ORIGINAL RESEARCH

Prediction of Epidemic Disease Dynamics on the Infection Risk Using Machine Learning Algorithms

Shanthi Palaniappan[1] · Ragavi V[1] · Beaulah David[2] · Pathur Nisha S[3]

Abstract
Accurate forecast for the public is more important to many organisations especially health organisations on infectious disease dynamics that prevails in prevention or decrease in disease transmission. With multiple data availability in healthcare and medical sectors, precise analysis of such data helps in disease detection and better health care of all individuals. With the existing computational power and big data, there are more chances in predicting an epidemic outbreak. The basic idea of this paper is to analyse and predict the spread of epidemic diseases mainly on the focus on infection risk. A machine learning model using Multivariate Logistic Regression on Modified SEIR has to be built to predict the epidemic disease dynamics on the infection risk.

Keywords COVID-19 · Mathematical model · Epidemics · Pandemic

Introduction

Over the last few days, an ongoing global effort has been driving advances in setting up a global communication network to tackle pandemics of emerging and reemerging infectious diseases. Mathematical modelling plays very essential role at forecasting, examining and analyzing future infections. The research on the study of controlling and modelling epidemic infectious disease dynamics are the major challenges in the current scenario.

✉ Shanthi Palaniappan
  shanthi@skcet.ac.in

  Ragavi V
  ragaviv@skcet.ac.in

  Beaulah David
  beaulahdavid@kce.ac.in

  Pathur Nisha S
  thanish05@gmail.com

1 Sri Krishna College of Engineering and Technology, Kuniamuthur, Coimbatore, India

2 Karpagam College of Engineering, Coimbatore, India

3 Nehru Institute of Technology, Coimbatore, India

The proposal leads to a study on effective modelling using identified machine learning algorithms for implementing for the betterment of society in predicting the infection risk of the disease.

Methodologies for Mathematical Modelling in Epidemics

Major three models used in ongoing epidemics are

I. Statistical methods for tracking outbreaks and for recognizing specific disease spatial patterns

II. Mathematical models for predicting the current epidemic spread within the context of Stat e-space simulations,

III. To predict current epidemic growth and using Machine learning or expert methods.

Machine learning has proven useful in several areas of risk prediction. It has major focus on infection risk, severity risk and outcome risk. Most of the infection risk is finding the chance of having COVID-19 for a particular person or a group. The risk also considers the medical treatment being unsuccessful for a particular person or group and how liable they die [7]. Based on the above risk stated, machine learning has several techniques like Naive Bayes, SVM, and

Neural network, random for prediction of epidemic diseases. Early evidence suggests that age, pre-existing conditions, hygiene practices in general, social behaviours, the number of human interactions, and the occurrence of contacts, socio-economic status, location, and climate are the key threat factors that determine whether an individual will contract COVID-19.

For instance, many authors have used technology mastering to build a preliminary COVID-19 vulnerability Index. Prevention steps like masking, hand washing and social distancing are also likely to reduce the overall risk. When more and more advanced information becomes accessible, and findings are provided by ongoing studies, there evolves additional realistic system mastering applications to predict the risk of infection. Many people undergo the strongest mild symptoms, even though others acquire severe disease on lungs or Acute Respiratory Distress Syndrome (ARDS) [3] that is without doubt deadly. Treatment and monitoring are not possible to anyone closely with minor signs and indications, so if more serious signs and symptoms are likely to occur it is much easier to start care early.

Predicting the procedure's result is an extension of the intensity prediction, which is also just a count of forecasting lifestyles and death. Obviously, with other symptoms, it is helpful to know what the chances a patient will survive. Even at this stage, it is important to remember that not all infected people are treated equally inside. If we can foresee the results of various care approaches, doctors will be able to deal with infected patients more effectively.

Scientists have also used tools that gain knowledge of predicting cancer immunotherapy responses. Nonetheless, because solution options for COVID-19 are emerging, particular treatments are being studied to predict outcomes. Additionally, outcome prediction remains an important aspect of hazard assessment, operating in tandem with the pollution and intensity predictions mentioned above.

The above Fig. 1, shows the framework of any machine learning model for prediction. A machine learning version is better able to test method on a scale. By translating the content material of public interactions on social media, a device-learning model assesses the risk of novel virus infection. The model may not be capable of classifying human beings on a character basis, even then, both of these documents may be used to monitor the pandemic's spread in real time and predict it in the coming weeks. Predicting the threat of recent pandemics and predicting correctly whether or not a coronavirus would move from one person to another will help doctors and health practitioners foresee pandemics of potential and bring together accordingly. The origin of coronavirus has not yet identified accurately. The researchers still undergo lot of studies on several aspects on the origin of COVID-19 and trying hard to justify their findings. While further research is needed to establish a forecast for direct transmission, knowing which influences are most likely to make a leap is a significant first step in pandemic preparation [8].

Mathematical Modelling Using Multivariate Logistic Regression Analysis and Modified SEIR Model

Initial prediction of infectious disease modelling is very important need for future analysis. The logistic model is used primarily in the field of epidemiology. Exploring the risk factors of a certain disease is normal, and predicting the probability of a certain disease occurring according to the risk factors. By logistic regression analysis, we can roughly predict the development and transmission law of epidemiology.

Multivariate logistic regression analysis is an addition of two or more (i.e., simple) regression in which two or more solitary variables (Xi) are taken into concern at the same time to predict a value of a reliant variable (Y) for each subject [6]. The dependent variable is dichotomized or categorical (i.e., multinomial or ordinal) variable when applying logistic regression models.

For example, data are derived from checked assets like John Hopkins University [4], WHO and Ding Xiang Yuan, a website approved through the use of the Chinese government. In addition to recovered cases, the recorded sites showed COVID-19 cases, and deaths for affected foreign locations and areas Computer getting to know has confirmed to be useful in predicting risks in many regions. Infection risk: How likely is it that a character or entity would contract COVID-19? Severity chance: How likely is it that a specific person or organisation will experience serious COVID-19 signs and symptoms or complications that will necessitate hospitalisation or intensive care? Risk of outcome: What are the chances that a particular strategy would be unsuccessful for a specific person or organisation, and how likely they are to die as a result? Age, pre-existing conditions, general hygiene patterns, social behaviours, number
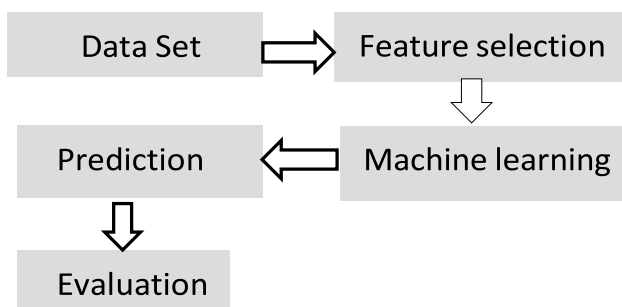

Fig. 1 Framework of Machine Learning model

of human contacts, frequency of contacts, socio-economic status, location, and climate, for example, may all be considered significant factors, using this dataset to estimate risk (odds ratios) of COVID-19. The general form of multivariate logistic regression model can be given as,

$$p = \frac{\exp\left(b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_p X_p\right)}{1 + \exp\left(b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_p X_p\right)} \tag{1}$$

The probability of the outcome given in the above Eq. (1) is present and coded as 1 or 0, X_1 to X_p are distinct independent variables with respect to the parameters that is needed to identify the probability outcome of the Susceptible and all the probability of the Exposed, Infected and Removed are also considered, respectively; and b_0 to b_p are the regression coefficients. The multiple logistic regression model is often written in a different way. The outcome is the predicted log of the chances that the outcome will be present in the form below. To forecast the outbreak using a multivariate logistic regression model, all variables pertaining to the outcome for Susceptible-Exposed-Infected-Removed (SEIR) are taken into account.

The above models provided in the Eqs. (2) to (4) have to be studied to identify the correctness of the prediction.

where S(t): The number of susceptible people in a province.

E(t): The number of exposed people (in a province).

I(t): The number of infected people in a province.

R(t): The number of the recovery or death (in a province).

$$p\left(\frac{ds}{dt}\right) = \frac{-\beta S(t)I(t)}{N} \tag{2}$$

$$p\left(\frac{dE(t)}{dt}\right) = \frac{-\beta S(t)I(t)}{N} - \sigma E(t) \tag{3}$$

$$p\left(\frac{dI(t)}{dt}\right) = \sigma E(t) - \gamma I(t) \tag{4}$$

$$p\left(\frac{dR(t)}{dt}\right) = \gamma I(t). \tag{5}$$

Conclusion

As per the result of this study, implementation on the prediction of the epidemic disease dynamics on the infection risk would be accurately achieved using multivariate logistic regression model over modified SEIR could be achieved. The study would also further enhance on the implementation of the work on data analytics tool to achieve better accuracy.

Declarations

Conflicts of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Abhijit G. (2019) https://medium.com/intel-student-ambassadors/prediction-of-epidemic-disease-dynamics-using-machine-learning-22cf3b7129f3.
2. Berne JD, Cook A, Rowe SA, Norwood SH. A multivariate logistic regression analysis of risk factors for blunt cerebrovascular injury. J Vas surg. 2010;51(1):57–64.
3. Force ADT, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E. Acute respiratory distress syndrome. JAMA. 2012;307(23):2526–33.
4. Johns Hopkins University. (2020). https://coronavirus.jhu.edu/
5. Li MY, Muldowney JS. Global stability for the SEIR model in epidemiology. Mathemat bioscie. 1995;125(2):155–64.
6. Van Wees JD, Osinga S, van der Kuip M, Tanck M, Hanegraaf M, Pluymaekers M. This paper was submitted to the Bulletin of the World Health Organization and was posted to the COVID-19 open site, according to the protocol for public health emergencies for international concern as described in Vasee Moorthy et al.
7. Markus S (2020). https://towardsdatascience.com/fight-covid-19-with-machine-learning-1d1106192d84.
8. Siettos CI, Russo L. Mathematical modeling of infectious disease dynamics. Virulence. 2013;4(4):295–306.