COVID – 19 SURVIVAL PREDICTION USING SUPERVISED MACHINE LEARNING MODELS

1T.SELVABANUPRIYA, 2D.THANMAYASREE NAGA MANIKANTA, 3A.VENKATA NAGA SRI SATYA SAI PAVAN KALYAN, 4M. MONIKA
1 priya8517@gmail.com, 2 thanmaya.manikanta@gmail.com, 3 pavankalyanattuluri@gmail.com, 4 mandepudimonika@gmail.com

1 2 3 4 Department of Computer Science and Engineering
BIST, BIHER

ABSTRACT

COVID-19 and related viruses have spread around the world, posing new threats to our society. There is a clear motivation to implement protective measures that aid in the prevention of outbreaks. The effect of the COVID-19 pandemic has prompted a flood of studies aimed at better understanding, tracking, and controlling the disease. Machine learning is increasingly becoming more prevalent in the area of medical diagnosis. This can be attributed largely to advancements in disease classification and identification systems, which can provide evidence that assists medical experts in the early detection of deadly diseases, resulting in a dramatic rise in patient survival rates. In this paper, we would like to introduce a method of predicting the probability of survival of an individual infected with COVID-19, which is troubling and widely distributed in the current case scenario, using recent algorithmic improvements that were created to predict death and recovery rates.

KEYWORDS: COVID – 19 , Data Analysis , Machine Learning , Supervised Machine Learning Algorithms.

I. INTRODUCTION

Machine learning (ML) has proved itself as a prominent field of study over the last decade by solving many very complex and sophisticated real-world problems. The application areas included almost all the real-world domains such as healthcare, autonomous vehicle (AV), business applications, natural language processing (NLP), intelligent robots, gaming, climate modeling, voice, and image processing.

ML algorithms’ learning is typically based on trail and error method quite opposite of conventional algorithms, which follows the programming instructions based on decision statements like if-else. One of the most significant areas of ML is forecasting, numerous standard ML algorithms have been used in this area to guide the future course of actions needed in many application areas including weather forecasting, disease forecasting, stock market forecasting as well as disease prognosis. Various regression and neural network models have wide applicability in predicting the conditions of patients in the future with a specific disease. There are lots of studies performed for the prediction of different diseases using machine learning techniques such as coronary artery disease, cardiovascular disease prediction, and breast cancer prediction. In particular, the study is focused on live forecasting of COVID-19 confirmed cases and study is also focused on the forecast of COVID-19 outbreak and early response. These prediction systems can be very helpful in decision making to handle the present scenario to guide early interventions to manage these diseases very effectively. The most challenging aspect of its spread is that a person can possess the virus for many days without showing symptoms. The causes of its spread and considering its danger, almost all the countries have declared either partial or strict lockdowns throughout the affected regions and cities. Medical researchers throughout the globe are currently involved to discover an appropriate vaccine and medications for the disease. Since there is no approved medication till now for killing the virus so the governments of all countries are focusing on the precautions which can stop the spread. Out of all precautions, “be informed” about all the aspects of COVID-19 is considered extremely important. To contribute
to this aspect of information, numerous researchers are studying the different dimensions of the pandemic and produce the results to help humanity.

II. EXISTING SYSTEM

A supervised learning model is built to make a prediction when it is provided with an unknown input instance. Thus in this learning technique, the learning algorithm takes a dataset with input instances along with their corresponding regressor to train the regression model. The trained model then generates a prediction for the given unforeseen input data or test dataset. This learning method may use regression techniques and classification algorithms for predictive models' development.

Four regression models have been used in this study of COVID-19 future forecasting:

- Linear Regression
- LASSO Regression
- Support Vector Machine
- Exponential Smoothing

The results of the study prove that ES performs best in the current forecasting domain given the nature and size of the dataset. LR and LASSO also perform well for forecasting to some extent to predict death rate and confirm cases. According to the results of these two models, the death rates will increase in upcoming days, and recoveries rate will be slowed down. SVM produces poor results in all scenarios because of the ups and downs in the dataset values. It was very difficult to put an accurate hyperplane between the given values of the dataset.

III. PROPOSED SYSTEM

In the proposed system, we have mainly used an algorithm named cox-regression to detect the survival rate. This is done mainly by the process of taking death and recovered data as input and estimating the final output of the survival respectively. In this project, we use censoring which is most common in survival analysis. We can know the time when it occurred. It is a form of instance-based learning where the function is only locally approximated and all the computation is postponed until prediction.

![System Architecture](image)

After getting the data which is required from the various data sources and the raw data is created as a Datasets, the data is pre-processed for removing the noise and extracting the features and then the data is divided into testing set and training set, in our case the test data contains 1141 instances and training data contains 2661 instances. This data is given to the algorithm namely Cox survival regression.

The idea behind Cox’s proportional hazard model is that the log-hazard of an individual is a linear function of their covariates and a population-level baseline hazard that changes over time. Mathematically:
In the above equation, the partial hazard is a time-invariant scalar factor that only increases or decreases the baseline hazard. Thus changes in covariates will only inflate or deflate the baseline hazard. After giving the data to the model and the following results are obtained.

From the above visualizations we can understand the accuracy of the model in predicting the death rate and survival rate. The more the data the accurate will be the results.

IV. FUTURE ENHANCEMENT

With technological development, the use of automated surveillance and management systems are growing in demand. The mechanism suggested can detect the disease previously as it happens. It is therefore feasible to save the loss and to reduce the reliance on the specialist to some extent. It can assist an individual with less understanding of the disease. We have to obtain the characteristics corresponding to the disease depending on these objectives. Using this system as framework a new concept of intelligent farming can be implemented where, using the self-operating systems, the field conditions are controlled and monitored.

V. CONCLUSION

With the capacity to differentiate, this advanced model is able to acknowledge survival taking distinct kinds of parameters into consideration from previous covid patients and their symptoms history. All the essential steps required to implement this model of survival recognition are fully described throughout this paper, starting with the importing of packages, data reading, create a database based on the insights drawn after data segregation. Regression is the projects main method which includes picture acquisition, ROI adjustment, classification,
function extraction and convolution. Lifeline’s module python library and python programming language is used to manipulate the raw data and to create a machine learning model that can predict the type of disease.

REFERENCES