; oxygen saturation level (PO2) includes two values including below and above 93%. We created three dummy variables for the diagnosis method (only positive PCR, only abnormal CT, positive PCR, and abnormal CT). Other variables have two values: yes or no.

For feature selection, we applied univariate analysis using Chi-square or Fisher exact tests for nonnumerical variables and Mann-Whitney *U* test for age and number of comorbidities (due to abnormal distribution). We created different feature sets to build the prediction models. The first set included all the 60 variables. The second set consisted of variables that were significant in univariate analysis (*P* value <0.05). The third feature set included the marginal variables based on univariate analysis (*P* value <0.2). To create the fourth feature set, we used the feature selection node in the IBM SPSS modeler. This node identifies important features based on univariate analysis as well as the frequency of missing values and the percentage of records with the same value.

IV Implementation

Covid-19 Deaths Prediction using Python

I hope you now have understood the problem statement mentioned above. Now I will import all the necessary Python libraries and the **dataset** we need for the task of covid-19 deaths prediction:

```
1import pandas as pd

2import numpy as np

3data = pd.read_csv("COVID19 data for overall INDIA.csv")

4print(data.head())
```

	Date	Date_YMD	Daily Confirmed	Daily Deceased
0	30 January 2020	2020-01-30	1	0
1	31 January 2020	2020-01-31	0	0
2	1 February 2020	2020-02-01	0	0
3	2 February 2020	2020-02-02	1	0
4	3 February 2020	2020-02-03	1	0

Before moving forward, let's have a quick look at whether this dataset contains any null values or not:

```
data.isnull().sum()
```

```
Date            0
Date_YMD        0
Daily Confirmed 0
Daily Deceased  0
```

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

dtype: int64

We don't need the date column, so let's drop this column from our dataset:

1

```
data = data.drop("Date", axis=1)
```

Let's have a look at the daily confirmed cases of Covid-19:

1

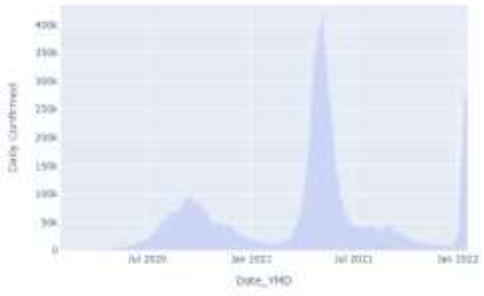
```
import plotly.express as px
```

2

```
fig = px.bar(data, x='Date_YMD', y='Daily Confirmed')
```

3

```
fig.show()
```



In the data visualization above, we can see a high wave of covid-19 cases between April 2021 and May 2021.

V Conclusion

So this is how we can predict covid-19 deaths with machine learning using the Python programming language. We can use the historical data of covid-19 cases and deaths to predict the number of deaths in future. You can implement the same method for predicting covid-19 deaths and waves on the latest dataset. I hope you liked this article on covid-19 deaths prediction with machine learning.

References

1. World Health Organization (2020). Coronavirus disease (COVID-19) Pandemic, WHO. Accessed from https://www.who.int/emergencies/diseases/novel-coronavirus-2019 on 31st March 2020
2. John Hopkins University (2020). Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE. Accessed from https://github.com/CSSEGISandData/COVID-19 on 6th April 2020
3. Sharma, N. (2020). India's swiftness in dealing with Covid-19 will decide the world's future, says WHO, Quartz India. Accessed from https://qz.com/india/1824041/who-saysindias-action-on-coronavirus-critical-for-the-world/ on 25th March 2020
4. 4Myers, J. (2020). India is now the world's 5th largest economy, World Economic Forum. Accessed from https://www.weforum.org/agenda/2020/02/india-gdp-economy-growthuk-france/ on 15th March 2020
5. Gupta, R., & Pal, S. K. (2020). Trend Analysis and Forecasting of COVID-19 outbreak in India. medRxiv. Accessed from https://www.medrxiv.org/content/10.1101/2020.03.26.20044511v1 on 3rd April 2020
6. Gupta, R., Pandey, G., Chaudhary, P., & Pal, S. K. (2020). SEIR and Regression Model based COVID-19 outbreak predictions in India. medRxiv. Accessed from

https://www.medrxiv.org/content/10.1101/2020.04.01.20049825v1 on 5th April 2020

7. Singh, R., & Adhikari, R. (2020). Age-structured impact of social distancing on the COVID-19 epidemic in India. arXiv preprint arXiv:2003.12055. Accessed from https://arxiv.org/pdf/2003.12055.pdf on 4th April 2020
8. Sahasranaman, A., & Kumar, N. (2020). Network structure of COVID-19 spread and the lacuna in India's testing strategy. Available at SSRN 3558548. Accessed from https://arxiv.org/ftp/arxiv/papers/2003/2003.09715.pdf on 3rd April 2020
9. Tanne, J. H., Hayasaki, E., Zastrow, M., Pulla, P., Smith, P., & Rada, A. G. (2020). Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. Bmj, 368.
10. Singhal, T. (2020). A review of coronavirus disease-2019 (COVID-19). The Indian Journal of Pediatrics, 1-6.
11. Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., ... & Agha, R. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). International Journal of Surgery.
12. Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., ... & Davies, N. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. The Lancet Infectious Diseases.