Research Question: What would be the number of deaths caused by COVID19 if the government have not imposed stringiest method such as lockdown?

The two prominently used machine learning models are SEIR [ 3 ] and Regression model [ 4 ] for the disease infec-tion forecasting around the world.

The SEIR model has mainly four components, viz., Susceptible ( S ), Exposed ( E ), Infected ( I ), and Recovered ( R ), as shown in Figure 1 . S is the fraction of susceptible individuals (those able to contract the disease), E is the fraction of exposed individuals (those who have been infected but are not yet in-fectious), I is the fraction of infective individuals (those capable of transmitting the disease), and R is the fraction of recovered individuals (those who have become immune). The most important part of this model is to calculate the R 0 value. The value of R 0 tells about the contagious-ness of disease. It is the fundamental goal of epidemiologists studying a new case. In simple terms R 0 determines an average of what number of people can be affected by a single infected person over a course of time. If the value of R 0 < 1, this signifies the spread is expected to stop. If the value of R 0 = 1, this signifies the spread is stable or endemic. If the value of R 0 > , 1 this signifies the spread is increasing in the absence of intervention. Regression models are statistical sets of processes that are used to estimate or predict the target or dependent variable on the basis of dependent variables.

The regression model has many variants, such as linear regression, ridge regression, stepwise regression, polynomial regression, and so on. This study has used linear regression and polynomial regression [ 5 ] for prediction of COVID-19 cases. Equation ( 1 ) shows the relationship between a dependent and independent variable in polynomial regression. In Equation ( 1 ), x is the independent variable and θ0 is the bias; also, the intercept and θ1 , θ2 , . . . . . . , θn are the weight or partial coefficients assigned to the predictors and n is the degree of polynomial.

There are two basic types of models used by epidemiologists to forecast the path of the COVID-19 pandemic: curve-fitting and structural. Fundamentally, the difference between the two is that curve-fitting relies on patterns in the data without any inference on or assumption about the underlying mechanisms that could shape the data, while the structural approach provides a theoretical framework to guide how the data and forecasts ought to behave in other circumstances, such as with out-of-sample data or alternative policies. Both approaches have advantages and disadvantages, and—depending on the context—different approaches suit different questions a forecaster might ask.

<https://www.clevelandfed.org/newsroom-and-events/publications/economic-commentary/2021-economic-commentaries/ec-202110-two-approaches-to-predicting-the-path-of-the-covid-19.aspx>

Approach discussed with Makeda

1) To create a prediction model relying on all the relevant variable (infected, recovered, deaths, severe, incubated?) using data at OSI < 80%.

2) To find a comparable country that had OSI <80% during all 2020 period.

3) To create linear regression model of infected people using R value and then to extract deaths according to mortality rate for COVID19.